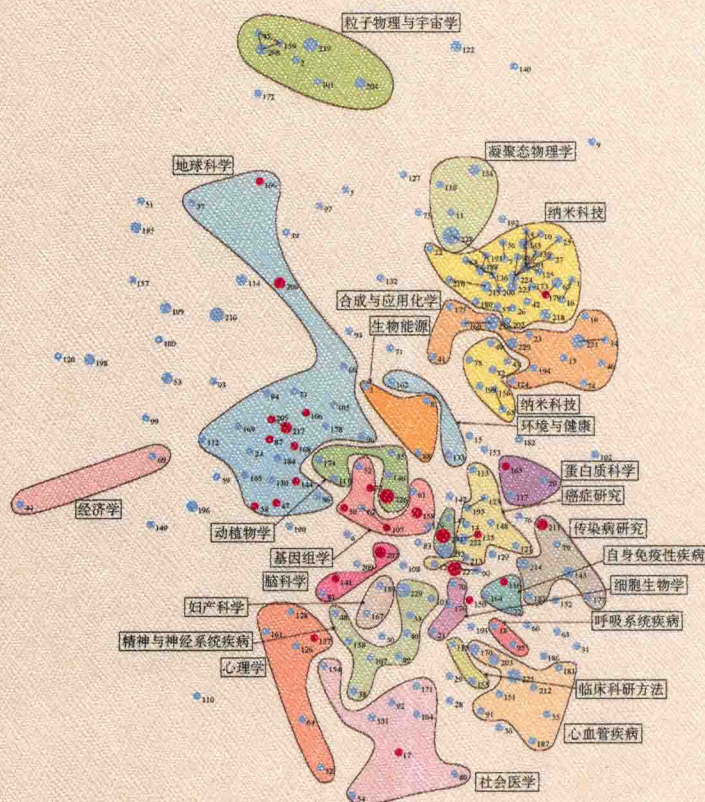


MAPPING SCIENCE STRUCTURE

科学结构图谱

2017

王小梅 韩涛 李国鹏 著
陈挺 张晓林



科学出版社

MAPPING SCIENCE STRUCTURE

科学结构图谱

2017

王小梅 韩 涛 李国鹏 著
陈 挺 张晓林

科学出版社

图书在版编目(CIP)数据

科学结构图谱.2017 / 王小梅等著. —北京: 科学出版社, 2017.10

ISBN 978-7-03-054369-1

I. ①科… II. ①王… III. ①科学体系学—研究 IV. ①G304

中国版本图书馆 CIP 数据核字 (2017) 第216392号

责任编辑: 邹 聪 乔艳茹 / 责任校对: 何艳萍

责任印制: 张克忠 / 封面设计: 有道文化

编辑部电话: 010-64035853

E-mail: houjunlin@mail. sciencep.com

科学出版社出版

北京东黄城根北街16号

邮政编码: 100717

<http://www.sciencep.com>

北京汇瑞嘉合文化发展有限公司印刷

科学出版社发行 各地新华书店经销

*

2017年10月第 一 版 开本: 787×1092 1/16

2017年10月第一次印刷 印张: 18

字数: 362 000

定价: 148.00元

(如有印装质量问题, 我社负责调换)

本书受中国科学院发展规划局“情报研究与平台建设”项目
及国家自然科学基金项目“科学结构特征及其演化动力学分析方法与应用研究”
(No.71173211) 资助

把握全球科学结构 揭示创新发展趋势

当前国际竞争形势瞬息万变，要求我们始终以全球视野谋划和推动科技创新，准确把握科技发展趋势，判断科技突破方向，抢占科技发展制高点，在重要科技领域跻身世界先进行列。面对多学科不断交叉融合会聚，把握科学研究总体结构及其不同领域的持续关联和演变，有助于我们及时探测和发现科学研究中的新问题、新领域及其萌生、崛起、发展的路径和动力因素，有助于我们认识科学规律、凝练研究重点、选择研究突破口。

