

基于多元弱关系融合的科学突破主题 早期识别研究

刘亚辉^{1,2}, 许海云^{3,4}, 武华维⁵, 刘春江⁶, 王海燕⁴

(1. 中国科学院文献情报中心, 北京 100190; 2. 中国科学院大学经济与管理学院图书情报与档案管理系, 北京 100190; 3. 山东理工大学管理学院, 淄博 255000; 4. 中国科学技术信息研究所, 北京 100038; 5. 西北师范大学档案馆, 兰州 730070; 6. 中国科学院成都文献情报中心, 成都 610041)

摘 要 科学突破是变革程度较高、对领域未来发展方向和趋势影响较为深远的科学研究, 旨在揭示早期尚未被认识到或重视起来的新现象、新规律。较早阶段识别科学突破主题能够为政策制定者和基金资助机构提供决策支持, 有助于优化资源配置。本文选取基因工程疫苗领域展开实证研究, 重点关注知识网络中的弱关联关系, 构建基于主题词共现-作者合著-参考文献同被引等关系的多层网络, 通过综合分析多个特征项之间的关联信息挖掘主题内容, 并借助专家判断和权威报告评估识别效果。与基于强关联关系的识别结果的对比分析, 验证了本文构建方法适用于科学突破主题的早期识别。在未来研究中可以借鉴网络表示学习, 引入时效网络精准定位到某项突破从产生到消退的时间点。

关键词 弱关系; 科学突破; 突破性创新; 多元关系融合; 早期识别

Scientific Breakthrough Topics Identification in an Early Stage Using Multiple Weak Linkage Fusion

Liu Yahui^{1,2}, Xu Haiyun^{3,4}, Wu Huawei⁵, Liu Chunjiang⁶ and Wang Haiyan⁴

(1. National Science Library, Chinese Academy of Sciences, Beijing 100190; 2. Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190; 3. Business School, Shandong University of Technology, Zibo 255000; 4. Institute of Scientific and Technical Information of China, Beijing 100038; 5. Archives of Northwest Normal University, Lanzhou 730070; 6. Chengdu Documentation and Information Center, Chinese Academy of Sciences, Chengdu 610041)

Abstract: By the term “scientific breakthroughs,” we mean innovations that are relatively transformative and have a profound impact on the future direction and trends of a disciplinary field. A scientific breakthrough aims to reveal new phenomena and laws that have emerged earlier but have not yet been recognized or investigated formally. Identifying scientific breakthroughs at an earlier stage can provide decision support to policymakers and granting agencies in optimizing resource allocation. This study selected the field of Gene Engineered Vaccine for empirical research—focusing on weak association linkages in knowledge networks—and constructed a multi-layer network based on subject term co-occurrence, author co-authorship, and reference co-citation relationships. Thereafter, we analyzed the association information between

收稿日期: 2021-11-10; 修回日期: 2022-06-05

基金项目: 国家自然科学基金项目“基于弱信号时效网络演化分析的变革性科技创新主题早期识别方法研究”(72274113); 国家重点研发计划项目课题“颠覆性技术地平线扫描系统”(2019YFA0707202-01); 山东省泰山学者工程项目(202103069)。

作者简介: 刘亚辉, 女, 1996年生, 博士研究生, 研究方向为情报计量学的理论与实践; 许海云, 女, 1982年生, 博士, 教授, 研究方向为情报计量学的理论与实践, E-mail: xuhaiyunnemo@gmail.com; 武华维, 男, 1985年生, 博士, 副教授, 研究方向为情报研究方法与技术; 刘春江, 男, 1984年生, 博士研究生, 副研究员, 研究方向为专利技术挖掘研究; 王海燕, 女, 1974年生, 博士, 副研究员, 研究方向为科学计量学。

multiple feature items to mine the subject content and evaluated the recognition effect with the help of expert judgment and authoritative reports. The comparative analysis with the identification results based on strong correlation relationships verified that the method constructed in this study is applicable to the early identification of scientific breakthroughs. Future research could draw on network representation learning and introduce temporal networks to pinpoint the moment when a breakthrough is made to fade away.

Key words: weak linkage; scientific breakthrough; breakthrough innovation; multi-relations fusion; early recognition

0 引言

21世纪以来,全球科技创新进入空前密集活跃的时期,重大的科学技术不断涌现,科技成果越来越多地应用于生产和实践。忽视前瞻性预测可能错失发展机遇,因此多数创新型国家、知名研究机构及科技智库竞相开展变革性科技创新主题的识别和预判^[1-4],争做世界主要科学中心和创新高地。我国在2020年11月发布了《中共中央关于制定国民经济和社会发展第十四个五年规划和二〇三五年远景目标的建议》^[5],强调创新在现代化建设全局中的核心地位,把科技自立自强作为国家发展的战略支撑,强调加强基础研究,注重原始创新。从各国科技战略部署来看,全球都将科技创新尤其是能带来根本性变革的一类突破性(原始)创新放置在非常重要的位置。此类创新常常导致学科范式的转变,将科学研究中的重大突破与产业技术上的技术革新相结合,挖掘最新科研成果的价值,为经济发展提供源动力。因此,如何在突破性创新萌芽时期关注其早期迹象,预测其未来的变革潜力,成为科技管理中至关重要又极具挑战的问题。

1 研究现状

科学突破主题的识别及预判是情报计量分析的重要工作。突破性创新是一个范围相对较广的概念,强调对未来发展产生“重大影响”,被认为是转折点的信号。因此,突破性创新的研究不仅涵盖科学基础研究中的突破性进展、重大发现,还包括技术改进、产品性能及市场价值的提升。本文重点关注科学基础研究中的突破性创新,聚焦于知识发现带来的科学进程中的突破性进展,对突破性主题进行早期识别,是后续在技术和产业等层面的创造性实践的先导。

1.1 科学突破的内涵与特征

Kuhn^[6]认为科学研究进展过程存在不同价值的科学发现,首先提出将科学研究分为“常规”科学

和“革命性”科学,这两种状态相互交替向前发展,科学发现常伴随着科学范式的转变。之后很多研究人员从突破性研究挑战现有的科学范式出发,论述突破性创新可能带来全新的理论或者方法,为当下科学理论无法解释的事实提供理论基础,为搁置已久的难题提供新的解决思路^[7-9]。此外,有研究人员进行了更深层次的思考,认为突破性创新可能会彻底改变某领域的知识结构,继而延伸出新的领域,领域知识结构的变化的观察、识别早期阶段的科学突破提供独特视角^[10-11]。

但也有研究人员并不认同“常规”科学和“革命性”科学的分类,认为科学发展不会存在界限分明的范式转变。例如,Galison^[12]运用“贸易区”类比科学系统,认为不同的物质文化可以在区域内交流共存,不必相互取代,通过粒子物理学的案例研究证明引起科学进展的一些小的突破性创新并没有引起范式的转变。同样地,van Raan^[13]将科学描述为一个自组织的“认知生态系统”,所有科学进展都是革命性的,根据复杂系统的典型统计,大多是微小的创新,科学突破数量很少。

实际上,多数研究发现科学突破并非凭空产生,而是由多个小的突破累积形成,最终促使研究范式革新^[14]。在这个量变到质变的过程中,突破性创新通常不会一次全部完成,因而短时间内并未产生明显的范式转变^[15],而是存在诸多新旧范式转化需要经历的“中间状态”^[16]。Hollingsworth^[17]和De Bellis^[18]等学者同样认为科学突破是一个过程,不是一个偶然成果,而是通过无数持续微小、渐进性进步累积而成的思考问题的新方式。因此,科学突破的产生归根结底是渐进、连续、有迹可循的,也是可以被早期识别的,关键在于如何有效捕捉早期碎片化的弱信号^[19]。

