情报学报　2025 年 3 月　第 44 卷　第 3 期
Journal of the China Society for Scientific and Technical Information, Mar. 2025, 44(3): 325-338

DOI: 10.3772/j.issn.1000-0135.2025.03.006

组合新颖性和学科新颖性能否提升论文的技术影响？基于直接和间接技术影响力的双重视角

杨　杰[1]，柳美君[2]，步　一[3]，赵　星[4,5]，邓三鸿[1]

（1. 南京大学信息管理学院，南京　210023；2. 复旦大学全球公共政策研究院，上海　200433；
3. 北京大学信息管理系，北京　100871；4. 复旦大学大数据研究院，上海　200433；
5. 复旦大学国家智能评价与治理实验基地，上海　200433）

摘　要　在发展新质生产力的背景下，理解新颖性的知识组合以及跨学科的知识融合能否提升论文的技术影响具有重要意义。为此，本文通过真实知识组合分布和模拟知识组合分布的相对高阶指标测度引文组合新颖性，基于学科共被引矩阵和距离分布测度学科新颖性。考虑“直接-间接”双重视角，通过融合“论文-专利”引用网络和“论文-论文”深层引用网络，提出论文的专利直接被引量和专利间接被引量等指标，以测度论文的直接技术影响和间接技术影响。通过对微软学术数据库中约3000万篇科学论文的固定效应回归分析，本文发现引文组合新颖性能够同时正向促进论文的直接技术影响和间接技术影响，并且对间接技术影响的促进效应更强；而学科新颖性仅能正向促进论文的间接技术影响。进一步地，本文还发现在STEM（science，technology，engineering，mathematics）领域和大型合作团队中，引文组合新颖性和学科新颖性能够更有效地提升论文的双重技术影响，且学科新颖性对论文的间接技术影响的促进效应随年份存在下降趋势。

关键词　新颖性；跨学科性；科学-技术；技术影响

Can Combinative and Disciplinary Novelty Enhance the Technological Impact of Scientific Papers? A Dual Perspective of Direct and Indirect Technological Impact

Yang Alex J.[1], Liu Meijun[2], Bu Yi[3], Zhao Star X.[4,5] and Deng Sanhong[1]

(1. *School of Information Management, Nanjing University, Nanjing*　210023; 2. *Institute for Global Public Policy, Fudan University, Shanghai*　200433; 3. *Department of Information Management, Peking University, Beijing*　100871; 4. *Institute of Big Data (IBD), Fudan University, Shanghai*　200433; 5. *National Institute of Intelligent Evaluation and Governance, Fudan University, Shanghai*　200433)

Abstract: Understanding whether combined novelty and interdisciplinary knowledge integration can enhance the technological impact of scientific papers is important for the development of new quality productive forces. To this end, this study employs Monte Carlo simulations to measure combinative novelty using higher-order metrics derived from real and

收稿日期：2024-04-22；**修回日期**：2024-12-27
基金项目：国家自然科学基金青年学生基础研究项目（博士研究生）“融合创新测度和生成模型的科学技术知识空间预测研究”（#724B2015）。
作者简介：杨杰，男，2000年生，博士研究生，主要研究领域为科学计量、创新评价、复杂网络分析；柳美君，女，1992年生，博士，青年副研究员，主要研究领域为创新评价、政策挖掘、大数据分析；步一，男，1994年生，博士，助理教授，博士生导师，主要研究领域为信息计量、学术文本挖掘、科学评价、知识发现；赵星，男，1983年生，教授，博士生导师，主要研究领域为信息计量、科学评价、元宇宙技术；邓三鸿，通信作者，男，1975年生，教授，博士生导师，主要研究领域为领域知识图谱、科学计量、信息分析与科学评价，E-mail：sanhong@nju.edu.cn。

simulated distributions of knowledge combinations. Disciplinary novelty is quantified using co-citation matrices and distributional distances across disciplines. Adopting a dual perspective of direct and indirect technological impact, this study integrates a "paper-to-patent" citation network with a deep "paper-to-paper" citation network to propose metrics such as patents directly and indirectly citing papers. These metrics capture the direct and indirect technological impact of scientific papers. Based on a fixed-effects regression analysis of 30 million scientific papers from the Microsoft Academic Database, the findings reveal that combinative novelty positively promotes both direct and indirect technological impacts of papers, with a stronger effect observed on indirect technological impact. By contrast, disciplinary novelty exclusively promotes indirect technological impact. Furthermore, the study identifies that, in STEM fields and large collaborative teams, both combinative and disciplinary novelty effectively enhance the dual technological impact of papers. Notably, the positive effect of disciplinary novelty on the indirect technological impact of papers exhibits a declining trend over time.

Keywords: novelty; interdisciplinarity; science–technology; technological impact

0 引 言

科学创新是技术进步的源泉。改革开放以来，中国对基础科学投资的力度不断加大，科学知识的增长与技术产业发展之间的密切联系带来了社会的长久繁荣[1]。基础科学的创新作为生产力系统的底层要素，是发展新质生产力的关键所在，也是中国在国际竞争新格局中取得高端产业链和价值链的基础[2]。习近平总书记指出，要"整合科技创新资源，引领发展战略性新兴产业和未来产业，加快形成新质生产力"。新颖的知识组合和跨学科的知识融合作为科学创新的事前表征[3]，理解其如何推动技术发明和产业进步，对于发展新质生产力和构建现代化创新发展体系有着重要的理论价值和政策意义。

一方面，在渐进性创新的范式下[4-6]，以专利为代表的技术发明往往并非孤立的创新成果，而是建立在相关科学知识基础上的拓展和应用。例如，药物的研发依赖于基因或蛋白质结构的发现与解析，大语言模型技术离不开 Transformer 和残差网络等深度学习框架的提出，纳米材料的制造依赖于低尺度下化学分子的结构发现，卫星通信技术的实现离不开信息论和黎曼几何的发展。另一方面，由于创新的不确定性与低回报性[7]，其在极大程度上依赖于政府、机构和社会的资助[8]。因此，探索科学的事前创新组合对技术产业的推动机制，有助于实现科学投入的最优资源配置，通过对基础科学研究的资源合理配置，最大程度地推动技术产业发展和经济繁荣，并反过来提高对基础科学的资助，构建正反馈的良性科研生态循环[9]。由此，从"科学-技术"关联视角出发，针对论文新颖性的技术促进效应进行定量分析具有较强的理论基础。

现有研究较少关注组合新颖性和学科新颖性是否能够提升论文的技术影响，分析这一问题面临多个挑战。一方面，科学新颖性的测度需要事前的、统一的、可靠的指标，被引量和颠覆性[10]等指标依赖于事后的引文评价维度，难以描述论文本身的知识组合创新程度，而基于领域主题词表和实体抽取技术的内容新颖性测度，则难以实现对多领域科学创新性的统一测度框架。另一方面，基础科学研究对技术产业的价值难以衡量，这是因为科学发现对技术发明的影响往往是隐式而漫长的[11]，当一项重要的基础科学发现得到发表，相关的技术通常需要数十年才能发展成熟并得以应用[12]，科学成果的社会价值通常不会在短时间内体现出来[13]，追溯科学与技术的关联充满挑战，科学到技术之间的距离也难以量化[14]。由此，分析科学创新的技术促进效应需要同时解决大数据下的新颖性测度和论文的技术影响测度问题。

本文旨在测度科学论文的新颖性和"直接-间接"双重技术影响，并分析二者之间的复杂关系。为此，基于微软学术数据库中 1960—2016 年的 29877430 篇论文和 38740313 个"论文-专利"引用链接，①通过真实知识组合分布和蒙特卡洛模拟下的随机知识组合分布，考虑相对高阶指标测度引文组合新颖性，基于学科共被引矩阵和距离分布测度学科新颖性，测度 29877430 篇研究论文的组合新颖性和学科新颖性；②从直接影响和间接影响双重视角出发，通过融合"论文-专利"引用网络和"论文-论文"深层引用网络，提出论文的专利直接被引量和专利间接被引量等指标，以测度论文的直接技术影响和间接技术影响；③基于固定效应回归模型，定量研究组合新颖性、学科新颖性与论文的技术影响的关系，并进一步分析在不同年份、团队规模和领域层次的异质性效应。本文的研究方法和结论扩展了知识组合创新和"科学-技术"等领域的

