

地理大数据为地理复杂性研究提供新机遇

程昌秀^{1,2,3}, 史培军^{1,2,3}, 宋长青^{1,3}, 高剑波^{1,3}

(1. 北京师范大学地表过程与资源生态国家重点实验室, 北京 100875; 2. 北京师范大学环境演变与自然灾害教育部重点实验室, 北京 100875; 3. 北京师范大学地理科学学部, 北京 100875)

摘要: 大数据之风自2010年席卷全球,已在科学、工程和社会等领域产生深远影响。本文首先从地理大数据、第四范式以及非线性复杂地理系统3组基本概念出发,剖析上述3组概念之间的科学联系与相互支撑作用,提出大数据和第四范式为地理复杂性研究提供新机遇。其后,探讨如何利用大数据和复杂性科学的理论方法开展地理复杂性研究。基于地理大数据,可以通过统计物理学的系列指标描述现实地理世界的复杂非线性特征,同时,还可利用深度学习、复杂网络、多智能体等方法,实现复杂非线性地理系统的推演和模拟。上述方法对认知地理现象和过程的复杂性,对复杂地理系统的分析、模拟、反演与预测有重要作用。最后,提出地理大数据和复杂性科学相互支撑可能成为21世纪地理学的主流科学方法。

关键词: 地理大数据;第四范式;非线性;地理复杂性

DOI: 10.11821/dlxb201808001

1 引言

自2010年以来,大数据之风以“迅雷不及掩耳之势”席卷全球。大数据已成为继云计算、物联网之后新一轮的技术变革热潮。过去8年,科学、工程和社会等领域分别围绕手机信令、社交媒体、智能刷卡、搜索引擎等数据开展了很多应用研究,上述数据多与空间位置相关,推动了地理大数据商业应用的发展。在过去一轮竞争和热炒中,收益最大的是拥有数据的单位。他们一边拥有宝贵的数据资源,一边研发着具有潜力的大众应用(例如百度路况、滴滴打车等),赚得个“钵满瓢盈”。而多数从事地理大数据研究的科研工作者不得不依附于数据拥有单位,只能从数据企业获得有限的的数据资源开展研究。运气好的做出不错的成果,很快被企业付诸于实践;运气差的则一边接受来自统计领域“样本偏性和误差”^[1]质疑,一边也在思考、甚至怀疑地理大数据分析的科学意义。因此,国内外反思大数据的文章不断出现。吴志峰等提出了地理学碰上“大数据”的热反应与冷思考^[2]。2014年4月,《纽约时报》发表题为“大数据带来的八个(不,是九个)问题”的文章。同样,英国重量级报刊《财经时报》也刊发了“大数据:我们正在犯大错误吗?”的文章。在过去一轮的大数据热潮中,大数据的价值是否被夸大?大数据背后是否还有其他温润的科学价值?值得地理学研究工作深思。

在大数据热炒之前,国内外曾掀起过一场继相对论和量子力学之后的科技革命——复杂性科学(Science of Complexity)。复杂性科学起源于20世纪80年代,是一门研究复

收稿日期: 2018-06-02; 修订日期: 2018-07-10

基金项目: 国家自然科学基金项目(41771537); 北京师范大学人才启动项目 [Foundation: National Natural Science Foundation of China, No.41771537; Talent Start Project of Beijing Normal University]

作者简介: 程昌秀(1973-), 女, 新疆人, 博士, 教授, 主要从事自然灾害时空分异规律与风险、地理大数据分析等方面的研究。E-mail: chengcx@bnu.edu.cn

1397-1406 页

杂性和复杂系统的前沿交叉科学,它打破了线性、均衡、简单、还原的传统范式,致力于研究非线性、非均衡和复杂系统带来的种种新问题。复杂性科学目前仍处于初级阶段,但已被霍金、成思危等科学家誉为“21世纪的主流科学”。地理系统是一个典型的、开放的复杂巨系统^[3]。地理复杂性逐步受到学者的关注,并成为地理学研究的重要特性。杨国安等思考了人地系统的复杂性^[4],李双成等提出了复杂性科学视角下的地理学研究范式和生态系统服务研究范式^[5-6],高江波等总结了LUCC研究中的实证主义范式、科学人文主义范式、结构功能主义范式以及复杂性范式^[7],但这些研究都极少涉及到地理大数据的相关概念。

复杂性科学与地理大数据之间是否存在潜在的联系?马振刚等简单提及了大数据可以为地理复杂性问题提供解决方案^[8],但尚未进行深度剖析。宋长青认为地理大数据科学范式有望为短时间尺度地理事件发生的监测和预测提供有力的科学和技术支持^[9],但尚未涉及复杂性的相关概念。本文从大数据、科学研究的第四范式以及复杂地理系统等概念出发,剖析上述概念间的相互支撑关系,结合复杂性分析方法和地理大数据实例,从非线性、复杂性的视角介绍地理现象或地理系统的复杂性及其应用,对地理大数据的研究本质提出一种新的解读,以期抛砖引玉,吸引更多的学者研究讨论。