通过上述概念梳理可知,科学突破的特征较为复杂,在学界尚未形成一致和全面的解读。本文将研究人员对科学突破的特征描述总结如下。①长期性:科学突破需要相对漫长的积累过程,需要科学自身、社会发展、学术思想及科研人员专业知识等

多方面的累积^[20-21]；②新颖性：带来新的科学原理或新的解决方案的科学突破具有原创性和前瞻性^[22-23]；③变革性：在一段时间内开创新的研究方向，往往导致与之相关的创新簇群或知识生产的“连锁反应”^[24]；④不确定性：科学突破发展过程中存在众多偶然性因素，高度不确定性导致预测难度较大^[16,25]；⑤不连续性：科学发展进程具有非线性

特征，科学突破的产生可能会引起不连续的断层现象^[26]。

1.2 科学突破主题识别方法

研究人员对突破性研究的主题识别进行了探索，从不同的角度和切入点出发，有不同的识别方法，可以总结为表 1。

表 1 科学突破主题识别方法

方法	说明	优势	局限	代表性研究
专家定性评价法	借助专家丰富的学识和经验识别并研判未来科技发展态势	专家有较强的专业知识和丰富的经验,适用于对单一学科突破性进展的判断	主观性强,面对非共识性研究会失效;融合研究背景下,耗时耗力成本高	《麻省理工科技评论》年度十大突破性技术;Science 年度十大科学突破
主题突变监测法	关注主题在某一时刻的突变程度以及离群数据,分析时序数据中突发词的变化	能够实时跟踪主题演化进程,增强了分析的时效性,可以识别非连续式创新主题	忽略文献之间和主题词之间的语义关联,不能体现学科领域的整体知识结构	Kleinberg ^[27] ; Yoon 等 ^[28] ; 张金柱等 ^[29]
知识网络结构变化测度法	测度知识网络结构的变化程度,重点关注引起结构显著变化的知识节点	从整体知识网络的视角出发分析主题的演化过程,可以推进早期识别	知识结构变化只能在一定程度上表征研究主题的演变,不能等同于突破创新	Chen 等 ^[30] ; 罗瑞 ^[31] ; Min 等 ^[11]
睡美人文献分析法	关注睡美人文献唤醒机制,挖掘此类文献中可能孕育的潜在突破性研究	以学术价值的延迟承认作为切入点,将突破性创新成果被认可的阶段前置	睡美人文献数量不多,绝大多数科学突破会很快被认可,因此该方法缺少覆盖面	杜建 ^[32] ; van Raan 等 ^[33]
技术演化视角分析法	参照技术轨道的转换论:在新的技术轨道未成型时,前瞻预判其转换期 ^[34]	科学突破常出现在新旧知识体系的更迭期,该视角识别结果更为准确	不同科技创新领域其突破性创新的发展模式不同,普适性的指标模型难获取	Sood 等 ^[35] ; Tellis ^[36] ; 杨国忠等 ^[37]
基于深度学习的语义分析方法	借助神经网络及深度学习等算法模型,从多源数据的分析中预测主题发展趋势,进而捕获突破性主题	能够处理面向自然语言领域的大数据,更擅长从早期碎片化信息中挖掘潜藏的特征和模式	缺乏垂直行业领域的标准实验数据集及评价方法,方法的有效性无法得到充分验证,离不开领域专家的参与	徐路路等 ^[38] ; 刘博文等 ^[39] ; Liang 等 ^[40]

综上,已有研究通过定性和定量识别方法探讨了科学突破的特征,并针对这些特征展开了专家咨询及定量识别。但大部分是回顾性或评估性研究,多基于高频词、强关联等强信号进行分析,识别结果通常是快速发展期甚至是成熟期的热点主题而非新兴前沿,前瞻性不足。突破性创新早期会以主题间微弱关联的形式存在,但弱关联关系目前仅作为一个附加关注对象,充当简单的辅助监测手段,未被充分挖掘利用,更缺乏明确地从弱关联关系分析的视角识别突破性创新的研究。

1.3 弱关联关系相关研究

在科学发展进程中,随着研究的持续推进会不断有线索和迹象显露出来,经过积累、序化,最终形成具有明确指向性的信息,这个过程一般时间跨度较长^[41]。因此,前瞻预判创新主题的未来发展趋势,需要研究人员主动地发掘、筛选“早期迹象”。

如何能在突破性主题的早期就预见其未来路径,需要借助早期的弱关联关系。弱关系携带更多元更丰富的信息,蕴含学科领域的未来发展趋势。可以对不同类型的弱关联关系进行捕获、分析并加以利用,挖掘更多有用的信息,为科学突破的早期识别提供可能性(图 1)。



图 1 弱关系的筛选利用过程

1.3.1 弱关系的类型

“弱关系”研究起源于社会学, Granovetter^[42]

指出弱关系常常作为连通不同团体间的桥梁,传递异质信息。“弱关系”与“强关系”相对应,是知识网络中节点之间的链接强度低于某一阈值的关系,其结构相比于强关系较为稀疏,但数量远多于强关联,对该类数据的忽略容易导致重要信息的遗漏^[19]。复杂网络中的弱关联具有跨学科和技术整合的早期特征,隐含领域的未来发展趋势^[43]。知识网络中的弱关联关系通常有如下三类:①低阈值关联关系类型。如图2a所示,外层网络中节点关系相较于内层网络,关联强度低,如共现频次较低的关键词之间是弱关联关系。②子网间关系节点。如图2b

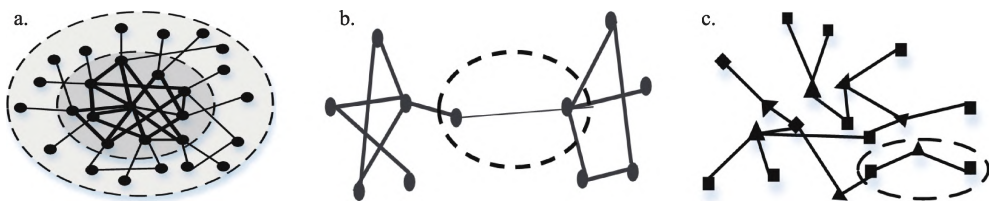


图2 知识网络中的弱关系

1.3.2 弱关系分析的应用

目前,研究人员已尝试运用弱关联来分析识别领域前沿与学科交叉主题。张英杰^[46]从主题词的低共现特征维度对科学前沿进行探测,从具有弱关系的凝聚子群、主题关键词网络中的弱桥等方面识别前沿主题。Wei等^[47]构建高频词弱共现矩阵和弱共现网络,进一步形成弱关系网络,实现对情报学的交叉前沿的识别。Yoo等^[48]重点关注未被选为文章关键词的词间弱关联关系,运用基于代理的模拟工具NetLogo探索纳米材料领域的技术创新动态行为,识别该领域的未来技术趋势。Griol-Barres等^[49]运用文本挖掘和自然语言处理技术“扫描”多源文本信息,捕获术语间的弱关联关系,为企业和政府提供决策支撑;在遥感领域的实证分析结果表明,该方法能够在较早阶段探测关键技术和领域新兴研究方向。牌艳欣等^[45]基于跨学科子网间、间接联系知识节点的弱引文关系,识别跨学科相关知识组合。岳增慧等^[50]研究发现,强连接与弱关联对学科引证中的知识扩散具有不同影响,如果能掌握强弱关联微观结构的演化规律,就可以更有效地预测网络的演化方向。上述研究表明,网络中节点间关系的强弱变化会传达出知识演化发展的信号,关注分析知识网络中的弱关联关系可以推进科学突破主题的早期识别。

