

# “科学学”视角下的科研工作者行为研究

贾韬<sup>1</sup>, 夏锋<sup>2</sup>

1. 西南大学计算机与信息科学学院, 重庆 400715;

2. 大连理工大学软件学院, 辽宁 大连 116620

## 摘要

科学的复杂性、数据的可用性、对科学研究中客观规律认知的现实需求使得“科学学”这一传统问题转变为新兴的交叉研究热点。科学研究的主导者是科研工作者, 研究科学中的规律首先要了解科研工作者的行为特征。梳理、介绍了几个相关的工作, 包括科研工作者研究兴趣的演化、科研合作的内在驱动因素、科研团队与科研team的异同、多维度科研评价与学术不端的监测。研究结果可以为科研管理等现实问题提供理论工具和实证支撑。

## 关键词

科学学 ; 科研合作 ; 科研评价 ; 学术画像

中图分类号 : TP39

文献标识码 : A

doi: 10.11959/j.issn.2096-0271.2019040

## *Quantifying patterns in the behavior of scientists in Science of Science study*

JIA Tao<sup>1</sup>, XIA Feng<sup>2</sup>

1. College of Computer and Information Science, Southwest University, Chongqing 400715, China

2. School of Software, Dalian University of Technology, Dalian 116620, China

## *Abstract*

The complexity of science, the data availability and the need for better understanding of science promote the research in an emerging field of “Science of Science”. To understand the regularities in science, we need to first understand the behavior patterns of scientists. A few related works were reviewed and introduced, including research interest evolution, internal microscopic driven force in scientific collaboration, the difference of research tuandui and research team, and the multidimensional measure of a scholar’s performance to prevent academic gaming. The results obtained will provide theoretical tools and empirical supports for practical issues related with research policy and research evaluation.

## *Key words*

Science of Science, scientific collaboration, research evaluation, academia profile

## 1 引言

只有认识科学研究中的客观规律,了解科学工作者的行为特征,才能有效地评价科研成果、奖励科研发现、资助科研工作、培养科研人员<sup>[1-5]</sup>。现代科学已经毫无疑问地成长为一个复杂系统:涉及各个学科领域之间越来越强的交叉关联,新兴学科不断涌现,创新与渐进的知识不断积累,科学工作者在知识空间和物理空间的迁移以及他们在全世界范围内的竞争和合作<sup>[6-9]</sup>。

“科学可以被描述为一个增长并且演化的网络”<sup>[9]</sup>,这个网络由学者与学者间的合作竞争关系、论文与论文间的引用关系、学者与不同研究主题间的参与关系、知识图谱中各个主题间的从属临近关系等多层异构网络构成,并随着相关主体的更新而演化。在过去的十几年间,科技论文、发明专利、项目申请书等与科研活动密切相关的资料呈现指数性的增长。这些资料以结构化、电子化的形式存储,提供了丰富的可供利用的“学术大数据”,为定量的、系统的、以科学系统为研究主体的研究工作开展提供了数据基础。“科学学(Science of Science)”这一传统问题<sup>[10]</sup>开始成为新兴的交叉研究热点。

利用大规模学术数据,结合信息科学领域的工具,是科学学研究中广泛采用的范式。区别于以往工作,当前的研究具有如下特点。

- 使用大规模高维度的数据,例如论文元数据(meta-data)往往包含数十万到上千万量级的论文数量,包含题目、摘要、作者、单位、相互引用情况等多维度结构信息。

- 具有交叉学科的鲜明特点,研究方法丰富多样,采用科学计量学中的指标衡

量科研的绩效产出,利用信息科学领域的可视化、网络科学、机器学习、信息挖掘、多智能体建模等工具,借鉴社会科学中的理论。

- 不再拘泥于算法性能的提升或评价指标的提出,而是以科学系统为研究对象,探索其中的普适性规律和基础演化机制。

近几年来,在*Nature*、*Science*、*Nature*和*Science*的相关子刊、*Proceedings of the National Academy of Sciences of the United States of America*上发表的一系列高水平工作<sup>[3,11-16]</sup>充分展现了科学学研究的创新性和前沿性,也充分说明了当前的研究范式在探索复杂问题时的可行性。