2 重要概念的范畴与约定

“大数据”除具有规模性(Volume)、多样性(Variety)、高速性(Velocity)和价值性(Value)的“4V”特征外,至今尚无统一定义。本文讨论的地理大数据是指通过各类传感器、物联网、连接用户与设备的网络,可自动实时获取、并且持续更新的、具有地理位置信息的长时间序列数据。地理大数据包括源于社交网络、公交刷卡、GPS定位、智能手机的用户位置数据^[10],包括基于无线传感器网络技术收集的地面台站观测数据^[8],也包括对地观测的遥感数据,甚至还包括某些组织或个人收集的长时间序列的专题数据库,例如,GDELT(<https://www.gdeltproject.org>)、Lexis(<https://www.lexisnexis.com>)等新闻媒体数据库,CBDB(<https://projects.iq.harvard.edu/chinesecbdb/cbdb-api>)等历史人物传记数据库,EMDAT(<https://www.emdat.be>)等自然灾害数据库^[11]。

至今对第四范式的解读不少见,但是本文仅从地理复杂性的角度,解读第四范式和地理大数据。视角不同对第四范式和地理大数据的理解可能存在差异,因此本文不排除其他视角的解读。

3 地理大数据与地理学研究范式

本文以宋长青^[9]和Jim Gray^[12]的范式为蓝本,融入了复杂性科学、地理科学、社会科学的相关背景和概念,形成了地理学科学研究的四个范式,旨在厘清地理大数据与复杂性研究的关系,突出数据密集性科学发现对地理复杂性研究的重要性。

3.1 地理学研究的范式

第一范式:地理经验范式。从脱离动物界进入文明时期,人类就开始用经验范式认识自然和社会。地理经验研究的假设是:地理空间不重复,即地理空间具有绝对的差异性和相对的近似性。研究目标是定义区划指标,刻画区域类型和区域差异;研究数据的特点是对地理现象的定性、定量记录和描述;研究方法是通过调查、测量和制图,建立地带性规律,形成地理区划。地理经验范式的缺点是受到认知能力和实验条件的限制,

难以对地理现象进行更精准的理解。

第二范式:地理实证范式。17世纪-19世纪末经典物理学的盛世,带动了地理实证研究的发展。地理实证研究的假设是:在封闭有边界系统中,地理要素演化遵守物质和能量守恒定律。研究目的是刻画格局与过程的联系,探求其动力学方程;数据特点是根据科学问题,通过采样获得精确的小数据;研究方法是基于还原论的思想,通过实验、统计,对地理学各要素进行动力建模;典型案例是计量地理学的相关研究。地理实证研究在一定程度上理解了大尺度的地理现象的动力过程;但是以还原论、经典牛顿力学理论为基础,受到了系统科学、非线性复杂系统的冲击挑战。

第三范式:地理系统仿真范式。该范式产生的时代背景是20世纪中叶计算机的产生,以及随后对科学实验进行模拟仿真的模式得到迅速普及。地理仿真研究的假设是:地理要素共存于同一个系统中,地理要素相互依存、协同演化。研究目的是刻画地理类型区和地域综合体多要素协同演化规律;数据特点是根据科学问题,获得种类丰富的地理数据;采用的方法主要是计算机仿真模拟,在自然地理中常采用自上而下的模型进行模拟,例如天气预报;在人地关系中常采用系统科学自下而上的方式模拟系统预测未来。地理仿真研究可以推演出越来越多的未知自然现象或人文现象。计算机虽然能够对部分自然系统进行精确的仿真,但难以模拟人地关系本身的复杂性。主要原因为:①虽然目前空间随机过程仿真取得了较好效果,但是地理现象中常见的非齐次泊松过程仿真效果尚不理想^[13];②仿真仅依赖于人类对系统推演规则的简单总结,缺乏对复杂现实世界的非线性认识及相关参数的获取;③由于缺少大数据案例修正仿真的参数和推演规则,此阶段的仿真结果总显得有些苍白。

第四范式:数据密集型地学发现。该范式产生的时代背景是21世纪初大数据时代的到来。第四范式的研究假设是:地理要素存在于一个开放的复杂地理系统中,要素相互依存、协同演化;研究目的是揭示现实世界中地理现象发生本质,监测和预测地理事件的发生和发展;数据特点是通过通信(讯)技术、互联网技术、物联网技术可实时自动获取、并持续更新具有地理位置信息的大数据。研究方法是数据驱动的复杂性科学方法,即可以基于地理大数据用幂律、分形、混沌等方法认识复杂现实世界的非线性特征参数,例如,地震的幂律与分形、水文序列在时间序列上的分形(长程记忆性);也可以基于地理大数据用深度学习、复杂网络、多智能体等方法,生成现实世界的复杂非线性推演规则,例如,疾病在复杂人群网络中的传播与演化等。第四范式改变传统封闭系统的假设,转向数据驱动的研究,因此可能会得出之前没有认识到的一些复杂现象或理论。但是地理学中数据驱动的复杂性分析还有很长的路要走,主要原因为:①复杂性科学的哲学思维还不够普及,且缺既懂地学又懂复杂性科学的人才去深入地研究地学问题;②虽然大数据可以提供大量训练样本修正仿真的参数和推演规则,但是在地学领域仍需积累相关大数据案例;③大数据案例与深度学习方法的结合能解决什么层次的科学问题也待探索。