所示,在社区识别结果中,子网内部节点关系紧密,子网间节点关联相对稀疏,形成弱关系^[44],如引用网络中跨学科的引用关系。③间接关联关系节点。如图2c所示,通过其他节点建立间接联系的间接关联,具有弱关联关系,比如,两位作者通过一篇引文建立联系^[45]。

本文将借助图2a和图2c两种类型的弱关系,即科学突破主题的识别除了关注主题词的直接弱共现关系,还需要融合基于作者、引文等中间实体产生的间接关联关系,获取更为全面的信息,增强早期主题语义关联。

2 基于弱关系的科学突破主题早期识别方法构建

2.1 研究路线

本文重点关注知识网络中的弱关联关系,以主题词共现关系为主,融合作者合著和同被引关系,尝试对科学突破主题进行早期识别。图3为本文的分析流程,下文将详细阐述不同模块的分析内容与分析方法。

首先,抽取主题词共现、作者合著、引文共现、作者-主题词共现和主题词-引文共现等5类关系,设置阈值构建基于多种弱关系的多层知识网络;之后,选取适合的多元关系融合算法对网络进行融合聚类,获取主题作为判断候选项;接着,借助专家和权威报告筛选出科学突破主题;最后,对比基于强关联关系识别出的主题,验证本文构建的方法是否做到了早期识别。

2.2 多元弱关系融合聚类

目前应用较为广泛的是线性融合、加权投影融合、非负矩阵分解融合、超网络融合及基于元路径的融合方法。相比于其他方法,基于元路径的融合方法能够表达出多元关系网络中更为丰富的语义信息,且不需要为选中的每种关系进行人为赋值。有研究证明,PathSelClus算法能够实现多元关系融合

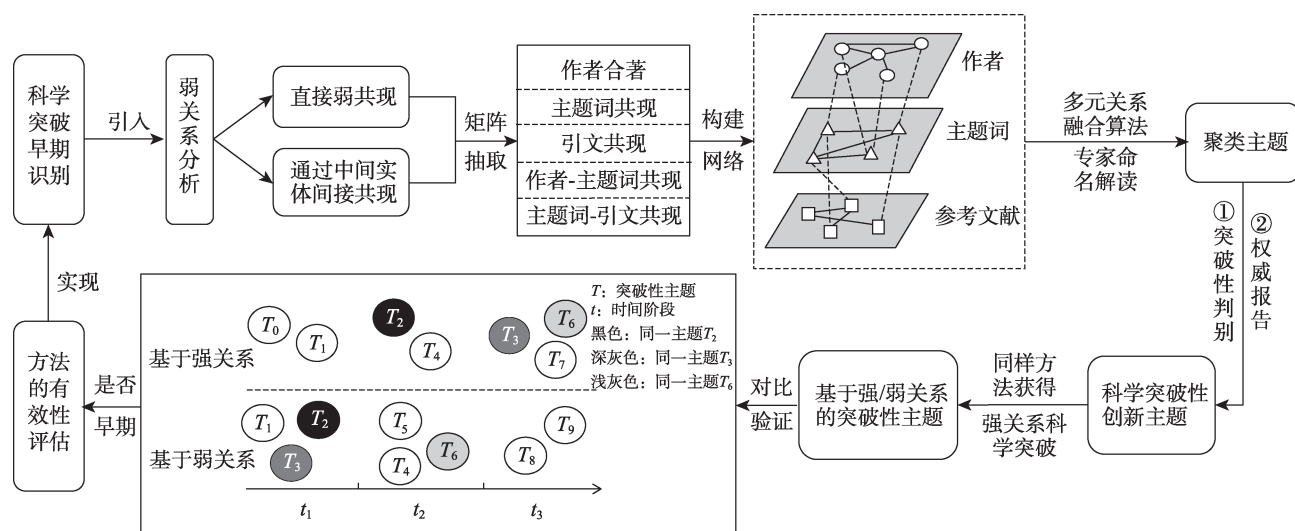


图3 分析流程

的主题聚类,可以有效提高文本主题聚类效果^[51]。因此,本文选取该算法对弱关系进行融合聚类。

PathSelClus 算法是运用概率图模型的方法定义问题，完整的聚类过程主要包含元路径的生成、标注种子对象引导聚类效果和确定元路径关系矩阵在聚类结果中的权重^[52]。该聚类方法通过不同类簇概率和每条元路径对聚类结果影响（即元路径权重）的迭代优化，在两类数值达到稳定状态时聚类任务完成，最终输出聚类结果以及相应每条元路径的权重^[53]。

对经过数据预处理导出的每一种关系矩阵分别设定阈值进行拆分, 得到强共现矩阵和弱共现矩阵两种, 之后分别利用 PathSelClus 算法对这两类关系矩阵进行融合聚类。

(1) 算法所需的输入信息为: ①聚类的目标类型 T 。 T 是需要聚类的顶点类型, 与该顶点连接的其他对象为特征类型。②聚簇的个数 K 和每个聚簇的对象种子 N_1, N_2, \dots, N_K 。 N_K 表示聚簇 K 的对象种子, 可以为空集, 即不提供聚类参照。③从目标类型 T 出发, 经过特征类型的 M 条元路径的集合。

此外，还需要计算元路径关系邻接矩阵，计算方法是将原路径中每一对顶点表示的关系矩阵依次相乘。

(2) 算法的输出信息包括两个结果：①每条元路径参与聚类的最佳权重；②基于强关系的聚类主题和基于弱关系的聚类主题。方便后续主题的解读和方法有效性的验证。

2.3 聚类结果比较分析

将融合各类关系聚类后获取的每一时段基于强

关系的聚类主题与基于弱关系的聚类主题进行对比分析, 评估方法的可行性。具体过程如下。

(1) 主题命名：在专家的指导下对聚类获取的主题簇进行规范化命名。

(2) 突破性判别：①从科学原理上来看，科学突破常建立在全新的知识理论或原理之上，攻克现存的难题或瓶颈；②从科学影响力上来看，科学突破通常指对学科领域产生重大影响的方法或作用机制等，促进学科发展甚至开启对新的分支的研究；③从未来前景上来看，科学突破受到科研机构及研发公司的关注，未来很可能被广泛应用并有巨大发展前景。

(3) 确定科学突破：借助权威报告和新闻报道，将经过突破性判别的主题与正式宣布的科学突破进行比较，最终确定科学突破。分辨利用弱关系比利用强关系能否在更早期识别突破性创新主题，以此证明有无关弱关系的必要性。

经过以上分析后,将基于强关系获取的科学突破和基于弱关系的科学突破分别按照时间阶段整理,若同一时段基于弱关系能比基于强关系识别出更多的突破性创新主题,或者利用弱关系能在较早阶段识别出正式宣布的科学突破,则可以此验证关注弱关系的必要性。

3 基因工程疫苗领域实证研究

3.1 实证领域选择

2020年传染性疾病新型冠状病毒肺炎发生并在全世界范围肆虐，引发此次疫情的新型冠状病

性强,传播速度快,致死率高。我国在“十四五”规划中明确指出,要从国家急需需要和长远需求出发,集中优势资源攻关新发突发传染病和生物安全风险防控^[5],创新疫苗和抗体药物等的研发制备被提到了重要位置。