研究边界，提供了新的理论框架和实证结果，并为相关科技政策的制定提供了启示。

1　相关研究

1.1　科学新颖性

新颖性是一个内涵丰富的术语，它没有统一且精确的定义[15]，但有着丰富多样的测度方法和框架[7]。科学论文的新颖性测度通常都以熊彼特的组合创新理论[16]作为理论基础，其认为创新的知识往往产生于广阔概念空间中较小概率的组合模式[17-18]。创新的知识组合通过整合不同的视角来重新定义问题空间，通过对既有知识的重组提出新的研究方法和模型[19]。在指数型增长的广阔文献中，如何对既有的科学知识进行取样和组合在一定程度上决定了论文的创新价值。事实上，大部分论文的知识组合都是基于最新、最热门的文献，在狭窄的知识口径中进行类似的组合[20]，而与之相反的、非典型的知识组合往往能产生更多的科学突破[21]。

如表 1 所示，目前，学界主要从内容新颖性、引文组合新颖性和学科新颖性等层次测度事前的新颖性。内容层次的新颖性通常以领域主题词表，如医学领域的 MeSH（medical subject headings）和物理学领域的 PACS（physics and astronomy classification scheme）或知识实体/知识元作为基础。例如，Azoulay 等[15]基于 MeSH 主题词计算其重组特征和首次出现的比例，以表征论文的新颖性；Liu 等[22-23]基于医学实体抽取和消歧技术（bio-entity extraction and disambiguation）从科学文本中抽取知识实体，并进一步基于 BERT（bidirectional encoder representations from transformers）模型计算知识实体的语义距离，以此表征论文的新颖性。内容新颖性测度方法具有细粒度、拓展性强、维度丰富、方法和模型灵活程度高等优势，而其局限性主要在于对主题词表数据/文本数据、实体抽取方法和 NLP（natural language processing）模型依赖程度高，通常局限于单一领域。

引文组合新颖性的测度方法则通常不考虑论文的内容和语义特征，而是以引文链接作为分析基础，基于复杂网络范式下的创新扩散和统计物理学模型进行新颖性的测度。其中最具代表性的工作是 Uzzi 等[21]的研究，其基于蒙特卡洛模拟计算出引文中期刊对组合的非典型性以表征论文的新颖性，这一方法得到了后续研究的验证[24]，被认为是最经典和可靠的新颖性指标之一[25-26]。除此以外，Wang 等[27]提出以参考文献“期刊对”的首次组合作为新颖性的测量标准，该方法更为直观和简洁。Bornmann 等[28]基于同行评议数据 F1000Prime 对两种引文组合新颖性的测度方法进行比较，发现非典型知识组合能更好地识别突破性研究。

学科新颖性通常以论文的后向引文或前向引文所属学科的距离分布为分析基础。Stirling[29]提出的跨学科性测度框架融合了多样性、均衡性和丰富性等多个维度的视角，能较为全面地捕获科学知识的学科距离分布。Fontana 等[30]发现非典型的知识组合（引文组合新颖性）和远距离的知识融合（学科新颖性）不仅在概念上高度相似，而且在实证上也体现出很强的相关性，二者从不同尺度表征了论文对既有知识的创新组合程度。在本文中，引文组合新颖性和学科新颖性的一大优势在于其能够较为便

表 1　科学论文新颖性的典型测度方法

层次	新颖性测度方法	优点	局限	参考文献
内容新颖性	基于领域主题词表的语义距离、分布特征或复杂网络特征	细粒度、主题词表质量通常较高、模型灵活	对主题词表数据和 NLP 模型依赖程度高，通常局限于单一领域	Azoulay 等[15]；Foster 等[31]
	基于标题/摘要/全文本的实体语义距离或组合分布	细粒度、维度丰富、方法和模型灵活	对文本数据、实体抽取方法和 NLP 模型依赖程度较高	Liu 等[22-23]
	基于外部词典的词频规则或结构相似度	效度高、外部词表质量通常较高、维度丰富	高质量的外部词典较为少见	李茂林等[32]；Krieger 等[33]
引文组合新颖性	参考文献“期刊对”的首次组合	方法统一、定义直观、计算简单	假设较为绝对、易受作者操控、无法解决跨学科期刊问题	Wang 等[27]
	参考文献“期刊对”组合的“非典型性”	应用广泛、收敛有效性强、易应用于大规模数据集	依赖引文耦合数据、无法解决跨学科期刊问题	Uzzi 等[21]；Shi 等[34]
学科新颖性	后向引文或前向引文所属学科的距离分布	与跨学科分析框架一致、易应用于大规模数据集	粗粒度、依赖于学科距离和引文数据	Stirling[29]；Fontana 等[30]

续表

层次	新颖性测度方法	优点	局限	参考文献
其他“事前”新颖性	参考文献的年龄结构特征（文献半衰期和普莱斯指数等） ……	方法统一、定义直观、涵盖纵向维度	粗粒度、易受作者操控	Onodera 等[35]
“事后”影响力或颠覆性	基于论文动态被引特征的模型	能够反映影响力创新扩散	依赖事后的被引数据、模型较为复杂且容易过拟合	Wang 等[36]
	前向引用和后向引用的复杂耦合关系 ……	能够反映范式变革理论	依赖事后的被引数据、易受作者操控、收敛有效性差、缺乏鲁棒性	Funk 等[10]；Park 等[26]

捷地应用于多学科的文献大数据，从而实现较为全面的分析。

1.2 论文的技术影响

由于创新的公益性质，竞争性市场对创新的激励不足，政府通过对科学领域的资助来解决对基础创新的需求[8]，分析科学创新是否能提升科学对技术的影响具有重要意义。为量化科学与技术的关联，学者们将视角转向了引用链接[37]。在科学和技术的各自领域，专利之间和论文之间的引用蕴涵是对既有知识的认可和继承，也是知识流动的主要方式[38]。而专利对论文的引用则代表知识跨越了科学和技术的边界[39]，反映知识从科学发现向技术发明的流动[14,37]。得益于最近“论文-专利”引用数据的可获得性，科学和技术的关联研究得以量化。专利对论文的引用提供了技术的科学基础，捕捉了基础科学如何影响技术应用的一种直接模式[14]。

然而，在实践中，基于这种直接的“论文-专利”引用数据难以直接追踪到技术的完整科学谱系[37]，有以下3点因素：其一，学术界的科学发现与产业界的技术发展之间存在较大程度的脱节[40]。科学的发展与社会需求往往并不同步[9]，基础科学具有公共产品的属性，很少能在短期的实际应用中取得回报[8]，从科学发现到技术应用通常需要长达数十年的时间，这导致了科学与技术关联的不稳定性[10]。其二，论文和专利有着不同的范式和引用模式。专利引用的审查非常严格，不太可能将其相关的科学论文全部引用。其三，科学知识的重复性和复杂性可能导致真正的科学知识源没有被直接引用[41]。为此，本文在专利对论文的直接引用基础上，进一步构建了间接影响力测度框架，通过考虑“论文-专利”引用的直接和间接双重视角分析科学对技术的影响。

2 研究框架

新颖性的科学研究往往都是风险与机遇并存[7]，这体现为新颖性指标与影响力指标之间错综复杂的关系[27]。现有研究从多个维度探索了科学新颖性和论文的学术影响力之间的关系[17,42]，目前还没有文献对科学创新的技术促进效应进行系统的研究。为解决此问题，本文通过融合多源大数据和前沿的分析模型，研究创新知识组合与科学对技术的影响的关系。

