情报学报　2024 年 9 月　第 43 卷　第 9 期
Journal of the China Society for Scientific and Technical Information, Sept. 2024, 43(9): 1003-1014

DOI: 10.3772/j.issn.1000-0135.2024.09.001

基于参考文献的论文跨学科性、跨时域性及其影响力研究

杨　杰[1,2]，王左戎[1,2]，邓三鸿[1,2]，王　昊[1,2]，章学周[3]

（1. 南京大学信息管理学院，南京　210023；2. 数据工程与知识服务省高校重点实验室（南京大学），南京　210023；
3. 江苏省质量和标准化研究院，南京　210004）

摘　要　本文在跨学科性概念的基础上，提出了跨时域性的测度框架，并将其量化为时域丰富性、时域非均衡性、时域差异性和时域深度4个维度。本文对微软学术数据库中1950—2020年38879575篇论文的跨学科性和跨时域性的时间趋势、分布特征以及科学家的生涯演化模式进行了细致分析，并进一步研究了跨学科性、跨时域性与科学影响力的关系。研究发现：①论文的跨学科性和跨时域性呈现一致的增长趋势和不一致的领域分布；②跨学科性和跨时域性指标具有无标度分布等异质性分布特征，且跨学科性和跨时域性的相关性并不强；③科学家生涯中后期的跨学科性和跨时域性明显增强，这种趋势在很大程度上可以归因于其引用行为的变化；④跨学科性和跨时域性对科学影响力有着相反的作用效应，即跨学科性显著促进科学影响力，跨时域性显著抑制科学影响力，较强的跨学科性和较弱的跨时域性组合成为颠覆性热点论文的概率最高。本文提出的跨时域性测度框架丰富了知识融合和跨学科理论，并为科技政策和学术评价提供了启示。

关键词　跨学科性；跨时域性；引文网络；科学家生涯

Interdisciplinarity, Temporal Diversity, and Scientific Impact: Perspective on References

Yang Alex J.[1,2], Wang Zuorong[1,2], Deng Sanhong[1,2], Wang Hao[1,2] and Zhang Xuezhou[3]

(1. *School of Information Management, Nanjing University, Nanjing*　210023; 2. *Key Laboratory of Data Engineering and Knowledge Services in Provincial Universities (Nanjing University), Nanjing*　210023;
3. *Jiangsu Institute of Quality and Standardization, Nanjing*　210004)

Abstract: Building on the concept of interdisciplinarity, this study proposes a measurement framework for temporal diversity, quantifying it into four dimensions: temporal richness, temporal imbalance, temporal disparity, and temporal depth. The study meticulously analyzes the temporal trends and distribution characteristics of interdisciplinarity and temporal diversity, as well as the career evolution patterns of high-level scientists, using 38,879,575 scientific papers from the Microsoft Academic Graph database spanning from 1950 to 2020. The relationship between interdisciplinarity, temporal diversity, and scientific impact is further explored. The findings reveal that: (1) The interdisciplinary and temporal diversity of papers exhibits a consistent growth trend that differs depending on the field. (2) Both interdisciplinarity and temporal diversi-

收稿日期：2023-10-08；**修回日期**：2024-07-12

基金项目：科睿唯安创新评价开放基金项目“基于科学颠覆性的多层次学术实体评价研究”（CXPJ2024004）。

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、复杂网络、创新测度；王左戎，女，1998年生，博士研究生，主要研究领域为信息组织与检索、科学计量；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；王昊，男，1981年生，博士，教授，博士生导师，主要研究领域为文本挖掘、知识图谱；章学周，男，1969年生，高级工程师，主要研究领域为标准分析、知识组织。

ty indicators show heterogeneous distribution characteristics, such as scale-free distribution, and a weak correlation exists between them. (3) In the careers of high-level scientists, both interdisciplinarity and temporal diversity show significant growth trends, largely attributable to an increase in the number of their referenced papers. (4) Interdisciplinarity and temporal diversity have opposite effects on scientific impact with interdisciplinarity significantly promoting scientific impact and temporal diversity significantly inhibiting it, suggesting that a combination of strong interdisciplinarity and weak temporal diversity has the highest probability of becoming a disruptive hotspot paper. The temporal diversity measurement framework proposed in this study enriches the theories of knowledge integration and interdisciplinarity, providing insights into science and technology policy and academic evaluation.

Keywords: interdisciplinarity; temporal diversity; citation network; scientific career

0 引 言

具有跨学科特征的研究模式被认为能够诞生更多的新发现、新见解、新概念和新技术[1]，跨学科性体现了科学知识领域分布的复杂性与丰富性[2]，反映了科学知识的横向维度[3]。与之对应的是知识基础的纵向维度，其表达了现有研究从广阔而丰富的知识网络中追根溯源的过程[4]。从横向和纵向双重视角来看，全面掌握从知识基础到科学影响力的内在规律，对于理解知识流动的内在机理具有重要意义，同时，对于指导学术评价工作和制定科技政策具有一定的参考价值。

在科学引文网络中，前向引用和后向引用能够反映科学知识的流动和继承。当研究主体发生变化时，引用关系的含义也会变化，即一篇论文的前向引用链接也能被视为另一篇论文的前向引用链接[5]。与引文相关的分析通常以论文的未来引文为主，如动态被引预测[6]、引文扩散、科学睡美人[7]、文献老化、科学家影响力[8]等，这些模型或方法都反映了科学知识的未来扩散或未来影响。与之对应地，对于科学知识的继承则可以基于论文的参考文献进行分析，其中一个典型的分析框架是跨学科的测度[1-2,9]，其充分考虑了参考文献在不同科学领域的分布，体现了科学知识的跨学科特征。本文将跨学科性视为一种横向分析视角，将参考文献的时间分布视为一种纵向分析视角。

当前，学界对科学知识跨学科性这一横向视角的理解较为深入[1-2,9]，但缺乏对科学知识纵向视角的系统性理解。一方面，参考文献是历史的和静态的，不具有被预测的价值，其分析潜力往往被低估；另一方面，现有研究通常只针对参考文献中蕴含的领域或知识主题等信息进行研究，除跨学科测度以外，科学学领域的创新组合理论通常只考虑了不同知识主题的组合[10]，往往忽略了参考文献在时间维度的组合。事实上，科学领域是非常注重延续性和历史继承的[11]，科学知识的纵向维度相当广阔。随着大规模文献数据库的发展和应用，科学家不仅需要考虑如何在不同领域的知识中进行取样和组合，也需要决定如何在不同年份的知识中进行取样和组合[12]。知识的横向分布和纵向分布对于论文的质量和未来影响都至关重要。然而，目前学界尚没有针对科学知识纵向组合的系统性研究，这在一定程度上是因为依然缺乏一个系统的分析框架，以量化论文参考文献的时间分布特征。

为此，本文以跨学科性概念为基准，开创性地提出跨时域性的测度框架，具体包含时域丰富性、时域非均衡性、时域差异性和时域深度四大维度。跨时域性的测度框架能够全方位刻画科学知识的纵向分布特征，并且与跨学科性框架一一对应、互为补充、有机结合。在此基础上，基于科学知识基础的横向分布和纵向分布双重视角，本文针对微软学术数据库（Microsoft Academic Graph，MAG）中1950—2020年19个领域的38879575篇论文进行大规模的实证分析，揭示了科学跨学科性和跨时域性的时间趋势、领域差异、分布特征、相互关系、科学家生涯演化模式；并进一步挖掘跨学科性、跨时域性与科学影响力的复杂关系，以期从理论研究层面丰富知识融合和创新理论，在现实实践层面为科技政策和学术评价提供启示。