基因工程疫苗 (gene engineered vaccine, GEV) 是现代生物工程新型疫苗的重要组成部分,研发时不依赖于病原体的培养,将受体体内无效的病原体 and 容易致病的成分去除,可以最高限度保证疫苗的

安全性^[54]。同时还具有成本低,可实现大规模生产的优势。基因工程疫苗至今已经具有一定规模的文献积累,这些文献大多数以规范化的形式存储在数据库中可供下载,考虑到发展前景及数据的可获取性,本文选择基因工程疫苗领域作为实证分析的对象。文献检索策略如表 2 所示。

最后获取 5453 条记录,初步清洗去除无作者署名的、撤稿的和重复的文献记录之后,共得到论文 4903 篇。

表 2 实证分析论文检索策略

数据库选取	Web of Science Core Collection
检索式构造	TI=((Genetic* adj engineer* or DNA adj engineer*) and (vaccine* or antigen*)) or TS=((Genetic* adj engineer* or DNA adj engineer*) and (vaccine* or antigen*)) or TI=((nucleic* adj acid* or RNA) and (vaccine* or antigen*)) or TI=((plasmid* adj DNA) and (vaccine* or antigen*)) not TI=(test or immunoassay adj detect* or detect*)
文献类型限定	Article、Proceedings Paper
时间窗口	出版年至 2020 年 12 月
导出格式	纯文本文件(全记录与引用的参考文献)

3.2 数据预处理

Derwent Data Analyzer (DDA) 是一款功能强大的文本分析类工具,可以对科技文献进行高效的文本分析。本文利用 DDA 软件,运用主题词簇方法^[55] (term clumping),对主题词、作者、引文等字段进行提取与预处理,为后续主题分析提供更佳的字段列表。

将初步筛选获得的 4903 篇文献数据导入 DDA 软件,分别对主题词、作者和引文等字段进行清洗,具体过程如表 3 所示。

清洗结束后,从 DDA 软件中导出各类关系每年的共现矩阵。以主题词共现矩阵为例,行字段和列字段都选择主题词,交叉字段选择出版年,最终获取 1991—2020 年这 30 年每一年的主题词共现频

表 3 数据预处理过程

字段类型	步骤	具体操作
主题词	主题词抽取	选择 keywords+phrase 字段中的 multi-word phrases 列表,作为主题词字段,共有 129640 个短语和词
	去停用词	采取 field—thesaurus—stopwords (general).the 的方法操作
	字形清洗	运用 field—list cleanup—general.fuz 来进行词形规范化处理,包括规范大小写和自动合并单复数
	人工清洗	逐条对应清理,去除如 previous study、first time、wide range、24 hours、6 months、key role 等表示时间及对后续主题识别来说没有意义的词;合并如 heterologous genes 与 foreign genes 等的异形同义词,以及 humam immunodeficiency virus 与 HIV 等全称及缩写词
作者	设置阈值	考虑到后续共现矩阵及共词网络的稀疏情况,设立阈值,将词频≥12 的主题词筛选出来,最终共得到 851 个主题词
	初步合并	抽取作者全称,运用 field—list cleanup—general.fuz 进行列表清理,获取初步合并的集合。该步骤匹配规则是系统自带的 50% 模糊匹配
	人工清洗	逐条验证。①外国人姓名:一般为三小截构成。若全部一致,则认为是同一作者;若出现类似于 Weiner, David B、Weiner, D B 和 Weiner, DB 的情况,则将对应的作者机构、通信地址及 ORCID 标识符等信息进行清洗。②中韩等亚洲作者常有同名或简称相同的情况,如 Wang, Yi、Wang Yi、Wang Y 等,需要依据机构地址、合著者、通信邮箱等进行验证
	设置阈值	合著网络不能太过稀疏,选取发文量≥4 的 652 位作者
引文	初步合并	选择引用的参考文献字段,运用 field—list cleanup—general.fuz 进行列表清理
	人工清洗	针对没有 DOI (digital object unique identifier) 号的引文,借鉴 CiteSpace 软件中的处理方式,选取参考文献中作者+期刊+出版年份信息进行比对,若三者一致,则认为是同一篇文献
	设置阈值	考虑到共被引矩阵的分布情况,选取引用频次≥12 的 642 条引文记录

次情况。由于后续需要分别利用高频强共现关系与高频弱共现关系进行聚类，将结果进行对比以验证弱关系能否在早期识别科学突破主题，因此需要设定阈值将每一种共现矩阵拆分为强共现矩阵和弱共现矩阵。由于每一年的共现频次区分度不明显，所以考虑将每 3 年作为一个时段进行分析，即将共现的节点对按每 3 年进行共现频次合并加总，获取每个时段的共现情况。以主题词共现情况为例，如表 4 所示。

跟高低频关键词的分布情况类似，每个类型的矩阵共现频次为 1 的节点对约占总节点对的 60%~80%，符合二八法则，80% 的关系属于尾部效应^[56]，包含大量噪声数据，因此在这里暂时不考虑共现频次为 1 的关系。设定阈值将每个时段中共现频次≥2 的节点对拆分为强关系和弱关系矩阵，以主题词共现关系为例，阈值设定的原则为每一个时段每种关系的共现网络中节点对为 150~200 个。

依据网络的分布情况，本文选择 5 条元路径表示节点之间不同的语义关联关系，具体如图 4

表 4 每个时段的主题词共现

term1	term2	count	pub_year
genetic engineering	fusion protein	7	1991-1993
genetic engineering	Escherichia coli	6	1991-1993
hepatitis delta virus (HDV)	genome replication	4	1991-1993
Escherichia coli	antibody fragments	4	1991-1993
genetic engineering	T cells	3	1991-1993
...
T cells	chimeric antigen receptor	28	2018-2020
chimeric antigen receptor	CAR T cells	16	2018-2020
oral vaccination	mucosal immunity	15	2018-2020
dendritic cells	infectious diseases	13	2018-2020

所示。

在图 4 中，①主题词直接共现关系：当两个主题词出现在同一篇文献中，产生共现关系。元路径表示为 term1-term2，这是最主要的语义关系。②基于作者的主题词强化关系：两个主题词没有直接共现关系，被同一作者使用后产生联系。元路径表示为 term1-author-term2。③基于作者的主题词新增关

