深度与广度技术搜索及创新影响的时机

Likun Cao[†*], James Evans[*‡1]

[†]普渡大学[*]芝加哥大学
[‡]圣菲研究所

预印本版本：2025年12月，已被《技术预测与社会变革》接受发表

摘要： 本研究通过对动态集体知识系统中的发明性搜索进行建模，并强调时机对技术影响的重要性，为创新战略中深度与广度之争提供了新的视角。我们利用前沿机器学习在双曲空间中预测专利引文网络，分析了490万项美国专利，考察搜索策略如何形成影响力累积的不同时间模式。研究发现，基于深度搜索的发明依赖于对复杂重组结构的专业化理解，能够通过在专业化社群中的早期采用带来更高的短期影响，但随着创新陷入“锁定”状态、扩散潜力有限，其收益会递减。相反，根植于跨越异质领域的广度搜索的发明，虽会遭遇初始阻力，却能触达认知多样化受众，实现更广泛的扩散并产生更大的长期影响。单项发明要实现稳定影响，需要兼具深度与广度。组织则可以在多项发明中有策略地平衡这两种路径：通过深度搜索构建可靠的技术基础设施，同时通过广度搜索拓展应用。我们通过揭示深度与广度搜索策略如何以不同方式塑造技术影响的时机与轨迹，以及个人发明者和组织如何利用这些机制平衡利用与探索，推动了创新理论的发展。

关键词： 技术领导力；组织学习；组合式创新；机器学习；创造力；获专利发明

[1] 通信请联系 芝加哥大学 Knowledge Lab 和社会学系的 James Evans，邮箱：jevans@uchicago.edu.

亮点

- 我们使用双曲嵌入，将技术演化建模为层次树。
- 这种方法将搜索路径的深度与广度视为相互独立的特征。
- 深度有助于在本地社区内获得短期认可，
- 广度通过触达更广泛的受众，预示长期影响。
- 组织可以将深度搜索中形成的稳健组件拓展至广泛应用。

1. 引言

源自复杂适应系统理论（Simon，1991 [1962]），的重组框架，已成为理解和预测技术演化的主流视角（Brian Arthur，2009年；Fleming，2001年；Weitzman，1998年；Xiao等人，2022年）。成功的重组需要将组件组装成适当的架构，以产生新的能力，例如恰当的模块化程度（Ethiraj等人，2008年）和多样性（Dhanaraj和Parkhe，2006年）。在重组文献中，最受关注且仍存在争议的问题之一，是深度与广度在技术搜索中的相对影响。此处，深度指在单一或少数几个紧密相关的领域内深入运用知识，以开发新技术；广度则指整合来自遥远且多样化来源的知识，从而创造新发明。以往研究对两者的影响得出了相互矛盾的证据。重组的“张力”视角认为，遥远或多样化的知识能够打破认知惯性，并激发突破性创意（Hall 等，2001；Trajtenberg 等，1997；Verhoeven等人，2016a年）。与之相对，“基础性”视角的研究则推崇深入钻研某一特定领域，认为这能让发明者获得必要的专业知识，从而识别并解决异常（Csikszentmihalyi，2009年；Fleming，2001年；Kaplan 和 Vakili，2015a年；Sternberg和Lubart，1995年）。迄今为止，实证研究的结果仍不一致，背后的机制也尚不清楚。本文重新审视这一核心议题，并通过纳入时间维度，提出一种解决方案，以调和这两种看似相互矛盾的观点。

深度与广度这两个概念最早受到关注，源于 March和Simon 的奠基性研究（1993 [1958]），并持续影响知识工作者的实际决策。尽管它们在理论和实践上都意义重大，既有研究却

对它们的影响得出了相互矛盾的证据。Kaplan 和 Vakili（2015b）指出，这些策略的益处可能因回报类型而异，无论是认知回报还是经济回报。尽管这一区分在一定程度上解释了相互矛盾的发现，但无法说明为何对同一结果变量，尤其是引文，仍会产生相反的影响（Fleming，2001年；Uzzi 等，2013年）。由于实证结果模糊不清，后续研究认为，深度与广度作为策略的重要性或多或少取决于情境，例如知识变化速度（Teodoridis 等，2019年）和制度环境（Zheng 等，2022年）。

为推进这一研究方向，我们转向计算方法，对知识组件之间的层级关系进行形式化，并将深度与广度建模为技术景观中相互独立的搜索维度，而这一景观会随时间演变。为此，我们采用 Poincaré 嵌入方法——一种机器学习算法（Nickel 和 Kiela，2017年），它有助于根据依赖网络，以经验为基础构建层级图。这个基于专利引文网络构建的树状技术图谱，捕捉了集体发明中技术类别之间的实际接近程度，并为度量深度与广度奠定了更坚实的基础。我们将深度 精确定义为知识组件所在位置到树根的距离，并将广度定义为树枝之间距离的总和。借助对深度 和广度进行的精确、直观且相互独立的度量，我们从集体智慧与创新的视角，探究这两种策略的影响及其背后的机制。

我们的结果表明，深度与广度的影响在不同的时间尺度上发挥作用。深度能够促进在本地社区中的短期认可，而广度则通过触达更广泛的受众来预测长期影响。基于这些实证模式以及对潜在机制的探索，我们认为，实证研究结果不一致，很可能源于

两个方面：一是依赖特定行业中规模小、范围有限的样本（Terjesen 和 Patel，2017年；Zhou 和 Li，2012年），二是普遍假设技术搜索是一种个人活动，忽视了集体发明的情境（Kaplan 和 Vakili，2015a年）。这些挑战共同导致当今文献在理论上存在模糊性：相互矛盾的研究发现不断增多，对相关机制的认识有限，也未能充分关注集体动态。传统的小数据研究方法难以克服这些局限，但计算支持的社会科学探索则为此带来了机遇。

我们的研究首次运用精确、高分辨率的技术映射，系统考察了深度与广度搜索及其随时间变化的结果（Aharonson and Schilling, 2016；Escolar 等人, 2023）。现有度量方法，如赫芬达尔多样性和技术距离（Hall 等，2001；Trajtenberg 等，1997），主要捕捉概念距离，而非技术在实践中的功能距离（Benson and Magee, 2015；Magee 等人, 2016）。概念分类体系具有静态性，因此难以识别我们在此观察到的时间模式。此外，先前研究将深度与广度视为单一连续谱的两端，而非两个相互独立的维度（Kaplan 和 Vakili, 2015a；Trajtenberg 等，1997）。与之相反，我们提出的新方法以整个知识系统的结构为参照框架，准确捕捉技术在实践中的功能，清晰区分深度与广度，并考虑集体发明的情境。该方法使我们能够以直观且可追踪的方式清晰观察知识扩散与积累的社会动态，从而揭示时间在调和深度与广度之争中的关键作用。除了有助于阐释之外，该方法还为预测创造了机会，因此对于寻求及时、可量化的前沿领域或技术进步见解的政策制定者和行业决策者尤其有价值。