相关研究在国内的开展具有更特殊的现实意义。中国的科学研究在过去的十多年间实现了飞速的发展<sup>[17]</sup>。中国每年有世界第二大的论文产出量和论文引用量,如图1(a)、图1(b)所示。然而体量的增长背后也隐藏着其他困境<sup>[18]</sup>。例如,中国科研论文在国际上仍然呈现“代表性不足”的特点,中国论文被国际同行引用的情况与论文的发表数量并不相称。通过分析Web of Science平台从1990年至2016年超过2 000万篇论文及其超过1亿次的引用关系,笔者计算了一个国家的所有论文在某一年的总引用数量。笔者发现,在中国论文每年的总引用数量中,国际引用(非本国引用,按论文的第一单位划分)占比自2010年来逐年下降。而世界上其他科研大国,如美国、德国、英国、法国、日本、韩国等,国际引用占比却逐年增加,如图1(c)所示(为了更好地表现变化趋势,数值按照各国2010年的数值进行了归一化)。中国论文在2016年的所有引用中,仅有42%是来自国际的,在全球15个科研大国中排名最后,如图1(d)所示。这一现象说明了我国科研管理

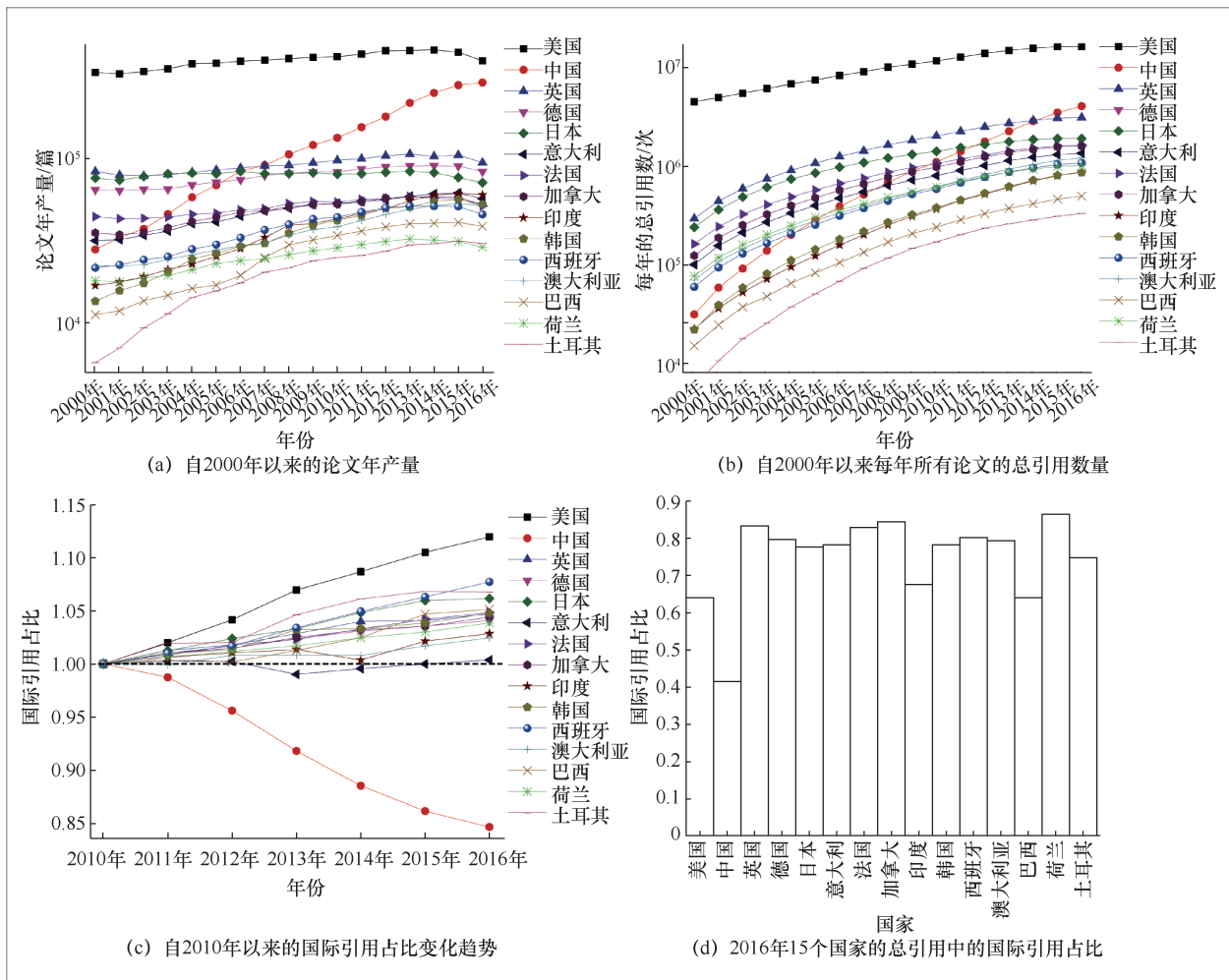


图1 全球SCI论文量较大的15个国家的论文产出和引用情况

面临的新挑战：如何从量的增产转换为质的突破。通过对科学本体的研究，发现其中的客观规律，理解存在不足的原因，是应对这一挑战的前提条件和理论准备。

综上所述，科学的复杂性、学术大数据带来的数据可用性、对科学研究中客观规律认知的现实需求，这3个要素共同推动“科学学”成为一个热点研究领域。科学研究的主导者是科研工作者，研究科学中的规律必然要先了解科研工作者的行为特征。基于人名消歧技术的进步，可以将科研工作者的研究成果进行较为精确的归集，

从而以科研工作者个体和群体为研究对象，挖掘他们在科研过程中的行为规律。在后文中，笔者将梳理、介绍几个相互关联、递进的研究问题：通过对科研工作者“研究兴趣”的表征，研究他们在职业生涯中研究方向改变的规律；利用研究兴趣这一特征，研究“科研合作”的内在机制；不同的科研工作者通过合作形成“科研team”，而其中一部分长期稳定的合作关系形成了“科研团队”，通过相对合作强度从科研team中挖掘科研团队，从而识别二者在科研产出、影响力方面的不同；通过对科研团队的识别，归类论文引用来源，实

现“研究成果多维量化”，同时帮助识别一些互引、自引、过度合作等可疑的“学术不端行为”。笔者从科学学的研究视角出发，聚焦于问题的提出与相应的科学发现。