1 研究现状与数据

1.1 研究现状

正式学术交流场景中，学者通常选择引用他人学术成果的形式以搭建自身研究的理论或者方法基础，其引用的参考文献反映了科学的知识基础[13]。以参考文献为视角，科学知识基础可以从领域/内容维度和时间维度两个方面展开，即横向扩散的知识领域分布特征和纵向演化的知识时间分布特征，

前者形成了跨学科性这一前沿主题的研究基础[1-2,9]，而后者可以被定义为论文的跨时域性。

论文的跨学科性主要表现为其知识基础领域分布的复杂性程度[14]。跨学科概念通常认为源自生物学的多样性概念，以知识融合为视角，从参考文献维度审视学科的多样性，使最初用于衡量生物多样性的框架能够应用于科学学的情境[1]。Stirling[2]提出跨学科研究理论框架和指标，并指出跨学科性至少应当涉及多样性、均衡性和差异性等维度。Leydesdorff 等[15]选择香农熵、中介中心性和距离矩阵用于评估论文的跨学科性，并提出单一指标可能无法满足文献跨学科性识别工作需要的结论；Rafols 等[16]进一步细化分析粒度，在多元化的基础上补充连贯性的分析维度，用于全面推进对新兴科学和技术领域的比较研究[17]。有些学者围绕研究参考文献内容维度的影响或应用展开讨论，如 Ren 等[18]面向自然科学领域，通过对获奖研究及其引用出版物在学科分布上的统计分析，进一步探究跨学科知识整合的重要影响；Ke[19]侧重于通过统计引用文献领域分布的多样性、差异性和均衡性来衡量引文学科分布，并在此基础上分析跨学科研究的社会影响。

本文认为，论文的跨时域性体现为其在时间维度覆盖了多个时间节点的知识。当前学界对于论文跨时域性的相关讨论较为薄弱，主要以参考文献的年龄[20]作为分析对象，通过模拟参考文献年龄分布规律或描述统计其分布特征[21]，关注分布特征与科学影响力之间的关系，以期实现对前瞻性研究的预测。例如，Gupta[22]针对遗传学领域专业发展不同阶段的原始文献中引用的参考文献年龄进行了统计分析，并探讨不同理论概率函数在引用参考文献的年龄密度中的适用性；Jarić 等[23]评估了参考文献的年龄与研究领域内发表率变化之间的关系，并提出了参考文献的相对年龄和前两年出版的参考文献比例两项指标，用于规避出版时间滞后性衍生的误差；Stacey[24]对多领域参考文献的年龄分布进行 γ 分布拟合分析，证实了 γ 分布模型在参考文献年龄分布中的稳定性，并提出派生指标作为预测知识增长的重要参考。此外，Reisz 等[25]也围绕参考文献的年龄分布特征与文献被引用情况的关系展开了统计分析。由此可见，当前关于知识基础纵向维度的特征要素研究较为单一，不具备系统性的科学知识跨时域性测度框架。

对于知识基础横向维度的研究，目前已有系统性的跨学科性分析框架和丰富的测度指标。与之相比，针对知识基础纵向维度，现有研究仍然停留在针对参考文献年龄进行描述统计等浅层次的分析。由此，本文借鉴跨学科的理论与方法，开创性地提出论文跨时域性的测度框架；并且在兼顾论文的跨学科性与跨时域性两项特征前提下，围绕知识融合与科学影响力的关系展开系统性实证分析，以期揭示两者之间的内在机制和作用路径，为科研管理和政策制定提供科学依据。

1.2　数据来源与清洗

本文采用微软学术数据库（MAG）作为论文引用数据[26]，其涵盖了发表于 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。为保证数据质量，筛选出 1950—2020 年发表的期刊论文并去除其他类型的文献，这是因为不同种类文献往往存在不一致的引用模式，并且微软学术数据库提供的期刊标签能够在一定程度上控制论文质量和论文类别的异质性影响。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 38879575 篇论文，以及 909136957 个引用链接。

微软学术数据库通过结合机器学习、众包和网络数据的混合方法解决了作者姓名消歧的问题[26]，具有较高的可信度。最新的一项研究通过人工匹配测试出样本中的准确率为 100%，召回率为 84%[27]。但是微软学术数据库对于亚裔学者姓名的消歧依然存在不少问题，这主要体现在发文量极高的作者中。在科学家生涯层次的分析中，本文通过设置发文量的阈值来控制消歧因素的影响。

微软学术数据库基于深度学习方法为领域分类提供了一个结构化框架，本文主要考虑两个层次：第一层有 19 个广泛的研究领域，包括医学、物理学和计算机科学；第二层进一步划分出 292 个二级领域。微软学术数据库中的每篇论文都通过机器学习模型生成一个或多个领域标签，并附有概率值以表明机器学习生成预测的置信度。本文的分析框架优先考虑每篇论文的二级标签，在多个标签中选择置信度最高的标签。本文的跨学科相关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。

2 测度框架、分布特征与生涯演化

2.1 跨学科性与跨时域性的测度框架

基于 Stirling 提出的框架[2]，本文将论文跨学科性划分为学科丰富性（disciplinary variety）、学科非均衡性（disciplinary imbalance）、学科差异性（disciplinary disparity）等维度，以 Rao-Stirling 指数作为总体跨学科性指标。将论文 i 的参考文献 $R_i = \{r_1, r_2, \cdots\}$ 领域映射为学科分布向量 $\boldsymbol{F}_i = (\text{field}_{r_1}, \text{field}_{r_2}, \cdots, \text{field}_{r_k}, \cdots)$，其中 field_{r_k} 反映了其中参考文献 r_k 的所属领域。学科丰富性由论文 i 参考文献覆盖的领域数量表示，定义为 $\sum_{r \in R_i} \text{field}_r$；学科非均衡性基于论文参考文献学科类别分布的标准化香农熵来表示；学科差异性由论文 i 参考文献中的最大学科距离表示，定义为 $\max\{d_{mn} | m,n \in F_i\}$，其中 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的全部论文及引证网络计算而得，即 $d_{mn} = 1 - \text{CosSim}(V_m, V_n)$；总体跨学科性指标 Rao-Stirling 指数[2]定义为 $\sum_{m \neq n} d_{mn} p_m p_n$，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。

本文首次给出科学知识跨时域性的理论与测度框架：跨时域性指的是科学知识在时空维度的广度、深度、均衡和差异分布情况，可以由论文参考文献的时间分布来具体测度。本文将论文跨时域性划分为时域丰富性（temporal variety）、时域非均衡性（temporal imbalance）、时域差异性（temporal disparity）三大维度，另外，本文考虑参考文献平均年龄作为时域深度（temporal depth）指标。对于论文 i 及其全部参考文献 $R_i = \{r_1, r_2, \cdots\}$，本文将其参考文献的年龄映射为时域分布向量 $\boldsymbol{A}_i = (\text{age}_{r_1}, \text{age}_{r_2}, \cdots, \text{age}_{r_k} \cdots)$，其中 age_{r_k} 代表参考文献 r_k 的年龄。