近期,Zimmermann等基于历史大数据总结了火焰系统混沌特征,再用机器学习推演混沌火焰系统在未来8个李雅普诺夫时间内的演化过程^[14]。这充分证明了大数据驱动的第四范式和非线性复杂性科学对传统第三范式(计算机仿真)的推动作用。

3.2 第四范式与其他范式的关系

上述四种范式是人类认识自然(社会)现象的历史演化过程,不是逐渐替代的过程;它们都是我们认识世界、进行地学研究的有效方式。四种范式从对立逐渐走向融合,逐步弥补各自缺陷,并在认识论、方法论上逐渐形成“通宏洞微”的连续谱。

地理实证研究的本质缺陷是用小数据来证明逻辑，即用简单的数量关系来应对复杂的自然或社会问题，用小数据、小样本来简单外推数据空缺区间的地理时空变化特征，由于统计回归的内生性问题和数据上无法匹配，有时会导致逻辑上无法自恰；而大数据的优势就在于用数据来发现逻辑。大数据分析技术的进步，也会促进了第一、第二范式的发展，海量数据的规模效应和全新特征使得定性研究和定量研究在资料获取和分析方法上互补。近年情报学领域的知识图谱研究^[15]证明了这种融合的可行性以及带来的惊人效果。

第三、第四范式在成果上都表现为：对未知趋势的推演和预测，但第四范式更注重现实世界复杂非线性特征的刻画。第三、第四范式的具体区别如表 1 所示。可见，第四范式在数据归纳、逻辑（理论）发现以及非线性建模等方面弥补了第三范式的缺陷。

表 1 第三、第四范式的区别
Tab. 1 Difference between the third Paradigm and the four Paradigm

	第三范式	第四范式
研究对象	仿真基于系统的数学建模	面向海量数据的知识发现
推理逻辑	仿真是依据模型演绎得出结果	依据大数据归纳得出数学模型
自动化程度	只有仿真实验是计算自动完成	从数据获取、建模到分析预测，都是计算机自动进行
解释力度	仿真基于假设的建模，为理论解释奠定了坚实的基础	缺乏假设的理论基础，解释力较低
基础设施	仿真可能涉及一台或多台计算机	大数据则涉及更多基础设施，包括自动获取数据的各类传感器、连接用户、物联网与电脑的网络设施

4 复杂性科学的相关概念

复杂性科学早期主要集中在概念和哲学理解阶段。近年来，随着物理学领域对复杂现象研究的深入，“新三论”（耗散结构论、协同学、突变论）^[16]为解决复杂性问题提供了新的方法和思路，促进了全新自然观和方法论的发展。

4.1 复杂性的特征概念与范畴

复杂性至今没有公认的定义。基于作者对国内外复杂性相关研究的理解，总结传统动力学系统（简单系统）与复杂系统的区别（表 2），可以帮助理解复杂性的相关概念。

这里重点介绍复杂性研究的非线性和不确定性两个重要特征。

表 2 传统动力学系统与复杂系统的对比
Tab. 2 Comparison of traditional dynamic system and complex system

	传统动力学系统(简单系统)	复杂系统
系统特点	机械的、没有进化功能的系统	有生命周期、可以进化的系统，会出现涌现现象
研究案例	机械结构，理想气体	细胞、生物、大脑、社会、城市、生态
可预测性	可预测的。已知初始值，通过方程表征其趋势规律，可预测系统未来时刻的状态。例如，钟表宇宙理论。	更侧重研究系统的不确定性。 ● 系统的测不准特性，开启了分形的研究； ● 系统的初值敏感、不可预测性，开启混沌中的秩序研究，例如倍周期、费根鲍姆常数等。
研究方法	经典牛顿力学	复杂性科学方法
数学/哲学思维	线性，可还原	非线性，不可还原
能量守恒定律	封闭系统，遵守	开放系统，不遵守
时间之矢	过程可逆	过程不可逆
统计分布	正态、泊松、指数	幂律、Zipf、Pareto、Gamma

(1) 非线性:地理实证研究沿袭的是经典牛顿力学的“还原论”,即将高层的、复杂的现象可以被清晰地分解为可重组的简单粒子或部分。地理学常将陆地表层系统拆分解水、土、气、生、人分别开展研究。还原论是迄今为止自然科学研究最基本的方法,人们习惯于以静止、孤立的观点考察组成系统诸要素的行为和性质,再将这些性质“组装”起来形成对整个系统的描述。随着科学研究的不断深入,科学家认识到系统中不同要素相互作用过程中会产生新的效益,即整体大于部分之和。因此,对于复杂地理系统,将系统各要素简单叠加的做法是受限的。文中的线性指的是线性系统,即整体是各部分的线性叠加;而非线性则指整体不再简单地等于各部分之和,可能出现不同于“线性叠加”的增益或亏损。正是由于意识到系统存在“线性叠加”的增益或亏损,黄秉维先生提出应加强陆地表层多要素相互作用的研究。

