

文章编号:1003-2053(2026)01-0094-07

复杂网络中的科学学:普适规律还是数字陷阱?

郑碧丽,侯剑华

(中山大学信息管理学院,广东广州 510006)

摘要:本研究探讨了复杂网络和计算科学学范式在科学学中的应用,聚焦于揭示科学活动背后的普适规律与规避数字陷阱的挑战。随着数智化时代的到来,大规模数据为科学学研究提供了全新的视角,复杂网络通过分析合作网络、知识传播路径等揭示科学研究中的隐性规律。而计算科学学则借助先进的算法和模型,推动了科学学研究的智能化。然而,数据驱动的方法也带来了“数字陷阱”的潜在风险,即过于依赖量化分析可能忽视科学活动中的复杂性和背景性因素。因此,本研究不仅探讨了复杂网络和计算科学学在科学学研究中的优势,还深入分析了如何平衡数据驱动与领域知识,以避免过度数字化带来的局限性。最终,本文旨在为科学学研究的未来发展提供新的思路,并呼吁更加谨慎和综合的方法以应对新时代的科学挑战。

关键词:科学学;复杂网络;计算科学学;数据驱动

中图分类号:G301

文献标识码:A

作为一门以科学自身为研究对象的学科,科学学始终与科学技术等发展实践同向同行,在推动和支撑我国科技创新与科技政策、科技管理与科技发展战略等方面发挥了重要作用^[1]。自20世纪初创立以来,科学学综合运用社会科学和自然科学的多学科理论和方法,对科学及其社会功能进行多视角研究和阐释,在揭示科学的本质及其社会嵌入性方面取得了重要成果。然而,随着科学规模的扩大和科学活动的复杂性不断增加,传统方法的局限性也日益凸显^[2]。其主要局限体现在以下几个方面:(1)样本有限。传统的科学学研究依赖手工搜集数据,往往只能覆盖有限的科学文献或科学家个体,难以全面把握科学活动的全貌;(2)研究的主观性较强,定性分析依赖研究者的解释和判断,可能带有较强的主观性,不易量化,也难以进行大规模的比较研究;(3)时间跨度有限,传统的研究方法无法及时追踪科学发展的动态变化,难以应对科学活动的迅速发展和复杂性的挑战。

进入数智化时代,科学活动产生了大量可量化的数据,诸如论文、专利、科研合作记录、政策文本、社交媒体文本等,构成了“科学大数据”的基础^[3]。

这些数据为科学学研究带来了前所未有的机遇,使得我们可以通过大规模数据分析更为全面和动态地理解科学活动的运行机制和发展趋势。在此背景下,复杂网络和计算科学学逐渐成为新的研究范式,为科学学的研究开辟了新的路径^[4]。在这样的研究背景下,英国剑桥大学出版社出版的 Science of Science 一书,系统地呈现复杂网络方法以及计算科学学范式对科学学研究的影响。这些新范式通过分析科学活动中的合作模式、知识传播路径以及科研成果的影响力,不仅推动了科学学的量化进程,还为科学管理、政策制定提供了新的工具。然而,这种数据驱动的研究方法是否能够揭示科学学的普适性规律,仍是一个值得深入探讨的问题。因此,本研究也关注另一个关键问题,即“数字陷阱”。美国信息计算学家 Cassidy R. Sugimoto 认为本书的观点是用数字堆砌出的“生涯成功”(Scientific success by numbers)^[5]。基于此,我们需要思考在数据主导的研究范式下,科学研究是否可能因为过度依赖数字化方法,而忽视了某些复杂性和背景性因素,从而陷入数字化分析的局限中?为了应对这一潜在的风险,本研究在概述复杂网络以及计算科学学的范式带来新

收稿日期:2024-10-23;修回日期:2025-02-27

基金项目:中国科协创新战略研究院老科学家学术成长资料采集工程项目“张碧晖学术成长资料采集”(CJGC2022-F-Z-XH12);复旦大学国家智能评价与治理实验基地与科睿唯安信息服务(北京)有限公司共同设立的创新评价开放基金项目“面向新文科的科研人员知识创造能力评价方法研究”(CXPJ2024002)

