

把握复杂科学结构 揭示创新动态趋势 《科学结构地图》系列专著介绍

文/王小梅* 韩涛 王俊
中国科学院国家科学图书馆 北京 100190

【关键词】 科学结构, 地图, 工具



中国科学院

继《科学结构地图 2009》后,《科学结构地图 2012》(图 1)近期即将出版。该书由潘教峰、张晓林等著,科学出版社出版。该书不仅可视化地展现了科学研究的宏观结构和复杂关系,还为观察、分析和探索科学发展趋势提供了新视角和新工具。

1 《科学结构地图》成为探索科学趋势的新工具

科技界一直非常重视对科学结构和科学发展趋势的分析。但是,在传统研究环境下,科学家们主要通过检索和分析相关文献来了解学科发展,通过追踪同行的研究活动来掌握学科趋势和研究热点,通过专家调研、研讨以及专门的规划研究来进一步判断可能的突破方向。例如,日本在历次技术预见中采用德尔菲法进行的大规模专家调研,中科院、国家基金委等组织系列专家组进行

学科态势分析。这些方法仍然是认知科学结构及其发展趋势的重要手段,但它们也日益受到新的科技创新需求的强烈挑战。

当前,科技创新已进入多学科交叉融汇的研究阶段,例如由纳米技术、生物技术、信息技术和认知技术交叉融汇所产生的NCBI(nano-bio-info-cogno)研究领域,以及麻省理工学院院长 Susan Hockfield 近年提出的物理、工程和生命学科融汇将产生的第三次

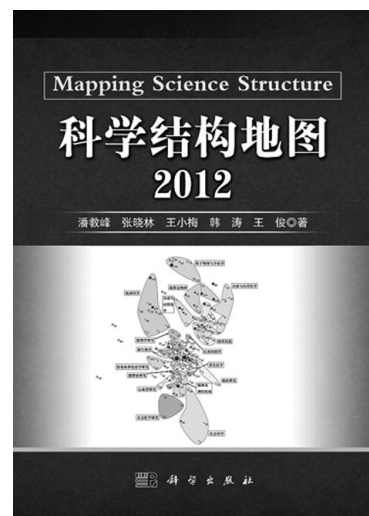


图1 《科学结构地图 2012》封面

* 中科院国家科学图书馆副研究员。E-mail: wangxm@mail.las.ac.cn

修改稿收到日期:2012年11月5日

生物科学革命。同时,贯穿从实验室到应用环境(From bench to bed)整个创新价值链的转换型研究也将基础科学问题、核心技术、应用技术、市场化个性化应用方法乃至应用中的政策、规则、伦理、管理等问题相互关联起来。而且,当前我们面临的许多重大问题,如全球气候变化、资源可持续发展、人类健康、环境保护与修复、人口老龄化等,都涉及到包括科学、技术、社会、经济乃至政治等多方面的复杂的相互动态影响的多重因素。在这样的情况下,重大创新突破的启动点往往来自所熟悉领域外的、复杂交叉关联的“其他”因素(那个没有注意到甚至没有意识到的“亚马逊蝴蝶”)。但是,限于传统的专业训练以及相应的认知坐标系,限于检索文献或跟踪业内同行的习惯方法,我们往往很难“观察到”不熟悉但相关的领域,更难把握它们之间的复杂结构和相互影响,往往难以发现隐藏在复杂关系下的正在积蓄颠覆性压力的结构化矛盾、可能的致变因素和潜在的发展趋势。

随着科技信息和科学研究的数字化,科技创新中的各类知识对象,例如文献、专利、机构、人员、项目、设施、产品,几乎全部信息化、数字化,形成反映科学活动的海量数据。人们逐步发展出利用这些数据来揭示大问题领域、大时间跨度、多维度关系上的科学结构,并将科学结构作为对科学布局、相互作用及演变趋势进行描述和分析的工具。起初,科学结构工具依赖对科研机构或科研产出(论文、专利等)的分类统计来描述,可以按领域类别、地域、项目投入或影响力的布局来展现;后来,人们更多利用一些内在指标来揭示内容关系,例如文献共词、作者共著、论文共引或共被引等,并发展出许多分析关联强度、集聚效应、关系性质的方法。

