

国外新兴研究话题发现研究综述

卢超¹, 侯海燕², Ding Ying³, 章成志^{1,4,5}

(1. 南京理工大学信息管理系, 南京 210094; 2. 大连理工大学WISE实验室, 大连 116024;
3. 印第安纳大学信息学、计算与工程学院, 布鲁明顿 47408; 4. 江苏省社会公共安全科技协同创新中心,
南京 210094; 5. 江苏省数据工程与知识服务重点实验室(南京大学), 南京 210093)

摘要 新兴研究话题发现一直以来都是学者们广泛关注的研究问题。对于新兴研究话题动向的把握, 不仅可从宏观层面为管理部门制定科技政策提供决策支持、以提高科研资金投入的效益、加速科学进步和发展, 还可为科研工作者投身富有前景的研究领域提供具体的研究方向建议。对国际核心期刊上关于新兴研究话题发现的研究现状进行梳理发现, 目前学术界对新兴研究话题及其相关概念尚无清晰的界定; 相应地, 新兴研究话题的探测方法也存在诸多的不足; 探测新兴研究话题中所依据的指标繁杂多样。综述现有研究, 本文给出新兴研究话题的概念及其特征的界定, 比较新兴研究话题及其相关概念的联系与区别, 并指出在新兴研究话题发现中应注意的具体问题并提出四个方面的工作展望。

关键词 新兴研究话题发现; 研究前沿; 共被引分析; 文献耦合分析; 直接引用分析

Review of International Studies on Discovering Emerging Topics

Lu Chao¹, Hou Haiyan², Ding Ying³ and Zhang Chengzhi^{1,4,5}

(1. Department of Information Management, Nanjing University of Science & Technology, Nanjing 210094;
2. WISE Lab, Dalian University of Technology, Dalian 116024;
3. School of Informatics, Computing, and Engineering, Indiana University Bloomington, Bloomington 47408;
4. Jiangsu Collaborative Innovation Center of Social Safety Science and Technology, Nanjing 210094;
5. Jiangsu Key Laboratory of Data Engineering and Knowledge Service (Nanjing University), Nanjing 210093)

Abstract: The detection of emerging topics is a research area that has been increasingly attracting numerous scholars' attention. An extensive knowledge of the dynamics of emerging topics in science and technology not only provides our government with support to make decisions on policies regarding guidelines for scientific research in order to improve profits from financial investments on scientific research and to accelerate the development of science, but can also offer scientific professionals directions for promising research areas that can help them dedicate themselves to scientific discovery. Although we used summaries of relevant literature on this topic that were published in international core journals, we found no clear definition for "emerging topic" or its related terms. The resulting methodology, therefore, has some limitations to conquer. Indicators for identifying emerging topics vary across the literature. Based on our investigation on the studies in this area, we proposed a definition for "emerging topic" and identified its attributes. Moreover, we also compared "emerging topic" with its related terms and identified some areas worth noting when detecting emerging topics in scientific research. Finally, we presented four aspects of this topic that require further study in the future.

Key words: discovering emerging topics; research front; co-citation analysis; bibliographic coupling analysis; direct citation analysis

收稿日期: 2018-09-21; 修回日期: 2018-10-23

基金项目: 国家社会科学基金重大项目“情报学学科建设与情报工作未来发展路径研究”(17ZDA291)。

作者简介: 卢超, 男, 博士研究生, 研究方向为文献计量、文本挖掘; 侯海燕, 女, 教授, 博士生导师, 研究方向为文献计量、研究前沿发现等; Ding Ying, 女, 教授, 博士生导师, 研究方向为数据驱动知识发现、创新扩散等; 章成志, 男, 教授, 博士生导师, 研究方向为信息组织、信息检索、文本挖掘与自然语言处理, E-mail: zhangcz@njust.edu.cn。

1 引言

创新是国家强盛之基，创新是民族进步之魂。党的十八大以来，创新驱动发展战略如火如荼地开展，硕果累累，党的十九大报告提出要坚定实施创新驱动发展战略，加快建设创新型国家。人民日报发表文章表示创新驱动不等同于简单的技术创新，创新驱动也不能简单地理解为“颠覆式创新”这一种认识^[1]。因此，厘清过去和现在是预测未来的基础，创新工作才能有的放矢。一个国家的长久发展离不开创新，一个学科的发展同样离不开创新。科学计量学领域的学者们一直注重对学科前沿发展态势的认识和探索。对学科未来发展的预测，既是重点、亦是难点。新兴研究话题的探测或发现旨在预测有潜力、有影响力的研究话题。从宏观层面来看，管理部门要制定科技政策、集中优势资源去发展新兴领域，如《2018年国务院政府工作报告》中就指出要大力发展人工智能的新技术新应用，推进多个领域实现“互联网+”战略。从微观层面来看，科研工作者亦需把握时代的脉搏，发现有潜力、有价值的研究领域，利用相对有限的资源完成卓有成效的科研工作。

系统梳理对新兴研究话题发现的相关研究，总结已有成果、指出存在问题、提出未来可能研究趋势，具有重要的理论与实践意义。鉴于此，本文以检索式（见附录）在Web of Science核心集中检索相关文献，时间范围为1900年1月到2018年4月，共检索到外文文献共177条记录，最新一条记录为2018年2月。通读这些文献后，首先人工剔出不相关的文献共80余条，筛选出余下直接相关的90余篇作为本文综述的文献来源。通读这些相关文献及其与本文相关的参考文献并归纳整理，本文从相关概念辨析、发现方法、探测指标和验证方法四个方面，对新兴研究话题发现的相关研究进行全面梳理和述评，以发现新兴研究话题发现研究的成果与局限以及实际研究中的难点，对未来该领域的研究进行展望，如图1所示。

2 新兴研究话题及其相关概念辨析

2.1 新兴研究话题

新兴研究话题（emerging research topic，简称emerging topic）通常又被称为新兴研究趋势（emerging trend）^[2-4]、新兴研究领域（emerging research

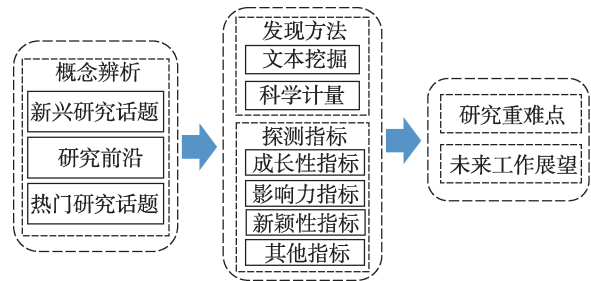


图1 研究思路图

field)^[5-7]或新兴研究主题（emerging theme）^[8]等。现有的研究仅有小部分研究对新兴研究话题进行定义^[3,9]，更多的则倾向于用特征描述的方式对新兴研究话题进行定义和概括^[3,8-10]。Tu等^[3]认为，“新兴研究话题是热门研究话题的前身，尚处在未被广泛关注的某个研究领域”。Wang^[9]在借鉴Rotolo等^[13]关于新兴研究技术的界定后，提出新兴研究话题的定义，即“新兴研究话题是极具新颖性并具有较快增长速度的研究话题，这样的话题通常具有一定程度的内在耦合性和科技影响力”，具体表现为强烈的新颖性、较快的增长速度、耦合性和学术影响力。新兴研究话题虽然具备较强的创新性（novelty），但由于并没有得到学者的广泛关注、影响力较小^[12]，这为发现和预测新兴研究话题增加了难度。Small等^[10]认为新兴研究话题应当具备两个基本的特征：成长性（growth）和新颖性（newness）。我们认为Small强调的这两个特征是相辅相成的关系。Glänzel等^[12]认为新兴研究话题作为话题本身除了需要新颖性（newness）之外，话题还需要一定的耦合性（coherence）^[9,14]、独立性（independence）以及自给自足（self-sustained）的关系结构。他强调新兴研究话题应当具备较高的未来影响力^[12,15]。这一特征和Small提出的成长性在逻辑上有一定的交叉：一个不断成长的研究话题必须有足够的理论基础维持其持续的发展，而不至于被其他研究话题所掩盖或吞噬。Yan^[16]、Dotsika等^[17]认为新兴研究话题具有突发性，是在给定的时间之前并未出现过的话题。这些研究对新兴研究话题的理解符合学界对新兴或涌现（emergence）这一词的理解，即强烈的创新性、耦合性、相关性、完整性、宏观性、动态性和可感知性^[11-12,18-19]。此外，以往研究也发现新兴研究话题论文通常具有较高的学科交叉属性^[20-21]。