作者简介:郑碧丽(1997-),女,博士研究生。侯剑华(1980-),男,教授、博士生导师,通讯作者,E-mail:hough5@mail.sysu.edu.cn。

机遇的基础上,探索如何平衡数据驱动与领域知识,从而避免“数字陷阱”,推动科学学研究范式转化的进一步完善和发展。

1 Science of Science 概述

Science of Science 由美国西北大学华人学者 Dashun Wang (王大顺) 与美国东北大学 Albert - László Barabási (艾伯特 - 拉斯洛·巴拉巴西) 两位杰出的复杂网络研究专家、计算社会学家共同撰写,2021年由剑桥大学出版社出版^[6]。中文版译名为《给科学家的科学思维》,2021年由天津科学技术出版社出版。该书的核心在于通过数据驱动的分析工具以及复杂网络的分析方法揭示科学研究的机制与规律,主要包括事业成功的科学新思维、团队组建的科学新思维、影响力升级的科学新思维、呼啸而来的新科学等主题内容。

1.1 事业成功的科学新思维

该书的第一部分聚焦于科学职业的生产力与影响力差异,作者通过大规模数据分析揭示了科学家职业发展的关键模式。例如,虽然爱因斯坦、达尔文等科学家因少数几项重大发现而被铭记,但生产力与影响力之间存在巨大的差异。一些科学家以少量作品成就非凡,而另一些则因大量作品保持高影响力。数据表明,科学家的职业发展并非线性,影响力的峰值往往出现在科学职业的不同阶段^[7,8]。通过对53万名科学家的职业生涯进行分析,书中指出,影响科学家生产力的因素不仅仅是时间投入,更与合作、资源获取等复杂因素相关。这部分内容对我们理解科学职业的周期性及其驱动因素提供了深入的见解^[9]。

1.2 团队组建的科学新思维

该书在第二部分探讨了科学合作对科学研究的推动作用。随着科学研究的复杂性增加,“单打独斗”的科学家越来越少,取而代之的是合作团队。通过分析近2000万篇科学论文和210万项专利发现,科学合作不仅成为主流,还极大地提高了研究的影响力^[10]。团队合作产生的论文往往比单独作者的论文引用次数更多,特别是在多学科合作中,这种影响尤为显著^[11]。

书中进一步指出,随着跨学科和跨国界的合作增多,科学研究的全球化趋势日益明显^[9]。研究表明,国际合作在推动科学突破方面发挥了重要作用,

尤其是在一些重大项目中,如粒子物理学的大型强子对撞机(LHC)项目^[6]。

1.3 影响力升级的科学新思维

书稿的第三部分探讨了科学影响力的形成与评估。作者通过对引用文献的研究,揭示了影响力的累积机制^[12]。例如,科学家们倾向于引用那些对自己研究有重要影响的论文,这使得某些高影响力的研究能迅速获得广泛的关注^[13]。然而,书中也指出,引用并不总能完全反映出研究的真实影响力,过度依赖引用计量学可能导致对科学价值的误读^[14]。此外,该书还探讨了科学中的“明星效应”,即少数科学家的工作往往得到比其他科学家更多的关注和认可,尽管这些工作并不一定比其他研究更有价值^[15,16]。这部分内容为理解科学影响力的形成机制和影响评价标准提供了新的视角。

1.4 呼啸而来的新科学

该书的最后一部分展望了未来科学学发展的趋势,特别是人工智能和自动化技术在科学研究中的应用。作者提出,随着数据规模的增加,科学家可以利用人工智能来生成假设、设计实验,甚至执行研究任务。这种自动化的科学研究流程可能会大大加速科学发现的进程^[17]。此外,该书强调了科学政策与科研管理的未来方向,认为随着科学学的进一步发展,科学政策的制定需要基于更为精确的数据分析,以提高科学研究的效率和产出^[18]。作者提出,通过跨学科的合作,科学学的未来研究将更加多元化和系统化,为全球科学界提供更加有效的支持和指导。