## 2 相关问题

### 2.1 研究兴趣

一个科研工作者会在其职业生涯中研究不同的科学问题，转换不同的研究方向。托马斯·库恩的著作《必要的张力》(The essential tension)是非常有影响力的研究成果之一。他开创性地描述了科学工作者在职业生涯中对探索(explore)与发掘(exploit)相互矛盾的需求<sup>[19]</sup>。一方面是职业发展的需求，不管是职位的晋升，还是基金项目的申请，都要求有稳定的科研产出，需要科研工作者在熟悉的领域中做出细小但是不间断的科学贡献；另一方面，虽然专注的研究可以带来稳定的产出，但却有可能影响成果的创新性，降低成果的影响力<sup>[20-21]</sup>。因此，一个成功的科研工作者需要巧妙地平衡探索与发掘这两种不同科研模式中的风险和收益，在职业生涯中谨慎但合理地改变研究问题和研究方向<sup>[22]</sup>。

研究兴趣的表征可以通过机器学习或者主题提取的方法进行，利用科研工作者在一定时间内的研究成果，建立科研工作者的“学术画像”。研究兴趣在工程类问题上有着直接的应用，例如用于推荐系统的优化，帮助更好地选取项目或论文的评审人，更准确及时地给研究人员提供当前领域内相关的成果<sup>[23-24]</sup>。

从理论研究的角度来看，对研究兴趣的分析可以帮助人们回答托马斯·库恩关注的问题：科研工作者如何选取下一个研究

问题？研究问题的选择直接反映在了论文的内容上，因此论文的研究内容变化可以被用来识别“探索”与“发掘”的行为。利用Web of Science平台上生物医学和化学领域的论文数据，Rzhetsky和他的合作者们<sup>[22,25]</sup>提取了论文研究的化合物。将一个化合物作为一个知识节点，利用不同化合物在论文中共同出现的关系，他们构建了化合物之间相互关联的知识网络。每篇论文对不同知识节点(不同化合物)的处理(如跳跃至新的知识点、链接不同的知识点、融合不同的知识点等)分别对应对传统领域的挖掘和对创新领域的探索。他们进一步结合知识网络的节点中心性，发现领域中主流的研究策略变得越发保守，研究内容往往包含被重点研究过的化合物(高中心性节点)。这种渐进的研究虽然对个人的职业发展是有利的(低风险)，但是降低了突破性科学发现的概率，这对学科的整体发展是不利的<sup>[26]</sup>。

区别于以论文为出发点的思路，笔者着重研究了科研工作者个体研究兴趣的变化<sup>[14]</sup>。利用人名消歧技术，笔者从美国物理学会(American Physical Society, APS)旗下期刊的论文数据中分离出14 715位作者和他们发表的所有论文。笔者将一个科研工作者的论文按照发表的时间排序，在序列的首尾各选取一组论文集合，每个集合各包含8篇论文。利用每篇论文的内容分类代码(PACS代码)构建每篇论文主题的多维表征，从而加权叠加为描述科研工作者研究生涯早期和晚期的研究兴趣向量 $\mathbf{g}_i$ 、 $\mathbf{g}_j$ ，最后利用余弦相似度计算两组论文对应研究方向的不同，量化出一个科研工作者职业生涯中兴趣转变的程度(如图2所示，示例中每组论文集合包含2篇论文)。笔者发现，14 715位作者的研究兴趣改变遵循一个指数分布，这意味着大部分人的研究兴趣变化很少，变化的概率随着