本文其余部分安排如下。我们首先回顾重组理论相关文献，重点讨论技术重组中有关深度搜索与广度搜索的争论。接下来，我们介绍表征学习框架，说明如何将其用于建模知识图景，并构建深度与广度的度量。随后，我们评估深度与广度如何影响专利技术随时间变化的影响。最后，我们探讨可能驱动所观察到的实证模式的机制。

2. 深度与广度

2.1 技术重组中的深度与广度搜索策略

众所周知，探索与利用的二分法（March，1991年）区分了两类人：一类人深入钻研单一领域，另一类人则四处探索并建立联系。以赛亚·柏林将他们形象地称为“刺猬型”和“狐狸型”（Berlin，2013 [1953]；另见 Dyson，2015年）。重组深度已与局部搜索以并整合构成某项技术基础的组件（Mannucci 和 Yong，2018年；Teodoridis 等，2019年）。相较之下，重组广度已与广泛全球性的探索性搜索联系起来，这种搜索

2015b；Traitenberg 等，1997年）。

在技术和商业发展的情境下，这一区分被重新表述为“基础”视角和“张力”视角，二者分别倾向于技术搜索的深度和广度（Kaplan 和 Vakili，2015a年）。“基础”视角推崇深入地研究一特定领域，因为这能让发明者形成专家思维，从而发现值得解决的异常（Csikszentmihalyi，2009年；Kaplan 和 Vakili，2015a年；

Sternberg和Lubart，1995年）。“张力”视角则认为，对遥远或多样化知识进行重组能够打破认知惯性，激发突破性创意（Fleming，2001年；Hall 等，2001年；Trajtenberg 等，1997年；Verhoeven 等人，2016b年）。

早期研究旨在确定深度搜索与广度搜索策略中，哪一种与成功的关联更强，由此引发了持续的争论（Fleming，2001年；Kaplan 和 Vakili，2015b；乌齐等，2013年）。为调和这些相互矛盾的发现，后续研究转而着重识别关键的权变因素。影响搜索深度与广度成效的情境包括知识领域内的变化速度（特奥多里迪斯等，2019年）、塑造信息流的网络结构（特尔·瓦尔等，2016年），以及更广泛的制度环境（Zheng 等，2022年）。企业内部也可以有益地平衡深度与广度，这一原则称为“双元性”（布拉姆等，2021年；Raisch 和 Birkinshaw，2008年；施泰特纳和拉维，2014年）。（Terjesen 和 Patel，2017年；Zhou 和 Li，2012年）

相互矛盾的证据可能既源于数据规模有限，也源于深度与广度的测量方式差异很大。定性信息（Boh 等，2014）和调查方法（Terjesen 和 Patel，2017；Zhou 和 Li，2012）能够更直接、更准确地评估搜索路径，但成本高昂，且覆盖的受访者数量有限。基于分类体系的定量度量，例如用于专利的美国专利分类（USPC）或国际专利分类（IPC）（Moorthy 和 Polley，2010；Nakamura 等，2015），可应用于大规模数据集，但仍只是对其底层概念不够完善且过度简化的代理指标。例如，深度通常被计算为广度的倒数（如 Kaplan 和 Vakili，2015a；Lodh 和 Battaggion，2015），或被视为对相同类别的使用（Chattopadhyay 和 Bercovitz，2020；Kim 等，未注明日期；Paruchuri 和 Awate，2017），但肤浅或重复都不必然意味着对

理论知识有深刻理解。此外，由于分类体系反映的是理论知识中技术组件的逻辑，而非这些组件在实践中的功能依赖关系（Benson and Magee，2015；Magee 等人，2016；Triulzi 等人，2020），基于这些分类体系的度量可能并未牢固扎根于发明的实证历史。

因此，由于数据和测量方面的限制，深度与广度的影响及其潜在机制仍未得到充分理解。这类实证简化也限制了现有研究推进理论发展的能力。为改进这些度量，并为深度与广度的讨论提供更坚实的实证基础，我们将集体发明引入模型之中。集体发明是一种知识生产形式，在学术界和产业界都日益凸显。

2.2 作为集体活动的技术发明

在19世纪末至20世纪初，发明的性质经历了系统性转变，从个人活动转向企业活动。在此期间，过去由个人发明者进行并依法归属于个人发明者的发明，逐渐主要由企业及相关研究机构提出目标、加以管理并持有（Lamoreaux and Sokoloff，1999年）。这些它们不仅开展新颖的研究并取得里程碑式专利，还申请了成组的防御性专利，以在法律上保护前者（Shapiro，2000年）。这三发展将发明转变为一种集体事业：如今，推动知识生产与知

尽管发明的集体性尚未得到充分建模或度量，但文献已经认识到这一点。例如，创新经济学领域的研究强调

在估算组织边界之内及跨组织边界的知识溢出时，面向公众的社会回报是发明回报的重要组成部分（Bernstein and Nadiri，1989年；Bloom 等人，2013年；Guillard 等人，不住明日期）。

我们的集体视角对理解深度搜索与广度搜索有两项重要启示。首先，从知识生产的角度来看，深度与广度应被视为集体实践中搜索路径的内在属性，现代发明主要是在与同事和客户持续交流信息与创意的过程中发现的，而非发明者回顾以往的技术记录或独别在工业实践和技术共识中的应用有多接近或多遥远（Benson和Magee，2015；Magee 等人，2016）。要在集体理解中捕捉这种实践上的接近程度，现有分类体系并不

我们需要一种能够反映技术实际用途的技术映射（Aharonson and Schilling，2016；Escolar 等人，2023；Linzhuo 等，2020；Teodoridis 等，2019）。本研究通过使用

其次，更重要的是，如果搜索会影响知识消费，这种影响便会通过集体评价和反应展现出来。后续研究的轨迹和时机对于理解发明的影响特征都至关重要，而二者都受到技术社群内部社会动态的塑造。这意味着引文数量相同、但发展模式不同的专利——一