(2) 不确定性:牛顿三大定律在科学研究中起了重要的作用。拉普拉斯曾断言:根据牛顿定律,只要知道宇宙中所有粒子的当前位置和速度,原则上就有可能预测粒子任何时刻的情况,即“钟表宇宙”的图景。地理实证研究阶段也基本基于的这个思路开展研究。钟表宇宙的图景原则上是可行的,但实际上是存在问题,因为现实世界显然不会像钟表一样沿着可预测的路径运行。20世纪“测不准原理”和“混沌”的发现,彻底击碎了钟表宇宙精确预测的梦想^[17]。既然系统初始状态测不准,且系统存在初值敏感的混沌现象,未来任何时刻的预测将无从谈起。当然“测不准原理”和“混沌”也不是复杂到不可认知的程度。复杂性科学则更注重研究这些复杂现象中的本质特征,例如用分形描述测不准背后的特征,用混沌运动、奇怪吸引子、通向混沌道路等描述混沌现象。

4.2 复杂性研究的方法特征

复杂性研究不存在孤立于其他学科的方法和理论。1999年“香山科学会议”总结了复杂性科学研究的主要特征:①研究对象是复杂系统,②研究方法是定性判断与定量计算相结合、微观分析与宏观分析相结合、还原论与整体论相结合、科学推理与哲学思辨相结合;③研究深度不限于对客观事物的描述,而是更着重于揭示客观事物构成的原因及其演化的历程,并力图尽可能准确地预测其未来的发展^[4]。

尽管复杂性研究采用整体论的思维开展研究,但其研究也是有边界的。首先,应把研究问题限定在某一个层次上。离开研究的层次,复杂性就是一个无法度量、具有无限深度的虚假问题。其次,要限定研究的粒度。第一个限定不仅涉及观察者对事物认识的深度,而且也涉及事物本身的结构层次问题;第二个限定则不仅涉及观察者的认识能力,也涉及事物可认识的理论极限等问题。

5 地理大数据中的复杂性分析方法

近年来,陆表系统复杂性的研究越来越受到国内外相关机构和学者的重视。美国地理学家协会(Association of American Geographers, AAG)2012年会提出地理信息技术应与地理学各专业领域结合开展各类复杂自然、经济和社会问题的研究^[18]。傅伯杰在新时代自然地理学发展的思考中,提出需要深化耦合自然与人文要素及过程,研究建立发展复杂系统模拟模型的研究^[19]。宋长青等提出复杂性是地理学继区域性、综合性之后的第三大重要特性,将成为地理学发展的新路径^[20]。

随着大数据时代的到来,地理观测数据的空间密度、时间密度以及数据种类的丰富程度得到极大提高,为地理复杂性研究提供新机遇。基于地理大数据,可以通过统计物理学的系列指标(例如,标度指数、分形维数、Hurst指数、李雅普诺夫指数、混沌的倍周期与吸引子等)描述现实世界的复杂非线性特征;同时,基于地理大数据,还可利用

深度学习、复杂网络、多智能体等方法,实现复杂非线性系统的推演和模拟。因此,从复杂系统理论出发,借助复杂性理论的基本数学工具,认识地理现象和过程的复杂性,对复杂地理系统的分析、模拟、反演与预测有重要作用^[21]。

本文重点介绍统计学派中地理复杂性相关的分析方法,对于系统学派的方法仅介绍复杂网络。此外,由于深度学习方法在解释地理复杂性机理方面还有不足,故暂不做介绍。

5.1 复杂的幂律分布及其标度指数

在大数定律和中心极限定律主导的年代,科学家一般都假定大部分自然现象服从正态分布。然而,在复杂非线性的现实世界中幂律分布比“正态分布”还要正态(normal),即数据的幂律分布是很正常的事情而不是意外。幂律分布特征体现了研究对象的复杂性。幂律分布具有如下性质:① 没有特征尺度:尾部概率密度高于正态,导致无数学期望;② 长尾分布特征:从高端到低端具有无限度延伸的趋势,形成长尾分布^[22];③ 自相关和偏自相关拖尾,代表记忆性和长程作用;④ 具有标度对称性,即伸缩变换下的不变性;⑤ 微分、积分结果依然为幂律。但是在传统计量地理的研究中,常常选择用对数正态分布函数来模拟幂律分布,因为对数正态分布具有特征参数,容易从数学上解析。这种替代会错误地将没有特征尺度的分布当作有特征尺度来研究,其解释和预测效果可想而知^[23]。因此,面对地理复杂分布,人们应该重新认识是地理学的计量化和理论的前因后果^[24]。