2.2 研究前沿

研究前沿（research front^[11,22-23]或research frontier^[24-25]）最早由Price^[26]提出。他认为，在一定时间

里，一个活跃的研究话题通常能够从被引文献中得到反映，因此一个引文网络中较早获得大量引用的一组被引文献共同组成富有活力的研究前沿。他认为研究前沿的特征是“生长的尖端”（growing tip）和研究的表皮层（epidermal layer）。Price 理解的研究前沿具备三个特征：活跃、影响力及新颖性。随后，Small 等认为研究前沿是由高被引论文构成的类簇（clusters）^[27]，并提出基于共被引网络^[28]探测研究前沿的方法，将整个共被引划分为两个部分：知识基础（knowledge base）和研究前沿^[10]。也就是说，较老的被引文献共同组成了知识基础，较新的被引文献及施引文献共同组成研究前沿。他们指出研究前沿应当具备两个特征：成长性（growth）和新颖性（newness）^[10]。Morris 等^[29]则认为一组论文能够被称为研究前沿，必须具备两个条件，即：这组论文的数量不能过小，它们由共同的相对固定的理论基础（或知识基础）。但 Garfield^[30]则并不区分知识基础和研究前沿，而是将所有共被引网络里的施引文献和被引文献共同称为研究前沿。Garfield 不区别较旧文献可能会导致发现研究前沿并不具备新颖性，所发现的话题可能是较为成熟的热门研究话题。逐渐地，后续的一部分研究开始认为研究前沿就是一组相似的高影响力的论文^[31]，忽视了研究前沿所具备的新颖性以及其不充分的学术影响力^[32]。由于持 Garfield 观点的这类文章缺乏对研究话题新颖性的考虑，因此，这种界定研究前沿的方式是不够充分的。

根据话题在不同时间窗口下的变化，Upham 等^[27]将研究前沿分为五类：生长中的（growing）、新兴的（emerging）、收缩的（shrinking）、稳定（stable）以及既存的（existing）研究前沿。我们认为他们所划分的五种前沿中，只有前两种是我们所理解的“研究前沿”，其余三种并不符合研究前沿的基本特征。同样地，在 Huang 等^[33]的研究中，只有其文章中界定为“growing fronts”和“emerging fronts”的话题才是本文所认可的符合研究前沿特征的研究话题。类似地，Shibata 等^[34]根据引文网络中各个话题的拓扑结构，将新兴研究话题分为两类：增量式创新话题和突破式创新话题。

2.3 热门研究话题

热门研究话题或研究热点（hot topic 或 hotspot）是另外一个和新兴研究话题、研究前沿经常放在一起比较的概念^[3,8,12,35]。一般认为，热门研究话题是所在学科被学者广泛关注的具有重要影响力的研究

话题^[3,12]。在方法层面，通过监测文献中的关键词或者受控词汇的词频便可以发现热门研究话题；热门研究话题通常由高频词来表征^[3-4,8]。事实上，在英文文献中有一个比较相近的词汇指代的却是“研究前沿”——“hot field”，比如，Toivanen^[24]写道，“We conceptualise research frontiers as 'hot fields', which capture the best and most influential new knowledge”。这里的 hot field 具备研究前沿的特征：新颖性和影响力。Small 等^[10]在其文章中也赞同了这一说法，“Small identified hot fields (i.e., what we might now call emergent topics) as those clusters with a high number of recent papers and a high mean publication year”。这两个出处共同指向了 Small 和 Griffith 在 1974 年发表的文章，探讨学科发展中的高端领域（specialty of high levels of activity）^[36]；后续学者认为这就是 Price 提出的“research front”的概念，如 Chen^[4]等。

针对我们检索所得文献的标题和摘要，我们进行 N-Gram 统计，设定 N 为 2~4，词频不小于 5，通过对概念单复数的归并，我们析出主要的相关概念如图 2 所示。我们发现研究前沿的使用率最高，其次是新兴研究话题的相关术语。有一小部分文献使用技术专利作为文献来源识别新兴研究话题，造成了新兴技术也成为高频词。

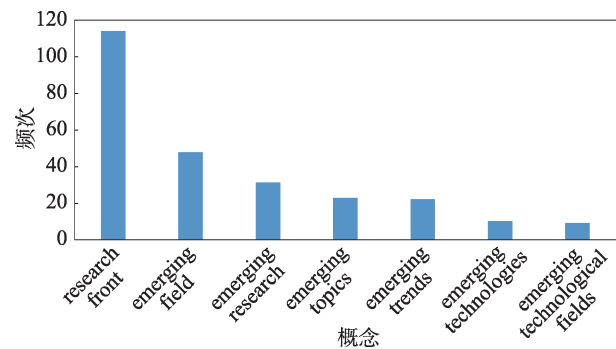


图2 新兴研究话题相关概念频次统计

2.4 概念辨析

比较这些概念的时候，我们会发现，新兴研究话题在概念上与研究前沿相似或等同^[4,10,27,37-42]。比如，Small 等^[10]理解的研究前沿就是笔者理解的新兴研究话题（emergent topic）。从他的表述来看，新兴研究话题和研究前沿只是不同时代下的不同表述，内涵基本一致。相似地，陈超美在比较 2006 年以前的关于研究前沿的主要工作后，也认为研究前沿就是新兴研究话题；只不过他对新兴话题的表述

是“Emerging thematic trends and surges of new topics”^[4]。比较新兴研究话题和研究前沿的特征时,我们不难发现新兴研究话题和研究前沿拥有共同的属性:成长性^[43]、新颖性^[3]以及影响力^[10]。因此,我们认为新兴研究话题在概念上与研究前沿等同,在下文用“新兴研究话题”统一指代“新兴研究话题”和“研究前沿”。

也有学者认为新兴研究话题和热门研究话题等同^[12,33,44-48]。Tseng等^[45]、Chen等^[46]认为热门研究话题(hot topic)和新兴研究话题(emerging trend)是同一个概念的不同称谓。由此可见,他们所理解的新兴研究话题(research front)指的是本文中的热门研究话题。类似地,Milman等^[49]利用高被引论文集探测研究前沿,默认发现所有的研究话题都是前沿,其中包括最大的几个类簇。然而,这些最大的类簇有可能已经是热门的研究话题。类似地,Huang等^[33]将文献耦合网络中识别的所有话题都称为“research front”,并提出“emerging front”作为新兴研究话题的一个子类更具有研究意义。

综述前人研究的观点,本文定义新兴研究话题如下:新兴研究话题,或称“研究前沿”、“新兴研究趋势”,是一个热门研究话题在其整个生命周期中(从产生到消亡)涌现阶段的称谓。在涌现阶段的研究话题具有一定的话题规模和话题影响力,研究话题在内容是具有较高的新颖性,研究话题的发展迅速,表现为吸引大量新的学者进入这个方向开展研究,完全成熟后该话题演变成为热门研究话题。如图3所示,新兴研究话题在兴起前尚不具备话题性,表现为文献发表基数过小、论文间话题的耦合性较小,但极具新颖性;新兴研究话题在快速发展后,进入缓慢的发展阶段,生长停滞或逐渐消亡,称为“热门研究话题”;热门研究话题的影响力进入巅峰状态,话题自身的新颖性也随之减弱,具体表现在话题的成长乏力、研究学者广泛、部分学者开始转移其研究重心。值得注意的是,任何一个研究话题都是存在其自身的生命周期,都会拥有自己的上升发展阶段。有些话题在上升阶段就是“新兴研究话题”并最终成长为具有重要影响力的“热门话题”,但是有些话题在其上升阶段并不能称为“新兴研究话题”,因为这些话题并不能最终发展成有重要影响力的“热门研究话题”。

本文拟定用“新兴研究话题”这一概念去综述相关研究是因为尽管“研究前沿”这一概念被大量使用(图2),但“新兴研究话题”这一概念在近5

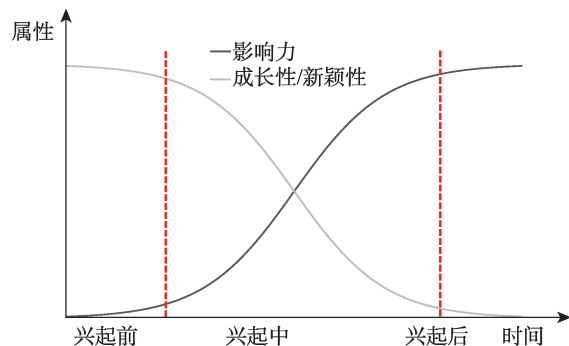


图3 新兴研究话题发展示意图(本图改编自文献^[13])