获得关注，但热度转瞬即逝（Kang 等人，2024年；Silva 等人，2020年）。这两种知识显然具有不同的

性质，在知识系统的构建中发挥着不同作用，对发明者和组织也各有不同价值。为更好地理解这种复杂性，我们将知识演化视为一个动态过程，并系统地追踪受众如何接纳一项发明，包括认可的速度和传播范围、受众关注的强度及其变化，以及扩散模式等。这

总之，集体通过竞争性搜索与协同搜索，在塑造知识的生产和消费方式方面发挥着核心作用。通过对集体知识的分布格局和搜索路径的结构进行建模，我们能够在实际情境中构建更准确的深度与广力度量，更重要的是，追踪二者的影响如何随时间动态累积。这种方法使我们能够将时间视为区分深度效应与广度效应的关键维度。接下来，我们将结合技术演化和复杂性理论的理论洞见，说明为何值得运用机器学习算法构建此类模型。

2.3 集体知识系统的双曲表示的理论基础

我们的方法基于技术演化理论，这些理论阐释了过去的发明如何塑造未来知识的出现。它们为从经验层面将知识作为集体系统进行分析提供了理论基础。

关于技术演化的理论讨论（Arthur and Polak，2006年；Brian Arthur，2009年）以及关于技术轨迹的实证研究（Fontana 等人，2008年；Huenteler 等人，2016年）都将技术描述为逐步且递归地模化。一些技术组件构成了由其他生产的更高层次系统的

应具有层级化的树状结构。这一观念呼应了从亚里士多德和希波的奥古斯丁，到构成狄德罗和达朗贝尔《百科全书》结构的“知识之树”等关于知识的哲学讨论（亚里士多德，1999 [350BCE]；奥古斯丁，2003 [426]；Weingart，n.d.）。

这一层级结构与 Simon（1991 [1962]）在复杂系统研究中提出、随后由阿瑟（2009年）在技术情境下进一步拓展的“模块化”概念密切相关。在层级化、模块化的知识系统中，较低层级的组件会整合成稳健且紧密连接的子系统，而这些子系统又以递归方成：涡轮转子、动力传动系统、安装结构和电网连接。转子又由叶片、轮毂和转子控制系统等更细粒度的组件构成（Huenteler 等人，2016年）。风力涡轮机领域的发明活动[illegible]之间的连接、重新设计某个子部件，或修改其中的低层级要素。这一过程可以向下追溯至实际技术变化发生的最低层级。

在这一情境下，深度局部搜索意味着深入探究技术基础设施，在子模块内部改进低层级知识组件。发明者探索的层级越低，他们参与搜索的程度就越深。这与“基础性观点”相一致；该观点认为，深度局部搜索能够帮助个人和企业更好地理解组件之间的底[illegible]年；Teodoridis 等，2019年）。

另一方面，广度搜索会融入来自其他领域的知识。在一个多层、递归的知识系统中，广度搜索涉及的不仅是不同领域，还包括

位于知识树不同分支上的领域。例如，将定量社会科学与统计学结合，其广度不及将其与生物学结合，尽管在概念分类体系中，这两者看起来可能同样遥远。这是因为统计学历来为应用社会科学的基础——即“知识根基”——作出了贡献，而生物学则属于一个独特的知识分支。

深度并不意味着狭窄，正如广度并不意味着浅薄。例如，一位电气工程师和一位材料科学家可以组成团队，将处理器和二维材料领域的深厚专业知识汇聚起来，发明下一代半导体。在这种情况下，团队对两个截然不同领域的集体深刻理解，可能结合成一项既有深度又有广度的发明。同时，一些科学家和发明者是通才，精通多个知识领域，并能创造性地跨领域开展工作（Milojević，2015年；Risha 等人，2023年；Stark，2011年）。因此，发明出的技术在知识重组方面可能具有广度、深度、兼具两者，或两者皆无。从这一理论视角出发，我们基于对技术的经验映射来构建深度与广度的度量，并将深度与广度操作化为技术重组的两个独立维度。我们将在第2.4节详细介绍这一方法。

2.4 构建技术树的双曲表示

创新理论需要一种超越先验分类体系的模型，使集体发明轨迹能够得到直接测量。理想情况下，这种模型既能捕捉技术知识的层级结构与实际邻近性，又能将深度与广度区分为独立维度。为此，我们采用双曲几何表示，这类表示能够自然地编码层级结构，从而对技术中的递归重组进行建模。

Hype双曲空间具有负曲率，数据点之间的距离

从中心到边缘呈指数级增长。设想一棵映射到平坦二维欧几里得空间中的层次树。随着层次层级（r）增加，每层的叶节点数量呈指数级增长（ ），而由于圆的面积相对于半径呈二次增长，欧几里得空间中每层叶节点之间的距离必须缩小。欧几里得空间中保度

度的操作化定义（Mueller 等．2021；Nakamura 等．2015；Papazoglou 和 Spanos，2018；Teodoridis 等，2019）。增加欧几里得空间的维度无法解决这个问题，因为仍然

幸运的是，庞加莱嵌入旨在捕捉层次结构，能够避免这种失真：它可以自然地表示水平距离随深度呈指数级增长。在双曲空间中投影层级结构的嵌入算法不会造成失真，并已成功捕捉到许多重要的文化和社会层级结构（Chamberlain 等，2017；

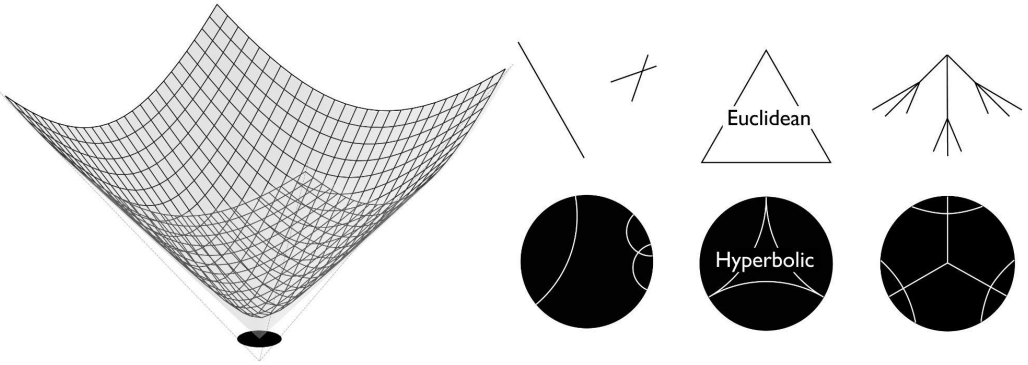
图1通过一种常见的双曲嵌入表示（二维庞加莱圆盘），从概念上阐释了双曲几何。将双曲面上的直线和点投影到这种圆盘表示中时，直线会变成圆弧，三角形的内角和小于180°，距离则呈指数级增长。该模型保保度与广度达万为两个独立维度。在第3节中，