本文的研究框架如图1所示，主要遵循数据驱

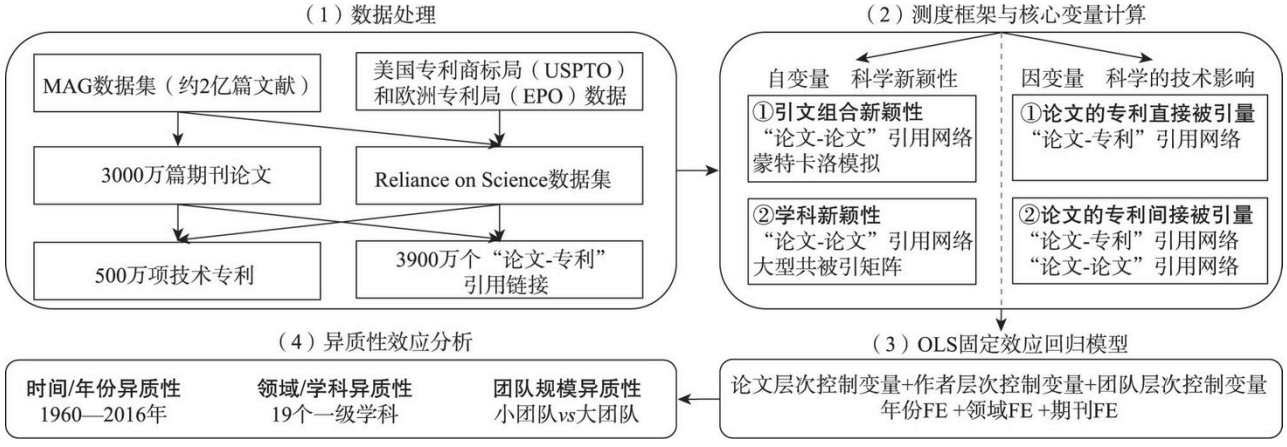


图1 研究框架

动的范式进行研究。首先，融合论文元数据、专利元数据、“论文-专利”引用数据以及“论文-论文”深层引用网络，得到涵盖292个领域的千万级论文数据和“论文-专利”引用链接。其次，设计核心变量的测度框架：基于组合创新的相关研究和理论，定义并计算科学新颖性（解释变量），其包含引文组合新颖性和学科新颖性两种不同层次的指标，均基于“论文-论文”引用网络计算；基于“科学-技术”的相关研究，定义并计算论文的技术影响（被解释变量），其包含论文的专利直接被引量和专利间接被引量，前者仅考虑“论文-专利”引用网络，后者则融合了“论文-专利”引用网络和“论文-论文”引用网络。在计算核心变量的基础上，进一步计算论文层次、作者层次和团队层次的控制变量，基于OLS（ordinary least squares）模型对千万级样本进行大数据固定效应回归分析。最后，分析回归效应在不同时间、不同领域以及不同团队规模中的异质性。

3　数　据

3.1　科学论文数据集

本文以微软学术数据库（Microsoft Academic Graph，MAG）作为科学论文数据[43]。MAG涵盖了发表于1800—2021年的2亿余篇文献和超过15亿个引用链接，类别包括期刊论文、会议论文、预印本、图书和各种其他形式的科技文献。本文选取发表于1960—2016年的论文，由此，每篇论文都至少有5年的被引时间窗口。由于不同种类文献往往存在不一致的引用模式，本文进一步将文献类型限定为期刊论文。另外，本文进一步选取至少有5篇参考文献的研究论文，以排除掉论文勘误、编辑导语及读者来信等非研究型的学术记录。基于以上数据筛选策略，本文最终得到29877430篇研究论文。

微软学术数据库基于深度学习模型为每篇论文提供了一个或多个领域类别标签，包含多个层次。本文主要考虑前两个层次：第一层，包括医学、物理学和计算机科学等19个广泛的领域；第二层，进一步划分出292个子领域。本文的指标计算和领域固定效应均依赖于292个微软学术数据库二级领域，且在多个标签中选择置信度最高的标签。

3.2　论文-专利引用数据

本文以“论文-专利”引用数据为基础衡量论文的技术影响，其中论文数据来自MAG数据集，而专利文献数据来自美国专利商标局（United States Patent and Trademark Office，USPTO）和欧洲专利局（European Patent Office，EPO）的专利文献，两者都提供了公开的专利文本及详细引用信息。Marx等[44]开发的Reliance on Science数据集将来自USPTO和EPO的技术专利与MAG中的科学文献关联起来。本文基于Reliance on Science数据集，匹配了与科学论文相关联的5087807项技术专利及38740313个“论文-专利”引用链接。

4　测度方法与框架

4.1　新颖性测度

本文基于Uzzi等[21]提出的方法，以期刊层次的知识粒度计算引文组合新颖性。期刊可以在广义上体现出各个科学领域的知识分布，MAG数据库提供了多达39893种学术期刊，这种方式相比于学科层次的知识领域（292个二级学科）具有更细的粒度和层次，并且可以实现对全部领域的论文进行新颖性计算。

如图2a所示，基于蒙特卡洛模拟[21,42,45]获得随机重组的引用网络，每组节点都以随机的方式重新分配边，并保留真实网络的时间分布、前向网络度分布和后向网络度分布。图2a中的边A和边B分别反映了不同论文的引用关系，因其前后节点的时间均一致，通过随机交换进行模拟，而边C由于存在不同的前后节点时间，则不进行交换。通过同时随机交换边的方式和控制边的节点时间，蒙特卡洛过程产生的模拟网络与真实网络相比，仅反映出个体知识分布的变化，而不影响整体知识分布。本文基于这种方式模拟出10组随机网络，作为网络零模型（null model）[46]，基于统计物理学[47]方法，将其应用于测算期刊对的随机分布概率。

为量化论文的引文组合新颖性，首先，基于引文期刊对表示真实知识嵌入网络，以表征科学知识组合的真实分布。其次，通过模拟的引用期刊对表示模拟知识嵌入网络，以表征科学知识组合的模拟分布。进一步地，通过引用期刊对组合的真实分布和模拟分布的对比，计算出任意组合的出现概率。具体而言，对于单篇论文引用的某一期刊对 m,n，

分别将其真实概率 $\mathrm{obs}(\mathrm{pair}_{mn})$ 与随机分布中的概率 $\exp(\mathrm{pair}_{ij})$ 进行对比，并转换为 Z 值，即

$$Z(\mathrm{Novelty})_{m,n}=\frac{\mathrm{obs}(\mathrm{pair}_{mn})-\exp(\mathrm{pair}_{mn})}{\sigma(\mathrm{pair}_{mn})} \quad (1)$$

其中，$\mathrm{obs}(\mathrm{pair}_{mn})$ 表示真实知识网络中期刊对的频率；$\exp(\mathrm{pair}_{mn})$ 和 $\sigma(\mathrm{pair}_{mn})$ 分别表示 10 次随机蒙特卡洛模拟引用网络中期刊对频率的均值与标准差。基于蒙特卡洛模拟得到样本中的每篇论文的全部期刊对的概率分布，其中某些期刊对与随机分布相比更罕见，即反映出引文知识组合新颖性。值得注意的是，一篇论文通常有多组引文期刊对，其真实分布和模拟分布的 Z 值会形成一组分布（向量）$\{Z\,\mathrm{score}_{i,j}|i,j\in J\}$。为了总结该分布中的信息，本文使用每篇论文引文期刊对 Z 值分布的第 10 百分位数作为高维统计量[21]，即 $Z\,\mathrm{score}_{10\mathrm{pt}}$。当科学知识组合的真实分布和模拟分布的 Z 值为较高的正值时，表明论文 i 的知识组合较为常见；而当 Z 值为较小的正值或者负值时，表明论文 i 的知识组合较为新颖。为此，本文将 $Z\,\mathrm{score}_{10\mathrm{pt}}$ 基于公式

$$\text{Combination novelty}=\begin{cases}-\ln(Z\,\mathrm{score}_{10\mathrm{pt}}+1), & \text{if } Z\,\mathrm{score}_{10\mathrm{pt}}>0\\ \ln(-Z\,\mathrm{score}_{10\mathrm{pt}}+1), & \text{if } Z\,\mathrm{score}_{10\mathrm{pt}}\leqslant 0\end{cases} \quad (2)$$

进一步转化，最终得到引文组合新颖性指标。

接下来考虑粗粒度的创新知识组合测度方法，即基于 MAG 数据库 292 个二级学科计算后向引文中的学科距离，测度论文的学科新颖性。学科新颖性反映了引文中学科的距离分布，当引文中的学科距离较远时，可以认为论文具有较高的学科新颖性。如图 2b 所示，d_{mn} 表示领域 m 和领域 n 之间的距离，基于微软学术数据库中发表于 1800—2021 年的全部论文及引证网络计算而得。具体而言，本文的数据涵盖了 15 亿余条引用链接构建大型学科共被引矩阵，考虑领域 m 的全部论文 $M=\{i|\mathrm{field}_i\in m\}$，基于论文引用关系构建学科共被引矩阵，在该矩阵中，领域 m 与其余 291 个领域的共被引分布向量 $\boldsymbol{V}_m$ 反映了不同学科之间的知识交互分布。在此基础上，基于 Rao-Stirling 指标[29,48]的测度理论，使用余弦距离计算领域 m 和领域 n 之间的距离 d_{mn}，即 $d_{mn}=1-\mathrm{CosSim}(\boldsymbol{V}_m,\boldsymbol{V}_n)$，学科新颖性指标[29]定义为

$$\text{Field novelty}=\sum_{m\neq n}d_{mn}p_mp_n \quad (3)$$

其中，p_m 和 p_n 分别表示论文 i 参考文献中的领域 m 和领域 n 的占比。值得注意的是，学科新颖性与跨学科综合性指标概念一致，本文之所以将其作为新颖性的一种测度指标，是借鉴了 Fontana 等[30]和 Yang 等[49]的研究，他们认为跨学科综合性指标表示远距离的知识融合，能够从学科层次表征新颖性。