年来被越来越多的学者所认可^[4,10]。此外,“新兴研究话题”从字面上更能够体现相关研究需要捕捉图3中“兴起中”这一阶段的研究话题。因此,综合考虑到概念名称的“新颖性”以及词语表达的精确性,我们采用“新兴研究话题”来统筹相关概念。

3 新兴研究话题的发现方法

通常,新兴研究话题发现的工作着力于对新兴研究话题的描述、可视化的个案分析^[50-52],仅有较少的研究提出或改进新兴研究话题发现的方法^[4,10-11]。参照前人的分类方法^[10,53-54],我们将发现(或探测)新兴研究话题的方法大致分为两类^[10]:文本挖掘方法和科学计量分析方法。

3.1 文本挖掘方法

一部分研究借助于自然语言处理技术对科技文献和专利文献中的话题^[5,16,55]、关键词或受控词^[8]进行挖掘分析,从而实现对新研究话题的识别和探测。Kreuchauff等^[107]利用15名领域专家,对其采集的“服务机器人”相关的228份专利全文进行影响力评判,筛选出新兴技术专利,然后利用支持向量机模型对这些文献进行分类,取得正确率为85%的预测结果。Robinson等^[56]通过挖掘“纳米技术”领域科技文献和科普杂志全文中专业术语之间的相互关系:科幻、远景、指引、期望、目标和证明六类发现该领域中的新兴研究话题。Mörchen等^[57]改进LDA(latent Dirichlet allocation)模型、构建“话题-概念模型”,对生物医学领域文献摘要中的概念术语及其相关文献所组成的话题进行追踪,通过该话题相关术语的MeSH词表收录情况,可以检测新兴研究话题的发展过程。Yan^[16,58]利用LDA对文献中的主题分布进行分析,识别包括新兴研究话题在内的六种话题。Ohniwa等^[84]精选作者关键词中的

MeSH 主题词构建共词网络探测新兴研究话题。

突发词检测 (burst detection) 是指利用文本挖掘技术检测出文本流 (如新闻) 中的突然涌现出来的单词, 这一算法最早是由 Kleinberg^[59] 提出的。Guo 等^[60] 综合利用突发词检测、新学者进入速率以及交叉学科参考文献数量, 分析新兴研究学科, 如纳米技术领域等四个领域, 他们分析发现, 新兴研究话题通常会吸引大量的新的作者进入, 然后引用大量的交叉学科文献, 最后出现大量该领域的突发词。后来陈超美将突发词检测改进为多个突发词检测, 并整合入 CiteSpaceII 中进行新兴研究话题的检测^[4], 在图情领域得到了的广泛运用^[38,60-62]。

3.2 科学计量分析方法

除了利用文本统计分析的方法外, 科学计量分析方法是更为常用的方法, 即利用以科技文献元数据的分析与挖掘为基础的方法, 主要包括各种引用关系网络、共现关系网络的分析方法。

直接引文网络分析方法 (direct citation network analysis) 借助学术文献之间的引证关系形成网络, 是对新兴研究话题进行发现和识别的一种常用方法^[11] (如图 4a 所示)。共被引网络分析方法 (co-citation network analysis) 最早由 Small 提出, 是利用学术文献之间被其他文献共同引用的关系建立网络, 并发现学术文献中的新兴研究话题^[36] (如图 4b 所示), 该方法亦是新兴研究话题发现的一种主要方法^[15,27,34,37-38,63-66]。文献耦合网络分析方法 (document bibliographic coupling analysis) 利用文献之间共同引用的参考文献为边构引文网路, 用以发现新兴研究话题, 是另外一种常见的新兴研究话题发现的引文网络^[9,11,40] (如图 4c 所示)。共词网络分析 (co-word network analysis) 是指利用关键词、术语之间的共现关系构建的关系网络。由于引文数据具有一定的时滞, 所以部分学者利用共词网络分析来克服这一问题^[42,67], 这种方法是国内研究前沿识别主要方法之一^[68-70]。An 等^[8] 利用术语被收录于工程索引 (EI) 主题词表的时间作为评估文献新兴程度

的指标, 区别于新兴研究话题和热门研究话题构建共词网络探测研究话题。An 等^[71] 利用共词网络分析干细胞领域的发展趋势时, 为了能够更好地对新兴研究话题的发现, 他们利用优化后的信息熵值优选主题词, 将文本信息处理技术和文献计量方法相结合。Li 等^[61] 将突发词检测和共词分析结合在一起, 实现新兴研究话题的发现。Wang 等^[108] 利用作者关键词及其论文的发表时间探测话题的演变规律。Hu 等^[72] 利用中文图书情报核心期刊构建共词网络, 借助网络中关键词的度中心度和话题密度识别中国图书馆学与情报学学科的热门研究话题和新兴研究话题。

以上提到的四种方法是科学计量学领域新兴研究话题发现研究的主要方法。同时我们也注意到, 有少部分的学者利用作者引用网络的方法探测新兴研究话题^[14,32,73-77]。Zhao 等^[74] 认为作者引用网络可以探测新兴研究话题的基础是: 活跃的研究前沿工作是由当时的活跃的作者所主导的; 一个类簇出现在作者耦合网络中却并未出现在作者共被引网络中, 这个类簇所代表的研究话题是一个新兴研究话题。Ma^[75] 指出作者耦合网络对新兴研究话题的探测, 相较于作者共被引分析, 有更好的结果。然而, 我们认为新兴研究话题应当在短时间内吸引大量的新的研究学者的进入^[12,60], 由于新兴话题本身具备较高的新颖性, 此时核心作者的数量应当是较少的^[14]。这会使得在构建作者耦合网络时, 人们很容易忽略新兴研究话题中的大量低产作者。

Shibata 等选取氮化镓、复杂网络和碳素纳米管三个不同的学科领域分别对三种引文网络 (直接引文网络、共被引网络、和文献耦合网路) 在可见度 (visibility)、速度 (speed) 和相关度 (relevance) 三个方面进行比较, 结果发现, 直接引用网络能够在网络中发现较新较大的研究话题, 在新兴研究话题发现中显示出突出的优势; 共被引网络发现新兴研究话题效果最差^[41]。Glänzel 等^[12] 表示在较短的时间片段里, 文献耦合网络比共被引网络有更好的新兴研究话题探测效果。Boyack 等^[11] 基于两百多万的

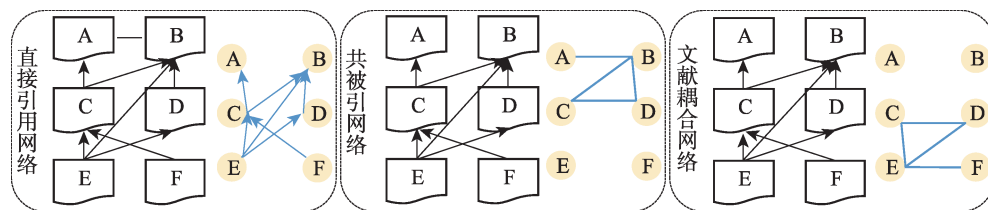


图 4 引用关系网络

文献数据对文献耦合、共被引、直接引用三种网络方法进行文本耦合度 (coherence) 和网络集中度 (concentration) 两个方面的准确性评估, 研究结果显示, 利用文献耦合方法的结果准确性高于共被引网络, 直接网络发现的新兴研究话题正确率最低。Huang 等^[40]在 OLED 领域下的测试结果也肯定了 Boyack 等^[11]的结论, 利用文献耦合可以发现更多更早的研究前沿。除此之外, Vladutz 等^[78]认为在大量的数据面前文献耦合方法就有具有更强的可用性, 对文献之间的语义上的相似性揭示得更好。而 Jarneving^[79]则认为利用共被引网络能保证网络中话题在内容上的耦合度和可描述性。Persson^[76]基于 JASIST 论文数据 (1986—1990 年) 表示作者耦合和共被引网络这两种方法都可以较好地展示数据集中的研究前沿。Zhao 等^[74]也指出将作者耦合方法与作者共被引方法相结合效果更好。

综上所述, 新兴研究话题的探测方法多种多样: 从文本挖掘到科学计量分析。由于对新兴研究话题的定义一直没有统一的认识和理解, 学者们使用新兴研究话题的方法也有很多的变化。随之而来

的结果是对新兴研究话题探测方法的性能评估也没有统一的认识。对新兴研究话题发现方法的评估指标尚停留在对研究话题的外部特征的描述上, 如研究话题类簇的大小、研究话题内部的相似度、话题的描述性等。这些指标还不能够很好地刻画新兴研究话题的本质特征。因此, 做好新兴研究话题发现方法的评估工作是亟须解决的问题。