这一算法的技术细节，以及它如何帮助产生新的理论见解。

图1 双曲庞加莱圆盘嵌入

2.5 深度与广度效应的时间假设及潜在机制

我们的方法能够全面分析由搜索深度与广度带来的专利引用增长所呈现的时间动态。我们假设



深度搜索产生短期影响，而广度搜索有助于产生长期影响，这一作用由先前研究提出的两种潜在机制驱动：锁定和桥接。

深度搜索可以通过锁定产生短期影响。文献中以多种形式讨论了这一机制，包括但不限于：认知锁定，即进行深度搜索的发明人会形成专业化的心智模型和惯例，这些模型和惯例使他们更容易发现所在领域内的异常，却更难识别领域之外的机会（Dougherty, 1992；Leonard-Barton, 1993；Levinthal and March, 1993；Nelson and Winter, 1990；Tripsas and Gavetti, 2000）；资源锁定，即投资于深度能力的组织会积累专业化的资产、技能和关系，从而产生转换成本（Christensen and Bower, 1996；Ghemawat, 1991；Henderson and Clark, 1990；Teece, 1986；Williamson, 1985）；以及网络锁定，即深度搜索会形成密集的本地网络，强化既有的发展轨迹

，但限制了接触远距离知识的机会（Granovetter, 1973；Rosenkopf and Nerkar, 2001；Stuart and Podolny, 2007；Uzzi, 1997）。

相反，广度搜索通过探索性桥接提升影响力。它使发明者能够在不同知识层级之间建立结构性连接（结构桥接）（Fleming and Waguespack，2007；Hargadon and Sutton，1997；Tushman and Scanlan，1981），发展有助于知识整合的转译能力（认知桥接）（*Carlile，2004*；*Hargadon，2002*；*Kellogg et al.，2006*；*Taylor and Greve，2006*），并开辟新的扩散路径，而这些路径可能需要时间才能成熟并获得认可（时间桥接）（Navis and Glynn，2010；Rogers et al.，2014；Santos and Eisenhardt，2009；Van de Ven，1986）。

我们讨论这些机制，并使用计算得出的代理指标对其进行实证探索。尽管深度与广度的作用在理论上被视为相互矛盾，我们的计算模型展示了这些机制如何在不同时间尺度上并存，并共同促成观测到的技术进步模式。

3. 运用双曲几何评估专利中的搜索策略

本节提出新的深度与广力度量，以反映潜在知识层级。我们首先利用美国专利数据构建庞加莱嵌入空间，展示这一方法；随后说明这种嵌入如何捕捉深度与广度，同时满足上述理论要求。

3.1 表示学习与嵌入方法

机器学习在社会科学研究中日益普及，主要用于：（1）协助人工编码员处理原始数据（Harrison 等，n.d.）；（2）从复杂的非结构化数据中提取细微特征，而这些特征原本难以量化（崔等，2021年；乔杜里等，2019年；Sgourev 等，2023年；维奇南扎等，2023年）；以及（3）揭示稳健的

模式，为理论构建提供支持，并作为传统统计估计的补充（乔杜里等，2021年；Kumar 等，2022年；Shrestha 等，2021年）。在机器学习方法中，嵌入算法已成为一种强大的工具，它能够自动将高维输入降至低维表达，并从原始数据中发现有意义且具有预测性的表示，从而支持后续分析中的特征提取（Aceves 和 Evans，2024年）。

词嵌入是社会科学文献中最突出的嵌入方法，这得益于文本数据和自然语言处理工具的广泛应用（Aceves and Evans，2024年）。它利用精确的词语邻近关系和序列信息来捕捉语义关系，并将词表示为高维向量（Joulin et al.，2017年；Kozlowski et al.，2019年；Mikolov et al.，2013年）。另一方面，庞加莱嵌入属于表示学习的另一个分支——网络嵌入。网络嵌入通过简化原始网络，使其便于分析和可视化。与词嵌入一样，网络嵌入旨在从复杂离散数据中学习简单且可解释的连续表示。词嵌入以词语或情境中的词语作为表示的基本单位，而网络嵌入则旨在将节点、边或整个网络表示为表征单位。

许多广受欢迎的网络嵌入方法，例如 node2vec，仍然将网络表示在扁平的欧几里得空间中，但需要很多维度（Grover 和 Leskovec，2016年；Hamilton 等，2017年；Perozzi 等，2014年）。只要知识扩散轨迹呈现出层级结构，其中核心技术类别是其他类别的基础和构建模块，我们便采用基于双曲几何的庞加莱嵌入，该方法旨在自然地编码层级结构（Chamberlain 等，2017年；Nickel 和 Kiela，2017年）。

3.2 技术演化、引文网络与庞加莱嵌入

量化技术演化轨迹的一种常见方法，是构建并分析追踪获专利发明之间知识扩散的引文网络（Chang 等人，2009年；Feldman 和 Yoon，2012年）。引文网络揭示了知识传承，因为引用专利会从其引用的专利中汲取灵感。

知识网络具有清晰的层级结构，这种结构源自技术演化的历史。在此类网络中，如果某一类别被多个不同类别引用并与之结合，那么它所代表的技术便已广泛应用于各种产品和流程中，这表明该技术具有更高的“一般性”（Feldman 和 Yoon，2012；Petralia，2020年；Sterzi，2013年；Trajtenberg 等，1997）。在极端情况下，此类技术会成为“通用目的”技术，例如电力、互联网，以及近来的人工智能技术；它们广泛应用于生产经济的各个领域（Crafts，2021年；Petralia，2020年）。相比之下，如果某一技术类别被引用较少，却引用其他类别较多，则表明其具有更高的特殊性，例如作为上游技术要素的实现方式或应用（Fontana 等人，2008年）。

知识树的这种层级结构可以通过庞加莱嵌入有效捕捉，该方法专为以最小失真对离散层级结构建模而设计。与引文原始网络相比，后者即使相距遥远的节点也仍由边相连，而庞加莱嵌入能提供更清晰、更简约的表示。相较于欧几里得嵌入，它只需少得多的维度，就能捕捉复杂得多的信息。在图 2 上方的三个面板中，我们对比了网络表示、双曲嵌入和欧几里得嵌入。在维度相同的情况下，双曲嵌入能更有效地捕捉数据中的层级关系。