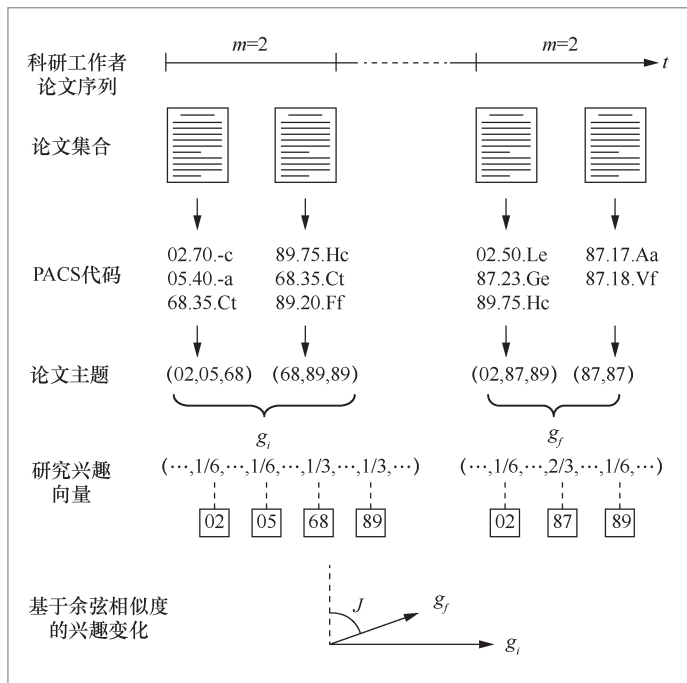


图2 对科研工作者研究兴趣演化的示意图

变化的程度呈指数下降,同时也存在一部分科研工作者,他们早期和晚期的研究内容完全没有重叠。笔者同时发现,尽管有很多因素影响科研工作者下一个研究问题的选择,但是研究兴趣演化的过程仍然可以用一个简单的随机游走模型进行描述。基于模型的模拟结果非常好地重现了实证数据中的观测结果。通过发现科研生涯中研究兴趣变化的宏观规律,可以进一步论证科研工作者职业发展的最佳策略,并进一步探索其兴趣演化中的微观机制。

## 2.2 科研合作

科研合作是现代科学研究中越来越普遍的行为。Wuchty S等人<sup>[7]</sup>考察了1975年至2005年发表的超过340万篇科学与工程领域的论文,发现在过去的几十年间单一作者的文章比例在稳步下降,而通过合作发表的科研论文比例却呈现稳定的上升趋势。

专业化被认为是科研合作的主要驱动因素,它带来两种不同的合作模式<sup>[27-28]</sup>:一方面,科学问题越来越复杂,产出率的要求使得科研工作者不得不将单一问题切割为不同模块,分配给相同领域的合作者,以加速研究进度;另一方面,不断增长的领域专业知识、不断加强的学科界限以及科学研究对创造性、前沿性问题的追求,使得一个科研工作者倾向于和具有不同专业领域知识的科研人员一同进行开创性的工作。

研究兴趣作为科研工作者基本的学术画像之一,可以帮助理解科研合作的内在机制。同样使用美国物理学会的论文数据,笔者把两个作者首次合作的论文作为原点,通过在时间轴上倒推,收集他们首次合作前的论文集合。通过比较两组论文对应的研究主题,测量两个科研工作者首次科研合作时研究领域的相似度。笔者在实证中发现,科研合作中趋同性(homophily)高于异质性(heterophily)。学者的合作模式可能随着学术年龄的变化而变化。笔者从微观角度入手,分析APS及DBLP(database systems and logic programming)的数据,对不同学术年龄的学者进行合作模式的分析,发现在同一时间段内,学术年龄分布符合长尾分布。此外,学术界的合作模式普遍存在“人以龄分”的趋同性现象<sup>[29]</sup>。