4 新兴研究话题的探测指标

在新兴研究话题发现研究中, 指标是揭示新兴研究话题存在、形态和发展趋势非常有力的手段。此外, 在研究中纳入指标有利于回避判定新兴研究话题时的主观偏见^[79], 利于研究结果的验证。虽然当前大量研究仍处于新兴研究话题的分析和描述的阶段, 但有一部分研究为新兴研究话题发现研究提供了较为丰富的指标工具。本研究依据新兴研究话题的特征属性, 将用来探测新兴研究话题的指标分为四类, 即成长性指标、影响力指标、新颖性指标以及其他类型指标。前三类主要指标见表 1。

表 1 新兴研究话题探测指标描述

类别	名称	特征参数	方法框架	研究对象	数据来源或领域
成长性指标	成长性指标 ^[9]	文献增长率	文献耦合网络	学术文献	多学科
	创新性 ^[9]	文献增长量	文献耦合网络	学术文献	多学科
	发表量指数 ^[3]	文献增长量	文献统计分析	学术文献	ACM 电子图书馆
	基尼系数 ^[82]	术语分布	共词网络	学术文献	分子生物学
	前景因子 ^[84]		共词网络	学术文献	生命科学
	互信息 ^[89]	论文标题、被引文献分布	文献统计分析	学术文献	富勒烯、纳米管、信息科学
影响力指标	学术影响力 ^[9]	话题被引数据	文献耦合网络	学术文献	多学科
	普适度 ^[20]	专利引用数据	共被引网络	专利文献	纳米技术
新颖性指标	新颖性 ^[3]	文献发表时间	文献统计分析	学术文献	ACM 数电子图书馆
	参考文献年龄 ^[7]	被引文献发表时间	共被引网络	学术文献	生态学与环境科学

4.1 成长性指标

成长性指标, 是用来揭示新兴研究话题在未来的发展情况的指标, 具体表现在“量”的层面。文献增长被大量研究者作为揭示新兴研究话题的一个指标^[9-10,12,15,45,80]。其中文献增长率被作为最为重要的指标之一, 用来解释待探测话题在未来的发展潜力^[9-10]。Wang^[9]指出新兴研究话题需要具备四个维度特征, 第一个就是新兴研究话题的成长性 (growth) ——给定时间段内, 该话题的文献增长率 (曲线斜率) 应当大于等于设定的该话题的最小增长率。其中的第二个指标创新性 (novelty) 也属于

成长性的指标范畴: 给定时间段内, 该话题的年平均发文量应当小于等于设定的最大年发文量。新进入学者的数量也被作为揭示新兴研究话题成长性的一个指标^[15,60,81]。Roche 等^[82]利用 $TF \times IDF$ 和经济学中的基尼系数 (Gini Index) 发现文献中领域词的隶属度和发展情况: 具备较高 $TF \times IDF$ 值和基尼系数值的词有较高隶属度的技术词; 相反, 该词具有较高的话题交叉性, 成长性不足。Shiebel 等^[83]利用基尼系数来筛选光电子领域不同阶段话题的关键词。Ohniwa 等^[84]利用关键词的增长率来筛选涌现的关键词来构建共词网络探测新兴研究话题, 并利用研究

话题中的年均涌现关键词的数量构建“前景因子”(perspective factor)来识别共词网络中的新兴研究话题。Lucio-Arias等^[89]将学科的发展理解为科学系统熵减的过程,即信息不确定性不断减小的过程,因此他们利用科学文献的标题、参考文献信息及发表时间计算三者之间的互信息,当他们的互信息值不断减少时,意味着学科发展逐渐成熟。

4.2 影响力指标

影响力指标,是用来解释新兴研究话题在研究价值和社会影响力方面的指标,体现该话题在学术上“质”的层面。Wang^[9]提出的判定新兴研究话题的第三个指标是学术影响力(scientific impact)—给定时间段内该话题的引文数量大于等于等时间窗口下设定的最小引文数量。Trajtenberg等^[85]提出利用普适度(generality)来测度某类专利文献与其他类别专利文献之间的互引关系,从而评估该类专利潜在经济的影响力。Schultz等^[20]利用这个指标来测度一个技术领域专利的影响力,发现高影响力的专利技术类目有较高的普适度。新兴研究话题通常被认为在结构上具有较高的中介中心度(betweenness centrality)和聚类系数(cluster coefficient),在内容上表现为较高的学科交叉度^[20-21,60,81,86],因此这些指标经常被用来作为发现核心文献^[12]的重要指标^[4,15,86-87]。但需要注意的是,“新兴研究话题”本身并不是一个已经具备足够影响力的研究话题,而是一个“未来”具备巨大影响力的话题,其在网络中只是较为稀疏的类簇^[14]。因此,在使用影响力的指标时需要辩证地看待具体的研究情境^[32]。

4.3 新颖性指标

新颖性指标用来揭示新兴研究话题在时间维度上的新旧程度。在对新兴研究话题新颖性进行揭示时,“被引文献的平均发表时间”和“类簇论文的平均发表时间”这两个指标经常被学者们使用^[7,10,15,27,60,81,88],比如,Tu等^[3]将话题第一篇论文的发表时间定为该话题的新颖性为100%的时间,然后该话题的新颖性逐年衰减;An等^[8]将作者关键词出现时间和此词成为受控词的时间差作为新颖性的判定指标。Jarić等^[7]利用被引文献发表的时间年限,构建两个指标来测度新兴研究话题:参考文献平均年龄和近两年参考文献比例。

4.4 其他类型指标

还有一些指标难以纳入到以上三个类别中,如

话题耦合性(coherence)^[9,12]。Wang^[9]、Glänzel等^[12]认为新兴研究话题必须具备一定的话题的耦合度。相反, Lee^[90]在研究信息安全领域的新兴话题时,总结未来发展成为热门研究话题的新兴研究话题的特征,指出这些话题通常具备较低的接近中心度(closeness centrality)和度中心度,即文献之间的联系还并不紧密^[90]。Jarneving^[79]利用平均耦合强度表示话题的耦合度,以评估文献耦合网络的聚类效果。Shibata等^[41]和Boyack等^[11]将研究话题包含的文献数目作为识别新兴研究话题方法性能的指标。Mund等^[81]指出期刊的大小、发行的年限和影响因子大小会有助于探测新兴研究话题。张丽华等^[91]发现期刊编委发表的文章能够帮助更早地识别新兴研究话题。

4.5 讨论

通常新兴研究话题发现需要同时利用到多个维度的指标。例如,Shibata等^[34]设计两个指标来描述研究话题中论文在其网络社区中的结构特征:类簇内的标准化度(z)和类簇间的参与系数(p);利用这两个系数他们识别出直接引用引网络中的两类新兴研究话题:增量式创新(大 z 和大 p)和突破式创新(大 z ,小 p)。Tu等利用论文发表情况(published volume index)和创新性(novelty index)共同探测新兴研究话题。Guo等^[60]利用三种不同的指标去发现新兴研究话题:突发词检测,新进入的学者数量与速率以及跨学科参考文献数量,分析四个不同领域的新兴研究话题的特征。

有时某些单个指标本身就反映新兴研究话题的多个属性。Small等^[10]提出新兴潜力指标(emergence potential, EP),用于综合评估待判定研究话题的成长性和新颖性。González-Alcaide等^[15]利用带判定话题新入学者数目来作为判定指标。该指标不仅符合判定成长性指标的特征,也具备判定影响力的特征。有趣性分值(interestingness score)是Mörchen等^[57]对待收录词在文献中的出现频次 y 与时间 t 进行指数拟合($\log y = at + b$)后的斜率 a 来揭示某一话题的影响力。事实上,该指标也同时具备了判定话题成长性的性质。

5 新兴研究话题发现结果的验证方法

按照研究对象,新兴研究话题的发现可大致分为两类:对既定的新兴研究话题进行描述分析^[22,29,92],以及在选定的领域识别新兴研究话

题^[9-10,74,76,93]。在第二类研究中,发现新兴研究话题后,很多研究缺乏对发现结果的验证^[40,74,76,94],仅有少数研究对发现结果进行了验证^[9-10,27,33,41,83]。目前,关于新兴研究话题验证的方法,可以分为三类:资料验证法、专家调查法以及指标验证法。