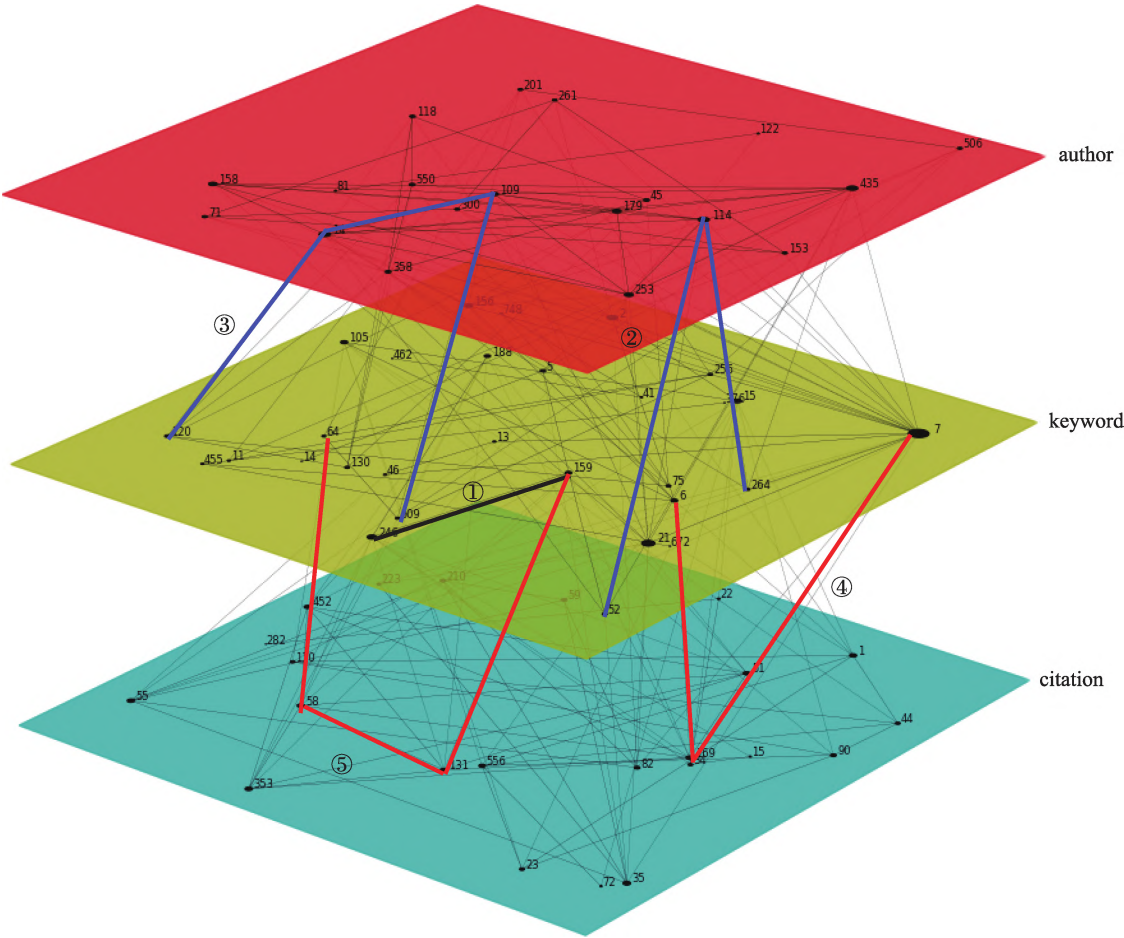


图 4 1994—1996 年时段的弱关系网络

系：两个主题词不在同一篇文献中，没有被同一作者使用，但主题词对应的作者之间存在合著关系。元路径表示为 term1-author1-author2-term2。④基于引文的主题词强化关系：两个主题词不在同一篇文献中，因为引用同一篇参考文献产生联系。元路径表示为 term1-reference-term2。⑤基于引文的主题词新增关系：两个主题词不在同一文献中，且没有共同参考文献，但主题词对应的两篇参考文献被同一篇文献引用后产生联系。元路径表示为 term1-reference1-reference2-term2。

据以上选定的元路径将 5 种语义关联关系抽取出来，并将每一个时段的主题词对齐后转化为邻接矩阵，为接下来的多元关系融合的输入信息做好数据准备。

3.3 多元关系融合聚类

PathSelClus 算法的优点在于通过标注种子对象能够引导聚类过程，由此获取更加符合预期的聚类主题。因此，在开始融合聚类前需要运用先验知识储备标注一批种子数据作为输入信息。以弱关系数据为例，分别抽取每一时段出现的主题词，依据已有的知识将明显存在语义关联的主题词概括为一个主题，每个时段标注 4 个种子对象，10 个时段共标注 40 个。这里的标注结果是经过专家检验并调整过的。将上述步骤里对齐处理后的元路径关系邻接矩阵、标记的种子对象作为输入信息分别进行强弱关系的融合聚类，获取聚类主题。

以弱关系融合聚类为例，融合算法首先设置每个阶段初始聚类个数 K 值为 4，然后每次循环递增 2 个，最终获取到 5 种聚类结果，每种聚类结果中每一年阶段的主题个数分别是 4、6、8、10、12。对照 5 种结果后发现， K 值太小时，每一个类簇包含太多的主题词，主题不够明确； K 值太大时，每一个类簇包含的主题词过少，无法辨别出主题词之间的语义关联。因此，选取可解释性较强的聚类结果，即每一时段聚类个数为 8 的主题进行分析，结果如下。

首先，依靠专家的指导对 72 个主题进行命名（由于前期关联关系较弱，1991—1993 年这一时段获取到的类簇无法解读出有意义的主题）。

其次，依据科学突破的特征及指标对这些主题进行判别，从识别出的 72 个主题中筛选出 37 个突破性创新主题。

最后，对应 *Science* 年度十大科学突破、《麻省

理工科技评论》年度十大突破性技术、中国科学院年度科学发展报告等各类权威报告及新闻，对上述筛选出的 37 个突破性创新主题进行验证，共有 20 个主题在不晚于正式宣布为科学突破时被识别出来（表 5）。

表 5 基于弱关系融合识别出的科学突破

时段(年)	主题	科学突破
1994—1996	#0	癌症免疫疗法的细胞表面抗原单克隆抗体制备
	#5	逆转录病毒作用于类风湿性关节炎的治疗
1997—1999	#2	以树突状细胞为基础的肿瘤免疫治疗
	#3	逆转录病毒载体的抗肿瘤疗法
2000—2002	#2	重组 DNA 乙肝疫苗的研发
	#4	单核细胞增生李斯特氏菌及检验
	#6	噬菌体展示技术用于癌症治疗
2003—2005	#0	CD28 共刺激与细胞免疫应答
	#4	肿瘤过继性免疫细胞治疗
2006—2008	#4	RNA 干扰在肿瘤治疗中的应用
2009—2011	#3	细胞毒性 T 淋巴细胞 (CTL) 对肿瘤的识别与治疗
2012—2014	#1	乳酸菌表达系统表达抗原蛋白制备口服疫苗
	#2	CAR-T 细胞免疫治疗治疗癌症
	#3	过继 T 细胞疗法治疗慢性白血病
2015—2017	#2	PD-1 阻断疗法激活 T 细胞治疗癌症
	#4	核心抗原酶联免疫检测诊断丙型肝炎病毒感染
	#7	T 细胞衰竭在癌症免疫功能障碍中的作用
2018—2020	#3	预防禽流感病毒的 DNA 疫苗研制
	#6	CAR-T 细胞过继治疗在实体瘤中的研究
	#7	免疫检查点调节剂疗法作用于肿瘤治疗

同样地，对基于强关系融合获取的 80 个聚类主题，在专家的指导下分别对其进行主题命名。后续依据上述突破性创新判断指标对 80 个主题进行突破性评定，筛选出 39 个突破性创新主题。最后，对照各类权威报告及新闻报道对上述 39 个突破性创新主题进行验证，共有 15 个主题在不晚于正式宣布为突破性创新时被识别出来（表 6）。

3.4 聚类结果分析

对比分析两类关系获取的科学突破主题结果，可以发现两类关系主题有重叠之处，但是差异也很大，尤其是科学突破主题被识别出的时段。选择重点主题进行对比分析，如表 7 所示。

因为本文是融合了多类型的关系进行聚类，所以基于强关系和基于弱关系获取的主题结果都表现得很好。从表 7 中的对比情况可以看出，同一时段基于弱关系比基于强关系识别出了更多的科学突