为揭示科学研究结构及寻找重点研究方向，科学结构课题组连续十年周期性（两年）监测科学研究结构及其演变规律，突破传统利用学科分类、主题聚类、共词聚类、专家评价等来分析科学结构的方法，利用高被引论文间的引用关系，尤其是论文同被引所揭示的文献之间潜在的关联性，分析科学研究者根据研究需要自发引用文献所产生的内容关联性，可在一定程度上排除用已有学科分类、常用词汇、项目、作者或机构等归类所存在的主观和先验影响，有助于发现尚未被科学界认识或承认的科学方向和前沿。《科学结构图谱 2017》的研究借鉴经济合作与发展组织（Organisation for Economic Co-operation and Development, OECD）《抓住科学演化本质，制定新的科学指标及学科分布图》的方法，利用科睿唯安公司的基本科学指标数据库（Essential Science Indicators, ESI）的高被引论文和科学引文索引（Science Citation Index, SCI）、社会科学引文索引（Social Science Citation Index, SSCI）中的施引论文，揭示出由同被引关系聚类形成的研究领域及

其相互关系，分析这些领域的学科交叉性，以及中国和其他代表性国家在各个交叉融汇领域的活跃程度。

连续五期的数据揭示，科学研究领域的规模继续扩展，学科交叉融汇所形成的新领域不断出现，本期新出现的研究热点很多与人类健康、环境与可持续发展息息相关。同时，中国在基础科学研究中持续快速崛起，正在深刻改变世界科技创新的版图。核心论文份额从第一期世界排名第五，到第三期世界排名第二，持续到 2015 年达到德国的 2.6 倍，TOP10% 施引论文份额世界排名第二，依然有着较强的上升后劲。中国在研究领域覆盖率、不同研究领域的均衡性、国际合著率等方面都在迅速进步，不断缩小与美国等科技发达国家的差距。

同时也要看到，该书采用的引用关系只揭示了科学研究中复杂关系的一个侧面，且所使用的数据比较偏重基础科学。因此，分析结果只是从一个侧面揭示了当前科技界关注的热点研究领域及重点研究方向，仍需不断完善。希望课题组在读者提出的宝贵意见的基础上，进一步丰富研究科学结构及其演变规律的角度和方法，提高科学结构分析的时效性，同时将科学结构分析与其他分析方法密切结合，更有力地支持科技决策和科技创新。

潘教峰

2017 年 7 月 6 日

当今时代，科学技术呈现迅猛发展的态势，对科学知识体系结构及其演化趋势的把握越来越重要，也越来越复杂。面对海量数据和日益交叉融汇的科学研究，传统的人工调研、研讨等方式受到强烈挑战。随着科技信息的数字化和科学研究的数字化，科技创新中的各类知识对象，如文献、专利、机构、人员、项目、设施等，几乎全部被信息化、数字化，形成反映科学活动的海量数据。文献计量界逐步发展出利用这些数据来揭示大问题领域及大时间跨度和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。科学结构发现及可视化方法比其他计量方法具有更独特的视角，揭示了科学领域间的内在联系及发展规律。

科学结构课题组从 2007 年起开展科学结构发现与可视化分析方法的研究，以直观形象的图谱形式展现高度抽象的科学，特别是自然科学基础研究的宏观结构，揭示科学热点前沿之间的关联关系与发展进程。由于可视化形式更符合人们的认知习惯，该方法越来越多地为各领域研究者所关注和使用。课题组基于 ESI 高被引论文研制科学结构图谱，每两年周期性监测科学研究结构及其演变规律，先后出版了《科学结构地图 2009》《科学结构地图 2012》《科学结构地图 2015》三部著作，可视化地展现了科学研究的宏观结构和复杂关系，揭示了全球范围内的热点或潜在热点研究领域，描绘了五个时期科学研究领域的演化变迁轨迹，为观察、分析和探索科学

现状及发展趋势提供了新的视角和工具。

《科学结构图谱 2017》是该系列的第四部著作。在英文中，“map”有图谱、映射的含义，但在中文中，“map”通常的译文“地图”有特定的含义。因此，“science map”翻译成“科学地图”造成一定的误解，很多读者会认为是和地理相关的地图。因此，为更好地适合读者对研究工作的理解，从本期开始，我们改名为“科学结构图谱”。