2 复杂网络与计算科学学范式带来的新机遇

Science of Science 呈现了复杂网络视角与计算科学学范式在科学生涯发展、科学团队构建、科学影响力预测等多个场景的优势。复杂网络分析通过研究科学活动中的多种关系结构,如科研合作网络、知识传播网络,揭示了科学活动中的隐性规律。计算科学学借助先进的算法和模型,能够自动化地处理海量的科学数据,并从中提取出有价值的信息。与传统的研究方法相比,数字化时代的新范式具有数据量大、分析精确、动态追踪等优势,使得科学学研究进入了一个新的量化和计算驱动的阶段。首先,复杂网络视角推动科学学研究的精细化转向,增强科学学研究的可解释性、可判识性与可推论性。其

次, AI 驱动的计算科学学范式赋能科学学研究的智能化, 包括观测单位、数据标识与处理、复杂关键的构建与解读。

2.1 复杂网络推动精细化科学学研究

虽然复杂网络是一门新兴的研究领域, 但复杂网络机制特点, 尤其是节点复杂性以及动力学过程为科学学研究提供了新的视角, 注入新鲜血液。与第一阶段的社会学分析与第二阶段的计量分析相比, 在动态过程的解释、复杂关系的判识、机制预测的推论三个方面推动科学学研究的精细化转向。

首先, 复杂网络为科学学研究提供动态过程的可解释性。科学发展是动态的过程。传统的社会学分析或基本的数理分析倾向于对于科学现象的静态描述, 并不能充分体现科学发展的动态过程。以基于引文网络的科学发展研究为例, 传统的计量分析研究局限在以滑动引文窗口内网络节点数量、点度中心性、中介中心性等基本网络指标反映引文分布的特征。

其次, 复杂网络为科学学研究提供复杂关系的可判识性。科学的发展本身就是一个复杂的、自组织的、不断进化的网络^[4], 是由社会结构、知识表征和自然世界之间的复杂关系相互作用驱动的一个动力系统。科学知识基于科学论文、著作、专利、政策文本等不同载体, 通过电视、广播、社交媒体等扩散媒介在正式交流渠道和非正式交流渠道中扩散至不同知识受体、学科领域、工业产业实践。复杂网络的动力学模型如随机游走、惰性随机游走、自回避行走、游客漫步、流行病模型等为科学发展的复杂动态过程提供丰富的解释模型。

最后, 复杂网络为科学学研究提供机制预测的可推论性。复杂网络将科学发展概念化为一个由社会发展、知识发展、自然世界发展的相互作用组成的复杂网络。此外, 复杂网络的演化模型及动力过程为科学学研究提供可判识的模型和可解释的关系。因此, 科学学研究中关于科学劳动、科学工作者、社会制度之间的关系可通过复杂网络模型进行结构理解与作用机制预测。复杂网络领域的研究通过海滩游走 (seashore walk)、开发-探索模型 (exploitation-exploration model) 呈现科学家选择研究问题、兴趣转移的过程及机制预测。科学家的研究兴趣与其生产率、被引率有关^[19], 研究主题的异质性和相似性促使科学家选择熟悉的研究领域, 而新近效应科学家探索新的研究问题^[20], 通过科学家的研究兴趣

轨迹可预测其未来研究问题的方向。

2.2 AI 驱动的计算科学学范式赋能智能化科学学研究

“AI for science” 基于海量的数据, 利用强大的数据归纳和分析能力学习科学规律和自然原理的, 总结面向大知识、大科学的计算及推演模型, 为科学学研究提供了重要的研究手段^[21], 促进科学学研究, 尤其是在观测单位、数据标识与处理、复杂关键的构建与解读方面的智能化转向。