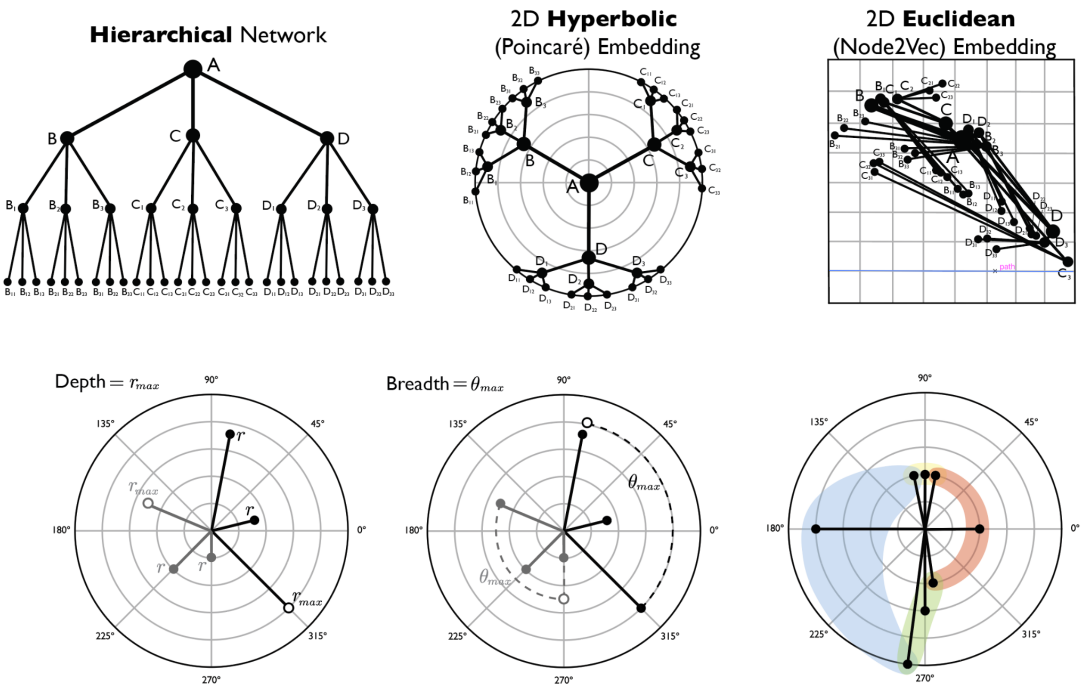


图 2 庞加莱圆盘上的层级深度与广度

在二维庞加莱嵌入圆盘中，节点具有两个维度：“一般性”维度从最居中的组件指向最不居中的组件；“相似性”维度则沿圆盘边缘延伸，相关组件在此处局部聚集，因此呈现出高度同质性。这一点在图 2 下方的三个面板中有所展示。左侧面板展示了嵌入的技术类别在庞加莱圆盘上的半径（r）差异。较大的r意味着技术类别及其相关专利“更深”或更具体。中间面板展示了跨越庞加莱圆盘的角度（θ）差异，用于衡量一项专利在技术锚定类别层级结构中的距离。深度与广度的不同组合形成四类发明：深度广泛（蓝色）、深度狭窄（绿色）、浅层广泛（红色）和浅层狭窄，如右侧面板所示。利用这种嵌入，我们重新审视深度与广度重组这一经典假说。

3.3 专利深度与广度度量的构建

我们基于在庞加莱圆盘中具有可识别现有技术类别的4,992,915项授权专利，考察深度与广度重组的影响。为便于计算未来五年引用次数等结果变量，我们的分析仅使用1976—2017年间授权的专利。首先，我们根据每个3年时期内专利中的技术类别引文构建加权有向网络（首个时期为1976—1978，最后一个时期为2015—2017），并将其嵌入庞加莱圆盘。随后，我们采用重组深度（r）和广度（θ）度量，并估计深度与广度重组对专利短期和长期影响的作用。

引文网络和庞加莱嵌入空间使用 Python 3 构建，回归分析则使用 Stata 18 进行。

3.3.1 美国专利的引文网络与庞加莱圆盘，1976—2017年

为构建第一步引文网络，我们采用 CPC分类体系，将每个四级分类（例如 A01B33）视为一个分析单元。由此得到约 20,000 个技术类别，作为引文网络中的节点。CPC分类体系为构建稳定的嵌入空间奠定了基础，相关验证测试见附录。

我们首先构建了 14 个时期的 CPC 技术类别之间的加权有向知识扩散网络：1976—1978、1979—1981、1982—1984 年，依此类推，最后一个时期为 2015—2017年。我们剔除了自环，因为它们与确定各类别之间的相对位置无关。在该网络中，被引用频率最高且引用来源最多样的类别位于层级结构的中心，表明其具有很高的下游应用潜力。

随后，我们使用庞加莱嵌入算法将该网络嵌入二维圆盘，训练参数设置遵循（Linzhuo 等，2020）[2]。

该嵌入空间以简洁的方式呈现了技术知识的层级结构。圆盘上的每个节点都被赋予极坐标（r，θ）。靠近中心的节点（r较小）会接收到来自多个领域的引文，并服务于许多应用。靠近边缘的节点（r较大）则应用于一组范围较窄的应用。与后续专利之间的θ差异较小，表明应用范围狭窄；θ差异较大，则表明多样性较广。由于我们的嵌入覆盖了14个时期且具有时间敏感性，节点位置可能会在不同嵌入之间发生变化。例如，随着整合型技术演变为未来发明的平台，它可能会从圆盘边缘向中心移动，反映其随时间变化的作用。

图 3（左图）展示了一个真实案例，呈现了2015年至2017年专利技术的庞加莱嵌入结果。节点颜色根据“部类符号”确定，即 CPC 分类系统中的最高层级类别。结果表明，物理学（G）和电力（H）是其他技术最普遍的基础，也构成了下游应用中的一个重要集群。除此之外，化学／冶金（C）、机械工程（F）和作业／运输（B）构成了另外三个主要的应用集群。作业／运输（B）遍布整个空间，起到黏合其他技术组件的作用。图3右图展示了技术树分支的三个示例，分别为小类族 C02F（黑色）、D01D（蓝色）和 G02B（红色）。各族中的节点散布