在这些努力的推动下,“科学结构”本身的作用也发生着重大变化。它不仅能够帮助我们把握大问题尺度上和复杂融汇环境下的知识结构和各种隐性关系,而且逐步帮助我们揭示演变趋势、鉴

别发展规律、挖掘结构矛盾、发现影响因子和潜在合作对象等。在动态的和可视化的试验与演绎基础上,可支持科学发展的预警和优先领域的遴选。这样,科学结构地图本身就成为研究科学发展的工具,成为探索甚至试验科学发展趋势工具。

2 《科学结构地图》所揭示的科学结构与发展趋势

《科学结构地图 2009》和《科学结构地图 2012》主要利用科学论文同被引关系分析和描述科学结构地图,并运用复杂网络分析法、社会网络分析法辅助确定重点研究领域和合作竞争趋势。

论文引证是科学家信息交流的直接表现,客观体现了知识的相互作用及其流动、融汇和演变。同被引是指一组论文共同被其他论文引用,反映了这组在学科分类、发表期刊、作者机构、项目甚至主题词等方面看似毫无关联的论文可能存在着某种关系,而且这种关系随着这组论文同时被其他论文引用的次数逐渐增加而不断被证实和加强。因此,同被引现象反映了科研内容和科研活动的聚合过程和聚合关系,反映了科学知识结构的演变与自组织,可以超越传统的分类方法来了解科学结构及其变化。日本科技政策研究所(NISTEP)从2002年起,利用同被引分析每两年制作一次科学地图,应用到日本各次技术预见活动中;SciTech Strategies公司也开始提供商业化的科学地图分析服务。我们以《科学结构地图 2012》为例,说明科学结构地图的主要应用。

通过表征科学知识流动的论文引证及其聚合关系揭示科学知识结构。该工作以汤森路透科技集团的基本科学指标库(ESI)为信息源,提取了从2004年1月至2010年4月的按学科分类的8 529个研究前沿,对这些研究前沿内的高被引论文进行同被引分析,生成了由同被引关系内在关联起来的各个高被引论文簇所代表的132个研究领域,并根据它们之间的关联关系绘制了整个科学结构地图(图2)。同时,通过对论文簇关键词的分

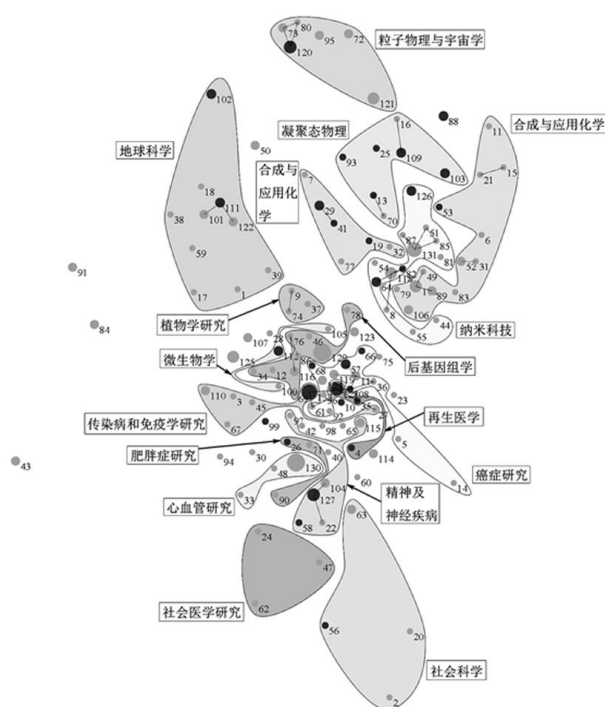


图2 科学结构地图

析以及领域专家的判读,确定了这些 隐性研究领域的名称及内涵。课题组还分析了《科学结构地图 2009》和《科学结构地图 2012》的动态变化,利用课题组开发的多时间窗聚类结果演化分析方法,判断两个不同结构中不同聚类之间的重叠演变情况,鉴别相关聚类(领域)的产生、消失、延续、分化、合并,体现了知识进化理论中的知识 遗传继承 现象,揭示科学知识演化趋势。