首先, 是观测单位从宏观单元到微观知识单元的智能抽取。在社会学分析阶段与计量分析阶段, 科学知识单元一般以中层知识单元的单篇论文、专利, 或高层知识单元的期刊、学科作为研究对象^[22], 因此, 科学学研究中对于科学发展的分析也局限在宏观知识单元层面^[23]。自 2002 年 LDA 主题模型提出以来, 基于科学论文及专利文本的科学知识单元和技术知识单元抽取的效率大大提高。2013 年后, Word2Vec、Top2Vec、BERT 等 AI 大模型, 拥有强大的上下文理解能力、语言生成能力、学习能力, 优化全文本分析与标注、语义识别与预测, 强化科学文本、技术文本的主题抽取效率和表征能力^[24], 推动科学学研究从微观的细粒度层次探讨科学的发展。

其次, 是数据标识从手工标注到机器学习推演。针对复杂的研究问题以及海量的研究样本数据, 手工收集、标注与统计分析难以满足日益增长的数据处理需求。机器学习、深度学习等人工智能技术解决多维度、多模态、多场景的数据采集与模拟仿真, 辅助科研人员在大量的实验中进行验证与试错^[25], 解决复杂的推演计算问题, 加速科学研究与技术开发的进程^[26]。

最后, 是关系构建从外在联系探索到内在机制推断与破解。对科学系统各要素之间的关系探讨一直是科学学研究的重要组成部分。社会学分析阶段的科学学研究常用演绎推理的方法构建要素与要素之间的关系链。计量分析阶段常用回归、相关性分析要素之间的相关关系。自 2010 年以来, 科学学领域借鉴计量经济学的因果模型 (双重差分法、断点回归法、格兰杰因果模型、倾向值匹配法、工具变量法), 以及机器学习中 BART、TMLE 和因果森林模型^[27]探讨的科学系统要素的内在因果机制。对于大规模、高维数据的因果推断需求, 机器学习方法可以快速处理大量数据, 预测因果间的非线性和高阶关系, 提高因果推断的准确性。此外, Python 与 R

提供基于领域知识的因果分析设置,包括随机森林、XGBoost 和超级学习器等即用型软件包,实现科学研究中科学系统要素内在机制的智能化推断与破解。

3 普适性规律还是数字陷阱?

尽管复杂网络与计算科学学范式的发展为科学学发展带来的广阔的机遇与前景,但也引发了一个核心的问题:这些方法是否真的可以揭示普适性规律?换句话说,这会令我们陷入所谓的数字陷阱吗?为了解答这一问题,本文深入分析揭示普适性规律的可能性以及数字陷阱的风险两个方面。

3.1 揭示普适性规律的可能性

Science of Science 在序言中提到,随着数字化数据的大量涌现,科学研究迎来了前所未有的机遇,复杂网络和计算科学学的范式为揭示科学研究背后的普遍规律提供了新途径^[6]。这些方法能够通过定量分析,揭示科学知识的生产与传播的深层结构和机制。复杂网络尤其擅长描绘科学家之间的复杂合作关系,跟踪科研成果的多样传播路径,从而帮助我们发现科学活动中潜在的模式。例如,通过分析大量的论文引用网络,研究者能够识别出高影响力的科学家和研究机构,从而揭示科研成就的分布规律。同时,AI 技术、算法极大程度地提高大规模、多模态、长跨度数据的计算效率,降低计算成本,增强预测准确性^[28, 29]。因此,基于海量数据的 AI 驱动复杂网络计算对于揭示普适性规律具有较高的可能性。然而,从海量数据中提取普适性规律或模式并非易事。例如,在 Science of Science 关于科学合作的部分中,作者指出,尽管复杂网络分析能够揭示出某些合作模式,但科学研究本身的多样性使得找到通用模式变得极具挑战性。再如,关于科学家生涯成功的机制探讨与预测中,作者提出 Q 因子(Q-factor, 研商),量化如何将运气转化为具有持续高影响力的职业生涯^[30]。然而,这个度量标准基于一个假设,即所有研究者拥有平等的资源,却忽略了不同国家或机构之间,不同性别之间,不同学科之间,不同科学家群体之间资源分配极不均衡的事实。不同学科、不同国家以及不同科研机构之间的文化、资源和方法各不相同,使得某些规律可能仅适用于特定的背景,而无法推广到所有科学领域。因此,科学学研究虽然能够提供有价值的见解,但要谨慎对待这