[2] 批量大小=30，学习率=0.2，模型每10个训练轮次评估一次。鉴于引文网络具有不对称性，我们将参数 symmetrize 设为 False。我们还将训练轮次设为1000，以确保充分训练，其中前20轮为“预热”轮次。

在圆盘上呈锥形分布，代表了通用组件逐渐组合成日益复杂应用的演化历程。例如，在G02B（光学元件）中，单片和复合透镜（G02B3）是显微镜（G02B21）和望远镜（G02B23）的基础。在 D01D（长丝、线和纤维制造的方法或设备）中，长丝成形（D01D5）是长丝成形处理（D01D1）和制造过程中的长丝物理处理（D01D10）的基础。

本方法的统计验证见附录A。这些检验表明，在全部14个建模时期中，嵌入空间保持稳定（使用普罗克拉斯提斯过程进行可视化；Hamilton等，2016）（附录A1），并且能够高效、准确地捕捉技术类别之间的距离（附录A2）。机器学习算法识别出的知识

Triulzi等，2020）。尽管我们的模型不像文本嵌入方法那样明确表示语义距离，但它考虑了领域之间的认知跨度，并能构建纳入这一维度的度量（附录 A3）。

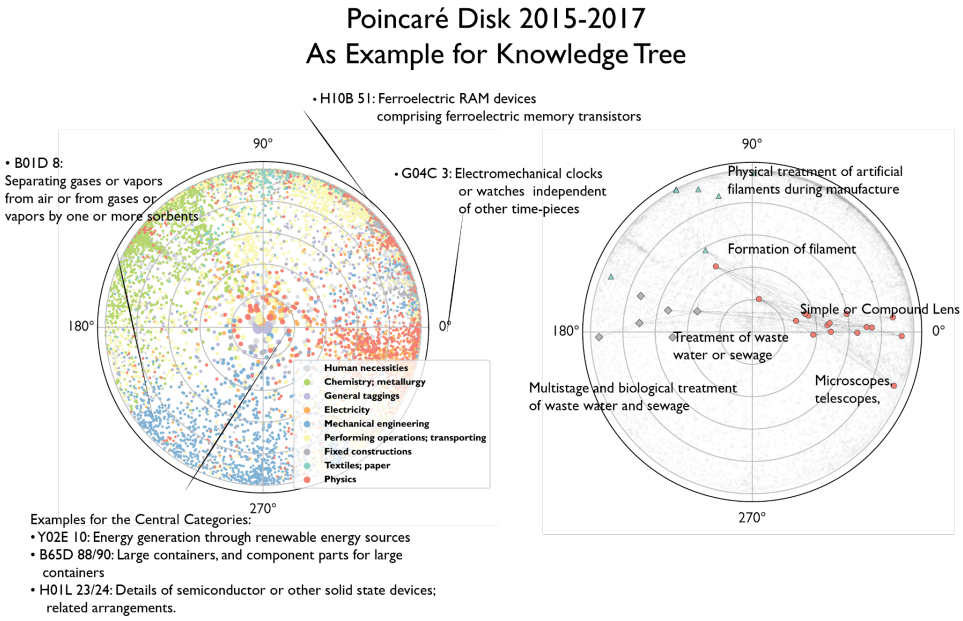


图 3 用于嵌入技术类别的庞加莱圆盘

3.3.2 构建深度与广度重组的度量

遵循既有文献的做法，我们将一步现有技术（被引专利）的技术类别视为每项焦点专利可能组合空间的代表（Kaplan 和 Vakili，2015a年；乌齐等，2013年；Verhoeven 等人，2016年c）。我们首先识别所有组合类别，并确定它们在庞加莱圆盘中的位置，然后根据这些位置及其在嵌入空间中的几何关系，计算深度与广度度量。

在二维庞加莱圆盘中，每个节点都用一个极坐标（r，θ）表示。由于 r表示知识树中的垂直深度，而θ 表示多样性，我们基于双曲嵌入构建深度与广度度量，以捕捉发明的最大深度和广度，具体如下：

$Depth = 90th\ Percentile(r_1, r_2, r_3, ...r_n)$，其中 $r_n = |r_n - r_{origin}|$

$Breadth = 90th\ Percentile(\theta_1, \theta_2, \theta_3, ...\theta_n)$，其中 $\theta_n = |\theta_i - \theta_j|, \forall i = i_1...i_k, \forall j = j_1...j_k, i \neq j$

深度度量知识树中知识元素的具体程度，而广度评估知识元素跨领域的跨度。我们使用第90百分位数作为阈值，以应对潜在的过度离散，但我们的结果对这一选择具有稳健性。使用第95百分位数和最大值作为替代阈值进行稳健性检验，所得结果相似（附录B4）。我们提供了两个专利示例来说明我们的度量，并在附录D2和D3中将其与经典度量进行比较。

不同于以往基于现有分类体系计算重组深度和宽度的研究，我们的度量将搜索风格视为极坐标二维空间中的联合分布。我们的数据中深度与广度之间的皮尔逊相关约为0.05，表明二者在概念上近乎独立。这使得高深度与高广度（或低深度与低广度）能够同时出现。它们与现有技术类别数量的相关性分别为0.09和0.20，表明二者仅存在较弱的关联。

基于这些稳健的、基于机器学习的指标，我们进行回归分析，以评估深度搜索与广度搜索对技术影响的作用。

4. 将深度与广度同短期与长期影响联系起来

4.1 变量

因变量。 我们使用专利前向引用这一最受认可的度量指标之一来衡量专利影响力。该指标指焦点专利在随后年份收到的引文数量，并且与专利许可收入相关（Sampat和Ziedonis，2005年）。引文数量越多，表明该专利所依据的技术被认为具有显著影响力，并且在后续发明活动中被频繁地重新使用、修订或扩展，其知识也在这些活动中传播（Fontana 等人，2008年）。专利引用中约有一半是

由专利发明人添加的，其余则由专利审查员在严格评估发明新颖性的过程中添加（Alcácer and Gittelman，2006年）。

由于引文增长具有右删失性质，技术影响通过专利获批后特定时期内获得的引文数量来衡量；实证研究中通常采用3–7年（Fleming，2001；Verhoeven 等人，2016c）。本文以5年作为短期引文的时间范围。同时，短期内积累大量引文未必能确保长期成功。集体发明机制——例如择优连接以及将创意融入后续研究——会在长期发挥影响（Wang 等人，2013）。为考量短期与长期影响之间的差异，我们构建了四个引文变量：0-5年、5-10年、10-15年和15-20年内的前向引用。这些变量追踪的影响从最即时到最持久不等。自变量。 我们采用基于双曲嵌入空间的深度和广度度量作为自变量，具体操作化方法见第3.3.2节。