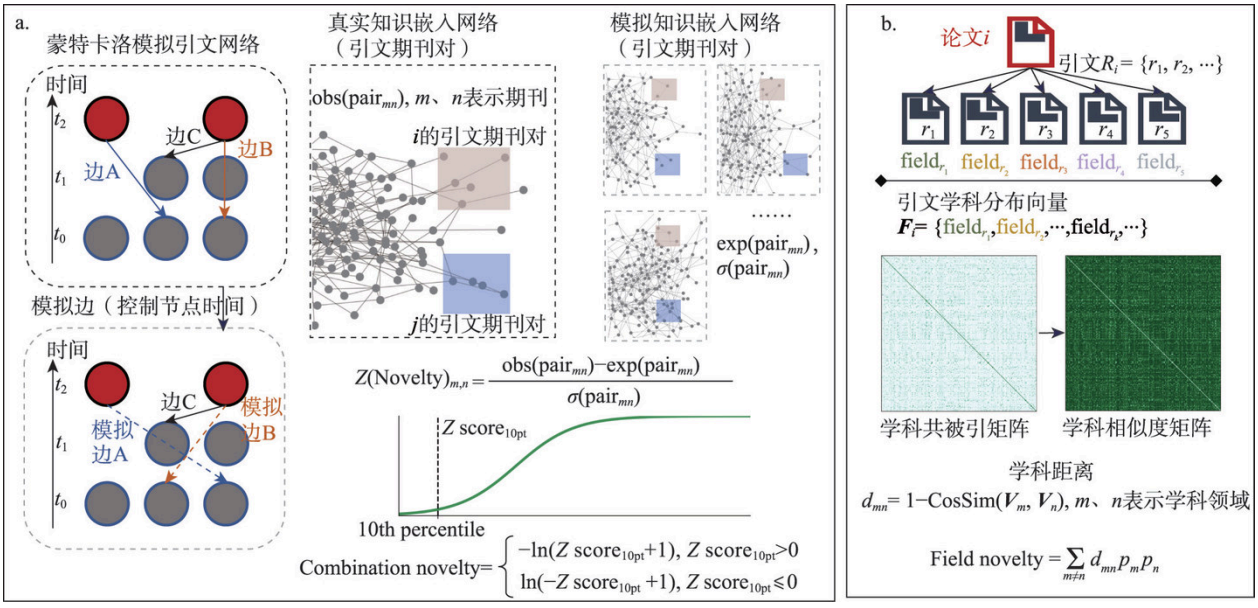


图 2 科学新颖性测度框架:引文组合新颖性与学科新颖性

4.2 双重技术影响测度

本文从引用关系来探究科学和技术领域的知识流动。如图 3a 所示，“论文-论文”引用反映了科学界中论文的知识流动，“论文-专利”引用则体现出论文的影响突破了科学边界，对技术发明的创造产生了直接影响。本文基于“专利-论文”

引用网络对科学对技术的直接影响进行测度[50]，并通过融合“论文-论文”引用网络和“专利-论文”引用网络提出科学对技术的间接影响测度指标。

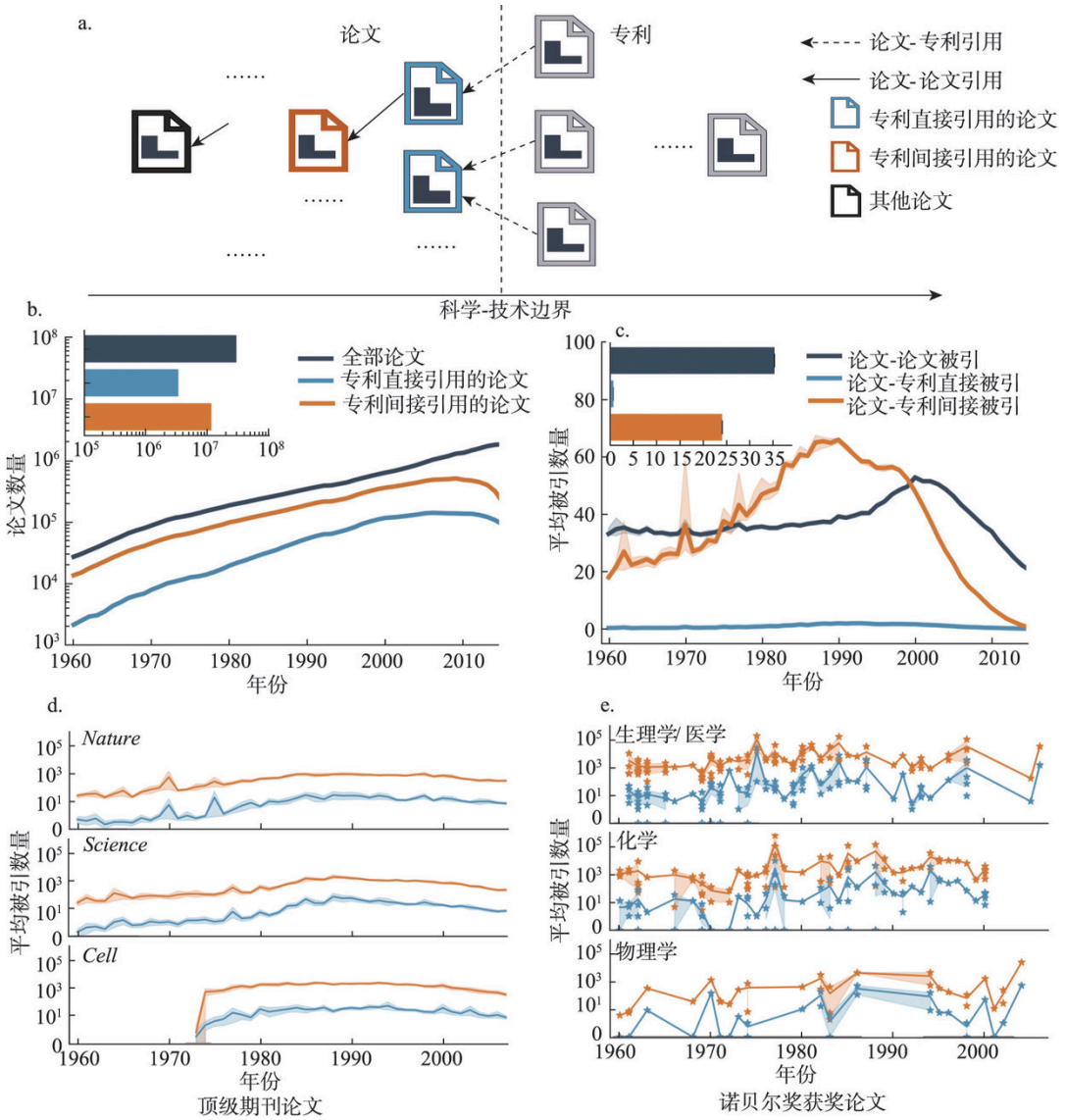


图 3　论文的技术影响测度框架:专利直接被引和专利间接被引（彩图请见 https://qbxb.istic.ac.cn）

误差棒和阴影区域表示 95% 置信区间（bootstrap）。

本文将论文的专利直接被引量定义为引用该论文的专利文献数量，它反映了论文的技术影响[51]。然而，由于科学到技术转变的过程存在明显滞后性，且科学家与技术开发人员所接触的知识有较大差异，很多创新的科学论文并没有被专利直接引用。为此，本文基于知识和技术扩散理论，提出论文的间接技术影响测度方法，将一篇论文的被引文献以其专利被引量求和。在“论文-论文”引用网络中，假设一篇论文 i 的入度（被引）节点集合为 $C_{\text{paper}}(i)$，则其被引量为 $|C_{\text{paper}}(i)|$；同样地，在和“专利-论文”引用网络中，假设论文 i 的入度（被引）节点集合为 $C_{\text{patent}}(i)$，则其专利直接被引量为 $|C_{\text{patent}}(i)|$，由此，论文的专利间接被引量 $C^2_{\text{patent}}(i)$ 为

$$C^2_{\text{patent}}(i) = \sum_{k \in C_{\text{paper}}(i)} (|C_{\text{patent}}(k)|) \tag{4}$$