些模式是否真正具备普适性。

3.2 数字陷阱的风险

随着计算科学学和复杂网络的兴起,数据驱动的分析方法成为主流,然而这种过度依赖数据的趋势也带来了潜在的风险。数据驱动的研究方法可能会让我们陷入“数字陷阱”,即过于关注数据中的模式而忽略了这些数据背后的社会、文化和历史因素。过度依赖算法或其他数据驱动的分析方法可能会对科学研究带来一些局限性及盲点。具体来说,包括以下几点:(1)“黑箱效应”。目前大多数算法,特别是深度学习模型,被视为“黑箱”,其内部工作机制不透明,难以解释分析结果的依据。(2)数据偏见。如果在具体运算中使用的数据集存在偏差,可能导致结果的失真。另外,在数据收集过程中,某些数据可能被忽视或未被记录,这会影响结果的准确性。例如,在健康医疗中使用机器学习算法的可能存在潜在偏见,尤其是在涉及不同种族和性别的患者时。研究发现,某些算法在不同群体的医疗影像分析中存在性能差异,可能导致不平等的医疗服务^[31]。(3)忽视定性因素。算法和数据分析往往难以捕捉复杂的人文社会背景和非量化的情感、动机等因素,这可能导致对现象的片面理解。某些现象的意义依赖于特定的社会和文化背景,数据驱动的方法可能忽略这些重要的上下文因素。

Science of Science 在科学影响力的部分,进一步阐述了数据驱动分析方法在不同学科中的适用性差异。例如,在自然科学领域,量化的网络分析和引用分析较为普遍且有效,但在社会科学和人文学科中,这些方法的适用性则受到更多质疑^[32]。人文社科领域的研究往往更依赖于定性分析和深度解读,单纯的数字化方法可能无法揭示出研究中的细微差别和文化背景。因此,在这些领域中,过度依赖数字工具可能会导致对研究对象的片面理解,忽略研究的复杂性。

3.3 如何避免数字陷阱

要避免陷入数字陷阱,关键在于将数据驱动的方法与领域知识相结合。正如 Science of Science 在最后部分中所指出的,尽管数据科学为科学学提供了强大的工具,但仅依赖这些工具是不够的^[6]。科学研究需要结合特定学科的背景知识和专业洞见,才能确保分析结果的准确性和有效性。数据科学家和领域专家之间的协作至关重要,他们需要共同分析数据背后的真实含义,避免过度简化复杂的科学

现象。这样,科学学的研究不仅可以从数据中提取有价值的信息,还能将这些信息与具体的科学背景结合,提供更加深入和全面的解释^[33]。

此外,跨学科的研究方法也是避免数字陷阱的有效途径。在 Science of Science 的序言中,作者提到了科学学的多学科特性。科学学的成功依赖于各学科研究者的共同努力,利用不同领域的知识和方法来解析科学活动的复杂性。通过结合社会科学、物理学、信息科学等多个学科的理论和方法,科学学可以更好地理解不同领域的特性,弥补单一学科方法带来的局限性^[4, 34]。因此,构建跨学科的研究框架,不仅能够提高数据分析的准确性,还能为科学政策制定者和管理者提供更加丰富的参考依据。

因此,避免数字陷阱的关键在于平衡数据驱动与领域知识,采用跨学科的合作方法。这种方式能够帮助我们从大量数据中提取出具有真正科学价值的见解,同时避免因过度依赖数字化工具而产生的误判。

4 关于未来科学学研究的思考

4.1 科学学的前沿趋势

在 Science of Science 的前沿展望部分中,作者提出了人工智能在科学学研究中的巨大潜力。人工智能不仅仅是用来辅助数据分析的工具,它逐渐成为科学研究的核心推动力量,尤其是在处理大规模数据时极大提高运算效率,减低运算成本。通过对科学出版物、专利、以及合作网络进行自动化分析,人工智能可以迅速发现潜在的研究热点、预测科学突破,并提出新的研究假设^[29]。这种能力为科学政策的制定者提供了更多的数据支撑,使得决策能够更加科学和透明。