控制变量。 我们纳入先前研究所确定的最重要控制变量，以控制潜在的混杂因素。控制变量包括：(1) 先前发明中组件和组合的熟悉度（Fleming，2001年）；(2) 焦点专利和现有技术中所包含的技术类别数量（Verhoeven 等人，2016年c）；(3) 发明人背景，包括组织受让人、团队协作和发明人经验的指标变量；以及 (4) 文献计量指标，包括焦点专利的非专利文献、权利要求数量和家庭规模（涵盖相同或相似技术内容的一组专利中的专利数量）。由于现有技术引用与现有技术中所代表的技术类别总数几乎完全相关（$r=0.93$），我们不

在回归中纳入前者。我们还控制年份固定效应以及所有专利类别的平均领域位置 θ，以解释不同技术领域之间的异质性。

4.2 数据与模型设定

在第一项分析中，我们使用1976—2017年间授权、且其先前类别可识别的所有美国专利，估计深度与广度对引文的影响。由此得到的样本共包含4,992,915项专利。引文呈现过度离散的非负整数值。遵循Fleming（2001年）和Verhoeven 等人（2016年）的做法，我们采用广义负二项回归来估计引文的平均值和离散度，并在附录B3中使用泊松极大似然估计量（“QMLE”）进行稳健性检验。第一组模型在专利层面进行估计：

$$log(Citations_i) = \beta_1 \times Depth_i + \beta_2 \times Breadth_i + \beta_3 \times Controls_i + \epsilon_i$$

，其中，表示焦点专利i的深度度量，表示专利广度。该模型估计可用最大样本中的组合效应。此处，包含、和，从而分别估计深度与广度对短期和长期影响的直接效应。在比较不同时期的影响时，我们仅纳入2002年以前授权的专利，以确保模型具有可比性。

$Depth_i$ $Breadth_i$ $Controls_i$ $Citations_i$ $Citations_{i,0-5}$ $Citations_{i,5-10}$ $Citations_{i,10-15}$ $Citations_{i,15-20}$

为解决潜在的内生性问题，我们还进行了两项额外分析。首先，我们采用了双胞胎专利设计（Bikard，2020年；Kovács等人，2021年）。在这项分析中，我们仅保留同时提交至美国专利商标局和欧洲专利局、并且在两个专利局均获授权的专利。这些专利的质量和专利特征完全相同，但在具有不同历史的不同知识系统中获得授权并接受评估，因而面临不同的预期。根据我们的理论预测，我们预期深度和广度效应仍将持续，

即使控制了发明的内在质量也是如此，因为引文是其他发明者通过集体智慧过程形成的认知和决策所塑造的。其次，我们使用工具变量——美国专利商标局分类体系的变化——来解决潜在的混杂效应。统计检验证实，该工具变量有效且具有较强解释力。双胞胎专利设计检验和工具变量分析的详情及结果见附录 B5。

在第二组分析中，我们旨在探究深度与广度通过何种机制分别与短期和长期影响相关联。在这部分分析中，我们基于完整的美国专利商标局样本（N=4,992,915）进行中介分析。我们从嵌入空间中构建了四个中介变量，以捕捉潜在的效应路径：（1）认知锁定，计算为一项发明在其自身领域内获得的认知可及引文（即，以$\Delta\theta$ 焦点专利与引用专利之间的标准差对引文进行除法运算）；（2）认知桥接，计算为后续研究在发明空间中的平均扩散范围（即，焦点专利与引用专利之间$\Delta\theta$ 的均值）；（3）网络锁定，以焦点专利的发明人与同一合作网络组件中后续研究的发明人之间的平均路径长度来度量，反映知识通过社会关系在本地社群中被识别和接受的程度；（4）结构桥接，以沿知识树各层级的后续研究平均扩散范围来度量（即，焦点专利与引用专利之间$|\Delta r|$ 的均值），反映跨知识层级形成结构连接的过程。尽管这些变量并非完美的代理指标，且可能部分受到外部因素影响，但它们使我们能够探究观测模式背后的智识与社会过程，为未来研究提出有意义且富有前景的假设。我们估计中介

回归，使用焦点专利获批后不同时期收到的引文，包括分别在0–5年、5–10年、10–15年和15–20年内获得的前向引用。

4.3 负二项回归结果

表1报告了不同样本中所有变量的基本描述性统计。表2呈现了相关矩阵。表3报告了负二项模型的结果。

我们的负二项模型表明，总体而言，深度对短期技术影响具有正向预测作用，而广度的影响则相反。深度增加一个单位（从理论最小值到最大值）会使五年前向引文数增加约0.196。乍看之下，这似乎不算显著。然而，考虑到前向引文分布高度偏斜，中位数为1，增加0.196意味着样本中超过一半的专利引文数上升约20%。相比之下，重组广度增加一个单位会使引文数减少0.043。鉴于庞加莱圆盘中广度的理论最大变化约为3.14弧度（180°），重组广度达到最大增幅将使短期引文减少约0.135。

模型2-5展示了基于2002年之前样本的回归分析，使我们能够利用20年（2002年—2022年）的时间跨度观察短期和长期引文。针对该样本，我们构建了四个因变量，分别统计不同时期的引文，并估算深度和广度随时间变化的效应：0–5年、5–10年、10–15年和15–20年引用次数。0–5年引用次数模型（模型2）的结果与一般模型（模型1）一致，但长期引文的深度和广度效应均出现反转（模型3-5）。随着时间推移，深度效应由正转负（从$\beta = 0.095$变为-0.434，两者均为$p<10^{-5}$），而广度效应则由负转正（从$\beta = -0.036$变为0.037，两者均为$p<10^{-5}$）。系数的变化

随着引文滞后期增长，深度和广度效应的变化均呈单调趋势，从模型2到模型5略有加速。我们在图4中直观展示了深度和广度效应随时间的变化。

我们还仅针对组织内引文以及其他子样本（2012 年之前的专利和 2007 年之前的专利）估计了模型1–5，以检验稳健性，结果呈现出相似的模式。这些结果见附录 B1 和 B2。在附录 E 中，我们还报告了离散度模型（表 E9），展示深度与广度对引文离散度的影响。总体而言，深度与广度都会降低引文的标准差（$\beta_{\text{depth}} = -0.117$, $\beta_{\text{breadth}} = -0.031$，如模型1所示）。不过，这两个变量随时间变化的模式不同：深度起初会提高方差，随后单调下降，最终转为负值；广度则始终保持稳定的负向影响。