（1）时域丰富性由论文 i 参考文献覆盖的年龄数量表示，定义为

$$\text{temporal variety} = \sum_{r \in R_i} \text{year}_r \tag{1}$$

（2）时域非均衡性基于论文参考文献学科年龄分布的标准化香农熵来表示，定义为

$$\text{temporal imbalance} = \text{Shannon entropy}_{\text{age}} = -\frac{\sum_{i=1}^{\max(\text{variety})} p_\alpha \ln(p_\beta)}{\ln(\max(\text{variety}))} \tag{2}$$

其中，$\alpha \in A_i$ 表示论文 i 参考文献的某一年龄；p_α 表示该年龄参考文献的概率；$\max(\text{variety})$ 表示时域丰富性最大值，本文中为 178。

（3）时域差异性由论文 i 参考文献中的最大年龄距离表示，定义为

$$\text{temporal disparity} = \max\{|\text{age}_{r_k} - \text{age}_{r_l}| \text{age}_{r_k}, \text{age}_{r_l} \in A_i\} \tag{3}$$

其中，$|\text{age}_{r_k} - \text{age}_{r_l}|$ 表示参考文献 r_k 和参考文献 r_l 之间的差异，等同于其发表年份差之绝对值。

（4）时域深度由论文 i 参考文献中的平均年龄表示，定义为

$$\text{temporal depth} = \text{mean}\{\text{age}_{r_k} | \text{age}_{r_k} \in A_i\} \tag{4}$$

值得注意的是，本文选取了“非均衡性”而非“均衡性”作为跨学科和跨时域测度框架的组成部分，这是主要考虑了标准化香农熵和跨学科性的定义。当一篇论文只引用了某一领域或某一年份的文献，其标准化香农熵为 0，反映出最低的跨学科性或跨时域性。此外，下文也证实了非均衡性与其他指标之间的一致性。

跨学科性和跨时域性的测度框架如图 1 所示，二者分别以知识的横向分布和纵向分布为视角，共同从丰富性、非均衡性、差异性等维度出发，以不同角度刻画科学知识基础的模式和特征。

2.2 跨学科性与跨时域性的时间趋势

本文对 38879575 篇论文的跨学科性指标和跨时域性指标的时间趋势进行分析，结果如图 2 所示。学科丰富性、学科非均衡性、学科差异性和总体跨学科性的 Rao-Stirling 指数均呈现明显的上升趋势。学科丰富性的增长趋势表明，论文呈现越来越强的知识丰富性和知识横向广度；学科非均衡性和差异性的增长趋势表明，尽管学科丰富性增强，少部分的学科知识依然占据较大的成分，且这种异质性在逐渐加强；总体跨学科性的上升趋势反映出科学领域正逐渐从独立细化研究模式走向跨学科的融合模式。

跨时域性呈现与跨学科性较为一致的趋势，即时域丰富性、时域非均衡性、时域差异性和时域深度均呈现明显的上升趋势。这表明随着科学知识的不断积累，论文的知识纵向深度和跨度越来越强，并且体现出越来越强的非均衡性，即少部分时间区间的知识占据了主导成分。

不同科学领域呈现不一致的跨学科性和跨时域性趋势。人文艺术领域的跨学科性指标明显弱于科学工程、生物医药和社会科学等领域，并且随时间上升的幅度较小；社会科学领域的学科丰富性和学

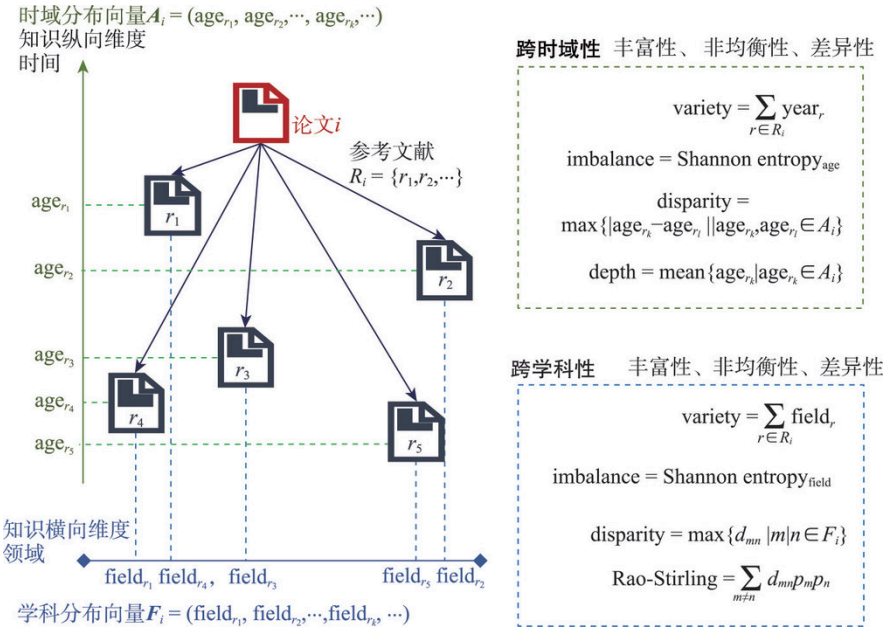


图 1　科学知识跨学科性和跨时域性测度框架

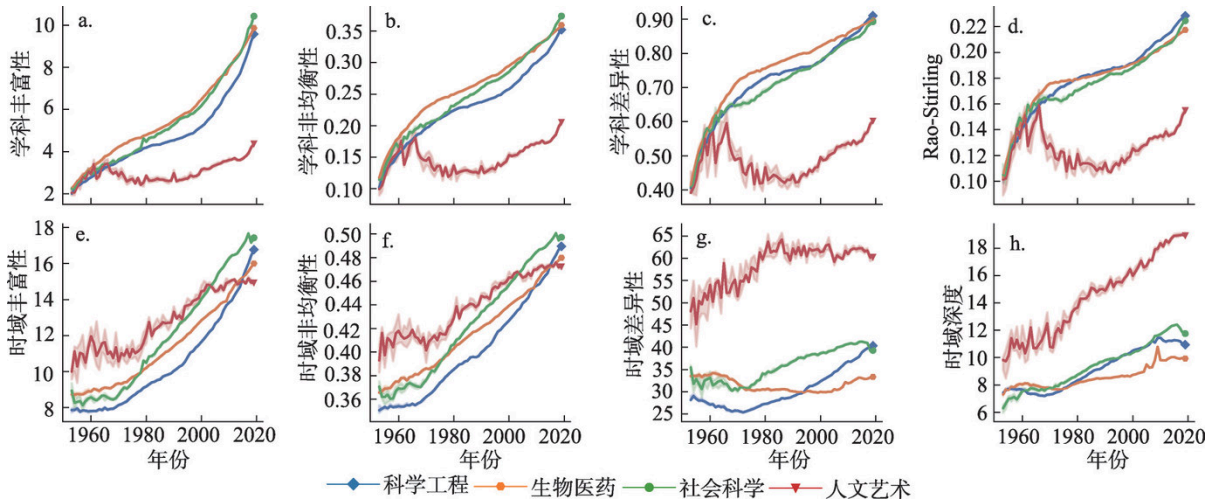


图 2　不同领域中跨学科性、跨时域性的时间趋势（彩图请见 https://qbxb.istic.ac.cn）

阴影区域表示 95% 置信区间。

科非均衡性增长最快，科学工程领域的学科差异性和总体跨学科性的 Rao-Stirling 指数增长最快，生物医药领域的跨学科性也呈现很强的增长趋势。人文艺术领域的时域差异性和时域深度明显强于其他领域，而其时域丰富性和非均衡性在早期很高，近 10 年则逐渐被其他领域超越，这体现出人文艺术领域侧重于引用经典文献和传统理论。生物医药等领域维持着较弱的时域差异性和时域深度，这体现了其知识迭代速度快，新的研究成果持续涌现。由此可见，不同学科的范式特征会导致其对于知识组合的偏好程度不同[28]，从而反映出不同领域的跨学科性和跨时域性模式。