5.1 资料验证法

本文说的资料验证法,是指研究者利用已发表的新兴研究话题发现或讨论的文章以及重要创新研究奖励(如诺贝尔奖)等对其研究结果的可靠性进行验证的方法^[84,95]。例如,Small等^[10]利用诺贝尔奖、权威潜力研究方向列表(如Science: Insights of the decade)或者领域相关的奖项对其发现的新兴研究话题进行验证。在其验证中,他们也发现1/3探测的研究前沿可以找到这样的验证;在探测为非新兴研究话题的话题中由25个话题找到了奖项的支持。Wang^[9]利用已发表的文献和研究新闻中对新兴研究话题的报道,验证自己的研究结果,其识别的19个研究话题都得到其他文献的支持。该种验证方式的优势在于验证的结果较为可靠、经济成本较低,便于研究结果的验证。但其劣势在于,利用已有资料进行验证无形中降低了该研究结果的价值性和时效性。

5.2 专家评估法

专家评估法是指研究者利用领域专家的知识 and 经验,对研究结果进行评估,以实现与研究结果进行可靠性验证的一种方法。比如,Upham等^[27]利用访谈的方式对其研究的领域中的30名学者进行访问。他们的访问结果显示,他们发现的3/4的研究前沿都是有意义和重要的研究方向,剩余的1/4研究前沿是一些特殊事件或者机构发布的数据。Jarneving^[79]组织专家对其构建的文献耦合网络提供优化建议,但并没有要求专家对其发现的类簇是否是新兴研究话题进行判断。Chen^[4]通过设计问卷的方式对探测的新兴研究话题中的核心文献作者(视其为领域专家)进行调查,这种验证方式存在较大的偏见。总体来看,专家评估法的优势在于研究结果较为可靠、验证后的研究结果可以具备相当的时效性和价值型。但是,由于研究结果的验证较多依赖专家意见,专家的选择可能会对验证结果带来较大的偏见^[45];同时,专家意见相左时,验证过程可能需要消耗较多的时间、精力与经济成本^[45,96]。

5.3 指标验证法

指标验证法,通常是在对新兴研究话题属性和

特征进行归纳的基础上,利用事先设定的指标对已识别的新兴研究话题进行验证或比较的方法。这种方法通常是用来比较不同的新兴研究话题发现方法的性能。比如,Shibata等^[41]利用可见度、速度和相关度等三个指标,对比了三种主流的引文网络方法在新兴研究话题发现任务上的性能,研究结果显示:直接引文能够发现最新最大的新兴话题、共被引表现最差。值得注意的是,此种验证方法并不是对发现结果的直接验证方法。一个话题的外部特征良好并不是一个新兴研究话题的充分条件。因此,该方法效果有赖于研究者对研究目标领域特征认知和理解的程度。

6 研究重难点与未来展望

6.1 研究重难点

在对新兴研究话题的发现研究中,存在着若干个研究的难点和重点。本文将按照新兴研究话题发现的一般流程阐明现有研究工作表现出的这些问题。首先,实验开始前研究者对新兴研究话题及其相关概念没有统一的理解。如上文所述,有的学者区别对待新兴研究话题和热门研究话题;有的学者将二者理解为同一概念。对新兴研究话题概念理解上的不同会直接影响到新兴研究话题发现实验的设计和实验结果的准确性。由于没有明确的定义,发现结果也没有统一的标准去评估和度量,导致了研究的评估方法多样。在热门领域中探测新兴研究话题。

在数据来源方面,目前绝大部分研究都在关注两大数据源:科技文献和技术专利。然而这两种文献存在其自身较大的弊端。举例来看,科技文献的出版周期越来越长,论文吸引引文也需要较长的时间。数据的不足会对新兴研究话题的发现,特别是预测提出了巨大的挑战。因此,引入学位论文^[97-98]、会议论文^[99]、网页新闻^[100]、社会化媒体^[101]、科研基金项目^[102]等数据会对新兴研究话题的发现可能会起到一定的辅助作用^[56,100]。另外,我们也注意到,利用全文数据及其中的引文内容数据能够提供更早更细节的影响力监测^[103],也是需要密切专注的一个数据源。

数据采集和清洗方面,相当一部分研究倾向于利用高被引论文数据集探寻新兴研究话题^[27,49,65-66,94]。Chang等^[94]利用图情领域的580篇高被引文献研究图情领域的学科发展动向。Upham

等^[27]用 top 1% 的文献去识别新兴探究话题。Huang 等^[33,104]收集 OLED 领域年度前 20 篇高被引论文探测该领域的新兴研究话题。利用高被引论文探测新兴研究话题或研究前沿有其较早的渊源，Small 在其 1974 年的文章中提出，利用共被引网络探测活跃的研究话题（specialty）就在构建共被引网络中利用“共被引强度”（co-citation strength）对网络中的节点与边进行过滤。这样的操作有利于发现较为有影响力的新兴研究话题，简化发现过程中的计算复杂度。但是，如果我们直接选择获取高被引的论文去构建引文网络可能引入两种偏见或误差：首先，在引文网络中发现的新兴研究话题实质上已经是热门研究话题或已经衰老的热门研究话题了；其次，真正的新兴研究话题由于还没有形成足够大的影响力会被主观的数据采集方式直接被过滤掉了。除此以外，由于脏数据造成的数据处理问题也值得关注^[69]。因此，如何制定研究的数据采集策略以及在对数据清洗的过程中选择合适的阈值过滤不重要的论文成为一个充满艺术性和技术性的难题^[12,79]。

数据处理方法方面，利用引文网络发现新兴研究话题时，学者们也倾向于选取引文网络中最大的组成部分进行新兴研究话题发现。这种处理方式是否合适，需要辩证地看待。一方面在绘制知识地图的时候，我们更多的任务是抓住主要问题，此时选取引文网络中的最大分图有利于可视化的效果。然而，面向新兴研究话题发现问题时，由于新兴研究话题是在成长中的有较大潜力的研究话题，新兴研究话题可能是完全独立于目标研究领域的研究范式和知识基础，而采纳其他学科的理论、知识和方法作为自己的知识基础以保持自身的独立性和可持续性。因此此种研究方法也存在忽视某些正在成长中的新兴研究话题的可能。我们认为应当利用整个引文网络进行新兴研究话题的发现。还有一部分学者在引文文献中去识别新兴研究话题^[23]。这种操作方式尽管回避了话题漂移的问题，但由于他们剔除了施引文献，也使得发现的新兴研究话题并未到最新的程度。

结果分析方面，目前我们的研究仍然更多地依赖于定性的表述或者比较浅层的定性描述。相当一部分的研究也或多或少地涉及了一些指标，但这些指标对新兴研究话题的研究还仅仅停留在简单的统计描述阶段，利用定量指标体系对新兴研究话题进行分析和展示的研究凤毛麟角^[9-10]。由于没有指标或规范的指导，识别出的新兴研究话题通常是研究

者的主观意见，缺乏足够的客观依据作为支撑^[79]，如文献[33,104-105]。具体来看，假设在某个领域 X，学者收集该领域的全部文献数据，构建共被引网络，通过聚类算法得到 m 个研究话题。这 m 个话题中，研究者可能并不区分哪几个是新兴研究话题，哪些是热门研究话题，哪些已经是“明日黄花”；由于缺乏甄别的指标，他们会认为所有的话题都是新兴研究话题。这种判断的方式无疑为新兴研究话题的发现结果引入了误差。因此，如何对现有文献中提及的指标进行系统性的归纳和筛选，结合新兴研究话题的特征属性将总结后的指标进行必要的增减，建立比较系统有效地识别新兴研究话题的指标体系也是一个亟须解决的问题。

实验结果验证方面，首先比较突出的问题是目前大量新兴研究发现研究缺乏验证手段，或者验证的方式存在缺陷。客观方面，我们对新兴研究话题的认识还不够清晰和统一，学者很难给出“金标准”；主观方面，新兴研究话题识别本身就需要识别者或领域专家的见解和认知，这为新兴研究话题发现的验证工作带来了挑战。因此，探析新兴研究话题内涵、外延、特征及其发展机制的研究需要进一步推进。在此基础上的标准化评价指标和数据集的构建也有可能成为另外一个需要突破的难题。

总体来看，现有的新兴研究话题的发现工作大多缺乏对新兴研究话题形成规律的探索，大部分的研究还停留在静态地查看引文数据和文献元数据的切片信息。一个话题的产生有其深厚的土壤（政治文化背景和错综复杂的社会关系），在科研领域，研究的地域性、时代性也都会在学者们的研究工作中充分体现。而现有新兴研究话题的识别研究中对这一更深厚的情报背景的解读工作还有很大的提升空间。对于新兴研究话题的发展演变的研究就更加稀有。这些问题都值得在未来的工作中不断解决。

6.2 未来工作展望

基于现有的新兴研究话题的研究现状和研究工作中出现的重难点问题，本文从四个方面对新兴研究话题发现的未来工作进行展望：①研究话题生命周期精准区分；②新兴研究话题标准化指标体系构建；③新兴研究话题规范化发现方案制定；④新兴研究话题发现研究从描述向探索转型。