通过结构关系及其变化揭示热点或潜在热点领域。热点或潜在热点领域一直是科技界和科技决策者关心的重要问题,《科学结构地图》系列专著利用多种方法来遴选热点或潜在热点。最直接的方法是找出那些论文被引频次及其年均增长率高于整体平均值的 热点 研究前沿和包含 热点 研究前沿数量超过整体平均分布的 热点 研究领域。但是,仅仅依靠当前活动频次不足以发现 潜在 的热点领域,课题组又引入了

枢纽研究领域 来帮助辨别某个领域对其他领域的作用力度。枢纽研究领域是指那些在其他研究领域间(尤其是在多个跨学科的研究领域间)起着知识关联桥梁作用的领域,一方面通过研究领域的中介中心性(可简要表示为通过该研究领域的最短关联路径数)来表示,另一方面通过计算所连接的其他领域的学科关联数(可简要表示为被同被引连接的其他领域分属的学科数量)来计算,那些中介中间性和学科关联数都居于前20%的研究领域就是枢纽研究领域。

通过学科多样性和网络凝聚性揭示知识整合的广度和强度。

研究领域的学科交叉性体现了认知的异质性、研究的多样化,可以通过分析研究领域核心论文所属不同学科数量、分布均衡度和学科相似性3个方面来描述研究领域的学科异质性。研究领域的网络凝聚性描述研究领域的论文关联网络的主题相似度和结构紧密度,反映该领域的知识整合程度,可通过同被引关系网络的平均链接强度来表示。我们利用学科多样性和网络凝聚性(图3)两个维度来反映学科交叉的程度以及潜在的知识整合能力,并通过形象的方式来展示各个研究领域的所属学科、是否为交叉领域以及各学科、各研究领域知识整合的广度和强度。

通过国家科学研究的结构和相互关系揭示知识合作与竞争趋势。各国在各研究领域中有不同的投入、成果和影响,可以反映国家在不同领域的相对优势。该系列专著通过不同国家在研究大类和研究领域中的核心论文份额与施引论文份额来表现这



中国科学院

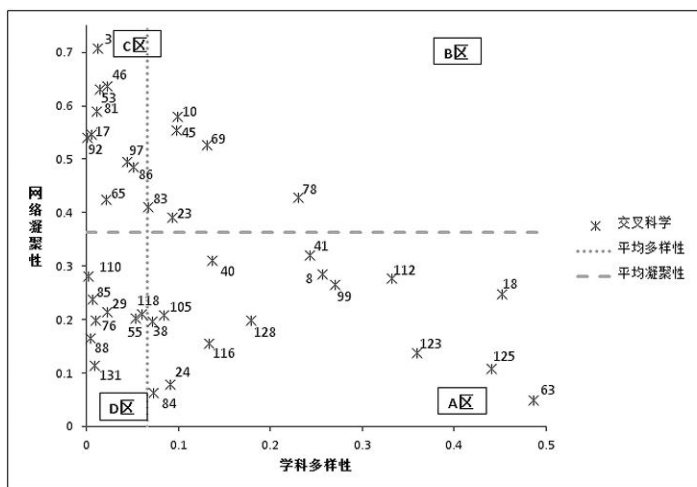


图3 交叉领域的学科多样性和网络凝聚性分布

些国家的研究活跃程度,列出了世界前30个国家在整个科学研究以及粒子物理与宇宙学、凝聚态物理、纳米科技、合成与应用化学、地球科学、生物学、医学等研究大类中的活跃程度,还根据在各个研究领域是否有高被引科研论文、这些高被引科研论文比例是否高于世界均值,描述了中国(图4)、美国、英国、德国、法国、日本、澳大利亚、韩国、俄罗斯、印度、巴西、南非在上述研究领域的相对活跃程度及其变化趋势。《科学结构地图2012》描