本书的研究和撰写得到了中国科学院发展规划局的指导和支持，以及中国科学院科技战略咨询研究院、中国科学院文献情报中心的大力支持。中国医学科学院医学信息研究所杜建，中国科学院科技战略咨询研究院黄龙光、张超星，中国科学院文献情报中心杨艳萍、谢秀芳、郭世杰撰写了相关学科的研究领域演变分析。于泳院士、李永舫院士、许智宏院士、曹雪涛院士、刘小汉研究员、魏志祥研究员对科学结构演化部分提出了指导性意见。中国科学院科技战略咨询研究院裴瑞敏、王海名、范唯唯，中国科学院文献情报中心陈欣、刘小平、吕晓蓉，中国科学院心理研究所王玮参与了研究领域的命名、分类和解读分析工作。中国科学院文献情报中心硕士研究生吕伟民、吴朋民参与了图表绘制、文字整理工作。中国科学院科技战略咨询研究院朱涛、王海霞老师提出了许多宝贵意见。在此对以上所有机构和人员及所有参与前几期科学结构图谱判读的专家、提出宝贵意见的专家致以衷心的感谢。

科学结构课题组

2017年6月6日

目 录

序	i
前言	iii
第一章 引言	001
第二章 研究方法 with 数据	004
一、利用同被引聚类确定研究领域	004
二、科学结构图谱可视化	005
三、科学结构演变轨迹	006
四、研究领域特征词抽取	008
五、研究领域学科交叉性度量	008
六、数据说明	010
第三章 科学结构及其演变	012
一、科学结构图谱 2010~2015	012
二、基于科学结构图谱观察科学研究的发展趋势	014
三、研究领域演变分析	019
第四章 研究领域的学科交叉性	057
一、在科学结构图谱中观察学科交叉性	057

二、学科交叉度	064
第五章 中国及代表性国家科学研究活跃度	068
一、中国及代表性国家整体科研活跃度时序发展	069
二、中国及代表性国家在各个研究领域的活跃度时序发展	075
第六章 中国及代表性国家的国际合作	092
一、基于科学结构图谱观察世界国际合作	093
二、中国及代表性国家国际合作时序变化	093
三、基于科学结构图谱观察中国及代表性国家国际合作的变化	097
第七章 科学结构图谱上的科学资助情况分析	108
一、中国及代表性国家政府资助核心论文在科学结构图谱上的分布	109
二、重要国家政府资助机构资助论文在科学结构图谱上的分布	124
第八章 结语	134
附录	137
附录 1 三维科学结构图谱	137
附录 2 科学结构图谱 2010~2015 研究领域详细信息	140
附录 3 研究领域演变轨迹	169
附录 4 各国在科学结构图谱 2010~2015 研究领域中核心论文份额 及国际合著率	236
附录 5 资助论文在科学结构图谱 2010~2015 中的分布	256

第一章

引言

科学结构图谱，通过可视化技术，以直观形象的图谱形式展现高度抽象的科学，特别是自然科学基础研究的宏观结构，揭示了科学热点前沿间的关联关系与发展进程。运用文献计量学的理论和方法周期性绘制科学结构图谱，可以快速、全面和形象化地把握科学总体态势，分析各个国家、机构在不同学科或主题上的优势领域和发展重点，监测科学发展的演变路径和变化趋势，追踪各个国家科研优势领域的动态变迁过程，鉴别主题交叉融汇趋势和潜在突破领域，辅助确定未来研发方向和创新机制，支撑科技发展规划和战略决策分析。因此，科学结构课题组自 2007 年起开展相关研究，每两年绘制一期科学结构图谱，周期性监测科学研究结构及其演变规律，监测科学发展趋势。

对科学知识体系结构及其演化趋势的研究一直是科技界发现研究热点、遴选优先领域、规划学科发展和制定科学政策的重要基础。传统的研究方法通过检索和分析相关文献以了解学科发展，追踪同行科学研究者的科研活动来掌握学科趋势，通过专家研讨、评议及专门的规划研究进一步判断可能的突破方向。但随着科技创新进入多学科交叉融汇的阶段，面对海量科技文献，限于固有的专业认知体系，科学研究者有时难以观察到不熟悉但相关联的领域，也难以把握它们之间的复杂结构和相互影响，更难以发现隐藏在复杂关系下的致变因素和潜在的发展趋势。

因此，本书研制的科学结构图谱利用大量科学论文之间同被引关系的聚类分析，超越传统的学科分类，直接体现科学研究者相互引证所表征的知识的相互作用及知识的流动、融汇和演变，帮助科学研究者了解隐藏在大规模的复杂关联的数据下面的科