首先，研究话题生命周期的精准区分。库恩在其《科学革命的结构》一书中陈明，学科的发展和变革有一定的规律性，在不同的发展阶段呈现出不

同的特征^[106]。研究话题的发展也不例外。如前文所述,我们认为新兴研究话题的出现必然存在着其发展前的萌芽阶段以及其涌现后的不断成熟伴随衰老的阶段。厘清不同阶段研究话题发展的规律和呈现的特征能够为研究新兴研究话题提供全景和蓝图。研究者清楚地认识研究话题的发展规律能够帮助做到:①在研究工作开展前有效规避混淆“新兴研究话题”和“热门研究话题”二者概念的问题;②针对新兴研究话题自身的特点制定合宜的识别方案和技术路径。这会为新兴研究话题发现的后续工作打好良好的理论基础。

其次,基于对新兴研究话题全面认识下的标准化指标体系的构建。新兴研究话题发现本身就存在着一定的主观性,研究者可能带入自己的理解和期待去识别话题。由于发现新兴研究话题研究一直以来缺乏普遍认可的指标体系或者没有指标辅助决策,加之验证手段的匮乏,新兴研究话题的识别结果难以保证其准确性和客观性。因此,标准化识别指标体系的构建能够为新兴研究话题的识别工作提供检验和评判的标准,不仅有利于相关研究发展和进步,还为新兴研究话题发现的应用工作铺平了道路。新兴研究话题的发现结果可以准确评估后,该成果将能真正意义上为组织或个人的发展规划建言献策。

同时,为了新兴研究话题发现研究能够更加有效地落到实处,研究设计新兴研究话题的识别方案和技术路径变得尤为重要。新兴研究话题识别的方法多种多样、数据处理的方式层出不穷,但对于新兴研究话题发现方法的评估研究还少之又少。在现有的研究方法的基础上,结合不同的数据源和发现目的,有针对性地制定出新兴研究话题的发现方案和技术路径能够发挥数据和方法的双重优势,有效为实践服务。

在此基础上,新兴话题发现研究从描述向探索转型。新兴研究话题发现的研究将不再着力于对新兴研究话题或学科的态势描述这种回溯型研究之上,因为回溯型研究成果的价值有限。虽然回溯型研究能为认识学科、话题发展提供素材,但回溯型研究的成果是过去的未来,对我们的未来发展没有更多的参考价值。更多的研究精力应当转向“沙里淘金”——从未知的学科、领域中识别出未来能成为卓有影响的研究话题。

值得注意的是,新兴技术的发展是提升新兴研究话题发现工作一个重要的支撑环节。近年来人工

智能技术、大数据处理技术、自然语言理解技术蓬勃发展,这些新的技术可以帮助我们不再被引文数据和文献元数据所局限,我们可以向学术全文挖掘的方向走得更远。同时,我们也看到,这些新的技术存在一定的门槛,对我们的学者还有一些小挑战。而克服这些困难需要我们更加紧密的合作、跨学科合作、不断推进的方式去实现。

7 结 语

科技创新是当代每个个体都需要关注的一个重要问题,大到从国家制定发展战略,保持持久的国际竞争优势,小到个体创造自身社会价值,跟上社会发展的潮流。创新与我们息息相关。本研究落脚于“科研创新”这一议题,系统梳理“新兴研究话题发现”相关的科学研究探索学者的科研创新的发现工作,并对这些研究从概念、方法、指标和研究难点等几个角度进行梳理和述评。同时,我们也注意到本研究的检索式也检出了较多不相关的文献,也可能存在一定的漏检。因此,我们利用人工阅读的方式对综述的文章加以剔除和补充。从综述结果来看,新兴研究话题发现具有重要的研究意义,它为提前识别有价值有意义的研究话题提供具体的方向和方法,为多方提供有价值的决策支持和导向。但是“新兴研究话题发现”本身却是一项极具研究难度的工作。特别当前的研究态势多变、新的研究对象和方法层出不穷,如何从浪里淘沙、沙里取金对现有的研究方法和工具都提出了重要的挑战。因此,需要引起学术界的关注。

参 考 文 献

- [1] 杨蕙馨,王军. 让创新驱动发展行稳致远[J]. 人民日报, 2018.
- [2] Goorha S, Ungar L. Discovery of Significant Emerging Trends [C]// Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2010: 57-64.
- [3] Tu Y N, Seng J L. Indices of novelty for emerging topic detection [J]. Information Processing & Management, 2012, 48(2): 303-325.
- [4] Chen C M. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature[J]. Journal of the American Society for Information Science and Technology, 2006, 57(3): 359-377.
- [5] Hansmann T, Niemeyer P. Big Data - Characterizing an Emerging Research Field Using Topic Models[C]// 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence and Intelli-

- gent Agent Technologies. IEEE, 2014, 1: 43-51.
- [6] Weismayer C, Pezenka I. Identifying emerging research fields: a longitudinal latent semantic keyword analysis[J]. *Scientometrics*, 2017, 113(3): 1757-1785.
- [7] Jarić I, Knezevic-Jaric J, Lenhardt M. Relative age of references as a tool to identify emerging research fields with an application to the field of ecology and environmental sciences[J]. *Scientometrics*, 2014, 100(2): 519-529.
- [8] An L, Lin X, Yu C, et al. Measuring and visualizing the contributions of Chinese and American LIS research institutions to emerging themes and salient themes[J]. *Scientometrics*, 2015, 105(3): 1605-1634.
- [9] Wang Q I. A bibliometric model for identifying emerging research topics[J]. *Journal of the Association for Information Science and Technology*, 2018, 69(2): 290-304.
- [10] Small H, Boyack K W, Klavans R. Identifying emerging topics in science and technology[J]. *Research Policy*, 2014, 43(8): 1450-1467.
- [11] Boyack K W, Klavans R. Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately?[J]. *Journal of the American Society for Information Science and Technology*, 2010, 61(12): 2389-2404.
- [12] Glänzel W, Thijs B. Using “core documents” for detecting and labelling new emerging topics[J]. *Scientometrics*, 2012, 91(2): 399-416.
- [13] Rotolo D, Hicks D, Martin B R. What is an emerging technology? [J]. *Research Policy*, 2015, 44(10): 1827-1843.
- [14] Fagerberg J, Verspagen B. Innovation studies - The emerging structure of a new scientific field[J]. *Research Policy*, 2009, 38(2): 218-233.
- [15] González-Alcaide G, Llorente P, Ramos J M. Bibliometric indicators to identify emerging research fields: publications on mass gatherings[J]. *Scientometrics*, 2016, 109(2): 1283-1298.
- [16] Yan E. Research dynamics: Measuring the continuity and popularity of research topics[J]. *Journal of Informetrics*, 2014, 8(1): 98-110.
- [17] Dotsika F, Watkins A. Identifying potentially disruptive trends by means of keyword network analysis[J]. *Technological Forecasting and Social Change*, 2017, 119: 114-127.
- [18] Goldstein J. Emergence as a construct: History and issues[J]. *Emergence*, 1999, 1(1): 49-72.
- [19] Ryan A J. Emergence is coupled to scope, not level[J]. *Complexity*, 2007, 13(2): 67-77.
- [20] Schultz L I, Joutz F L. Methods for identifying emerging General Purpose Technologies: a case study of nanotechnologies[J]. *Scientometrics*, 2010, 85(1): 155-170.
- [21] Meyer M, Persson O. Nanotechnology-interdisciplinarity, patterns of collaboration and differences in application[J]. *Scientometrics*, 1998, 42(2): 195-205.
- [22] Anegón F D, Contreras E J, Corrochano M D. Research fronts in library and information science in Spain (1985 - 1994)[J]. *Scientometrics*, 1998, 42(2): 229-246.
- [23] Astrom F. Changes in the LIS research front: Time-sliced cocitation analyses of LIS journal articles, 1990-2004[J]. *Journal of the American Society for Information Science and Technology*, 2007, 58(7): 947-957.
- [24] Toivanen H. The shift from theory to innovation: the evolution of Brazilian research frontiers 2005 - 2011[J]. *Technology Analysis & Strategic Management*, 2014, 26(1): 105-119.
- [25] Morris M, Herrmann O J. Beyond surveys: The research frontier moves to the use of administrative data to evaluate R&D grants [J]. *Research Evaluation*, 2013, 22(5): 298-306.
- [26] Price D J D S. Networks of scientific papers[J]. *Science*, 1965, 149(3683): 510-515.
- [27] Upham S P, Small H. Emerging research fronts in science and technology: patterns of new knowledge development[J]. *Scientometrics*, 2010, 83(1): 15-38.
- [28] Small H G. A co-citation model of a scientific specialty: A longitudinal study of collagen research[J]. *Social studies of science*, 1977, 7(2): 139-166.
- [29] Morris S A, Yen G, Wu Z, et al. Time line visualization of research fronts[J]. *Journal of the American Society for Information Science and Technology*, 2003, 54(5): 413-422.
- [30] Garfield E. Scientography: Mapping the tracks of science[J]. *Current Contents: Social & Behavioural Sciences*, 1994, 7(45): 5-10.
- [31] Huang M H, Chen S H, Lin C Y, et al. Exploring temporal relationships between scientific and technical fronts: a case of biotechnology field[J]. *Scientometrics*, 2014, 98(2): 1085-1100.
- [32] Larsen K. Knowledge network hubs and measures of research impact, science structure, and publication output in nanostructured solar cell research[J]. *Scientometrics*, 2008, 74(1): 123-142.
- [33] Huang M H, Chang C P. A comparative study on three citation windows for detecting research fronts[J]. *Scientometrics*, 2016, 109(3): 1835-1853.
- [34] Shibata N, Kajikawa Y, Takeda Y, et al. Detecting emerging research fronts based on topological measures in citation networks of scientific publications[J]. *Technovation*, 2008, 28(11): 758-775.
- [35] Liu C, Gui Q. Mapping intellectual structures and dynamics of transport geography research: a scientometric overview from 1982 to 2014[J]. *Scientometrics*, 2016, 109(1): 159-184.
- [36] Small H, Griffith B C. The structure of scientific literatures I: Identifying and graphing specialties[J]. *Science Studies*, 1974, 4(1): 17-40.
- [37] Ma V C, Liu J S. Exploring the research fronts and main paths of literature: a case study of shareholder activism research[J]. *Scient-*