学术会议是交流科研成果的重要学术活动。共同参会的学者往往具有相似的研究兴趣,这种相似性会推动合作进程,延长合作的持续时间。考虑到学术会议的社会功能,笔者提出学术会议闭包的概念,并基于此提出SCORE(sustainable collaborator recommendation)推荐方法,以合作关系和合作强度对学术网络进行重构,最后以重启型随机游走(random walk with restart)方法进行合作者推荐<sup>[30]</sup>。好奇心和观察力是新奇性

(serendipity) 科研发现的驱动因素和重要的前提条件, 而学者的合作行为也存在一定的意外性、相关性及价值性。笔者基于这3个指标定义和量化学者的新奇合作者, 提出新奇合作者识别模型, 并设计新奇合作者推荐算法。以新奇合作者的量化分数重构合作网络, 改进网络表示学习算法的节点序列生成过程, 增加随机游走访问新奇合作者节点的概率, 最终将得到的节点序列输入skip-gram模型。通过计算目标学者向量与其他学者向量的余弦相似度, 向学者推荐合适的合作者, 从而更有效地产生高影响力、高创新性的科研成果。

### 2.3 科研team与科研团队

科研合作促成不同的科研工作者共同完成一项科研工作, 形成一个team<sup>[7,16,31]</sup>。当前, 国际上“team science”研究对team有明确的定义, 即完成一个具体工作的人员集合, 例如完成一篇论文的共同作者team、申请一个项目的项目基金team等。与科研团队相关的研究在国内始于21世纪初<sup>[32]</sup>, 更多借鉴了组织学中的概念, 描述具有一定稳定结构和管理层级的人员集合。显而易见, 多个科研工作者在研究一个具体问题、发表一篇科研论文的过程中必然形成科研team, 而科研team却不一定对应一个科研团队。科研团队是科研team的子集, 两者并不等价。对科研team已有的研究发现, team具有流动性, 没有稳定的边界和固定的成员, 其形成具有自发性, 管理具有自治性<sup>[33]</sup>。这些特点也从一个侧面印证了科研团队与科研team的不同。

科研team有明确的研究对象, 这使得量化的、实证的研究得以广泛开展, 大量基于数据的结论被提出和证实。与之相反, 科研团队的定义非常模糊, 这使得相关研究要么缺少实证数据支撑, 要么简单

地把团队和team等同, 将team的结论照搬到科研团队。可以说, 科研团队是当前国际、国内研究中鲜有涉及的部分。虽然科研团队的概念极大地影响着我国的科学研究(年轻的科研工作者被要求加入一个科研团队, 有潜力的科研工作者被要求建立和领导一个科研团队, 进而各个高校研究机构需要评估和管理科研团队), 但是对科研团队的理解还停留在简单直观的水平, 缺少量化的、基于实证数据的研究。

当前对科研team的研究已经从一个侧面暗示出科研团队在某些层面上的不足。例如, 不同的研究均发现, 稳定的合作成员会降低合作者之间的异质性, 进而降低team研究工作的创新性<sup>[33-34]</sup>, 这暗示着科研团队在对创新性要求较高的研究领域可能存在不足。Wu L等人<sup>[16]</sup>发现, 大team能对已有的研究进行挖掘, 但是小team更容易带来颠覆性的创新成果。这意味着科研团队可能在不同的任务中存在优势和劣势, 科技政策中需要平衡两者间的关系。这个现象对中国科研界有着更重要的启示。我国小team的产出在整个科技论文的总量中占比偏少, 远远低于美国和国际均值。例如在2017年, 我国只有37%的论文是由小team(论文作者数小于5位)完成的, 而美国为58%, 国际均值为53%。虽然大team的论文往往更有可能获得更多的引用, 但是Dong Y等人<sup>[35]</sup>发现, team中成员单位的异质性才是决定高水平研究成果产生的因素, 来自同一单位的大team并不具有产生高水平成果的更高概率, 这意味着大型科研团队在高水平成果的创造中可能并无优势。

从技术上来讲, 科研team不需要识别, 任何一个项目、一篇论文, 只要参与的人员多于1人, 就形成了一个team。而识别科研团队则困难得多。首先需要有准确的人名消歧数据, 才能追踪科研人员的长期合作关系, 构造科研合作网络。但即便给