学研究结构及其变化,努力帮助科学研究者把握大问题尺度上和交叉融汇机制下的知识结构、新兴领域及其相互关系,逐步帮助科学研究者揭示演变趋势、预警新兴领域、发掘潜在合作对象、遴选优先领域等,辅助决策者对科学发展的规划。

本书以科睿唯安(Clarivate Analytics)公司的ESI为信息源,提取了2010~2015年9546个研究前沿中包含的高被引论文,通过同被引聚类分析,得到了232个研究领域,形成了全球视野的科学结构图谱,可视化地展现了2010~2015年的科学研究宏观结构及其内在关系,揭示了国际社会普遍关注的热点研究领域。在此基础上,通过2002~2007年、2004~2009年、2006~2011年、2008~2013年和2010~2015年五个时期科学结构的演化变迁轨迹,分析了各个学科研究领域的演变情况。通过引入生物学第三代多样性计量方法,度量了各国研究领域的学科多样性。同时,基于科学结构图谱,从国家科学研究的结构上反映了中国及代表性国家在不同研究领域的活跃程度及其变化趋势,通过国际合著率描述了中国及代表性国家国际合作的总体趋势。通过可视化展现中国及代表性国家政府科学资助基金在科学结构图谱上的资助分布,对比分析不同国家科学资助或同一国家不同资助机构的资助布局。

本书的结构如下:第一章,引言;第二章,研究方法 with 数据,介绍了绘制科学结构图谱的理论和方法;第三章,科学结构及其演变,绘制了232个研究领域组成的科学结构图谱2010~2015,分析其主要特点,通过五期科学结构图谱的时序分析展示了科学结构的演变;第四章,研究领域的学科交叉性,通过分析232个研究领域与14个学科之间的关系揭示了研究领域的学科交叉性及在科学结构图谱上的分布;第五章,中国及代表性国家科学研究活跃度,通过观察中国及代表性国家在每个研究领域中的论文份额分布,讨论各国在不同研究领域中的相对优势;第六章,中国及代表性国家的国际合作,通过国际合著分析度量知识的国际流动;第七章,科学结构图谱上的科学资助情况分析,基于科学结构图谱分析了中国及代表性国家的政府科学资助基金对各个研究领域的资助布局;第八章,结语。

本书术语解释

科学结构图谱:或称为科学知识图谱,是一种显示科学知识结构关系与发展进程的图形,反映了科学知识之间的结构、互动、交叉、演化等诸多关系。

高被引论文(highly cited paper):ESI对过去10年SCI论文被引频次进行统计,将22个学科领域中被引频次TOP1%的论文遴选为高被引论文。

研究前沿(research front, RF):ESI以SCI近6年的高被引论文为基础,利用论文之间的同被引关系聚类产生的一系列论文集合。

研究领域(research area, RA):在研究前沿基础上的再次聚类得到的一系列高被

引论文集合。

同被引 (co-cited): 一组论文共同被其他论文引用。

核心论文 (core paper): 研究领域中的高被引论文。

施引论文 (citing paper): 引用核心论文的论文。

重力模型 (gravity model): 对象之间关联强度的可视化方法, 基本思想是把对象间的关联强度看作原子之间的引力和斥力, 当力平衡时就形成各个对象之间的相对位置。

平均年 (mean year): 一组论文的出版年的平均值。

国家核心论文份额: 该国发表的核心论文数占世界核心论文数的比例。

国家施引论文份额: 该国引用核心论文的论文数占世界引用核心论文的论文数的比例。

国际合著率: 一国有多国著者的论文数占该国总论文数的比例。

国家论文计数方法: 由于存在国际合著论文, 一个国家的论文量因为不同的计数方法而不同, 计算方法包括: ①全计数法 (complete counting); ②分数化全计数法 (complete-normalized counting), 也称分数计数法 (fractional counting); ③第一作者直接计数法 (straight counting); ④整体计数法 (whole counting); ⑤分数化整体计数法 (whole-normalized counting)。采用方法②③⑤计算时所有国家份额之和等于 100%, 采用方法①④计算时所有国家份额之和超过 100%。本书中论文份额统计采用分数计数法, 国际合著率统计采用整体计数法。