- ometrics, 2016, 109(1): 33-52.
- [38] Li M. An exploration to visualise the emerging trends of technology foresight based on an improved technique of co-word analysis and relevant literature data of WoS[J]. *Technology Analysis & Strategic Management*, 2017, 29(6): 655-671.
- [39] Chen C, Dubin R, Kim M C. Emerging trends and new developments in regenerative medicine: a scientometric update (2000 - 2014) [J]. *Expert Opinion on Biological Therapy*, 2014, 14(9): 1295-1317.
- [40] Huang M H, Chang C P. Detecting research fronts in OLED field using bibliographic coupling with sliding window[J]. *Scientometrics*, 2014, 98(3): 1721-1744.
- [41] Shibata N, Kajikawa Y, Takeda Y, et al. Comparative study on methods of detecting research fronts using different types of citation[J]. *Journal of the American Society for Information Science and Technology*, 2009, 60(3): 571-580.
- [42] Liu X, Jiang T, Ma F. Collective dynamics in knowledge networks: Emerging trends analysis[J]. *Journal of Informetrics*, 2013, 7(2): 425-438.
- [43] Wagner C S, Leydesdorff L. Seismology as a dynamic, distributed area of scientific research[J]. *Scientometrics*, 2003, 58(1): 91-114.
- [44] Jensen S, Liu X, Yu Y, et al. Generation of topic evolution trees from heterogeneous bibliographic networks[J]. *Journal of Informetrics*, 2016, 10(2): 606-621.
- [45] Tseng Y H, Lin Y I, Lee Y Y, et al. A comparison of methods for detecting hot topics[J]. *Scientometrics*, 2009, 81(1): 73-90.
- [46] Chen K, Guan J. A bibliometric investigation of research performance in emerging nanobiopharmaceuticals[J]. *Journal of Informetrics*, 2011, 5(2): 233-247.
- [47] Aris A, Shneiderman B, Qazvinian V, et al. Visual overviews for discovering key papers and influences across research fronts[J]. *Journal of the American Society for Information Science and Technology*, 2009, 60(11): 2219-2228.
- [48] Natale F, Fiore G, Hofherr J. Mapping the research on aquaculture. A bibliometric analysis of aquaculture literature[J]. *Scientometrics*, 2012, 90(3): 983-999.
- [49] Milman B, Gavrilova Y. Analysis of citation and cocitation in chemical-engineering[J]. *Scientometrics*, 1993, 27(1): 53-74.
- [50] Batagelj V, Ferligoj A, Squazzoni F. The emergence of a field: a network analysis of research on peer review[J]. *Scientometrics*, 2017, 113(1): 503-532.
- [51] Shapira P, Kwon S, Youtie J. Tracking the emergence of synthetic biology[J]. *Scientometrics*, 2017, 112(3): 1439-1469.
- [52] Nederhof A J, Vanwijk E. Mapping the social and behavioral sciences world-wide: Use of maps in portfolio analysis of national research efforts[J]. *Scientometrics*, 1997, 40(2): 237-276.
- [53] 陈仕吉. 科学研究前沿探测方法综述[J]. *数据分析与知识发现*, 2009(9): 28-33.
- [54] 许晓阳, 郑彦宁, 赵筱媛, 等. 研究前沿识别方法的研究进展[J]. *情报理论与实践*, 2014, 37(6): 139-144.
- [55] Liu Z, Yin Y, Liu W, et al. Visualizing the intellectual structure and evolution of innovation systems research: a bibliometric analysis[J]. *Scientometrics*, 2015, 103(1): 135-158.
- [56] Robinson D K R, Ruivenkamp M, Rip A. Tracking the evolution of new and emerging S&T via statement-linkages: Vision assessment in molecular machines[J]. *Scientometrics*, 2007, 70(3): 831-858.
- [57] Mörchen F, Dejori M, Fradkin D, et al. Anticipating annotations and emerging trends in biomedical literature[C]// *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 2008: 954-962.
- [58] Yan E. Research dynamics, impact, and dissemination: A topic-level analysis[J]. *Journal of the Association for Information Science and Technology*, 2015, 66(11): 2357-2372.
- [59] Kleinberg J. Bursty and hierarchical structure in streams[J]. *Data Mining and Knowledge Discovery*, 2003, 7(4): 373-397.
- [60] Guo H, Weingart S, Börner K. Mixed-indicators model for identifying emerging research areas[J]. *Scientometrics*, 2011, 89(1): 421-435.
- [61] Li M, Chu Y. Explore the research front of a specific research theme based on a novel technique of enhanced co-word analysis [J]. *Journal of Information Science*, 2017, 43(6): 725-741.
- [62] Swar B, Khan G F. Mapping ICT knowledge infrastructure in South Asia[J]. *Scientometrics*, 2014, 99(1): 117-137.
- [63] Li X, Qiao H, Wang S. Exploring evolution and emerging trends in business model study: a co-citation analysis[J]. *Scientometrics*, 2017, 111(2): 869-887.
- [64] Zitt M, Bassecouard E. Development of a method for detection and trend analysis of research fronts built by lexical or cocitation analysis[J]. *Scientometrics*, 1994, 30(1): 333-351.
- [65] Schwechheimer H, Winterhager M. Mapping interdisciplinary research fronts in neuroscience: A bibliometric view to retrograde amnesia[J]. *Scientometrics*, 2001, 51(1): 311-318.
- [66] Schwechheimer H, Winterhager M. Highly dynamic specialities in climate research[J]. *Scientometrics*, 1999, 44(3): 547-560.
- [67] Kontostathis A, De I, Holzman L E, et al. Use of term clusters for emerging trend detection[J]. Preprint, 2004.
- [68] 郑彦宁, 许晓阳, 刘志辉. 基于关键词共现的研究前沿识别方法研究[J]. *图书情报工作*, 2016, 60(4): 85-92.
- [69] 潘玮, 牟冬梅, 李茵, 等. 关键词共现方法识别领域研究热点过程中的数据清洗方法[J]. *图书情报工作*, 2017, 61(7): 111-117.
- [70] 许晓阳, 郑彦宁, 刘志辉. 论文和专利相结合的研究前沿识别方法研究[J]. *图书情报工作*, 2016, 60(24): 97-106.
- [71] An X Y, Wu Q Q. Co-word analysis of the trends in stem cells field based on subject heading weighting[J]. *Scientometrics*,