表 6 基于强关系融合识别出的科学突破

时段(年)	主题	科学突破
1991—1993	#0	乙型肝炎基因工程疫苗研制
1994—1996	#1	HIV 的合成肽治疗和传统的抗逆转录病毒疗法
	#5	亚单位疫苗的研制
1997—1999	#0	核酸疫苗技术及抗寄生虫感染研究
	#6	艾滋病高效抗逆转录病毒疗法
2000—2002	#3	CD28 共刺激与细胞免疫应答
2003—2005	#1	基因重组乙肝疫苗的研制
2006—2008	#0	过继性免疫细胞治疗癌症
2009—2011	#4	过继性 T 细胞免疫疗法
	#5	减毒活疫苗的研发配制
2012—2014	#3	过继 T 细胞疗法治疗肿瘤
2015—2017	#2	嵌合抗原受体(CAR)T 细胞疗法
	#5	PD-1 免疫检查点阻断疗法
2018—2020	#2	基于双免疫联合疗法的肿瘤精准治疗
	#4	CAR-NK 细胞用于癌症治疗

破，或者利用弱关系能在较早阶段识别出正式宣布的科学突破，证明了关注弱关联关系是有必要的，具体分析如下。

(1) 树突状细胞疫苗作用于肿瘤治疗。树突状细胞最初是由 Steinman 在 1973 年提出并进行阐述的，1998 年在美国召开的“树突状细胞的细胞与分子生物学 Keystone 会议”上讨论了体外致敏树突状细胞(DC)回输疗法在国外应用于 4 种晚期肿瘤的早期临床研究^[58]。Kuglar 等^[68]运用 DC 融合疫苗对晚期肾癌患者进行治疗，取得很好的效果。后续研究人员纷纷进行临床试验，证明了 DC 疫苗能使胃癌患者、皮肤癌患者产生抗肿瘤特异反应。2010 年，美国食品药品监督管理局首次批准名为 Provenge 的 DC 疫苗上市，用于治疗前列腺癌。

(2) 噬菌体展示技术。噬菌体多肽或抗体展示文库在 1990 年构建，之后对该项技术的研究层出不穷。2002 年获得美国食品药品监督管理局认证，将其用于人体试验以探索癌症的治疗^[59]。2018 年噬菌体展示技术获得诺贝尔化学奖^[69]，开始被广泛用于药物开发和抗体筛选制备。基于弱关系聚类结果在 2000—2002 年这一时段就识别出了该项突破，相比于强关系的识别至少早了 4 年时间，此时该技术已经发展成为研究热点。

表 7 基于强弱关系获取的科学主题对比分析

弱关系聚类		重大科学突破	强关系聚类	
主题名称	时段(年)		时段(年)	主题名称
未识别	—	1993 年,中国生产出第一批重组基因工程乙肝疫苗(非自主研发),并开始在全国接种 ^[57]	1991—1993	乙型肝炎基因工程疫苗研制
以树突状细胞为基础的肿瘤免疫治疗	1997—1999	1998 年,美国召开“树突状细胞的细胞与分子生物学 Keystone 会议”上讨论了体外致敏树突状细胞(DC)回输疗法应用于 4 种晚期肿瘤的早期临床研究 ^[58] ,之后开启了树突状细胞对肿瘤治疗的临床研究	2000—2002	树突状细胞用于肿瘤免疫治疗
噬菌体展示技术用于癌症治疗	2000—2002	2002 年,美国食品药品监督管理局批准将噬菌体展示技术用于人体试验 ^[59]	2006—2008	基于噬菌体展示技术的肿瘤识别
肿瘤过继性免疫细胞治疗	2003—2005	2004 年,实验运用 CIK 细胞等效应细胞实现肿瘤过继性细胞免疫治疗 ^[60]	2006—2008	过继性免疫细胞治疗癌症
重组 DNA 乙肝疫苗的研发	2000—2002	2005 年,我国自主研发的新一代基因重组乙肝疫苗上市 ^[61]	2003—2005	基因重组乙肝疫苗的研制
RNA 干扰在肿瘤治疗中的应用	2006—2008	2010 年,英国 Nature 杂志发表研究报告,首次实现在人体中利用 RNA 干扰来治疗癌症 ^[62]	—	未识别
CAR-T 细胞免疫疗法治疗癌症	2012—2014	2013 年,美国 Science 年度十大科学突破之首 ^[63] ,之后两年 FDA 分别授予 Juno Therapeutics 的 JCAR015 和 Kite Pharma 的 KTE-C19 两种 CAR-T 细胞免疫治疗方法“突破性疗法”资格 ^[64]	2015—2017	嵌合抗原受体(CAR)T 细胞疗法
PD-1 阻断疗法激活 T 细胞治疗癌症	2015—2017	2017 年,美国食品药品监督管理局批准利用阻断 PD-1 通路的抗癌药物(免疫检查点抑制剂)治疗某些癌症 ^[65]	2015—2017	PD-1 免疫检查点阻断疗法
未识别	—	2020 年,德克萨斯大学 M.D. 安德森癌症中心等机构的研究人员研发新型细胞疗法——CAR-NK 疗法 ^[66]	2018—2020	CAR-NK 细胞用于癌症治疗
CAR-T 细胞过继治疗在实体瘤中的研究	2018—2020	2021 年,Science 发表研究,找到了将 CAR-T 细胞疗法应用于实体瘤的解决方案 ^[67]	—	未识别

(3) 人体中利用RNA干扰治疗癌症。1998年, Mello和Fire在*Nature*期刊中发表了有关RNA干扰法的文章, 该技术在2002年被评为美国*Science*年度十大科学突破之一^[70], 两位学者在2006年获得诺贝尔生理学或医学奖^[71]。2010年, 美国加州理工学院的研究人员将特定RNA作用于人体癌变部位, 干扰癌细胞基因进而起到治疗作用。实验证明该项技术可以抑制各种基因, 可以被广泛用于治疗各种癌症^[62]。基于弱关系的聚类结果在2006—2008年这一时段识别出该项突破, 基于强关系的聚类结果中没有解读出这一主题。

(4) 细胞免疫疗法。随着医学技术的不断进步, 癌症的治疗方式由传统的化疗、放疗, 转向免疫治疗、靶向治疗等。CAR-T细胞免疫疗法是一种新型免疫疗法, 此类免疫疗法在2013年被*Science*评为年度十大科学突破之首, 并于2018年获得诺贝尔生理学或医学奖。基于弱关系的聚类结果在2012—2014年这一时段解读出来该项突破, 基于强关系的聚类结果识别要晚一个时段。

(5) 新型癌症免疫疗法。细胞免疫疗法逐渐成熟, 现阶段更多的研究聚焦于肿瘤的精准治疗: 重组CAR-T疗法、NK联合PD-1疗法、TILs疗法及新型的TCR-T疗法。联合免疫疗法被更多地考虑。2020年德克萨斯大学M.D.安德森癌症中心研发出新型CAR-NK疗法并开始进行临床试验, 制备周期更短, 成本更低^[66]。在血液恶性肿瘤的研究取得一定进展后, 研究人员开始集中攻克实体肿瘤。2021年, *Science*发表了将CAR-T细胞疗法应用于实体瘤的解决方案^[67], 实现细胞免疫疗法质的飞跃。对比分析可知, 2018—2020年这一时段基于弱关系和基于强关系的融合聚类方法都实现了对未来新型免疫疗法的预判。

4 结论与展望

4.1 结论

本文通过梳理科学突破的概念内涵和现有的识别方法, 发现当前没有做到科学突破主题的早期识别预判。本文将弱关系分析引入科学突破的识别中, 能够缩短时滞, 并将识别阶段前置。首先, 构建了基于弱关系的方法模型, 将作者合著-主题词共现-同被引多种弱关系的知识网络作为识别科学突破主题的依据。选取PathSelClus算法对网络进行融合聚类, 获取主题作为判断候选项, 借助专家和