此方法将论文被引文献的技术影响汇聚到中心文献，结合“论文-论文”引用和“论文-专利”引用的高阶指标，从广度和深度拓展了对论文的技术影响的测度模式。值得注意的是，论文的专利间接被引量不仅包含了论文的技术影响，也考虑了学术影响，即当一篇论文被较多论文引用时，其有更大

的概率得到更多的专利间接被引量。

图 3b 呈现了被专利直接引用的论文数和被专利间接引用的论文数。在本文的样本中，仅有 11.4%（3399237 篇）的论文被专利直接引用，而被专利间接引用的论文则达到了 39.7%（11724598 篇）。从平均专利被引量这一角度来看，结果更为显著（图 3c）：论文的平均专利直接被引量仅为 0.84，而专利间接被引量的平均值达到了 24.19，后者约为前者的 29 倍。

考虑发表于顶级期刊（*Nature*、*Science*、*Cell*）的研究论文，对比其专利直接被引量和专利间接被引量，结果如图 3d 所示。其中，*Nature*、*Science*、*Cell* 论文的平均专利直接被引量分别为 10.03、13.98、20.46，而其平均专利间接被引量分别为 415.79、515.24、1188.85，后者约为前者的 37~58 倍。进一步地，本文基于 Li 等[52]提供的自然科学领域诺贝尔奖获奖论文数据，共匹配出 205 篇论文，分析这些论文的专利直接被引量和专利间接被引量，结果如图 3e 所示。物理学、化学和医学领域诺贝尔奖获奖论文的平均专利直接被引量分别为 60.48、359.03、570.16，而其平均专利间接被引量分别为 1637.87、17863.76、11333.87，后者约为前者的 20~50 倍。以上结果表明，专利间接被引量指标能够有效测度论文的高阶技术影响，并且对于高水平论文和突破性研究的技术影响测度更为有效。

5 回归分析

5.1 控制变量与模型设定

为分析新颖性对于论文的技术影响的关系，本文采取回归分析的方法，将引文组合新颖性和学科新颖性作为自变量，将论文的专利直接被引量和专利间接被引量作为因变量，基于 OLS 进行回归分析。为排除干扰因素的影响，回归模型需要考虑相关的控制变量。首先，加入年份和领域的固定效应，本文针对 1960—2016 年的 57 个年份和 292 个 MAG 二级领域分别生成哑变量，将其作为固定效应加入模型，以控制论文在不同时间维度、学科领域的异质性。其次，本文还考虑期刊层次的固定效应，基于 MAG 数据库中 39893 种学术期刊生成哑变量，将其作为固定效应加入模型，以控制论文质量、类别和风格对于变量相关性的干扰。

论文层次的控制变量包括参考文献数量[46]和基金资助情况[3]。其中，参考文献数量是计数变量，在模型中进行对数化处理。基金资助是一个 0-1 变量，定义为该论文是否受美国国家科学基金会（National Science Foundation，United States，NSF）或美国国立卫生研究院（National Institutes of Health，NIH）基金资助[3]。

作者层次的控制变量包括作者的最高机构排名、作者的平均学术年龄、作者的平均发文量、作者的平均“高被引”论文发表数量[53]。其中，机构排名基于 MAG 中机构的总被引量进行排名；作者的学术年龄、发文量和“高被引”论文发表数量等变量均基于截至论文发表当年的数据计算，“高被引”论文指的是在论文所属学科（292 个 MAG 二级学科）和年份发表的论文中被引量排名前 5% 的论文，作者层次的控制变量在模型中均进行对数化处理。

团队层次的控制变量包括团队规模（即作者人数）、团队中的国家数量、团队中的学者领域数量[54]。其中，学者领域依据每个作者学术生涯中的主要领域（292 个 MAG 二级学科）确定[54]，团队层次的控制变量在模型中均进行对数化处理。

本文基于以下 OLS 回归模型

$$\text{Patent citation}_i = \alpha + \beta_1\left(\text{Novelty}_i\right) + \sum_{k=2}^{K}\beta_k\left(\text{Control}_{k,i}\right) + \gamma_t Y_t + \delta_f F_f + \varepsilon_j J_j + \varepsilon_i \quad (5)$$

分析新颖性与论文的技术影响的关系。其中，i 代表论文；t 代表论文发表年份；f 代表论文所属领域；j 代表论文所属期刊；Patent citation_i 包含专利直接被引量和专利间接被引量；Novelty_i 包含引文组合新颖性和学科新颖性；Y_t 代表年份固定效应；F_f 代表学科固定效应；J_j 代表期刊固定效应；ε_i 代表误差项。

5.2 新颖性与论文的技术影响

各项指标的描述性统计如表 2 所示。其中，引文组合新颖性的均值为 –0.40，标准差为 2.64，这表明大部分论文的科学知识组合真实分布和模拟分布的 Z 值的第 10 百分位数为正值，反映出大部分论文的知识组合是常见的，新颖性的知识组合则是较为罕见的。学科新颖性的均值为 0.20，标准差为 0.08，值分布在 0~0.43。

为研究新颖性与论文的技术影响的关系，本文基于 OLS 模型将引文组合新颖性对专利直接被引量和专利间接被引量进行回归，结果如表 3 所示。模型(1)和模型(5)仅考虑控制年份固定效应和领域固

表 2　各项指标的描述性统计

类别	变量	样本量	均值	标准差	最小值	中位数	最大值
自变量	引文组合新颖性	29877430	−0.40	2.64	−11.54	−0.33	5.92
	学科新颖性	29877430	0.20	0.08	0.00	0.21	0.43
因变量	ln(专利直接被引量+1)	29877430	0.16	0.54	0.00	0.00	10.28
	ln(专利间接被引量+1)	29877430	0.97	1.53	0.00	0.00	13.56
控制变量	ln(参考文献数量)	29877430	3.06	0.75	1.61	3.09	7.51
	基金资助	29877430	0.08	0.27	0.00	0.00	1.00
	ln(机构排名)	29877430	4.08	2.95	0.00	4.86	10.20
	ln(学术年龄)	29877430	2.00	0.96	0.00	2.25	4.09
	ln(发文量+1)	29877430	2.72	1.47	0.00	2.98	6.64
	ln(高被引数量+1)	29877430	1.01	1.03	0.00	0.69	5.63
	ln(团队规模)	29877430	1.14	0.68	0.00	1.10	8.54
	ln(国家数量)	29877430	0.10	0.28	0.00	0.00	4.36
	ln(团队作者领域)	29877430	0.58	0.53	0.00	0.69	4.04

定效应的回归，模型(2)和模型(6)代表加入期刊固定效应和论文层次控制变量的回归，模型(3)和模型(7)代表进一步加入作者层次控制变量的回归，模型(4)和模型(8)进一步考虑了团队层次控制变量，即包含全部控制变量的回归。

由于引文组合新颖性和学科新颖性的值域和分布有较大差异，本文基于标准差的变动解释模型结果。引文组合新颖性的标准差为2.64，因此，在模型(1)和模型(5)中，引文组合新颖性提升1个标准差，会促进专利直接被引量提升约3.72%，促进专利间接被引量提升约17.53%。加入期刊固定效应、论文层次控制变量和作者层次控制变量后，在模型(3)和模型(7)中，引文组合新颖性提升1个标准差，会促进专利直接被引量提升约2.06%，促进专利间接被引量提升约6.55%。考虑全部控制变量后，在模型(4)和模型(8)中，引文组合新颖性提升1个标准差，会促进专利直接被引量提升约1.93%，促进专利间接被引量提升约6.13%。