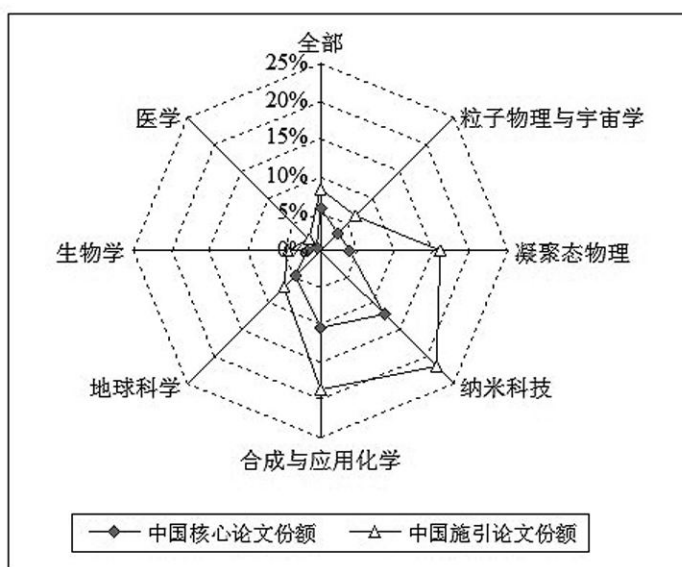


图4 中国核心论文和施引论文

述了代表性国家在全部核心论文中的国际合著率,反映国际合作的总体趋势;进一步地,描述了代表性国家在科学结构地图中各个研究领域的国际合著率及其变化。将国家在各个领域的核心论文数和国际合著率结合起来,将其与国际平均值比较,可以发现哪些核心论文数高于平均值但国际合著率低于平均值的国家(自主创新强但国际依赖性低)和哪些核心论文数低于平均值但国际合著率高于平均值的国家(自主创新能力可能较弱,国际依赖性较强),还新增了机构合著网络分析,从更微观的

角度展现科学研究合作的特征。

3 发挥科学结构工具作用,预警科技创新的潜在可能

目前科学结构地图相关工作仅仅是对科学结构分析能力的初步应用。如前文所述,《科学结构地图》可以作为深入研究科学发展趋势的工具,帮助我们揭示规律、发现矛盾、预警演变、发掘重要

影响因子、发掘潜在合作对象等。除进一步解决大数据量、稀疏关系和大跨度间接连接等影响全科学空间分析精度和效率的问题,支持全谱段多层次贯通分析外,我们还将科学结构地图工具化,从静态、独立的分析发展到动态、关联的分析,从描述性分析深入到解释性、揭示性、评价性应用。

所谓科学结构的动态性关联化分析,包括两个方面的动态展开。一方面,将多年度科学结构地图关联起来进行分析,着重考察科学结构的演变趋势,考察在这些演变当中的趋势以及突变、异变、阻滞等现象,分析这

些现象背后的影响因素,揭示科学发展中的规律。另一方面,突破目前科学结构地图利用引文数据所导致的相对滞后影响,将依据过去引文数据的科学结构地图与当前论文产出分析结合起来,考察科学结构的现期变化;可以将科学结构地图与当前科研项目分析结合起来,考察科学研究产出结构与科学研究投入结构的关系;还可将科学结构地图与基于专利的技术结构分析结合起来,考察科研成果转移转化的趋势和问题;这样,更深入和更全面地反映科技结构变化,探查各方面因素的影响。

所谓科学结构的解释性、揭示性分析,是希望《科学结构地图》不仅能帮助我们看到是什么,还帮助我们回答为什么。例如,当我们发现科学领域发生、分化或重合时,我们希望鉴别引起、推动这些变化的深层因素,而这些推动因素可能是某个新概念、新方法、新工具、新的应用场景、新的参加机构或人员等。《科学结构地图》所依据的文献

内容中已经包含了这些信息,如何挖掘和描述这些信息,如何分析判断这些因素对于科学结构中各个领域及其关系的作用性质,对我们发现甚至预警推动因素及其导致或可能导致的变化,都非常重要。在科学结构的评价性分析上,希望能在目前的按照产出情况及其影响力进行基本水平评价的基础上,尝试根据各个研究领域的产出与影响差异性,根据在科学结构中引领新领域、引领新合作的能力,来评价国家和机构的科学竞争力和发展潜力,并与其他分析方法结合尝试未来突破性和未来竞争力的评价。

当然,《科学结构地图》作为一种数据密集型研究工具,我们对它的了解和应用刚刚起步。未来,它将是一个开放的科学结构地图知识服务平台,连接描述科学技术各个环节的各种海量数据,连接多元化的需求、方法与工具,形成科学结构、科学发展趋势的分析、探索、试验和预警能力。



中国科学院