在开放的复杂地理系统中,符合幂律分布的现象比比皆是。例如,城市位序—规模、地震规模、月球表面上月坑直径、战争伤亡人数、国家GDP、复杂网络中节点度、行星间碎片大小、太阳耀斑强度、城市居民收入等分布。这些分布中我们不能简单地用大数定律和中心极限思考问题,无法找到具有特征尺度的测度描述现象。例如,市民收入是典型的幂律分布,很难想象通过简单平均得到的工资均值竟能用于描述市民的平均收入。标度研究的意义在于:提供了一个新指标对无特征尺度的复杂现象进行解释和预测^[23]。例如,Clauset等开创性地发现恐怖袭击伤亡人数服从标度值为2的幂律分布,且无论是整体还是不同类武器的分布均为标度不变的幂律分布^[25];Clauset等进一步确认了恐怖袭击中死亡人数普遍存在的幂律分布及其时空稳定性,其研究结果有助于预测恐怖袭击的死亡人数^[26-27]。陆松发现火灾中死亡人数、直接经济损失或过火面积都满足幂律分布,然后,运用标度指数分析了6种影响因素对频率—死亡人数分布的影响^[28]。计算社会科学的Cioffi教授系统分析了灾害和人道主义风险中存在的系列幂律分布规律和相关的灾害风险分析方法^[29]。

5.2 空间分形与时间分形

分形是指局部与整体之存在着一种自相似性,其本质也是一种标度量。这种自相似性也是复杂系统自组织特性的一种体现。地理现象存在空间分形和时间分形两种类型。

所谓空间分形可理解为地理现象在不同空间尺度上表现出来的一种形态自相似性。例如,海岸线形态在不同空间尺度下存在自相似性,森林火灾的过火面积在不同尺度的形态上存在自相似性,地震断裂带具有分形特征。这些不同尺度上的自相似性常用分形维数表示,通常分形维数越高表示系统驱动因素、关系、层次越复杂。例如,不同地区海岸线的分维数可以表征该海岸线的复杂程度。陆松利用深林火灾过火面积的分形指数,分析了中日两国火灾的分异规律^[28]。在地震案例数据有偏的情况下,Parvez等利用地震的幂律分布与分形规律,推演出无偏的地震危险性分布图^[30]。Tao等利用降雨的分形规律,对卫星遥感降雨产品进行降尺度,提高了水文模型输入变量的分辨率^[31]。分形过程是一个简单的规则加上一点随机输入,重复演化,生成复杂的现象。如何找到这个规则是地理学研究的重要内容之一。

所谓时间分形可理解为地理现象在不同时间尺度上表现出来的一种形态自相似性。基于长时间序列的高频监测数据可以探寻这种时间分形规律(即长程相关性或长程记忆)。早在20世纪40年代,英国水文学家就提出了用于表征非季节性河流水文过程长程记忆特征的Hurst指数。长程相关性的概念起源较早,目前在金融领域的应用较多^[32]。地理学者也开始尝试利用长程相关性和互相关性,解释不同要素间的相互驱动、响应、反馈和耦合规律,从而理解地理过程与地理系统的非线性特征。Bowers通过树轮的长程相关性,研究气候的记忆性^[33]。刘祖涵利用塔里木河流域气候—水文观测数据研究了该地区气候—水文过程的长程记忆性及其强度空间分布^[34]。史凯等利用成都及其周边地区空气污染指数,分析不同地区该指数的长程相关性和互相关性,探讨了该区污染的耦合与运移规律^[35-36]。

5.3 地理中的混沌现象

简单地讲混沌是用来描述对初始条件敏感的动力系统,即在动力系统中如果初始值稍有偏差就可能导致系统朝着截然不同的方向运行,例如蝴蝶效应。

当然复杂性科学研究中的混沌也不是完全不可预测的。尽管混沌存在着初值敏感、难以预测的特点,但这些混沌现象其实也有踪迹可寻。李雅普诺夫指数(Lyapunov)是用于识别地表系统混沌运动常用的特征指数之一。王卫国等利用李雅普诺夫指数,发现北半球臭氧层系统是一种耗散混沌的运动,不同纬度臭氧层系统的相空间总体上是收缩的,同时提供了存在奇怪吸引子的证据^[37]。高志球等基于黑河实验站的水平风速数据,计算并分析关联维数、李雅普诺夫指数和科尔莫哥罗夫熵等混沌特征量,表明干旱地区大气边界层湍流是一种混沌运动^[38]。Gao等用尺度依赖的李雅普诺夫指数(SDLE)研究了科罗拉多河和安普夸河的间歇性混沌特征^[32]。刘祖涵利用李雅普诺夫指数分析了塔里木地区气候—水文系统中蝴蝶效应的强弱,以及年径流量系统蝴蝶效应的强度及其空间分布^[34]。Zimmermann等基于大数据和机器学习,可以预测出混沌火焰系统在未来8个李雅普诺夫时间内的演化过程^[14]。上述混沌现象的研究为我们更深刻地理解我们赖以生存的地理环境提供了新的视角。

5.4 复杂网络

复杂网络是具有自组织、自相似、吸引子、小世界、无标度中的部分或全部性质的网络。无标度网络中节点的度数服从幂律分布,而交通网络节点的度分布通常为泊松分布。在地理研究中,国家间关系^[39]、疾病传播^[40],甚至于空气温度^[41]、水文波动^[34]等都可以被构造为复杂网络开展研究。