2.3　跨学科性与跨时域性的分布关系

在本文中，跨学科性和跨时域性测度框架包含 8 个具体指标，理解这些指标的分布特征和相互关系显得尤为重要。为此，本文分析了 38879575 篇论文的跨学科性和跨时域性指标的概率分布，结果如图 3a~图 3h 所示。根据 2.1 节的定义，学科丰富性、时域丰富性、时域差异性以及时域深度这 4 个指标都可以被视为计数变量。结果表明，其分布特征符合无标度的幂律特征，基于幂律函数 $N(x)\sim kx^{-\gamma}$ 对

这 4 个指标的分布进行拟合，幂律拟合结果 γ 均为较大值。除此之外，学科非均衡性、Rao-Stirling、时域非均衡性等指标呈现对数正态分布的特征，其值大多集中于中位数附近。这些结果表明，跨学科性和跨时域性的指标都不是均匀分布的，而是呈现明显的异质性分布特征。

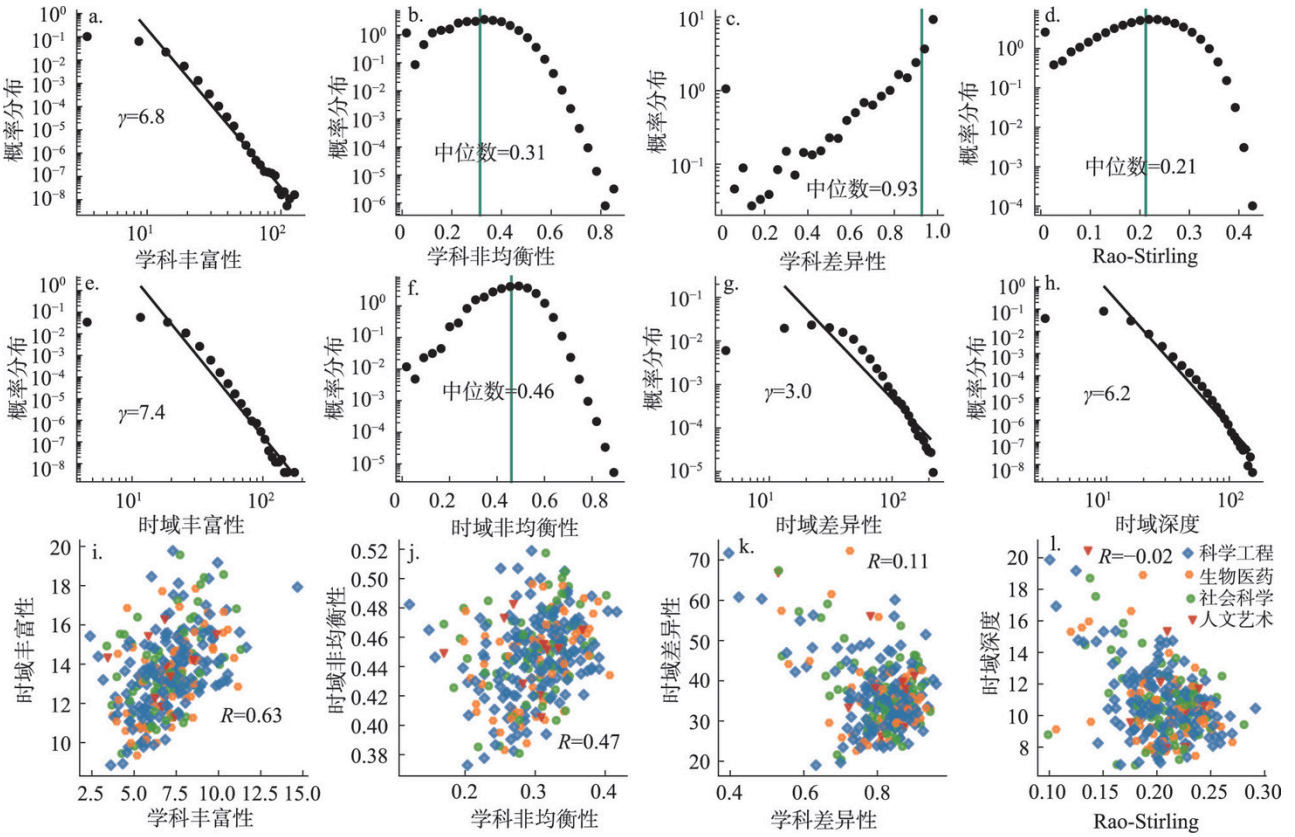


图 3 跨学科性、跨时域性的分布特征与相互关系（彩图请见 http://qbxb.istic.ac.cn）

图 i~图 l 中的点表示 292 个 MAG 二级领域。

图 3i~图 3l 呈现了跨学科性指标和跨时域性指标之间的相互关系，其中的点表示微软学术数据库中 292 个二级领域。时域丰富性和学科丰富性的皮尔逊相关系数为 0.63，时域非均衡性和学科非均衡性的皮尔逊相关系数为 0.47，这两组变量呈现较明显的正相关性；而时域差异性和学科差异性的皮尔逊相关系数为 0.11，Rao-Stirling 和时域深度的皮尔逊相关系数为 −0.02，呈现很弱的相关性。这些结果表明，尽管跨学科性指标和跨时域性指标采用了类似的定义框架，且在时间上呈现类似的趋势，它们的相关性却并非很强。

2.4 基于科学家生涯的跨学科性与跨时域性演化

为进一步理解跨学科性与跨时域性的演化机制，本节从科学家生涯层次进行分析。已有研究表明，在完整的科学家生涯中，其论文影响力分布是相对随机的[8,29]，而其合作模式和研究主题则常常会发生变化[30]。与此同时，科学家生涯层次的研究跨学科性特征演化和研究跨时域性特征演化的模式并不清楚。此外，作者层次的分析也有助于控制选择偏误问题。为了减轻不完整生涯数据、高影响力合作者等因素对科学家的跨学科、跨时域特征的影响[31]，本节仅考虑高产科学家。由于微软学术数据库的学者消歧精度在发文量极高的作者中表现不佳，为保证数据质量，本节筛选出被数据库收录了 100~200 篇论文的 119880 位高产科学家作为样本。

对于科学家 a 生涯中发表的全部 N 篇论文，本文先将其全部论文按照发表日期依次排序，得到真实的生涯论文序列 $I_a(\text{real})$；然后，对每位科学家的生涯论文序列进行多次随机重组，得到一系列随机的生涯论文序列 $I_a(\text{random})$ 作为零模型（null model）。在图 4a~图 4h 中，绿色线代表科学家生涯

的真实数据 I_a(real) 的每篇论文 i(real) 的指标值随论文序列的演化，黑色线代表零模型随机序列 I_a(random) 的每篇论文 i(random) 的指标值随论文序列的演化。研究结果表明，跨学科性与跨时域性的全部 8 个指标都随着科学家的生涯呈现明显的增长趋势，而零模型则不存在明显的生涯趋势。