人工智能的发展还使得科学学有可能进入一个全新的阶段:不仅仅是加速科研过程,还能通过自动化算法,逐步实现科学发现的自动化。这意味着,未来的科研流程可能从“人机协作”走向“全自动化”,使得人工智能能够从数据中自动生成假设、设计实验并进行分析,从而大幅提高科研效率^[28]。当前,许多实验室已经开始利用机器人和人工智能技术自动进行数据处理和实验操作。该书中的“Adam”机器人就是一个典型的例子,它能够从假设的生成到实验设计,完全由算法进行自主操作^[35]。这种科学研究流程的自动化趋势,不仅使研究效率大幅提高,

同时也减少了人为错误的干扰。未来,这种自动化流程有望进一步扩展到更复杂的科学领域,如基因组研究、药物筛选以及材料科学等^[36]。

4.2 跨学科合作的必要性

Science of Science 提出,科学学的未来发展需要依赖于跨学科合作的深化。科学研究的复杂性和大数据的迅猛增长使得单一学科已经难以应对当今科学中的挑战。复杂网络科学、数据科学、社会科学等不同领域的结合,能够为科学学研究提供新的视角。例如,在科研影响力的研究中,网络科学与社会科学的结合使得研究人员能够更好地理解科学影响力的传递与扩散^[4]。

科学学本质上是一个多学科交叉的领域,融合了信息科学、图书馆科学、社会科学等多个领域的理论与方法。该书指出,未来科学学的发展将越来越依赖于各个学科的互补优势,如工程学与生物学的结合将推动新药研发的创新,而信息科学与社会科学的结合则能够提供更深刻的科研合作模式分析。

数据科学与社会科学的结合被认为是未来科学学的重要发展方向。科学学不仅仅是对科学数据的分析,更需要理解这些数据背后的人类行为和社会结构^[37]。通过数据科学中的算法,研究人员可以解析庞大的科学合作网络、引用模式等,而社会科学的视角则可以帮助解释这些模式背后的动因和影响。因此,数据科学与社会科学的结合有望推动科学学更为全面的发展,能够更好地解释科学发现的社会背景和动因。

4.3 对科学政策的影响

Science of Science 强调,随着科学学研究的深入发展,科学政策的制定也需要借助更多新工具和新思维。通过对科学活动的全面量化分析,政策制定者可以更好地理解科学资源的配置情况,并据此做出更为精准的决策。例如,复杂网络分析可以帮助科学政策制定者识别科研合作中的关键节点,进而优化科研资助的分配策略^[38]。同时,人工智能与数据科学的结合使得政策制定者可以实时追踪科研进展,及时调整战略。科学学的研究对提升科学效率具有重要贡献。它不仅为科研人员提供了关于如何提高科研产出的见解,还帮助政策制定者优化科研管理的方式。通过对科学合作、研究产出与引用模式等的量化分析,科学学能够揭示出哪些合作模式更具效率,哪些研究策略更易产生重大突破,这些成果为提高整体科研效率提供了重要的理论

支持^[39]。

参考文献:

- [1] 陈劲. 不断探索科学学中国化时代化新规律[J]. 科学学研究, 2024, 42(1): 1–2. Chen J. Continuously explore new principles for the localization and modernization of science of science in China [J]. Studies in Science of Science, 2024, 42(1): 1–2.
- [2] 郑碧丽, 侯剑华. 数智驱动的科学学: 国内理论科学学研究综述(2017–2023) [J]. 科学学研究, 2025, 43(3): 514–522. Zheng B L, Hou J H. Data-driven science of science: Review of domestic theoretical science of science (2017–2023) [J]. Studies in Science of Science, 2025, 43(3): 514–522.
- [3] Lin Z, Yin Y, Liu L, et al. SciSciNet: A large-scale open data lake for the science of science research [J]. Scientific Data, 2023, 10(1): 315.
- [4] Fortunato S, Bergstrom C T, Börner K, et al. Science of science [J]. Science, 2018, 359(6379).
- [5] Sugimoto C R. Scientific success by numbers [J]. Nature, 2021, 593(7857): 30–31.
- [6] Wang D, Barabási A-L. The Science of Science [M]. Cambridge: Cambridge University Press, 2021.
- [7] Liu L, Wang Y, Sinatra R, et al. Hot streaks in artistic, cultural, and scientific careers [J]. Nature, 2018, 559(7714): 396–399.
- [8] Liu L, Dehmamy N, Chown J, et al. Understanding the onset of hot streaks across artistic, cultural, and scientific careers [J]. Nature Communications, 2021, 12(1).
- [9] Dong Y, Ma H, Shen Z, et al. A Century of Science: Globalization of Scientific Collaborations, Citations, and Innovations [C]. Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Halifax, NS, Canada; Association for Computing Machinery. 2017: 1437–1446. 10.1145/3097983.3098016.
- [10] Wu L, Wang D, Evans J A. Large teams develop and small teams disrupt science and technology [J]. Nature, 2019, 566(7744): 378–382.
- [11] Wuchty S, Jones B F, Uzzi B. The increasing dominance of teams in production of knowledge [J]. Science, 2007, 316(5827): 1036–1039.
- [12] Price D D S. A general theory of bibliometric and other cumulative advantage processes [J]. Journal of the American Society for Information Science, 1976, 27: 292–306.
- [13] Barabási A-L. Network Science [M]. Cambridge: Cambridge University Press, 2016.
- [14] Wang D S, Song C M, Barabasi A L. Quantifying long-term scientific impact [J]. Science, 2013, 342(6154): 127–132.
- [15] Merton R K. The Matthew effect in science, II: Cumulative advantage and the symbolism of intellectual property [J]. Isis, 1988, 79(4): 606–623.
- [16] Simcoe T S, Waguespack D M. Status, quality, and attention: What's in a (Missing) Name? [J]. Management Science, 2011, 57(2): 274–290.
- [17] Evans J, Rzhetsky A. Machine Science [J]. Science, 2010, 329(5990): 399–400.
- [18] Boudreau K, Brady T, Ganguli I, et al. A field experiment on search costs and the formation of scientific collaborations [J]. The Review of Economics and Statistics, 2017, 99(4): 565–576.
- [19] Zeng A, Shen Z S, Zhou J L, et al. Increasing trend of scientists to switch between topics [J]. Nature Communications, 2019, 10: 3439.
- [20] Jia T, Wang D, Szymanski B K. Quantifying patterns of research-interest evolution [J]. Nature Human Behaviour, 2017, 1(4): 0078.
- [21] Zhang X, Wang L, Helwig J, et al. Artificial intelligence for science in quantum, atomistic, and continuum Systems [J]. ArXiv, 2023, abs/2307.08423.
- [22] 赵蓉英, 魏绪秋. 引证视角下的作者知识扩散规律探析[J]. 情报理论与实践, 2016, 39(8): 12–17. Zhao R Y, Wei X Q. Analysis of the law of author knowledge diffusion from citation perspective [J]. Information Studies: Theory & Application, 2016, 39(8): 12–17.
- [23] Liu Y, Rousseau R. Knowledge diffusion through publications and citations: A case study using ESI-fields as unit of diffusion [J]. Journal of the American Society for Information Science and Technology, 2010, 61(2): 340–351.
- [24] Reimers N, Gurevych I. Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks; proceedings of the Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), F, 2019 [C]. Association for Computational Linguistics.
- [25] Peng B, Wei Y, Qin Y, et al. Machine learning-enabled constrained multi-objective design of architected materials [J]. Nature Communications, 2023, 14