权威报告进一步筛选科学突破, 对比基于强关联关系识别出的科学突破, 验证了本文方法的可行性和有效性。之后, 选取基因工程疫苗领域做实证分析。最后, 结合领域专家对实证结果的判断及基因工程疫苗领域的研究进展和科学突破, 证明利用多元弱关系可以有效追踪感知某一学科领域科学研究中的突破性进展。本文聚焦弱关联关系的分析, 丰富了科学突破的识别方法, 为科学突破主题的早期识别提供了新视角。识别结果可以为政策制定者和基金资助机构提供决策支持, 有助于及时调整决策方向, 优化资源配置。

4.2 不足与展望

本文的研究存在以下不足之处: ①选取的多元关系融合算法存在主观因素。选取元路径时两个主题词直接共现和通过作者或引文等中间实体间接共现关系可能被多次重复计算; 同时, 该算法聚类结果的好坏很大程度上依赖于专家指导下标注的种子对象的质量, 缺乏一定的客观性。②研究时间阶段的划分有待改进。3年一个时段的划分标准无法精确地定位到突破初现的时间点。想要实现依据新知识或者新连接的出现对突破性创新主题进行监测, 需要在保证每个时段关系数据充足与精准的识别时间点之间寻求平衡。

科学突破早期识别的难点在于, 萌芽之初可获得的预警信号较少。未来我们将从以下两个方面开展进一步研究: ①采用网络表示学习方法。融合多元关系更全面地考虑主题词间的语义关联是未来研究方向之一, 网络表示学习通过对网络中的节点、连边等学习出低维的表征向量, 能够最大限度保留原网络中的节点关联信息, 有效融合、平衡多种关系信息。②引入时效网络分析。本文的研究是将时间切片内的各种关联信息聚集在一起构成静态网络, 无法及时捕捉反映网络节点及节点之间关联关系的动态变化等信息。时效网络是用时间标记网络节点的出现与消失以及节点之间连通时间等动态变化的网络, 未来研究可以将时效网络引入科学突破的识别中, 精准地定位到某项突破从产生到消退的时间点。

参考文献

- [1] 周斐辰. 日本科技创新战略重点及施策方向解析——基于日本《科学技术创新综合战略2020》[J]. 世界科技研究与发展, 2021, 43(4): 440-449.

- [2] 任声策. 从美国智库报告看美国科技创新战略新趋势[J]. 上海质量, 2021(1): 13-16.
- [3] A bill: to authorize appropriations for fiscal years 2022, 2023, 2024, 2025, and 2026 for the National Science Foundation, and for other purposes[R/OL]. (2021-03-26) [2021-08-10]. https://science.house.gov/imo/media/doc/NSF-FORTHEFUTURE_01_xml.pdf.
- [4] 科技部. 欧盟委员会发布报告列出 100 项对全球经济具有重大影响的突破性创新[EB/OL]. (2019-08-27) [2021-08-29]. http://www.most.gov.cn/gnwkjdt/201908/t20190827_148433.html.
- [5] 新华社. 中华人民共和国国民经济和社会发展第十四个五年规划和 2035 年远景目标纲要[EB/OL]. (2021-03-13) [2021-09-05]. http://www.gov.cn/xinwen/2021-03/13/content_5592681.htm.
- [6] Shapere D. The structure of scientific revolutions[J]. The Philosophical Review, 1964, 73(3): 383-394.
- [7] Merton R K. Singletons and multiples in scientific discovery: a chapter in the sociology of science[J]. Proceedings of the American Philosophical Society, 1961, 105(5): 470-486.
- [8] Brad Wray K. Kuhn and the discovery of paradigms[J]. Philosophy of the Social Sciences, 2011, 41(3): 380-397.
- [9] 杜建, 孙铁楠, 张阳, 等. 变革性研究的科学计量学特征与早期识别方法[J]. 中国科学基金, 2019, 33(1): 88-98.
- [10] Andersen H, Barker P, Chen X. The cognitive structure of scientific revolutions[M]. New York: Cambridge University Press, 2006.
- [11] Min C, Bu Y, Sun J J. Predicting scientific breakthroughs based on knowledge structure variations[J]. Technological Forecasting and Social Change, 2021, 164: 120502.
- [12] Galison P. Reflections on *Image and logic: a material culture of microphysics*[J]. Perspectives on Science, 1999, 7(2): 255-284.
- [13] van Raan A F J. On growth, ageing, and fractal differentiation of science[J]. Scientometrics, 2000, 47(2): 347-362.
- [14] Smalheiser N R. Rediscovering don swanson: the past, present and future of literature-based discovery[J]. Journal of Data and Information Science, 2017, 2(4): 43-64.
- [15] Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks[J]. Nature, 1998, 393(6684): 440-442.
- [16] Winnink J J. Early-stage detection of breakthrough-class scientific research: using micro-level citation dynamics[D]. Leiden: Leiden University, 2017.
- [17] Hollingsworth J R. Scientific discoveries: an institutionalist and path-dependent perspective[M]// Biomedicine in the Twentieth Century: Practices, Policies, and Politics. Amsterdam: IOS Press, 2008: 317-353.
- [18] De Bellis N. Bibliometrics and citation analysis: from the science citation index to cybermetrics[M]. Lanham: Scarecrow Press, 2009.
- [19] 刘亚辉, 许海云. 突破性创新早期识别与弱信号分析综述[J]. 图书情报工作, 2021, 65(4): 89-101.
- [20] 付玉秀, 张洪石. 突破性创新: 概念界定与比较[J]. 数量经济技术经济研究, 2004, 21(3): 73-83.
- [21] 于缓生. 论基础研究原始创新的特点[J]. 技术与创新管理, 2017, 38(4): 354-360.
- [22] 蒋军锋, 李孝兵, 殷婷婷, 等. 突破性技术创新的形成: 述评与未来研究[J]. 研究与发展管理, 2017, 29(6): 109-120.
- [23] 张金柱, 张晓林. 基于专利科学引文的突破性创新识别研究述评[J]. 情报学报, 2016, 35(9): 955-962.
- [24] 陈劲, 谢靓红. 原始性创新研究综述[J]. 科学学与科学技术管理, 2004, 25(2): 23-26.
- [25] 路甬祥. 从诺贝尔奖与 20 世纪重大科学成就看科技原始创新的规律(摘要)[J]. 中国科学院院刊, 2000, 15(5): 370-376.
- [26] 梁正, 邓兴华, 洪一晨. 从变革性研究到变革性创新: 概念演变与政策启示[J]. 科学与社会, 2017, 7(3): 94-106.
- [27] Kleinberg J. Bursty and hierarchical structure in streams[J]. Data Mining and Knowledge Discovery, 2003, 7(4): 373-397.
- [28] Yoon J, Kim K. Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks[J]. Scientometrics, 2011, 88(1): 213-228.
- [29] 张金柱, 张晓林. 利用引用科学知识突变识别突破性创新[J]. 情报学报, 2014, 33(3): 259-266.
- [30] Chen C M, Chen Y, Horowitz M, et al. Towards an explanatory and computational theory of scientific discovery[J]. Journal of Informetrics, 2009, 3(3): 191-209.
- [31] 罗瑞. 基于熵值的领域科学突破主题的识别与预测[D]. 北京: 中国科学院大学(中国科学院文献情报中心), 2020.
- [32] 杜建. “睡美人”文献的识别方法与唤醒机制研究[D]. 南京: 南京大学, 2017.
- [33] van Raan A F J, Winnink J J. Do younger Sleeping Beauties prefer a technological prince?[J]. Scientometrics, 2018, 114(2): 701-717.
- [34] 曹艺文, 许海云, 武华维, 等. 基于引文曲线拟合的新兴技术主题的突破性预测——以干细胞领域为例[J]. 图书情报工作, 2020, 64(5): 100-113.
- [35] Sood A, Tellis G J. Technological evolution and radical innovation[J]. Journal of Marketing, 2005, 69(3): 152-168.
- [36] Tellis G J. Disruptive technology or visionary leadership? [J]. Journal of Product Innovation Management, 2006, 23(1): 34-38.
- [37] 杨国忠, 陈佳. 企业突破性技术创新行为研究——基于前景理论的演化博弈分析[J]. 工业技术经济, 2020, 39(5): 57-64.
- [38] 徐路路, 王芳. 基于支持向量机和改进粒子群算法的科学前沿预测模型研究[J]. 情报科学, 2019, 37(8): 22-28.
- [39] 刘博文, 白如江, 周彦廷, 等. 基金项目数据和论文数据融合视角下科学研究前沿主题识别——以碳纳米管领域为例[J]. 数据分析与知识发现, 2019, 3(8): 114-122.
- [40] Liang Z T, Mao J, Lu K, et al. Combining deep neural network and bibliometric indicator for emerging research topic prediction [J]. Information Processing & Management, 2021, 58(5): 102611.