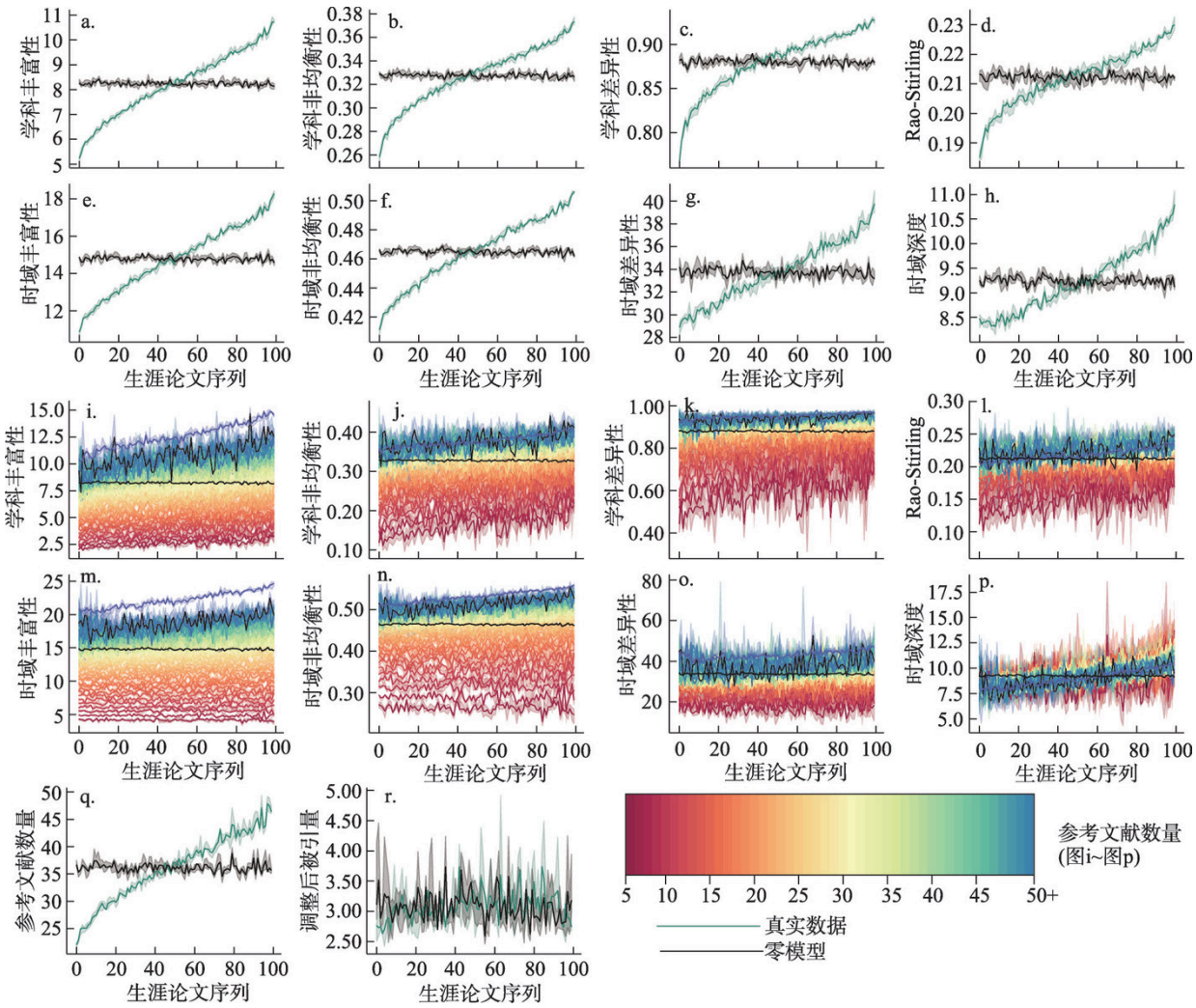


图 4　科学家层次的跨学科性与跨时域性生涯演化特征(彩图请见 https://qbxb.istic.ac.cn)

阴影区域表示 95% 置信区间。

同时，本文观察到论文的参考文献数量也存在随作者生涯上升的趋势（图 4q），这表明科学家在其生涯中倾向于引用越来越多的文献，而这势必会影响到跨学科性与跨时域性。为此，本文控制了论文的参考文献数量，单独分析相同参考文献数量论文在作者生涯中的演变，结果如图 4i~图 4p 所示。在控制了论文的参考文献数量后，科学家跨学科性与跨时域性的演化趋势明显减弱。这表明在科学家的生涯中，跨学科性和跨时域性呈现显著的增长趋势，这种趋势很大程度上可以归因于论文参考文献数量的增长。值得注意的是，这一结论在丰富性指标中最显著，而在时域深度中并不成立，因为其反映了参考文献的年龄。

本文进一步分析了在科学家生涯中论文影响力的变化趋势，结果如图 4r 所示。纵轴以调整后的被引量 $C_{f,t}$ 代表论文影响力，$C_{f,t}$ 定义为论文的原始被引量除以同一发表年份和领域的论文被引量均值，这种方法可以很好地解决原始被引量的时间窗口和领域异质性等问题[32]。图 4r 中的结果表明，真实数据中的论文影响力与随机零模型中的论文影响力并无显著区别，反映出科学家生涯中论文影响力是相对随机的，这与既有研究结论[8,29]保持一致，例如，Liu 等[8]发现科学家生涯中的论文影响力可以用多阶段正态分布模型拟合。这些发现引发了新的研究问

题：既然在科学家生涯中，论文影响力是相对随机的，而跨学科性与跨时域性存在受引文模式调节的增长模式，那么跨学科性、跨时域性与论文影响力之间的关系是什么？为此，下文将构建固定效应回归模型来研究这个问题。

4 跨学科性、跨时域性与论文影响力

4.1 变量介绍与实证模型

论文的原始被引量具有可解释性强的特征，因此，本文将其作为科学影响力的主要表征。被引量反映了论文在学术界的认可程度与贡献份额[33]，但是论文的原始被引量也会受到领域、年份和时间窗口等因素的影响，本文将通过固定效应模型减轻其对回归结果的影响。同时，论文的原始被引量是典型的长尾分布变量，本文将对其进行自然对数处理。

本文构建了“颠覆性热点论文”这一高阶指标，用于表征论文的颠覆性影响和整体影响力，其同时包含了论文的颠覆性和学术影响力两个维度。其中，颠覆性指数是 Funk 等[34]提出的创新指标，其从引文网络视角出发，评估论文影响科学网络结构的程度，如果核心论文（focal paper，FP）的后续引文也引用了 FP 的参考文献，那么可以认为 FP 巩固了知识范式；反之，如果 FP 的未来引文仅引用 FP 本身而忽略其知识基础，那么这些未来的引文就承认了 FP 的颠覆性价值。

颠覆性指数定义为

$$\text{CD index} = p_{\text{D}} - p_{\text{C}} = \frac{n_{\text{DC}} - n_{\text{CC}}}{n_{\text{DC}} + n_{\text{CC}} + n_{\text{RC}}}$$