由此可见，引文组合新颖性对专利直接被引量和专利间接被引量均有着显著的正向促进效应，而且对专利间接被引量的促进效应更强。另外，从控制变量中可以发现，参考文献数量、基金资助、机构排名、“高被引”论文发表数量、团队规模和团队作者领域数量对论文的技术影响有显著的促进作用，这表明了政策资助、作者声誉、跨学科合作等因素对科学产业化应用的积极作用。

表4呈现了学科新颖性与论文的技术影响的关系。基于标准差的变动解释模型结果，学科新颖性的标准差为0.08，因此，在模型(1)和模型(5)中，学科新颖性提升1个标准差，会促进专利直接被引量提升约1.09%，促进专利间接被引量提升约9.25%。加入期刊固定效应、论文层次控制变量后，在模型(2)和模型(6)中，学科新颖性提升1个标准差，仅能促进专利直接被引量提升约0.03%，促进专利间接被引量提升约1.45%。加入作者层次控制变量后，在模型(3)和模型(7)中，学科新颖性与专利直接被引量的关系变为不显著，且系数变为负值。另外，学科新颖性提升1个标准差，会促进专利间接被引量提升约1.24%。考虑全部控制变量后，在模型(4)和模型(8)中，学科新颖性提升1个标准差，会导致专利直接被引量下降约0.12%，促进专利间接被引量提升约0.94%。

加入作者层次或团队层次的控制变量后，学科新颖性对专利直接被引量并不存在显著的正向关系，因此，缺乏足够的证据表明学科新颖性与论文的直接技术影响存在正向相关性。但学科新颖性与论文的间接技术影响则存在显著的正向关系，这体现出跨学科性的研究往往需要更长的影响路径才能体现出其价值。

比较引文组合新颖性和学科新颖性对论文的技术影响的效应强度，发现引文组合新颖性不仅呈现更一致和稳健的正向促进关系，而且在效应上更强。因此，在一定程度上可以认为引文组合新颖性比学科新颖性更能够反映论文的质量和潜在价值。

5.3　时间、团队规模和领域异质性

为进一步对不同年份、不同团队规模以及不同

表3 引文组合新颖性与论文的技术影响的正向关系

模型	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	预测变量:ln(专利直接被引量+1)				预测变量:ln(专利间接被引量+1)			
引文组合新颖性	0.0141^{***} (3.94×10^{-5})	0.0078^{***} (4.06×10^{-5})	0.0078^{***} (4.05×10^{-5})	0.0073^{***} (4.06×10^{-5})	0.0664^{***} (9.87×10^{-5})	0.0248^{***} (9.23×10^{-5})	0.0248^{***} (9.16×10^{-5})	0.0232^{***} (9.17×10^{-5})
ln(参考文献数量)		0.0441^{***} (0.0002)	0.0398^{***} (0.0002)	0.0397^{***} (0.0002)		0.2718^{***} (0.0004)	0.2469^{***} (0.0004)	0.2460^{***} (0.0004)
基金资助		0.0624^{***} (0.0005)	0.0453^{***} (0.0005)	0.0422^{***} (0.0005)		0.2281^{***} (0.0011)	0.1445^{***} (0.0011)	0.1331^{***} (0.0011)
ln(机构排名)			0.0004^{***} (3.44×10^{-5})	0.0003^{***} (3.44×10^{-5})			0.0047^{***} (7.95×10^{-5})	0.0043^{***} (7.93×10^{-5})
ln(学术年龄)			-0.0028^{***} (0.0002)	-5.60×10^{-5} (0.0002)			0.0019^{***} (0.0005)	0.0137^{***} (0.0005)
ln(发文量+1)			-0.0153^{***} (0.0002)	-0.0206^{***} (0.0002)			-0.0517^{***} (0.0004)	-0.0722^{***} (0.0004)
ln(高被引数量+1)			0.0454^{***} (0.0002)	0.0488^{***} (0.0002)			0.2074^{***} (0.0004)	0.2199^{***} (0.0004)
ln(团队规模)				0.0334^{***} (0.0002)				0.1383^{***} (0.0005)
ln(国家数量)				-0.0317^{***} (0.0004)				-0.1086^{***} (0.0008)
ln(团队作者领域)				0.0074^{***} (0.0002)				0.0107^{***} (0.0006)
年份FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
领域FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
期刊FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
样本量	29585145	29585145	29585145	29585145	29585145	29585145	29585145	29585145
R^2	0.06015	0.17155	0.17412	0.17546	0.24194	0.46660	0.47432	0.47674

注:所有模型均基于OLS回归,括号内报告了稳健标准误;*表示$p<0.05$,**表示$p<0.01$,***表示$p<0.001$;FE表示固定效应(fixed effects)。

学科的论文进行分析，本文进行细粒度的分样本回归分析。如图4所示，每一个点均表示OLS回归模型中引文组合新颖性或学科新颖性的估计系数，模型中均加入全部控制变量和固定效应，误差棒表示基于稳健标准误的95%置信区间，蓝色回归点表示对论文的专利直接被引量的回归结果（取对数），橙色回归点表示对论文的专利间接被引量的回归结果（取对数）。

图4a~图4b呈现了时间维度的分样本回归结果。以10年为间隔，将样本划分为6个阶段，分别对各个阶段的样本进行回归。引文组合新颖性对专利直接被引量和专利间接被引量的效应在所有时间区间均显著为正，并且在1990年之后存在效应减弱的趋势。学科新颖性对专利直接被引量的效应较小且体现出不显著或负的系数，学科新颖性对专利间接被引量的效应则一直存在减弱的趋势，2000年之前学科新颖性对专利间接被引量的效应显著为正，而2000年之后学科新颖性对专利间接被引量的效应显著为负。

图4c~图4d展现了基于团队规模的分样本回归结果。引文组合新颖性对专利直接被引量和专利间接被引量的效应在所有团队规模中均显著为正，且效应大小随着团队规模而增加。学科新颖性对专利直接被引量和专利间接被引量的效应大小也随着团队规模而增加。值得注意的是，尽管学科新颖性对专利直接被引量的总体回归系数不显著或为负，但

表 4　学科新颖性与论文的技术影响的关系

模型	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	预测变量：ln(专利直接被引量+1)				预测变量：ln(专利间接被引量+1)			
学科新颖性	0.1357***	0.0038**	−0.0002	−0.0152***	1.156***	0.1813***	0.1546***	0.1169
	(0.0011)	(0.0012)	(0.0012)	(0.0012)	(0.0029)	(0.0028)	(0.0028)	(0.0028)
ln(参考文献数量)		0.0527***	0.0485***	0.0483***		0.2939***	0.2698***	0.2682***
		(0.0002)	(0.0002)	(0.0002)		(0.0004)	(0.0004)	(0.0004)
基金资助		0.0638***	0.0468***	0.0435***		0.2325***	0.1492***	0.1370***
		(0.0005)	(0.0005)	(0.0005)		(0.0011)	(0.0011)	(0.0011)
ln(机构排名)			0.0004***	0.0003***			0.0048***	0.0043***
			(3.44×10^{-5})	(3.44×10^{-5})			(7.96×10^{-5})	(7.94×10^{-5})
ln(学术年龄)			−0.0030***	−0.0004*			0.0009*	0.0126***
			(0.0002)	(0.0002)			(0.0005)	(0.0005)
ln(发文量+1)			−0.0150***	−0.0207***			−0.0508***	−0.0722***
			(0.0002)	(0.0002)			(0.0004)	(0.0004)
ln(高被引数量+1)			0.0453***	0.0489***			0.2069***	0.2199***
			(0.0002)	(0.0002)			(0.0004)	(0.0004)
ln(团队规模)				0.0326***				0.1375***
				(0.0002)				(0.0005)
ln(国家数量)				−0.0324***				−0.1107***
				(0.0004)				(0.0008)
ln(团队作者领域)				0.0118***				0.0204***
				(0.0003)				(0.0006)
年份 FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
领域 FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
期刊 FE	No	Yes	Yes	Yes	No	Yes	Yes	Yes
样本量	29585145	29585145	29585145	29585145	29585145	29585145	29585145	29585145
R^2	0.05608	0.17046	0.17303	0.17451	0.23327	0.4653	0.473	0.47557

注：所有模型均基于 OLS 回归，括号内报告了稳健标准误；*表示 $p<0.05$，**表示 $p<0.01$，***表示 $p<0.001$；FE 表示固定效应。