- [41] 刘亚辉. 科学突破性创新早期识别中的弱关系分析[D]. 北京: 中国科学院大学(中国科学院文献情报中心), 2021.
- [42] Granovetter M. The strength of weak ties: a network theory revisited[J]. *Sociological Theory*, 1983, 1: 201-233.
- [43] Xu S, Hao L Y, An X, et al. Review on emerging research topics with key-route main path analysis[J]. *Scientometrics*, 2020, 122(1): 607-624.
- [44] 林向义, 罗洪云, 李秀成. 企业个体从社交媒体网络吸收异质性知识的过程机理: 弱连接关系视角[J]. *情报理论与实践*, 2019, 42(3): 65-71.
- [45] 牌艳欣, 李长玲, 徐璐. 弱引文关系视角下跨学科相关知识组合识别方法探讨——以情报学为例[J]. *图书情报工作*, 2020, 64(21): 111-119.
- [46] 张英杰. 科技领域前沿计量探测方法研究[D]. 北京: 中国科学院研究生院(中国科学院文献情报中心), 2011.
- [47] Wei L, Xu H Y, Wang Z M, et al. Topic detection based on weak tie analysis: a case study of LIS research[J]. *Journal of Data and Information Science*, 2016, 1(4): 81-101.
- [48] Yoo S, Won D. Simulation of weak signals of nanotechnology innovation in complex system[J]. *Sustainability*, 2018, 10(2): 486.
- [49] Griol-Barres I, Milla S, Cebrián A, et al. Detecting weak signals of the future: a system implementation based on text mining and natural language processing[J]. *Sustainability*, 2020, 12(19): 7848.
- [50] 岳增慧, 许海云, 赵敏. 强弱连接对学科引证知识扩散动态链路预测的影响研究[J]. *图书情报工作*, 2021, 65(13): 66-76.
- [51] 许海云, 武华维, 罗瑞, 等. 基于多元关系融合的科技文本主题识别方法研究[J]. *中国图书馆学报*, 2019, 45(1): 82-94.
- [52] Sun Y Z, Norick B, Han J W, et al. PathSelClus: integrating meta-path selection with user-guided object clustering in heterogeneous information networks[J]. *ACM Transactions on Knowledge Discovery from Data*, 2013, 7(3): Article No.11.
- [53] 傅俊英, 彭喆, 郑佳, 等. 基于专利异构网络的中小企业潜在合作伙伴研究——以石墨烯领域为例[J]. *情报学报*, 2019, 38(4): 391-401.
- [54] 李伟, 刘冬冬, 吴丹, 等. 基因工程疫苗在动物疫病防治中的应用[J]. *畜牧兽医科学(电子版)*, 2019(22): 90-91.
- [55] Zhang Y, Porter A L, Hu Z Y, et al. "Term clumping" for technical intelligence: a case study on dye-sensitized solar cells[J]. *Technological Forecasting and Social Change*, 2014, 85: 26-39.
- [56] 刘奕杉, 王玉琳, 李明鑫. 词频分析法中高频词阈值界定方法适用性的实证分析[J]. *数字图书馆论坛*, 2017(9): 42-49.
- [57] 为了一个没有乙肝的中国[EB/OL]. (2018-04-06) [2021-09-05]. <http://www.hbver.com/Article/ygzf/ygym/201804/8593.html>.
- [58] 吴硕, 徐秀玉. 树突状细胞与喉癌的基因免疫治疗[J]. *临床耳鼻咽喉科杂志*, 2006, 20(8): 381-384.
- [59] Krag D N, Fuller S P, Oligino L, et al. Phage-displayed random peptide libraries in mice: toxicity after serial panning[J]. *Cancer Chemotherapy and Pharmacology*, 2002, 50(4): 325-332.
- [60] Nagaraj S, Ziske C, Schmidt-Wolf I G. Human cytokine-induced killer cells have enhanced *in vitro* cytolytic activity via non-viral interleukin-2 gene transfer[J]. *Genetic Vaccines and Therapy*, 2004, 2(1): 12.
- [61] 我国自主研发的新一代基因重组乙肝疫苗上市[J]. *肝博士*, 2005(5): 63.
- [62] 科技部. 核糖核酸(RNA)干扰机制首次在人体中获得证实[EB/OL]. (2010-03-26) [2021-09-06]. http://www.most.gov.cn/gnwkjdt/201003/t20100325_76448.html.
- [63] 中国科学报. 《科学》评出2013年十大突破[EB/OL]. (2013-12-25) [2021-09-06]. http://www.cas.cn/xw/kjxm/gjdt/201312/t20131225_4005535.shtml.
- [64] Carroll J. CAR-T player Juno picks up 'breakthrough' status for lead cancer therapy[EB/OL]. (2014-11-24) [2021-09-06]. <https://www.fiercebiotech.com/regulatory/car-t-player-juno-picks-up-breakthrough-status-for-lead-cancer-therapy>.
- [65] Kamphorst A O, Wieland A, Nasti T, et al. Rescue of exhausted CD8 T cells by PD-1 - targeted therapies is CD28-dependent[J]. *Science*, 2017, 355(6332): 1423-1427.
- [66] Robbins Y, Greene S, Friedman J, et al. Tumor control via targeting PD-L1 with chimeric antigen receptor modified NK cells[J]. *eLife*, 2020, 9: e54854.
- [67] Hernandez-Lopez R A, Yu W, Cabral K A, et al. T cell circuits that sense antigen density with an ultrasensitive threshold[J]. *Science*, 2021, 371(6534): 1166-1171.
- [68] Kugler A, Stuhler G, Walden P, et al. Regression of human metastatic renal cell carcinoma after vaccination with tumor cell - dendritic cell hybrids[J]. *Nature Medicine*, 2000, 6(3): 332-336.
- [69] 中国科学院. 2018年诺贝尔奖[EB/OL]. (2018-10-03) [2021-09-08]. <http://www.cas.cn/zt/sszt/2018nobelprize/>.
- [70] 李新荣, 毛磊, 赵路. 《科学》评出2002年十大科学突破[J]. *科学咨询*, 2003(1): 7-8.
- [71] 李旭丰, 张涛. RNA干扰: 控制遗传信息流动的机制[J]. *国外科技动态*, 2006(10): 18-28.

(责任编辑 王克平)