定合作网络,当前也缺少合适的工具从中识别出团队信息。从网络科学的视角来看,作者与论文之间形成的是一个二分图,由两组节点(作者和论文)构成,组间连接,而组内没有连接。作者合作关系网络通过将二分图投影获得,包含许多致密的团(clique)结构,每个团内的节点相互完全连接(如图3(a)所示)。最终形成的科研合作网络连接非常紧密,社团结构非常弱化,现有的工具难以准确识别其中的群团信息。与此同时,团队识别不能简单依靠合作网络的拓扑结构,合作强度带来的连边权重也极大地影响着团队的识别。如图3(b)所示,如果不考虑连边权重,仅仅依靠拓扑结构对合作网络进行划分,极有可能将合作关系非常紧密的节点3和节点4划分在两个不同的团队中。为解决这一问题,笔者提出了一套基于相对合作强度的指标,对现有的合作连边进行切割,从而获得相对稀疏的合作网络,帮助获取团队信息<sup>[36]</sup>。进一步与团队的科研绩效数据结合,帮助从实证上分析团队在科研上的优势与劣势。同时也进一步拓宽当前国际上team science的研究,获得团队与team在科研绩效成果上的异同。

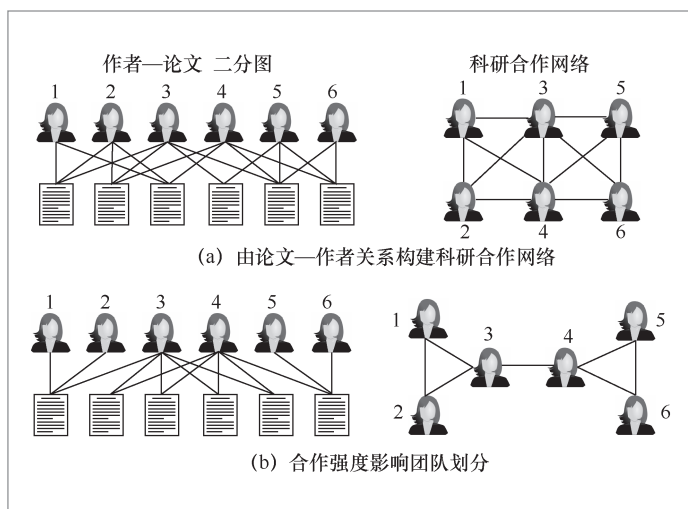


图3 从论文构建科研合作网络的示例

## 2.4 成果多维量化与学术不端行为识别

反应行为(reactivity)在心理学上指的是个体因为知道被观测而改变行为的现象。当前各种科学计量指标被广泛地应用于评价一个科研工作者的科研绩效,并且在很大程度上与项目经费、职位晋升、个人收入等相关。古德哈特定律(Goodhart's law)精辟地指出,一项指标一旦成为政策制定的依据,便会逐渐失效。科研领域也不例外,科研工作者并不是科研评价体系下的被动接受者,他们可以主动地改变自己的行为,以迎合相应的评价指标。科研领域的反应行为对科学研究有着深刻的影响,在个体层面,导致目标置换(goal displacement),在评价系统中获得更高的分数成为科研的目标;在群体层面,影响科研的基础氛围,例如带来跟风性质的研究,或避开困难的研究问题而青睐于短平快的方向,进一步损害整个科学共同体的利益。

我国科研管理结构已经意识到这一问题的重要性,当前大力推进的破“五唯(唯论文、唯帽子、唯职称、唯学历、唯奖项)”正是对这一现象的纠正。然而需要注意的是,量化指标仍然是衡量科研工作者科研绩效的重要工具。“五唯”的问题不在于指标,而在于“唯”,只有提供科研工作者成果绩效的多维度量,才能有效地实现破“五唯”。基于学术大数据的多维特征和数据挖掘方法,可以实现这一目标。例如,通过对论文作者进行人名消歧可以更好地移除自引,获得它引数量;通过科研合作网络分析,利用复杂网络的各类节点中心性指标,识别网络中的重要节点,描述学者在相关领域中的影响力;通过分析论文的引用信息,结合学术团队的识别,获得一个科研工作者影响力的来源,尤其可以采用Top  $k$ 方法描述引用异质性特征和影响力范围。

如两个科研工作者的它引数量均为1 000次,但是其中一人30%的引用来自自己的科研团队,20%的引用来自科研团队1,15%的引用来自科研团队2;另一人本团队引用仅占15%,引用来源占比最高的其他团队仅为8%和6%。这说明虽然两人引用数量相同,但是后者的引用分布更加均匀,引用来自于更多的科研团队,影响力范围更加广泛。这些多维度的信息能更丰富地描述科研工作者的成果和研究能力,避免对单一指标的过度依赖。