其在大团队科学中对于专利直接被引量存在着显著的正向促进作用；在小团队中，学科新颖性对专利直接被引量和专利间接被引量的效应均显著为负。

图 4e~图 4f 呈现了领域层次的分样本回归结果。本文分别对 MAG 中一级领域分类进行分样本回归，结果呈现明显的领域异质性。引文组合新颖性对专利直接被引量的影响效应在艺术、哲学、历史等领域并不显著，而在 STEM（science, technology, engineering, mathematics）领域则显著为正（除了计算机科学领域中，效应为负），而引文组合新颖性对专利间接被引量的影响效应在所有领域均显著为正，且在 STEM 领域的效应更强。学科新颖性对专利直接被引量的影响效应仅仅在环境科学、地质学、心理学和医学领域显著为正，而其他领域均不显著或为负值，而学科新颖性对专利间接被引量的影响效应在大部分领域显著为正。

6　结论与启示

本文以“科学-技术”为切入点，通过融合“论文-专利”引用网络和“论文-论文”深层引用网络，从“直接-间接”技术影响的双重视角出发，回答了组合新颖性和学科新颖性是否能提升论文的技术影响这一问题。本文得出了以下主要结论：①引文组合新颖性能够显著促进论文的直接技术影响和间接技术影响；而学科新颖性仅能显著促进论文的间接技术影响，其与专利直接被引量并不存在一致的正向关系。②学科新颖性对论文的间接技术影响的促进效应随年份存在降低的趋势，表明科学领域

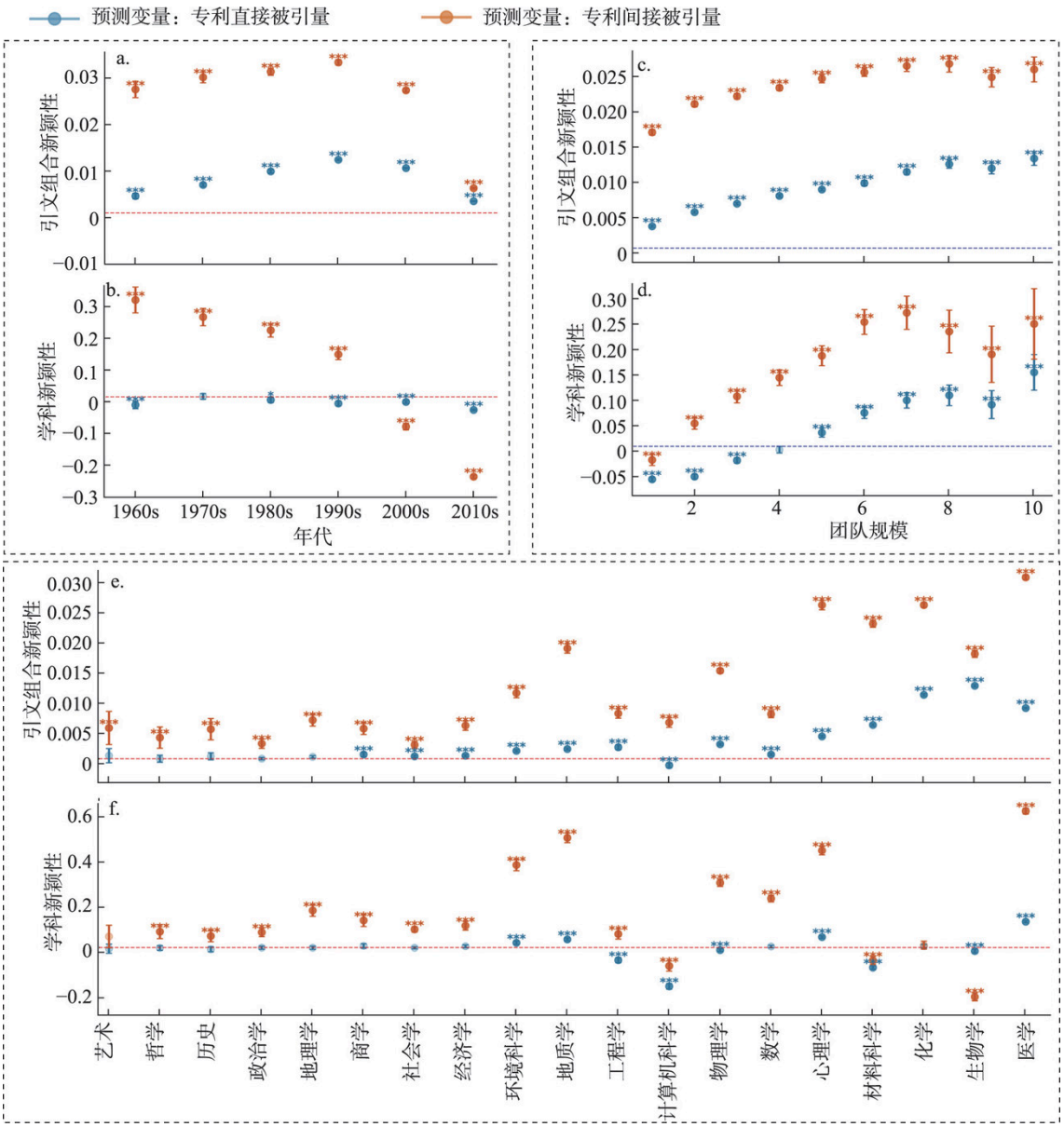


图4 时间-领域-团队规模异质性分析(彩图请见 https://qbxb.istic.ac.cn)

误差棒表示95%置信区间;*表示p<0.05,**表示p<0.01,***表示p<0.001;

1960s表示20世纪60年代,1970s表示20世纪70年代,以此类推。

逐渐可能更难从新颖的知识组合中受益，也可能是由于新颖的科学研究突破技术边界需要更长的时间窗口。③在STEM领域和大型合作团队中，引文组合新颖性和学科新颖性能够更有效地提升论文的双重技术影响，这可能是由于STEM领域和大团队往往更擅长利用跨学科的优势，激发更有价值和高质量的研究，而人文社科领域和小团队则较难从新颖性的知识组合或跨学科的知识融合中受益，专注而狭窄的知识领域更可能是这种科学群体的典型特征。

基于本文的结论，提出如下政策建议：①非典型的科学知识组合或远距离的融合能够带来技术的可持续发展。因此，政府和科研指导机构要重点支持原创性的研究，同时在政策上进一步鼓励跨学科的合作研究，加快建设新兴学科，以期在未来促进技术产业的持续进步。②大团队科研中，引文组合新颖性和学科新颖性能够更有效地提升论文的双重技术影响，因此，大科学团队要更加重视事前的创新，争取提升知识搜索的广度和深度。③要实现技术领域的可持续创新和突破性发展，应当从政策上支持和鼓励基础科学，尽管科学创新在短期难以体现出明显的技术影响，其往往能在将来成为关键技

术突破的核心要素。

最后，对本文的不足之处进行总结，并为未来的研究指明方向。①论文的专利间接被引量可以进一步拓展为专利 n 阶被引量，即 $C^{n}_{\text{patent}}(i)$，尽管其对应的计算量会显著提高，但扩充高阶被引形成的引用级联可以进一步丰富网络信息，未来研究可对此进行分析。②本文的结论并不能体现变量间的因果性，未来可以进行更严谨的研究设计，基于因果推断方法分析科学创新对技术的促进效应。③本文没有考虑内容层次的科学新颖性，未来研究可以结合实体抽取技术和大语言模型进行更细粒度的研究。④新颖性作为事前指标测量科学的创新性，而非事后分析的指标，未来研究可以基于颠覆性等指标[55-56]进一步分析科学创新对技术应用的影响。

参　考　文　献

[1] Azoulay P, Graff-Zivin J, Uzzi B, et al. Toward a more scientific science[J]. Science, 2018, 361(6408): 1194-1197.

[2] 刘伟. 科学认识与切实发展新质生产力[J]. 经济研究, 2024, 59(3): 4-11.

[3] Yang A J. Unveiling the impact and dual innovation of funded research[J]. Journal of Informetrics, 2024, 18(1): 101480.

[4] Green J R, Scotchmer S. On the division of profit in sequential innovation[J]. The RAND Journal of Economics, 1995, 26(1): 20-33.

[5] Lee F. Recombinant uncertainty in technological search[J]. Management Science, 2001, 47(1): 117-132.

[6] Furman J L, Stern S. Climbing atop the shoulders of giants: the impact of institutions on cumulative research[J]. American Economic Review, 2011, 101(5): 1933-1963.