- 2011, 88(1): 133-144.
- [72] Hu C P, Hu J M, Deng S L, et al. A co-word analysis of library and information science in China[J]. *Scientometrics*, 2013, 97(2): 369-382.
- [73] Zhao D, Strotmann A. The knowledge base and research front of information science 2006 - 2010: An author cocitation and bibliographic coupling analysis[J]. *Journal of the Association for Information Science and Technology*, 2014, 65(5): 995-1006.
- [74] Zhao D, Strotmann A. Evolution of research activities and intellectual influences in information science 1996 - 2005: Introducing author bibliographic-coupling analysis[J]. *Journal of the American Society for Information Science and Technology*, 2008, 59(13): 2070-2086.
- [75] Ma R. Author bibliographic coupling analysis: A test based on a Chinese academic database[J]. *Journal of Informetrics*, 2012, 6(4): 532-542.
- [76] Persson O. The intellectual base and research fronts of JASIS 1986 - 1990[J]. *Journal of the American Society for Information Science*, 1994, 45(1): 31-38.
- [77] Bettencourt L M A, Kaiser D I, Kaur J. Scientific discovery and topological transitions in collaboration networks[J]. *Journal of Informetrics*, 2009, 3(3): 210-221.
- [78] Vladutz G, Cook J. Bibliographic coupling and subject relatedness[J]. *Proceedings of the American Society for Information Science*, 1984, 21: 204-207.
- [79] Jarneving B. Bibliographic coupling and its application to research-front and other core documents[J]. *Journal of Informetrics*, 2007, 1(4): 287-307.
- [80] Bonaccorsi A, Thoma G. Institutional complementarity and inventive performance in nano science and technology[J]. *Research Policy*, 2007, 36(6): 813-831.
- [81] Mund C, Neuhaeusler P. Towards an early-stage identification of emerging topics in science - The usability of bibliometric characteristics[J]. *Journal of Informetrics*, 2015, 9(4): 1018-1033.
- [82] Roche I, Besagni D, Francois C, et al. Identification and characterisation of technological topics in the field of Molecular Biology [J]. *Scientometrics*, 2010, 82(3): 663-676.
- [83] Schiebel E, Hörlesberger M, Roche I, et al. An advanced diffusion model to identify emergent research issues: the case of optoelectronic devices[J]. *Scientometrics*, 2010, 83(3): 765-781.
- [84] Ohniwa R L, Hibino A, Takeyasu K. Trends in research foci in life science fields over the last 30 years monitored by emerging topics[J]. *Scientometrics*, 2010, 85(1): 111-127.
- [85] Trajtenberg M, Henderson R, Jaffe A. University versus corporate patents: A window on the basicness of invention[J]. *Economics of Innovation and New Technology*, 1997, 5(1): 19-50.
- [86] Goldman A W. Conceptualizing the interdisciplinary diffusion and evolution of emerging fields: The case of systems biology[J]. *Journal of Informetrics*, 2014, 8(1): 43-58.
- [87] Bettencourt L M A, Kaiser D I, Kaur J, et al. Population modeling of the emergence and development of scientific fields[J]. *Scientometrics*, 2008, 75(3): 495-518.
- [88] Comins J A, Leydesdorff L. RPYS i/o: software demonstration of a web-based tool for the historiography and visualization of citation classics, sleeping beauties and research fronts[J]. *Scientometrics*, 2016, 107(3): 1509-1517.
- [89] Lucio-Arias D, Leydesdorff L. An indicator of research front activity: Measuring intellectual organization as uncertainty reduction in document sets[J]. *Journal of the American Society for Information Science and Technology*, 2009, 60(12): 2488-2498.
- [90] Lee W H. How to identify emerging research fields using scientometrics: An example in the field of Information Security[J]. *Scientometrics*, 2008, 76(3): 503-525.
- [91] 张丽华, 曲建升. 期刊编委比非编委论文作者能更早探测出研究前沿吗[J]. *情报杂志*, 2017, 36(8): 113-119.
- [92] Kim M C, Chen C. A scientometric review of emerging trends and new developments in recommendation systems[J]. *Scientometrics*, 2015, 104(1): 239-263.
- [93] Takeda Y, Kajikawa Y. Optics: a bibliometric approach to detect emerging research domains and intellectual bases[J]. *Scientometrics*, 2009, 78(3): 543-558.
- [94] Chang Y W, Huang M H, Lin C W. Evolution of research subjects in library and information science based on keyword, bibliographical coupling, and co-citation analyses[J]. *Scientometrics*, 2015, 105(3): 2071-2087.
- [95] Rohrbeck R, Battistella C, Huizingh E. Corporate foresight: An emerging field with a rich tradition[J]. *Technological Forecasting and Social Change*, 2015, 101: 1-9.
- [96] Laurens P, Zitt M, Bassecouard E. Delineation of the genomics field by hybrid citation-lexical methods: interaction with experts and validation process[J]. *Scientometrics*, 2010, 82(3): 647-662.
- [97] 付鑫金, 方曙, 庞弘燊. 基于共词分析的我国情报学博士学位论文研究热点分析[J]. *情报科学*, 2011, 29(11): 1722-1725.
- [98] 魏群义, 侯桂楠, 霍然. 近 10 年国内情报学硕士学位论文研究热点统计分析[J]. *图书情报工作*, 2012, 56(2): 35-81.
- [99] Freyne J, Coyle L, Smyth B, et al. Relative status of journal and conference publications in computer science[J]. *Communications of the ACM*, 2010, 53(11): 124-132.
- [100] Zhao D, Strotmann A. Can citation analysis of Web publications better detect research fronts?[J]. *Journal of the American Society for Information Science and Technology*, 2007, 58(9): 1285-1302.
- [101] Glassey O. Exploring the weak signals of starts-ups as a folksonomic system[J]. *Technology Analysis & Strategic Management*, 2009, 21(3): 321-332.
- [102] 王效岳, 刘自强, 白如江, 等. 基于基金项目数据的研究前沿主题探测方法[J]. *图书情报工作*, 2017, 61(13): 87-98.

- [103] Lu C, Ding Y, Zhang C. Understanding the impact change of a highly cited article: a content-based citation analysis[J]. *Scientometrics*, 2017, 112(2): 927-945.
- [104] Huang M, Chang C P. A comparative study on detecting research fronts in the organic light-emitting diode (OLED) field using bibliographic coupling and co-citation[J]. *Scientometrics*, 2015, 102(3): 2041-2057.
- [105] Miguel S, Moya-Anegon F, Herrero-Solana V. New approach to institutional domain analysis: Multilevel research fronts structure [J]. *Scientometrics*, 2008, 74(3): 331-344.
- [106] Kuhn T S. *The structure of scientific revolutions*[M]. Chicago: University of Chicago Press, 1963: 2.
- [107] Kreuchauff F, Korzinov V. A patent search strategy based on machine learning for the emerging field of service robotics[J]. *Scientometrics*, 2017, 111(2): 743-772.
- [108] Wang X, Cheng Q, Lu W. Analyzing evolution of research topics with NEViewer: a new method based on dynamic co-word networks[J]. *Scientometrics*, 2014, 101(2): 1253-1271.

附录

1. TS= ("emerg* topic*" OR "emerg* research* topic*" OR "emerg* scien* topic*" OR "emerging* academic* topic*")
2. OR "emergence of* topic*" OR "emergence of* scien* topic*" OR "emergence of* research* topic*" OR "emergence of* academic* topic*"
3. OR "topic* emergence" OR "research* topic* emergence" OR "scien* topic* emergence" OR "academic* topic* emergence"
4. OR "research* front*" OR "scien* front*" OR "academic* front*"
 5. OR "emerg* field*" OR "emerg* scien* field*" OR "emerg* research field*" OR "emerg* academic* field*"
 6. OR "emergence of* field*" OR "emergence of* scien* field*" OR "emergence of* research field*" OR "emergence of* academic* field*"
 7. OR "field* emergence" OR "research* field* emergence" OR "scien* field* emergence" OR "academic* field* emergence"
 8. OR "emerg* trend*" OR "emerg* research* trend*" OR "emerg* scien* trend*" OR "emerg* academic* trend*"
 9. OR "emergence of* trend*" OR "emergence of* research* trend*" OR "emergence of* scien* trend*" OR "emergence of* academic* trend*"
 10. OR "trend* emergence" OR "research* trend* emergence" OR "scien* trend* emergence" OR "academic* trend* emergence"
 11.) AND SO=(
 12. SCIENTOMETRICS
 13. OR TECHNOLOGICAL FORECASTING AND SOCIAL CHANGE
 14. OR TECHNOLOGY ANALYSIS STRATEGIC MANAGEMENT
 15. OR JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY
 16. OR RESEARCH POLICY
 17. OR JOURNAL OF INFORMETRICS
 18. OR JOURNAL OF THE ASSOCIATION FOR INFORMATION SCIENCE AND TECHNOLOGY
 19. OR RESEARCH EVALUATION)

(责任编辑 魏瑞斌)