复杂性科学研究的核心概念之一就是深刻理解从微观基本单元到宏观复杂结构和统计规律之间的涌现机制。可以说,当前有关复杂网络的众多研究都与该问题相关,例如宏观动力学过程、宏观结构和宏观统计规律的涌现等。在有些研究中,直接找到从微观个体到全网的宏观结构比较困难,两个端点之间尚需要一个中间过渡,也就是以模体(motif)和群落(community)为代表的中观结构。这些中观结构的分布、在动力学过程中扮演的作用,以及从微观到中观,从中观到宏观的涌现过程,都是值得高度关注的。复杂网络在地理研究中的应用主要集中在3个方面:①复杂网络的静态特征,度分布、小世界特征、模体和群落的特征等,例如,胡小兵等针对复杂网络提出了凝聚度指标,用于描述与测度社会生态系统抗干扰能力^[42];②复杂网络的形成机理,例如Wang等用中国34个城市的气温构建了个地区关系网络,将34个城市集中在4个模块中,不同模块的灾害发生的行为各有不同,研究结果有助于编制灾害应急预案^[41];③复杂网络上的动力学机制,鲁棒性、疾病传播等,例如何大韧等介绍了基于复杂网络上的传播动力学的SIS模型和SIR模型^[40]。

5.5 方法小结

尽管上面给出了一些非线性研究的地理案例,但以上这些研究还是比较粗浅,受限的主要原因是:数据本身的限制和对地理系统结构的认识不足。当今,地理学进入大数据时代,地理大数据为开展地表复杂非线性研究提供了机遇,为认识地理复杂性的本质提供了可能。通过地理复杂性研究可开启认识地理现象和过程的新视角,以便更深刻的认识地理现象的结构、功能和过程。在后续地理复杂性研究中,应充分利用地理大数据,结合复杂地理现象的问题,融合或发展更多的复杂非线性分析方法。

6 结论

地理大数据驱动的第四范式将突破传统封闭系统的假设,以真实、开放的地理复杂系统为研究对象,打破了第二范式中经典牛顿力学的建模方法。运用大数据挖掘工具进行统计和计算,进而对内容进行分析,避免了第三范式因数据有限而导致的系统功能结构划分的主观性,或智能主体行为刻画的缺陷性,可能会得出之前没有认识到的一些系统特征或理论。

地理大数据驱动的第四范式为复杂系统研究奠定基础。随着大数据时代的到来,地理系统的相关属性数据、时空数据和行为数据能全面、真实反映地理各要素的状态以及演化过程,为捕捉复杂地理系统中的标度指数、长程相关性、李雅普诺夫指数以及混沌的倍周期与吸引子等特征提供的机遇。此外,基于地理大数据还可以刻画智能主体的非线性参数和推演规则,实现复杂现实世界的推演和仿真。

地理复杂性方法的研究已经取得了一定成果,表明复杂性方法逐渐受到地理学领域的重视,且抓住了地理复杂现象和过程的本质。当然,复杂性认识和方法十分年轻,复杂性方法的运用和创新任重道远,还有很多困难和未知等待科学家解决和认识。但是无论采用何种方法,复杂性研究方法的本质是抓住地理系统中非线性和不确定的特征,找到了描述复杂性本质的简化方法,而不是简单采用线性化方法。

地理大数据的积累推动了地理复杂性科学研究的发展,同时地理复杂性的研究又加固了地理大数据在科学范式演化革命中的地位。因此,地理大数据和复杂性科学相互支撑可能将成为21世纪地理学的主流科学方法。

参考文献(References)

- [1] Wang J F, Stein A, Gao B B, et al. A review of spatial sampling. *Spatial Statistics*, 2012, 2(1): 1-14.
- [2] Wu Zhifeng, Chai Yanwei, Dang Anrong, et al. Geography interact with big data: Dialogue and reflection. *Geographical Research*, 2015, 34(12): 2207-2221. [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. *地理研究*, 2015, 34(12): 2207-2221.]
- [3] Qian Xuesen, Yu Jingyuan, Dai Ruwei. A new discipline of science: The study of open complex giant system and its methodology. *Nature Journal*, 1990, 13(1): 3-10. [钱学森, 于景元, 戴汝为. 一个科学新领域: 开放的复杂系统及其方法论. *自然杂志*, 1990, 13(1): 3-10.]
- [4] Yang Guo'an, Gan Guohui. Thoughts on complexity of human-earth system. *Science and Technology Review*, 2002(3): 58-60. [杨国安, 甘国辉. 人地系统复杂性思考. *科技导报*, 2002(3): 58-60.]
- [5] Li Shuangcheng, Wang Yang, Cai Yunlong. The paradigm transformation of geography from the perspective of complexity sciences. *Acta Geographica Sinica*, 2010, 65(11): 1315-1324. [李双成, 王羊, 蔡运龙. 复杂性科学视角下的地理学研究范式转型. *地理学报*, 2010, 65(11): 1315-1324.]
- [6] [Li Shuangcheng, Liu Jinlong, Zhang Caiyu, et al. The research trends of ecosystem services and the paradigm in geography. *Acta Geographica Sinica*, 2011, 65(11): 1618-1630. [李双成, 刘金龙, 张才玉, 等. 生态系统服务研究动态及地理学研究范式. *地理学报*, 2011, 65(11): 1619-1630.]
- [7] Gao Jiangbo, Cai Yunlong. Paradigm shift in the study of land use cover change. *China Population, Resources and*