反应行为的过度延伸可形成学术不端。创新管理领域期刊*Research Policy*在2019年推出一期主题为“Academic Misconduct, Misrepresentation, and Gaming”的专刊<sup>[37]</sup>,对各类学术不端行为进行了分析总结,这从一个侧面说明了这一问题的重要学术价值。我国当前在大力呼吁加强科研诚信、严惩学术不端,然而基于大规模数据的科研工作者反应行为探测、挖掘的工作仍然不多。虽然科研工作者学术不端的参考标准数据难以获得,但是从数据中挖掘可疑的、不寻常的行为却非常可行。例如一些工作中总结了常见的一些学术不端行为,如过度合作(或名誉作者)现象、通过拆分工作获得更多的论文数、通过大量自引获得更多的引用数、互引圈子(citation cliques)间的相互引用。这些现象均可以利用机器学习的方法,通过设定的阈值在学术大数据中进行识别,再通过人工筛查,可对一些早期的学术不端行为提出预警。

### 3 结束语

基于大数据的在线用户行为分析已经成为一个热点研究问题,梳理其发展的脉络,不难发现其中的一些特点。首先用户行

为具有复杂性,需要理论和实证的研究;其次是大量用户各类行为被记录和保存,使得数据驱动的研究工作得以开展;最后是用户行为在推荐、信息传播、安全等各领域均有实际的应用。对照基于学术大数据的科研工作者行为研究,不难发现类似的因素:科学的复杂性、学术大数据的可用性、对科学规律认识的需求分别对应了在线用户行为分析领域的驱动力量。当前国际上科学研究的兴起也充分说明了这是一个大有可为的领域。相关的研究问题很多,本文只是列举了一部分,其余部分可由读者进一步发掘探索。

### 参考文献:

- [1] KUHN T S. The structure of scientific revolutions[M]. Chicago: University of Chicago Press, 2012.
- [2] DE SOLLA P D J. Little science, big science and beyond[M]. New York: Columbia University Press, 1986.
- [3] SINATRA R, WANG D, DEVILLE P, et al. Quantifying the evolution of individual scientific impact[J]. *Science*, 2016, 354(6312): 5239.
- [4] DA SILVA J A T. Does China need to rethink its metrics—and citation-based research rewards policies [J]. *Scientometrics*, 2017, 112(3): 1853-1857.
- [5] AZOULAY P, GRAFF-ZIVIN J, UZZI B, et al. Toward a more scientific science[J]. *Science*, 2018, 361(6408): 1194-1197.
- [6] JONES B F, WUCHTY S, UZZI B. Multi-university research teams: shifting impact, geography, and stratification in science[J]. *Science*, 2008, 322(5905): 1259-1262.
- [7] WUCHTY S, JONES B F, UZZI B. The increasing dominance of teams in production of knowledge[J]. *Science*, 2007, 316(5827): 1036-1039.

- [8] ZENG A, SHEN Z S, ZHOU J L, et al. The Science of Science: from the perspective of complex systems[J]. *Physics Reports–Review Section of Physics Letters*, 2017(714–715): 1–73.
- [9] FORTUNATO S, BERGSTROM C T, BÖRNER K, et al. Science of science[J]. *Science*, 2018, 359(6379).
- [10] 钱学森. 希望[J]. *科学学与科学技术管理*, 1981(1): 7.  
QIAN X S. Hope[J]. *Science of Science and Management of S.&T.*, 1981(1): 7.
- [11] WANG D, SONG C, BARABÁSI A L. Quantifying long-term scientific impact[J]. *Science*, 2013, 342(6154): 127–132.
- [12] HSIEHCHEN D, ESPINOZA M, HSIEH A. Multinational teams and diseconomies of scale in collaborative research[J]. *Science Advances*, 2015, 1(8): e1500211.
- [13] RZHETSKY A, FOSTER J G, FOSTER I T, et al. Choosing experiments to accelerate collective discovery[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, 112(47): 14569–14574.
- [14] JIA T, WANG D, SZYMANSKI B K. Quantifying patterns of research–interest evolution[J]. *Nature Human Behaviour*, 2017, 1(4): 78.
- [15] WAY S F, MORGAN A C, CLAUSET A, et al. The misleading narrative of the canonical faculty productivity trajectory[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2017, 114(44): 9216–9223.
- [16] WU L, WANG D, EVANS J A. Large teams develop and small teams disrupt science and technology[J]. *Nature*, 2019, 566(7744): 378–382.
- [17] XIE Y, ZHANG C, LAI Q. China’s rise as a major contributor to science and technology[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2014, 111(26): 9437–9442.
- [18] HUANG F. Quality deficit belies the hype[J]. *Nature*, 2018, 564(7735): 70.
- [19] KUHN T S. The essential tension[J]. *American Journal of Physics*, 1979, 47(6): 568–569.
- [20] AZOULAY P, GRAFF ZIVIN J S, MANSO G. Incentives and creativity: evidence from the academic life sciences[J]. *The RAND Journal of Economics*, 2011, 42(3): 527–554.
- [21] TIAN X, WANG T Y. Tolerance for failure and corporate innovation[J]. *The Review of Financial Studies*, 2014, 27(1): 211–255.
- [22] FOSTER J G, RZHETSKY A, EVANS J A. Tradition and innovation in scientists’ research strategies[J]. *American Sociological Review*, 2015, 80(5): 875–908.
- [23] TANG J, ZHANG J, YAO L, et al. Arnetminer: extraction and mining of academic social networks[C]//The 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 24–27, 2008, Las Vegas, USA. New York: ACM Press, 2018: 990–998.
- [24] XIA F, WANG W, BEKELE T M, et al. Big scholarly data: a survey[J]. *IEEE Transactions on Big Data*, 2017, 3(1): 18–35.
- [25] RZHETSKY A, FOSTER J G, FOSTER I T, et al. Choosing experiments to accelerate collective discovery[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2015, 112(47): 14569–14574.
- [26] COKOL M, IOSSIFOV I, WEINREB C, et al. Emergent behavior of growing knowledge about molecular interactions[J]. *Nature Biotechnology*, 2005, 23(10): 1243–1247.
- [27] KABO F W, COTTON–NESSLER N, HWANG Y, et al. Proximity effects on the dynamics and outcomes of scientific collaborations[J]. *Research Policy*, 2014, 43(9): 1469–1485.
- [28] LEAHEY E. From sole investigator to team scientist: trends in the practice and study of research collaboration[J]. *Annual Review of Sociology*, 2016, 42(1): 81–100.