第二章

研究方法 with 数据

科学结构图谱的主体分析单元是热点研究领域，它通过对高被引论文的同被引关系聚类产生。本期科学结构图谱的构建，首先对高被引论文的同被引关系进行聚类分析，产生若干研究领域；其次根据各个研究领域间的关联强度利用重力模型计算其相对位置并可视化，通过平行映射保持与前期图谱位置的稳定性和连续性；接着通过文本分析对研究领域中论文的题目和摘要抽取特征词以标识各个研究领域的 content；最后参考在研究领域之上的三层同被引聚类，由科技情报研究人员审核、修改以确定该研究领域的命名及所属的研究大类。

一、利用同被引聚类确定研究领域

同被引指一组论文共同被其他论文引用，反映了在学科分类、发表期刊、作者机构、研究项目等方面看似毫无关联的该组论文可能存在着某种关系。当该组论文同时被引用的次数逐渐增加时，它们之间的内在关联不断加强。因此，同被引现象通过作者自发的引用行为反映了科学研究内容和科学研究活动的聚合关系，可以超越传统的学科分类限制，反映了科学研究内容的自组织与科学结构的演变。

本书沿用前四期科学结构图谱的同被引聚类方法，并沿用《科学结构地图 2012》中的聚类参数。研究前沿取自 ESI 于 2016 年 3 月公布的 2010 年 1 月至 2015 年 12 月的研究前沿，共 9546 个，其中包含 45 657 篇高被引论文。施引论文集选自 SCI 和 SSCI，论文发表时间范围为 2010~2015 年。通过同被引聚类，形成 232 个研究领域，其中包含 3464 个研究前沿，19 850 篇高被引论文（核心论文）。

科学论文间的引用反映了科学研究的动态交互。同被引是指一组论文同时被其他论文引用，如图 2-1 所示，论文 A、B、C 同时被论文 1、2、3 引用。如果论文 A、B、C 频繁同被引，可以推测它们拥有相同或相近的研究主题。

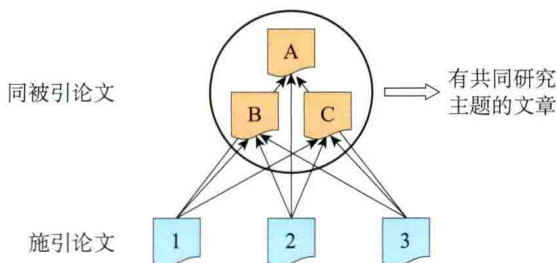


图 2-1 通过同被引分析确定研究领域

A、B、C 为核心论文

使用同被引的方法，计算高被引论文两两之间的同被引关系，并根据同被引关系对高被引论文进行聚类形成若干论文簇，称为“研究前沿”；在此基础上利用同被引关系对上述研究前沿再次聚类，得到的若干论文簇，称为“研究领域”。高被引论文、研究前沿及研究领域之间的关系如图 2-2 所示。

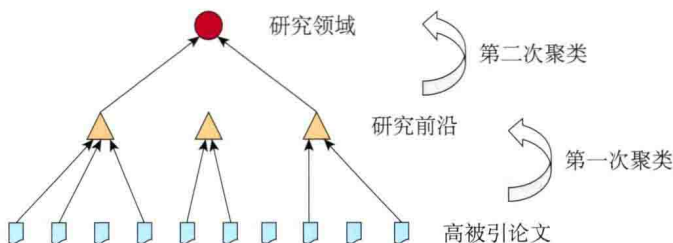


图 2-2 高被引论文、研究前沿、研究领域的关系

二、科学结构图谱可视化

科学结构图谱采用重力模型算法^①形象描述了研究领域之间的相互关系，确定各个研究领域在二维空间中的布局位置，绘制成图谱。为了使两次科学结构图谱纵向可比，本书采用平行映射算法对重力模型进行了改进。各研究领域除受本期其他研究领域的关系（引力和斥力）影响外，还受前一期科学结构图谱中研究领域的引力，即后一期科学结构图谱中研究领域的相对位置参考了前一期科学结构图谱的位

^① Quinn N R, Breuer M A. 1979. A forced directed component placement procedure for printed circuit boards. IEEE Transactions on Circuits and Systems, 26(6): 377-388.