其中，n_{CC}、n_{DC} 和 n_{RC} 分别代表引用 FP 及其参考文献、引用 FP 而不引用其参考文献和引用 FP 的参考文献而不引用 FP 本身的论文数量。颠覆性指数更高的论文被认为破坏了其局部知识结构，而较低的颠覆性指数则表示该论文巩固了既有的范式[35]。

由于颠覆性指数存在分布不均匀、测度不准确等局限[36-38]，本文以 0 为分界点，将其设定为二元变量，进一步将颠覆性热点论文定义为颠覆性指数大于 0 的热点论文（该领域和年份被引量位居前 10%）。这种定义方式综合了科学影响力和颠覆性指数，能够更准确地反映论文的颠覆性影响力。在本文的数据中，仅有 1.9% 的论文为颠覆性热点论文，这些颠覆性热点论文涵盖了数据集中 3/4 以上的诺贝尔奖论文。

为控制干扰因素对论文跨学科性、跨时域性、被引量等特征的影响，本文考虑以下控制变量：①参考文献数量[36]；②团队合作规模，即论文作者人数[39]；③是否跨国合作[39]；④是否跨学科团队[27]；⑤团队平均学术年龄[8]；⑥截至论文发表当年的团队平均发文量[8]；⑦是否受 NIH（National Science Foundation）和 NSF（National Science Foundation，United States）等基金资助[40]；⑧年份固定效应[41]（1950—2020 年）；⑨领域固定效应[28]（292 个领域）；⑩期刊固定效应（39893 种期刊）。

为分析跨学科性、跨时域性对论文影响力的作用效应，本文使用 OLS（ordinary least squares）回归模型进行估计，即

$$\text{citation}_{iyfj} = \alpha + \beta_1(\text{diversity}_i) + \beta_2(\text{controls}) + Y_y + F_f + J_j + \varepsilon_{iyfi} \quad (5)$$

其中，i 表示论文；y 表示其发表年份；f 代表领域；j 代表期刊；ε 代表误差项。模型控制了年份、领域和期刊三重固定效应，以排除时间趋势[41]、领域分布[28]、论文类型及主题的影响，从而能够较为准确地估计跨学科性、跨时域性对论文影响力的作用效应。

4.2 跨学科性、跨时域性与论文被引量

表 1 展示了跨学科性指标对论文被引量的 OLS 回归结果。可以发现，在控制了参考文献数量、团队规模、团队平均学术年龄、团队发文量、跨国合作、跨学科合作、基金资助情况，以及年份、领域和期刊三重固定效应等干扰因素后，学科丰富性、学科非均衡性、学科差异性和 Rao-Stirling 均能显著正向促进论文的被引量。

具体而言，学科丰富性增加一个单位，会促进被引量增加 0.25%；学科差异性增加一个单位，会促进被引量增加 6.24%；Rao-Stirling 增加一个单位，会促进被引量增加 7.75%；学科非均衡性增加一个单位，会促进被引量增加 6.10%。进一步基于变量标准差解释其含义：学科丰富性增加一个标准差，会促进被引量增加 1.15%（0.25%×4.61）；学科差异性增加一个标准差，会促进被引量增加 1.50%（6.24%×0.24）；Rao-Stirling 增加一个标准差，会促进被引量增加 0.62%（7.75%×0.08）；学科非均衡性增加一个标准差，会促进被引量增加 0.73%（6.10%×0.12）。

表 2 展示了跨时域性指标对论文被引量的 OLS 回归结果。可以发现，在控制了参考文献数量、团队规模、团队平均学术年龄、团队发文量、跨国合

表 1　跨学科性指标与论文被引量的 OLS 回归结果

模型	ln(被引量)			
	(1)	(2)	(3)	(4)
学科丰富性	0.0025^{***} (6×10^{-5})			
学科非均衡性		0.0610^{***} (0.0018)		
学科差异性			0.0624^{***} (0.0009)	
Rao-Stirling				0.0775^{***} (0.0024)
ln(参考文献)	0.5115^{***} (0.0004)	0.5169^{***} (0.0003)	0.5137^{***} (0.0003)	0.5198^{***} (0.0003)
ln(团队规模)	0.1937^{***} (0.0003)	0.1935^{***} (0.0003)	0.1934^{***} (0.0003)	0.1937^{***} (0.0003)
ln(学术年龄)	-0.0577^{***} (0.0005)	-0.0579^{***} (0.0005)	-0.0580^{***} (0.0005)	-0.0581^{***} (0.0005)
ln(发文量)	0.0911^{***} (0.0003)	0.0911^{***} (0.0003)	0.0911^{***} (0.0003)	0.0912^{***} (0.0003)
跨国合作	0.0315^{***} (0.0005)	0.0318^{***} (0.0005)	0.0319^{***} (0.0005)	0.0317^{***} (0.0005)
跨学科合作	-0.0342^{***} (0.0004)	-0.0338^{***} (0.0004)	-0.0340^{***} (0.0004)	-0.0339^{***} (0.0004)
基金资助	0.1368^{***} (0.0006)	0.1372^{***} (0.0006)	0.1376^{***} (0.0006)	0.1372^{***} (0.0006)
期刊FE	yes	yes	yes	yes
年份FE	yes	yes	yes	yes
领域FE	yes	yes	yes	yes
样本量	38469142	38469142	38469142	38469142
R^2	0.49740	0.49739	0.49745	0.49739

注：所有模型均基于 OLS 回归，括号表示稳健标准误。本表展示了包含全部控制变量的结果，若不考虑其中部分控制变量或固定效应，结论仍成立。*表示 $p<0.05$，**表示 $p<0.01$，***表示 $p<0.001$。

作、跨时域合作、基金资助情况，以及年份、领域和期刊三重固定效应等干扰因素后，时域丰富性、时域非均衡性、时域差异性和时域深度均显著抑制论文的被引量。

具体而言，基于变量标准差解释其含义：时域丰富性增加一个标准差，会导致被引量降低 18.00%（2.62%×6.87）；时域非均衡性增加一个标准差，会导致被引量降低 0.26%（2.62%×0.10）；时域差异性

表 2　跨时域性指标与论文被引量的 OLS 回归结果

模型	ln(被引量)			
	(1)	(2)	(3)	(4)
时域丰富性	-0.0262^{***} (4.81×10^{-5})			
时域非均衡性		-1.539^{***} (0.0031)		
时域差异性			-0.0012^{***} (7.95×10^{-6})	
时域深度				-0.0195^{***} (3.35×10^{-5})
ln(参考文献)	0.7317^{***} (0.0005)	0.6792^{***} (0.0004)	0.5355^{***} (0.0003)	0.5225^{***} (0.0003)
ln(团队规模)	0.1859^{***} (0.0003)	0.1938^{***} (0.0003)	0.1912^{***} (0.0003)	0.1829^{***} (0.0003)
ln(学术年龄)	-0.0355^{***} (0.0005)	-0.0369^{***} (0.0005)	-0.0543^{***} (0.0005)	-0.0264^{***} (0.0005)
ln(发文量)	0.0798^{***} (0.0003)	0.0796^{***} (0.0003)	0.0895^{***} (0.0003)	0.0729^{***} (0.0003)
跨国合作	0.0338^{***} (0.0005)	0.0307^{***} (0.0005)	0.0323^{***} (0.0005)	0.0325^{***} (0.0005)
跨学科合作	-0.0322^{***} (0.0004)	-0.0319^{***} (0.0004)	-0.0328^{***} (0.0004)	-0.0325^{***} (0.0004)
基金资助	0.1388^{***} (0.0006)	0.1376^{***} (0.0006)	0.1369^{***} (0.0006)	0.1370^{***} (0.0006)
期刊FE	yes	yes	yes	yes
年份FE	yes	yes	yes	yes
领域FE	yes	yes	yes	yes
样本量	38469142	38469142	38469142	38469142
R^2	0.50159	0.50130	0.49767	0.50231