- [29] WANG W, YU S, BEKELE T M, et al. Scientific collaboration patterns vary with scholars' academic ages[J]. *Scientometrics*, 2017, 112(1): 329-343.
- [30] WANG W, LIU J, YANG Z, et al. Sustainable collaborator recommendation based on conference closure[J]. *IEEE Transactions on Computational Social Systems*, 2019, 6(2): 311-322.
- [31] National Research Council. Enhancing the effectiveness of team science[M]. Pittsburgh: National Academies Press, 2015.
- [32] 方文东. 关于科研团队组建的一些认识[J]. *科技管理研究*, 2002, 22(4): 42-43.  
FANG W D. The attitude about organizing scientific research group[J]. *Science and Technology Management Research*, 2002, 22(4): 42-43.
- [33] WANG J, HICKS D. Scientific teams: self-assembly, fluidness, and interdependence[J]. *Journal of Informetrics*, 2015, 9(1): 197-207.
- [34] LEE Y N, WALSH J P, WANG J. Creativity in scientific teams: unpacking novelty and impact[J]. *Research Policy*, 2015, 44(3): 684-697.
- [35] DONG Y, MA H, TANG J, et al. Collaboration diversity and scientific impact[J]. *Computer Science*, 2018, arXiv:1806.03694.
- [36] YU S, XIA F, ZHANG K, et al. Team recognition in big scholarly data: exploring collaboration intensity[C]// 2017 IEEE 15th International Conference on Dependable, Autonomic and Secure Computing, September 17-21, 2017, St. Petersburg, USA. Piscataway: IEEE Press, 2017: 925-932.
- [37] BIAGIOLI M, KENNEY M, MARTIN B, et al. Academic misconduct, misrepresentation and gaming: a reassessment[J]. *Research Policy*, 2018.

## 作者简介



贾韬(1982- ),男,博士,西南大学计算机与信息科学学院教授、副院长,主要研究方向为大数据、复杂网络等,以第一作者(兼共同通信作者)在*Nature Human Behaviour*、*Nature Communications*、*Physical Review Letters*等学术期刊发表多篇论文。获2010年年度国家优秀自费留学生奖学金。



夏锋(1980- ),男,博士,大连理工大学教授、博士生导师,软件学院副院长、学术委员会副主任,阿尔法实验室(The Alpha Lab)创始人兼主任,ACM高级会员、IEEE高级会员,主要研究方向为数据科学、大数据、知识管理、系统工程等。出版英文学术专著2部(Springer出版),在国际期刊和国际会议上发表论文290余篇,其中IEEE和ACM期刊论文70余篇。连续5年(2014—2018年)入选“中国高被引学者”榜单。

收稿日期: 2019-06-17

基金项目: 国家自然科学基金资助项目(No.61603309)

Foundation Item: The National Natural Science Foundation of China (No.61603309)