置。有学者研究表明,全领域的科学结构具有“共识性科学结构”(consensus map of science)^①,本书作者通过试验也发现新一期科学结构与科学结构2002~2007、科学结构2004~2009的布局存在一致性。为了减少算法的参数对科学结构布局的影响,通过平行映射,可以得到时间相异但形态更加一致的科学结构图谱,从而支持后续科学结构的演化分析。

本书采用三种可视化方法展现科学结构中研究领域的布局,每种方法各有利弊,可以互补:①点线图,用圆圈表示研究领域,圆圈半径表示研究领域的扩展度(即研究领域中包含核心论文的数量),圆圈之间的连线表示研究领域的关联关系。这种方法可清晰地展现研究领域的具体位置及关联关系,缺点是当研究领域核心论文数相差悬殊时,圆圈大小难以反映此悬殊情况。②密度图,使用了高斯函数表示研究领域扩展度在二维平面上的密度分布。这种方法无法直观反映研究领域之间的关联关系,但揭示出研究领域核心论文密度的分布情况。③三维地形图,使用地理信息系统的三维可视化模块进行展示(附录1)。这种方法将研究领域扩展度分布在三维空间中,使科学结构图谱具有立体感,更加直观,但与密度图一样无法直观反映研究领域之间的关联关系,而且研究领域的具体位置不如二维图清晰。

三、科学结构演变轨迹

研究领域的演变可以归纳为新增、消失、分化、合并、延续五种模式,但是在知识的演变过程中,分化和融合具有相互转化、相互渗透的辩证统一关系,融合往往意味着另一种形式的分化,再精细的分化也总是伴随着不同学科知识的交叉和融合,由此形成一种演变模式综合交错的演变路径。本书采用图2-3所示的演变轨迹流图展现研究领域演变路径。研究领域的演变关系基于两个时期科学结构共同时间窗内(4年)的重叠度(重叠论文),重叠论文越多,表明研究领域之间的继承关系越强。



图 2-3 研究领域演变轨迹

^① Klavans R, Boyack K W. 2009. Toward a consensus map of science. Journal of the American Society for Information Science and Technology, 60(3): 455-476.

实现方法如下：

1. 计算研究领域间的重叠度

研究领域的演变分析基于两次聚类结果公共时间窗内的重叠度（重叠论文）判断演变关系，重叠知识在两个时间窗的聚类结果中传承，以前聚合在一起的知识以后可能分离，或者现在分散的知识可能后来聚合起来，这种现象正如知识进化理论提出的“遗传继承”。

科学结构图谱相邻两期的公共时间窗口是 4 年。如图 2-4 所示，在公共时间窗口，前一期科学结构图谱中的研究领域 P 有核心论文 N_P ，后一期科学结构图谱中的研究领域 Q 有核心论文 N_Q ，两个研究领域有共同的核心论文 N_{PQ} ，定义两个研究领域的重叠度为

$$N_{CO} = N_{PQ} / \sqrt{N_P N_Q}$$

重叠度反映了研究领域重叠关系的强弱。重叠度越大，表明研究领域之间的继承关系越强。

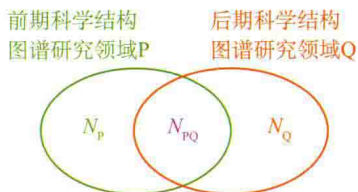


图 2-4 研究领域的重叠

2. 确定演变轨迹流

根据不同时期研究领域的重叠度对研究领域进行单链接聚类，聚在一起的研究领域形成若干个演变轨迹流。经过试验，最终选定每个演变轨迹流最多包含 40 个研究领域，最少包含 1 个研究领域。

3. 可视化展示演变轨迹

使用如图 2-3 所示的河流图可视化展示演变轨迹。图中圆圈代表研究领域，从左到右分别表示科学结构 2002~2007、科学结构 2004~2009、科学结构 2006~2011、科学结构 2008~2013、科学结构 2010~2015 五个时期。圆圈的面积与所代表研究领域核心论文数成正比。圆圈右方对该研究领域进行了标识和描述：第一行的数字代表研究领域 ID 号，后面跟着的是研究领域所属的学科分类；第二行是研究领域名称。圆圈之间的连线表示研究领域之间有论文重叠，红色连线代表重叠度在 0.2 以上，灰色连线代表在 0.2 以下，线条粗细和重叠度成正比。圆圈的颜色根据中国在各个研究领域的份额确定，蓝色：0；绿色：(0, 1%)；黄色：[1%, 3%)；橙色：[3%, 7%)；紫色：[7%, 12%)；红色：[12%, 100%]。