注：所有模型均基于 OLS 回归，括号表示稳健标准误。本表展示了包含全部控制变量的结果，若不考虑其中部分控制变量或固定效应，结论仍成立。*表示 $p<0.05$，**表示 $p<0.01$，***表示 $p<0.001$。

增加一个标准差，会导致被引量降低2.77%（0.12%×23.06）；时域深度增加一个标准差，会导致被引量降低11.19%（1.95%×5.74）。

由此可见，跨时域性对论文影响力的作用效应与跨学科性对论文影响力的作用效应截然相反，即科学知识跨学科性显著促进论文被引量，而科学知识跨时域性则显著抑制论文的被引量。

4.3 跨学科性、跨时域性与颠覆性热点论文

为分析知识基础横向分布特征（跨学科性）以及纵向分布特征（跨时域性）组合对科学颠覆性影响力的作用机制，本文将跨学科性和跨时域性的各个指标进行了分位数划分，并计算出跨学科性特征和跨时域性特征的二元组合与颠覆性热点论文的概率分布，结果如图5所示。

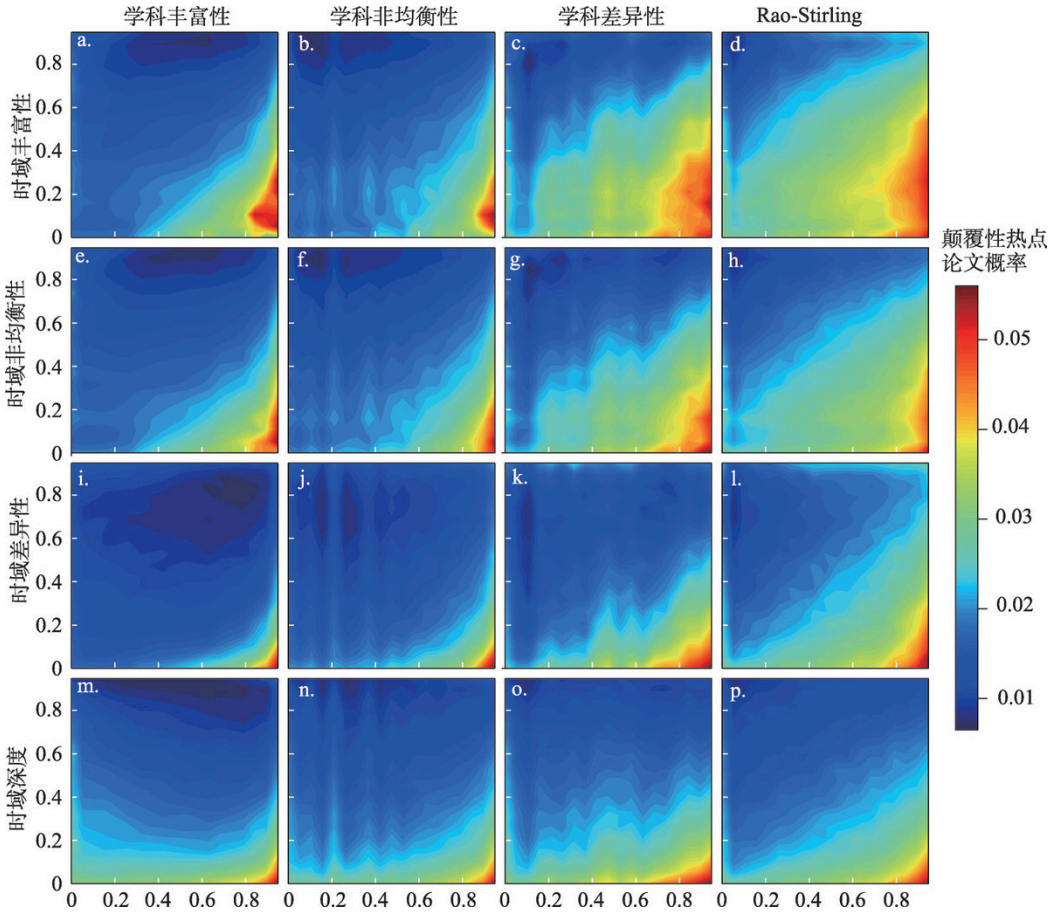


图5 跨学科性、跨时域性与颠覆性热点论文概率分布

横纵坐标刻度值均线性映射为分位数。

图5揭示了跨学科性特征和跨时域性特征对科学颠覆性影响力的内在作用机制：颠覆性热点论文的高概率空间位于时域丰富性、时域非均衡性、时域差异性和Rao-Stirling较弱的位置，同时位于学科丰富性、学科差异性、Rao-Stirling较强的位置。若一篇论文兼具较弱的跨时域性和较强的跨学科性，则其有最高的概率成为颠覆性热点论文。

由此可见，知识基础的横向离散性和纵向聚焦性是产生突破性研究的关键所在。在纵向维度，年轻而聚焦的知识基础有助于论文提升影响力；在横向维度，整合多个学科的知识可以显著提高论文的影响力。具有颠覆性的、高影响力的研究往往需要在前沿的知识基础上进行拓展，而非选择老旧的知识基础，即“站在巨人的肩膀上”；另外，跨学科的知识融合是提升科学影响力的关键，科学知识在横向离散和纵向聚焦的有机结合，能诞生更多高影响力的研究成果。

5 总结与展望

本文通过对微软学术数据库中1950—2020年的38879575篇论文知识基础横向和纵向维度的分析，

探讨了跨学科性、跨时域性的时间趋势、领域差异、分布特征、相互关系、科学家生涯演化模式，并进一步挖掘了跨学科性、跨时域性与科学影响力的复杂关系。

本文的研究具有深刻的理论意义。其一，本文在跨学科性概念的基础上，提出了跨时域性理论与测度框架，并将论文的跨时域性量化为时域丰富性、时域非均衡性、时域差异性和时域深度4个维度。跨时域性理论丰富了科学学领域的测度框架和方法，为学术评价领域提供了新的研究思路。跨时域性研究框架不仅是对现有测度理论的拓展和补充，也是首次针对论文参考文献纵向分布的系统性测度框架，弥补了现有研究在分析科学知识纵向维度的空白。其二，本文基于大样本的实证研究，分析了科学知识基础的趋势和特征。跨学科性和跨时域性均呈现增长的时间趋势，一方面，体现出细分领域知识的融合程度在不断提高；另一方面，反映出知识过载可能对未来创新的负面效应[28,42]。跨学科性和跨时域性等指标呈现以无标度为代表的异质性分布特征，并且跨学科性和跨时域性并无很强的相关性，这一发现丰富了对相关指标的理解。在科学家的学术生涯中，跨学科性和跨时域性呈现明显的增长趋势，这种趋势很大程度上可以归因于其论文参考文献数量的增长，这一结论揭示了科学家层次的演化特征，丰富了本文的贡献领域。