- Environment, 2011, 21(10): 114-120. [高江波, 蔡运龙. 土地利用/土地覆被变化研究范式的转变. 中国人口·资源与环境, 2011, 21(10): 114-120.]
- [8] Ma Zhengang, Li Lili, Xu Xuegong. Big data research of physical geography. *Geography and Geo-Information Science*, 2015, 31(3): 54-64. [马振刚, 力黎黎, 许学工. 自然地理学的大数据研究. 地理与地理信息科学, 2015, 31(3): 54-64.]
- [9] Song Changqing. On paradigms of geographical research. *Progress in Geography*, 2016, 35(1): 1-3. [宋长青. 地理学研究范式的思考. 地理科学进展, 2016, 35(1): 1-3.]
- [10] Zhen Feng, Qin Xiao, Wang Bo. Human geography research and practical application in big data era. *Human Geography*, 2014, 29(3): 1-6. [甄峰, 秦萧, 王波. 大数据时代的人文地理研究与应用实践. 人文地理, 2014, 29(3): 1-6.]
- [11] Shen S, Cheng C X, Song C Q, et al. Spatial distribution patterns of global natural disasters based on biclustering. *Natural Hazards*, 2018, 92(3): 1809-1820.
- [12] Hey T. The fourth paradigm: Data-intensive scientific discovery. Microsoft Research, 2009: 8-10.
- [13] Sullivan D, George L W P. *Spatial Simulation: Exploring Pattern and Process*. John Wiley & Sons, 2013: 213.
- [14] Zimmermann R S, Parlitz U. Observing spatio-temporal dynamics of excitable media using reservoir computing. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2018, 28(4): 043118.
- [15] Shen S, Cheng C X, Yang J, et al. Visualized analysis of developing trends and hot topics in natural disaster research. *PLoS One*, 2018, 13(1): e0191250.
- [16] Zhou Chenhu. *Researches of Geographical Cellular Automata*. Beijing: Science Press, 1999. [周成虎. 地理元胞自动机研究. 北京: 科学出版社, 1999.]
- [17] Tang Luyi. *Complexity*. Changsha: Hunan Science and Technology Press, 2018: 1, 22-23. [唐璐译. 复杂. 长沙: 湖南科学技术出版社, 2018: 1, 22-23.]
- [18] Guo Huiquan, Zhang Guoyou, He Shujin. American geography research hot spots in recent years. *Geographical Research*, 2013, 32(7): 1375-1377. [郭慧泉, 张国友, 何书金. 近年来美国地理学研究热点问题. 地理研究, 2013, 32(7): 1375-1377.]
- [19] Fu Bojie. Thoughts on the recent development of physical geography. *Progress in Geography*, 2018, 37(1): 1-7. [傅伯杰. 新时代自然地理学发展的思考. 地理科学进展, 2018, 37(1): 1-7.]
- [20] Song Changqing, Cheng Changxiu, Shi Peijun. Geography complexity: New connotations of geography in the new era. *Acta Geographica Sinica*, 2018, 73(7): 1189-1198. [宋长青, 程昌秀, 史培军. 新时代地理复杂性的内涵. 地理学报, 2018, 73(7): 1189-1198.]
- [21] Wei Yiming. Complexity theory of flood disaster. *Nature Journal*, 1999, 21(3): 139-142. [魏一鸣. 洪水灾害研究的复杂性理论. 自然杂志, 1999, 21(3): 139-142.]
- [22] Jiang B. Head tail breaks for visualization of city structure and dynamics. *Cities*, 2015, 43(3): 69-77.
- [23] Chen Yanguang. Simplicity, complexity, and mathematical modeling of geographical distributions. *Progress in Geography*, 2015, 34(3): 321-329. [陈彦光. 简单、复杂与地理分布模型的选择. 地理科学进展, 2015, 34(3): 321-329.]
- [24] Philo C, Mitchell R, More A. Reconsidering quantitative geography: Things that count. *Environment and Planning A*, 1998, 30(2): 191-201.
- [25] Clauset A, Young M. Scale invariance in global terrorism. *EPrint Arxiv physica*, 2005: 0502014.
- [26] Clauset A, Shalizi C R, Newman M E. Power-law distributions in empirical data. *SIAM Review*, 2009, 51(4): 661-703.
- [27] Clauset A, Wiegand F W. A generalized aggregation-disintegration model for the frequency of severe terrorist attacks. *Journal of Conflict Resolution*, 2010, 54(1): 179-97.
- [28] Lu Song. Spatial, temporal distribution and influencing factors of high-casualty fire in China [D]. Hefei: University of Science and Technology of China, 2012. [陆松. 中国群死群伤活在时空分布规律[D]. 合肥: 中国科技大学, 2012.]
- [29] Cioffi R C. Seeing it coming: A complexity approach to disasters and humanitarian crises. *Complexity*, 2014, 19(6): 95-108.
- [30] Parvez I A, Nekrasova A, Kossobokov V. Earthquake hazard and risk assessment based on unified scaling law for earthquakes: State of Gujarat, India. *Pure and Applied Geophysics*, 2017, 174(3): 1441-1452.
- [31] Tao K, Barros A P. Using fractal downscaling of satellite precipitation products for hydro meteorological applications. *Journal of Atmospheric and Oceanic Technology*, 2010, 27(3): 409-427.
- [32] Gao J B, Cao Y H, Tung W W, et al. *Multiscale Analysis of Complex Time Series: Integration of Chaos and Random Fractal Theory, and Beyond*. John Wiley & Sons, 2007.
- [33] Bowers M C, Gao J B, Tung W W. Long range correlations in tree ring chronologies of the USA: Variation within and across species. *Geophysical Research Letters*, 2013, 40(3): 568-572.
- [34] Liu Zuhan. A study of the complex and nonlinear investigation of climatological-hydrological process in the Tarim River