- (1): 6630.
- [26] 王飞跃, 缪青海. 人工智能驱动的科学新范式: 从 AI4S 到智能科学[J]. 中国科学院院刊, 2023, 38(4): 536 – 540. Wang F Y, Miu Q H. Novel paradigm for AI – driven scientific research: From AI4S to intelligent science [J]. Bulletin of Chinese Academy of Sciences, 2023, 38(4): 536 – 540.
- [27] Wager S, Athey S. Estimation and inference of heterogeneous treatment effects using random forests [J]. Journal of the American Statistical Association, 2018, 113(523): 1228 – 1242.
- [28] Messeri L, Crockett M J. Artificial intelligence and illusions of understanding in scientific research [J]. Nature, 2024, 627(8002): 49 – 58.
- [29] Gao J, Wang D. Quantifying the use and potential benefits of artificial intelligence in scientific research [J]. Nature Human Behaviour, 2024.
- [30] Sinatra R, Wang D, Deville P, et al. Quantifying the evolution of individual scientific impact [J]. Science, 2016, 354(6312): aaf5239.
- [31] Seyyed – Kalantari L, Zhang H, Mcdermott M B A, et al. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under – served patient populations [J]. Nature Medicine, 2021, 27(12): 2176 – 2182.
- [32] Radicchi F, Fortunato S, Castellano C. Universality of citation distributions: Toward an objective measure of scientific impact [J]. Proceedings of the National Academy of Sciences of the United States of America, 2008, 105(45): 17268 – 17272.
- [33] 陈悦, 王智琦. 数据迷雾下渐行渐近的科学学[J]. 科学学与科学技术管理, 2022, 43(10): 70 – 82. Chen Y, Wang Z Q. The gradually asymptotic science of science under the fog of data [J]. Science of Science and Management of S. & T., 2022, 43(10): 70 – 82.
- [34] Makinen E I, Evans E D, Mcfarland D A. The patterning of collaborative behavior and knowledge culminations in interdisciplinary research centers [J]. Minerva, 2020, 58(1): 25.
- [35] King R D, Rowland J, Oliver S G, et al. The automation of science [J]. Science, 2009, 324(5923): 85 – 89.
- [36] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold [J]. Nature, 2021, 596(7873): 583 – 589.
- [37] 侯剑华. 计算科学学: 数智驱动的科学学新发展[J]. 科学学研究, 2025, 43(7): 1395 – 1402. Hou J H. Computational science of science: New developments in science of science driven by digital and intelligent technologies[J]. Studies in Science of Science, 2025, 43(7): 1395 – 1402.
- [38] Giannakas F, Troussas C, Krouska A, et al. Multi – technique comparative analysis of machine learning algorithms for improving the prediction of teams’ performance [J]. Education and Information Technologies, 2022, 27(6): 8461 – 8487.
- [39] Yoo H S, Jung Y L, Jun S – P. Prediction of SMEs’ R&D performances by machine learning for project selection [J]. Scientific Reports, 2023, 13(1): 7598.

Science of Science in complex networks: General patterns or numerical game?

ZHENG Bi – li, HOU Jian – hua

(School of Information Management, Sun Yat – sen University, Guangzhou 510006, China)

Abstract: This study explores the application of complex networks and computational science of science paradigms in the field of Science of Science, focusing on uncovering universal laws behind scientific activities while addressing the challenges of avoiding the “numerical game.” In the era of big data, large – scale data has provided new perspectives for science studies, where complex networks help reveal hidden patterns in scientific research by analyzing collaboration networks and knowledge dissemination pathways. Computational science of science, powered by advanced algorithms and models, has further enabled intelligent analysis in science studies. However, this data – driven approach also poses potential risks of falling into the “numerical game,” where over – reliance on quantitative analysis may overlook the complexities and contextual factors inherent in scientific activities. This study not only examines the strengths of complex networks and computational science of science in related studies but also delves into how to balance data – driven methods with domain knowledge to avoid the limitations of excessive data dependence. Ultimately, this paper aims to offer new perspectives for the future development of science studies and calls for more cautious and integrative approaches to address the challenges of modern science.

Key words: Science of Science; complex networks; computational science of science; data – driven analysis