[7] Fortunato S, Bergstrom C T, Börner K, et al. Science of science [J]. Science, 2018, 359(6379): eaao0185.

[8] Yin Y A, Dong Y X, Wang K S, et al. Public use and public funding of science[J]. Nature Human Behaviour, 2022, 6(10): 1344-1350.

[9] Yin Y A, Gao J, Jones B F, et al. Coevolution of policy and science during the pandemic[J]. Science, 2021, 371(6525): 128-130.

[10] Funk R J, Owen-Smith J. A dynamic network measure of technological change[J]. Management Science, 2017, 63(3): 791-817.

[11] Wang D S, Barabási A L. The science of science[M]. Cambridge: Cambridge University Press, 2021.

[12] Lane J, Bertuzzi S. Measuring the results of science investments [J]. Science, 2011, 331(6018): 678-680.

[13] Baruffaldi S, Simeth M, Wehrheim D. Asymmetric information and R&D disclosure: evidence from scientific publications[J]. Management Science, 2023, 70(2): 1052-1069.

[14] Ahmadpoor M, Jones B F. The dual frontier: patented inventions and prior scientific advance[J]. Science, 2017, 357(6351): 583-587.

[15] 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.

[16] Schumpeter J A. Business cycles: a theoretical, historical, and statistical analysis of the capitalist process[M]. New York: McGraw-Hill, 1939.

[17] Hofstra B, Kulkarni V V, Munoz-Najar Galvez S, et al. The diversity-innovation paradox in science[J]. Proceedings of the National Academy of Sciences of the United States of America, 2020, 117(17): 9284-9291.

[18] Trapido D. How novelty in knowledge earns recognition: the role of consistent identities[J]. Research Policy, 2015, 44(8): 1488-1500.

[19] Bush V. Science: the endless frontier[M]. Washington: United States Government Printing Office, 1945.

[20] Evans J A. Electronic publication and the narrowing of science and scholarship[J]. Science, 2008, 321(5887): 395-399.

[21] Uzzi B, Mukherjee S, Stringer M, et al. Atypical combinations and scientific impact[J]. Science, 2013, 342(6157): 468-472.

[22] Liu M J, Bu Y, Chen C Y, et al. Pandemics are catalysts of scientific novelty: evidence from COVID-19[J]. Journal of the Association for Information Science and Technology, 2022, 73(8): 1065-1078.

[23] Liu M J, Xie Z H, Yang A J, et al. The prominent and heterogeneous gender disparities in scientific novelty: evidence from biomedical doctoral theses[J]. Information Processing & Management, 2024, 61(4): 103743.

[24] Kim D, Cerigo D B, Jeong H, et al. Technological novelty profile and invention's future impact[J]. EPJ Data Science, 2016, 5(1): Article No.8.

[25] Wu L F, Wang D S, Evans J A. Large teams develop and small teams disrupt science and technology[J]. Nature, 2019, 566 (7744): 378-382.

[26] Park M, Leahey E, Funk R J. Papers and patents are becoming less disruptive over time[J]. Nature, 2023, 613(7942): 138-144.

[27] Wang J, Veugelers R, Stephan P. Bias against novelty in science: a cautionary tale for users of bibliometric indicators[J]. Research Policy, 2017, 46(8): 1416-1436.

[28] Bornmann L, Tekles A, Zhang H H, et al. Do we measure novelty when we analyze unusual combinations of cited references? A validation study of bibliometric novelty indicators based on F1000Prime data[J]. Journal of Informetrics, 2019, 13(4): 100979.

[29] Stirling A. A general framework for analysing diversity in science, technology and society[J]. Journal of the Royal Society Interface, 2007, 4(15): 707-719.

[30] Fontana M, Iori M, Montobbio F, et al. New and atypical combinations: an assessment of novelty and interdisciplinarity[J]. Research Policy, 2020, 49(7): 104063.

[31] Foster J G, Rzhetsky A, Evans J A. Tradition and innovation in scientists' research strategies[J]. American Sociological Review, 2015, 80(5): 875-908.

[32] 李茂林, 王子路, 何光辉, 等. 银行业金融科技创新、结构性普惠效应与创业活力[J]. 管理世界, 2024, 40(6): 195-224.

[33] Krieger J, Li D, Papanikolaou D. Missing novelty in drug development[J]. The Review of Financial Studies, 2022, 35(2): 636-679.

[34] Shi F, Evans J. Surprising combinations of research contents and contexts are related to impact and emerge with scientific outsiders from distant disciplines[J]. Nature Communications, 2023, 14 (1): 1641.

[35] Onodera N, Yoshikane F. Factors affecting citation rates of research articles[J]. Journal of the Association for Information Science and Technology, 2015, 66(4): 739-764.

[36] Wang D S, Song C M, Barabási A L. Quantifying long-term scientific impact[J]. Science, 2013, 342(6154): 127-132.

[37] Azoulay P, Li D, Graff Zivin J S, et al. Public R&D investments and private-sector patenting: evidence from NIH funding rules[J]. The Review of Economic Studies, 2019, 86(1): 117-152.

[38] Waltman L. A review of the literature on citation impact indicators[J]. Journal of Informetrics, 2016, 10(2): 365-391.

[39] 梁镇涛, 毛进, 李纲. 融合"科学-技术"知识关联的高颠覆性专利预测方法[J]. 情报学报, 2023, 42(6): 649-662.

[40] Li D, Azoulay P, Sampat B N. The applied value of public investments in biomedical research[J]. Science, 2017, 356(6333): 78-81.

[41] Williams H L. Intellectual property rights and innovation: evidence from the human genome[J]. The Journal of Political Economy, 2010, 121(1): 1-27.

[42] Lee Y N, Walsh J P, Wang J. Creativity in scientific teams: unpacking novelty and impact[J]. Research Policy, 2015, 44(3): 684-697.

[43] Wang K S, Shen Z H, Huang C Y, et al. Microsoft Academic Graph: when experts are not enough[J]. Quantitative Science Studies, 2020, 1(1): 396-413.

[44] Marx M, Fuegi A. Reliance on science by inventors: hybrid extraction of in-text patent-to-article citations[J]. Journal of Economics & Management Strategy, 2022, 31(2): 369-392.

[45] Gates A J, Barabási A L. Reproducible science of science at scale: pySciSci[J]. Quantitative Science Studies, 2023, 4(3): 700-710.

[46] Zeng A, Shen Z S, Zhou J L, et al. The science of science: from the perspective of complex systems[J]. Physics Reports, 2017, 714: 1-73.

[47] Cimini G, Squartini T, Saracco F, et al. The statistical physics of real-world networks[J]. Nature Reviews Physics, 2019, 1: 58-71.

[48] 刘嘉明, 孙建军. 参考文献跨学科性与论文学术影响力的关系研究[J]. 情报学报, 2023, 42(5): 525-536.

[49] Yang Y, Tian T Y, Woodruff T K, et al. Gender-diverse teams produce more novel and higher-impact scientific ideas[J]. Proceedings of the National Academy of Sciences of the United States of America, 2022, 119(36): e2200841119.

[50] 孔嘉, 邓三鸿, 张佳锐, 等. 技术创新中的知识融合: "科学——技术"二元知识对专利影响研究[J]. 情报学报, 2022, 41(11): 1161-1173.

[51] Lin Z H, Yin Y A, Liu L, et al. SciSciNet: a large-scale open data lake for the science of science research[J]. Scientific Data, 2023, 10: Article No.315.

[52] Li J C, Yin Y A, Fortunato S, et al. A dataset of publication records for Nobel laureates[J]. Scientific Data, 2019, 6(1): 33.

[53] Brogaard J, Engelberg J E, Eswar S K, et al. On the causal effect of fame on citations[J]. Management Science, 2024, 70(10): 7187.

[54] Lin Y L, Frey C B, Wu L F. Remote collaboration fuses fewer breakthrough ideas[J]. Nature, 2023, 623(7989): 987-991.

[55] 杨杰, 邓三鸿, 王昊. 科学研究的颠覆性创新测度——相对颠覆性指数[J]. 情报学报, 2023, 42(9): 1052-1064.

[56] 杨杰, 孔嘉, 张艺炜, 等. 融合论文颠覆性与巩固性的学者二元影响力测度[J]. 情报学报, 2023, 42(12): 1412-1423.

（责任编辑 王克平）