- Basin [D]. Shanghai: East China Normal University, 2014. [刘祖涵. 塔里木河流域气候—水文过程的复杂性与非线性 [D]. 上海: 华东师范大学, 2014.]
- [35] Shi Kai, Liu Chunqiong, Wu Shenghu. Long range correlation of air pollution between Chengdu city and its surrounding towns. *Resources and Environment in the Yangtze River*, 2014, 23(11): 1633-1640. [史凯, 刘春琼, 吴生虎. 基于 DCCA 方法的成都市市区与周边城镇大气污染长程相关性分析. *长江流域资源与环境*, 2014, 23(11): 1633-1640.]
- [36] Wu Shenghu. Fractal feature and risk evaluation of PM_{2.5} evolution during typical haze periods [D]. Jishou: Jishou University, 2016. [吴生虎. 重度灰霾期间高浓度 PM_{2.5} 的时空分形演化及预测预警 [D]. 吉首: 吉首大学, 2016.]
- [37] Wang Weiguo, Zhao Huazhu, Xie Yingqi, et al. A study on chaotic attractor character of zonal average ozonosphere in Northern Hemisphere. *Acta Geophysica Sinica*, 1997, 40(3): 317-324. [王卫国, 赵华柱, 谢应齐, 等. 北半球不同纬度臭氧层系统混沌吸引子的特征研究. *地球物理学报*, 1997, 40(3): 317-324.]
- [38] Gao Zhiqiu, Wang Jiemin. A study on the chaotic characteristics of the atmospheric boundary layer turbulence over oasis and desert in HEIFE area. *Plateau Meteorology*, 1998, 17(4): 397-402. [高志球, 王介民. HEIFE 绿洲和沙漠地区大气边界层湍流混沌特性研究. *高原气象*, 1998, 17(4): 397-402.]
- [39] Ding Rongjun. The complexity turn in international relations theory: "Complex systems" as a research program [D]. Beijing: China Foreign Affairs University, 2012. [丁榕俊. 国际关系理论的复杂性转向: “复杂系统”研究 [D]. 北京: 外交学院, 2012.]
- [40] He Daren, Liu Zonghua, Wang Binghong. *Complex Systems and Complex Networks*. Beijing: Higher Education Press, 2012: 198-207. [何大初, 刘宗华, 汪秉宏. 复杂系统与复杂网络. 北京: 高等教育出版社, 2012: 198-207.]
- [41] Wang J, Yang H. Complex network-based analysis of air temperature data in China. *Modern Physics Letters B*, 2009, 23(14): 1781-1789.
- [42] Hu Xiaobing, Shi Peijun, Wang Ming. Consilience degree: A new network property to evaluate system's performance against disturbances. *Scientia Sinica (Informationis)*, 2014, 44(11): 1467-1481. [胡小兵, 史培军, 汪明. 凝聚度: 描述与测度社会生态系统抗干扰能力的一种新特性. *中国科学: 信息科学*, 2014, 44(11): 1467-1481.]

Geographic big-data: A new opportunity for geography complexity study

CHENG Changxiu^{1,2,3}, SHI Peijun^{1,2,3}, SONG Changqing^{1,3}, GAO Jianbo^{1,3}

(1. State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China; 2. Key Laboratory of Environmental Change and Natural Disaster, Beijing Normal University, Beijing 100875, China; 3. Faculty of Geographical Science, Beijing Normal University, Beijing 100875, China)

Abstract: Since 2010, big data has played a significant role in various fields of science, engineering and society. The paper introduces the concepts of geographic big-data, the fourth paradigm and nonlinear complex geographic system, and discusses interactive relationships of these concepts. It is proposed that geographic big-data and the fourth paradigm would become a new opportunity to research on geography complexity. Then the paper discusses how to use the methods of geographic big-data and complexity science to examine geography complexity. For example, based on big-data, a series of indicators of statistical physics fields could be constructed to describe the complex nonlinear characteristics of the real geographic world. Deep learning, complex network and multi-agent methods can be used to model and simulate the complex nonlinear geographic systems. These methods are important for a better understanding of the complexity of geographic phenomena and processes, as well as the analysis, simulation, inversion and prediction of complex geographic systems. Finally, the paper highlights that the combination of geographic big-data and complexity science would be the mainstream scientific method of geography in the 21st century.

Keywords: geographic big-data; the fourth paradigm; nonlinearity; geography complexity