本文的研究结论还具有实践意义。科学知识的跨学科性能够促进论文影响力增强，而跨时域性会抑制论文的影响力，这一结论为科研人员在规划研究方向和选择知识基础时提供了有益的指导。此外，本文发现颠覆性热点论文同时具备横向离散性和纵向聚焦性的知识基础，年轻的知识基础更能产生高影响力乃至颠覆性的研究。基于此，政策制定者应鼓励科研人员参与前沿性的、跨学科的研究项目，为其提供更多的研究经费。跨学科性的知识基础对科研的影响力具有积极作用，可以通过鼓励跨学科合作、支持多样性研究团队以及提供科研机会等方式来促进跨学科性的知识基础发展。

本文也存在一些局限性。其一，本文仅分析了跨学科性、跨时域性与科学影响力的相关性，而未深入因果效应；其二，本文假定所有的引文都是一致的，没有考虑无效引用和引用极性的问题[43]；其三，科学影响力及科学创新具有复杂性和多面性，本文仅从被引量和颠覆性这两个视角进行分析，未考虑社会影响等其他维度；其四，本文假设全部学科的论文都符合“范式变革”的创新理论[44]，然而，这一理论在人文社会科学中未必成立，基于“范式交替”等理论[45]视角能够为这些学科提供更精确的研究结论。

未来的研究可以结合准随机试验等实证方法，探讨跨学科性、跨时域性与科学影响力的因果机制；还可以考虑异质性的科学创新类型，针对不同领域的机制进行细粒度分析。此外，本文主要针对论文的参考文献进行跨学科性和跨时域性的分析，未来可以采取更细粒度的测度方法，从文本或主题中挖掘知识融合的模式，也可以结合知识的后向流动，基于引文级联，分析知识基础跨学科性、跨时域性对整个科学体系更为广泛的影响机制。

参　考　文　献

[1] Gates A J, Ke Q, Varol O, et al. Nature’s reach: narrow work has broad impact[J]. Nature, 2019, 575(7781): 32-34.

[2] 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.

[3] Zhang Y, Wang Y, Du H F, et al. Delayed citation impact of interdisciplinary research[J]. Journal of Informetrics, 2024, 18(1): 101468.

[4] Sheng L B, Lyu D Q, Ruan X M, et al. The association between prior knowledge and the disruption of an article[J]. Scientometrics, 2023, 128(8): 4731-4751.

[5] Zeng A, Shen Z S, Zhou J L, et al. The science of science: from the perspective of complex systems[J]. Physics Reports, 2017, 714-715: 1-73.

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

[7] Yang A J, Zhao S X, Deng S H. Revisiting delayed recognition in science: a large-scale and comprehensive study[J/OL]. Journal of Information Science, (2024-05-29). https://doi.org/10.1177/01655515241244462.

[8] Liu L, Wang Y, Sinatra R, et al. Hot streaks in artistic, cultural, and scientific careers[J]. Nature, 2018, 559(7714): 396-399.

[9] 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.

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

[11] Bloom N, Jones C I, van Reenen J, et al. Are ideas getting harder to find?[J]. American Economic Review, 2020, 110(4): 1104-1144.

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

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

[14] 章成志, 吴小兰 . 跨学科研究综述[J]. 情报学报, 2017, 36(5): 523-535.

[15] Leydesdorff L, Rafols I. Indicators of the interdisciplinarity of journals: diversity, centrality, and citations[J]. Journal of Informetrics, 2011, 5(1): 87-100.

[16] Rafols I, Meyer M. Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience[J]. Scientometrics, 2010, 82(2): 263-287.

[17] Ding Y, Rousseau R, Wolfram D. Measuring scholarly impact: methods and practice[M]. Cham: Springer, 2014.

[18] Ren J J, Wang F, Li M L. Dynamics and characteristics of interdisciplinary research in scientific breakthroughs: case studies of Nobel-winning research in the past 120 years[J]. Scientometrics, 2023, 128(8): 4383-4419.

[19] Ke Q. Interdisciplinary research and technological impact: evidence from biomedicine[J]. Scientometrics, 2023, 128(4): 2035-2077.

[20] Pan R K, Petersen A M, Pammolli F, et al. The memory of science: inflation, myopia, and the knowledge network[J]. Journal of Informetrics, 2018, 12(3): 656-678.

[21] Ghaffari F, Wilson M C. A model for reference list length of scholarly articles[J]. Scientometrics, 2023, 128(9): 5335-5350.

[22] Gupta B M. Analysis of distribution of the age of citations in theoretical population genetics[J]. Scientometrics, 1997, 40(1): 139-162.

[23] Jarić I, Knežević-Jarić 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.

[24] Stacey A G. Ages of cited references and growth of scientific knowledge: an explication of the gamma distribution in business and management disciplines[J]. Scientometrics, 2021, 126(1): 619-640.

[25] Reisz N, Servedio V D P, Loreto V, et al. Loss of sustainability in scientific work[J]. New Journal of Physics, 2022, 24(5): 053041.

[26] 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.

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

[28] Chu J S G, Evans J A. Slowed canonical progress in large fields of science[J]. Proceedings of the National Academy of Sciences of the United States of America, 2021, 118(41): e2021636118.

[29] Sinatra R, Wang D S, Deville P, et al. Quantifying the evolution of individual scientific impact[J]. Science, 2016, 354(6312): aaf5239.

[30] Zeng A, Shen Z S, Zhou J L, et al. Increasing trend of scientists to switch between topics[J]. Nature Communications, 2019, 10 (1): Article No.3439.

[31] Petersen A M, Fortunato S, Pan R K, et al. Reputation and impact in academic careers[J]. Proceedings of the National Academy of Sciences of the United States of America, 2014, 111(43): 15316-15321.

[32] Radicchi F, Fortunato S, Castellano C. Universality of citation distributions: toward an objective measure of scientific impact[J]. Proceedings of the National Academy of Sciences of the United States of America, 2008, 105(45): 17268-17272.

[33] Yang A J, Wu L W, Zhang Q, et al. The k-step h-index in citation networks at the paper, author, and institution levels[J]. Journal of Informetrics, 2023, 17(4): 101456.

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

[35] Yang A J, Deng S H, Wang H, et al. Disruptive coefficient and 2-step disruptive coefficient: novel measures for identifying vital nodes in complex networks[J]. Journal of Informetrics, 2023, 17 (3): 101411.

[36] Yang A J, Gong H C, Wang Y H, et al. Rescaling the disruption index reveals the universality of disruption distributions in science[J]. Scientometrics, 2024, 129(1): 561-580.

[37] 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.

[38] Leibel C, Bornmann L. What do we know about the disruption index in scientometrics? An overview of the literature[J]. Scientometrics, 2024, 129(1): 601-639.

[39] Jones B F, Wuchty S, Uzzi B. Multi-university research teams: shifting impact, geography, and stratification in science[J]. Science, 2008, 322(5905): 1259-1262.

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

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

[42] Jones B F. The burden of knowledge and the “death of the renaissance man”: is innovation getting harder?[J]. The Review of Economic Studies, 2009, 76(1): 283-317.

[43] Bao H L, Teplitskiy M. A simulation-based analysis of the impact of rhetorical citations in science[J]. Nature Communications, 2024, 15: Article No.431.

[44] Kuhn T S. Historical structure of scientific discovery[J]. Science, 1962, 136(3518): 760-764.

[45] 赵鼎新 . 社会科学研究的困境: 从与自然科学的区别谈起[J]. 社会学评论, 2015, 3(4): 3-18.

（责任编辑 王克平）