我们的模型表明，专利的深度与广度会影响发明者对其重要性的判断，以及他们是否会在自己的工作中采用该专利。与跨类别理论（Zuckerman，1999年）一致，我们的模型显示，跨越许多彼此相距较远类别的发明会在短期内受到惩罚。不过，我们的研究还揭示了以往研究较少探讨的时间动态：跨类别发明起初可能因含义模糊而未受到充分认可，但随着时间推移往往会获得更多注意力；而深度较高的发明早期便脱颖而出，却很快耗尽其潜力。为探究这些模式背后的流程，我们转而分析其作用机制。

表1 描述性统计

	完整样本 (N=4,992,915)				双胞胎专利：美国样本 (N=84,916)				双胞胎专利：欧洲样本 (N=84,916)			
	均值	**标准差**	**MIN**	**MAX**	**MEAN**	**SD**	**MIN**	**MAX**	**MEAN**	**SD**	**MIN**	**MAX**
0–5年引用次数	3.443	10.156	0	1142	1.087	1.789	0	45	0.521	1.592	0	45
5–10年引用次数	5.320	19.205	0	2555	1.144	2.059	0	125	0.888	3.025	0	365
10–15年引用次数	4.115	18.029	0	2542	0.639	1.476	0	82	0.709	4.011	0	379
15–20年引用次数	2.926	16.013	0	2468	0.242	0.935	0	47	0.357	2.880	0	319
深度	0.817	0.160	.034	.999	0.680	0.172	0	0.999	0.671	0.245	0	1
广度	1.052	0.995	0.000	3.141	0.923	0.908	0	3.140	1.004	0.984	0	3.14
水平位置（平均 theta）	0.201	1.551	-3.139	3.138	-0.072	1.653	-3.313	3.129	0.211	1.549	-3.127	3.134
自然对数（组件熟悉度）	11.426	2.546	0	17.494	7.850	1.483	0	11.614	10.389	1.376	4.359	14.299
自然对数（组合熟悉度）	4.279	3.358	0	13.496	2.723	2.223	0	8.782	5.852	2.661	0	12.174
技术类别数量	5.372	4.352	0	17	5.345	4.040	0	16	4.586	3.096	1	12
现有技术类别数量	56.605	75.070	2	295	47.200	50.063	0	196	206.335	159.531	9	626
分配给组织	.897	0.304	0	1	0.973	0.162	0	1	0.973	0.162	0	1
团队发明	.628	0.483	0	1	0.709	0.454	0	1	0.709	0.454	0	1
Ln（发明人经验）	2.662	1.854	0	12.219	1.975	1.179	0	7.510	1.850	1.295	0.693	6.685
非专利文献	3.081	5.999	0	23	1.997	3.211	0	12	0.481	0.911	0	3
权利要求数量	15.272	8.907	1	36	15.724	8.982	1	37	4.43	6.461	0	20
家庭规模	.439	0.611	0	2	1	0	1	1	1	0	1	1

注：1976—2017年间由美国专利商标局授予的所有专利中，=4992,915 项；2001—2012年间由美国专利商标局和欧洲专利局共同授予的专利中，*N*=84,916 项。对于其引用的专利中包含 < 2 个现有技术类别的专利，由于无法构建有意义的自变量，故予以排除。

表2 主要变量的相关矩阵（全样本，N=4,992,91 5)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
（1）五年引用次数	1.000													
（2）深度	0.017	1												
	[0.000]													
（3）广度	0.034	-0.160	1											
	[0.000]	[0.000]												
（4）水平位置（平均 theta）	0.018	-0.060	0.010	1										
	[0.000]	[0.000]	[0.000]											
(5) 自然对数（组件熟悉度）	0.059	-0.421	0.003	0.037	1									
	[0.000]	[0.000]	[0.000]	[0.000]										
(6) 组合熟悉度的自然对数	0.013	-0.223	-0.122	0.016	0.314	1								
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]									
(7) 技术类别数量	0.086	-0.118	0.122	0.029	0.337	-0.371	1							
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]								
(8) 现有技术类别数量	0.169	0.043	0.274	0.016	0.254	-0.063	0.382	1						
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]							
（9）分配给组织	0.025	-0.094	-0.049	0.017	0.228	0.079	0.094	0.084	1					
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]						
(10) 团队发明	0.035	-0.067	0.004	0.002	0.174	0.008	0.128	0.113	0.291	1				
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]					
(11) Ln（发明人经验）	0.040	-0.139	0.036	0.005	0.305	0.088	0.201	0.223	0.224	0.178	1			
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]				
(12) 非专利文献	0.113	0.050	0.014	0.007	0.235	-0.030	0.235	0.456	0.112	0.149	0.165	1		
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]			
(13) 权利要求数量	0.098	-0.001	0.040	-0.011	0.154	0.039	0.097	0.214	0.093	0.098	0.089	0.183	1	
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]		
(14) 家庭规模	-0.071	-0.086	-0.042	0.014	-0.000	0.008	0.002	-0.184	0.158	0.056	0.081	-0.145	-0.166	1
	[0.000]	[0.000]	[0.000]	[0.000]	[0.544]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	

注：*p* 值位于方括号内。所有检验均为双尾检验。

表3的影响　　重组深度与广度对专利短期和长期影响的作用

	所有专利模型：1976—2017年		早期专利模型：1976—2002年		
	模型1：预测0-5年引文	**模型2 预测0-5年引文**	**模型3 预测5-10年引文**	**模型 4 预测10-15年份引文**	**模型5 预测15-20年份引文**
主要模型					
深度	0.196***	0.095***	-0.000	-0.164***	-0.434***
	[0.000]	[0.000]	[0.950]	[0.000]	[0.000]
	(0.005)	(0.006)	(0.007)	(0.008)	(0.009)
广度	-0.043***	-0.036***	-0.019***	-0.001	0.037***
	[0.000]	[0.000]	[0.000]	[0.549]	[0.000]
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
水平位置（平均 theta）	0.035***	0.077***	0.073***	0.069***	0.087***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
自然对数（组件熟悉度）	0.033***	0.029***	0.033***	0.034***	0.023***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
自然对数（组合熟悉度）	0.030***	0.035***	0.025***	0.019***	0.015***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
技术类别数量	0.028***	0.032***	0.040***	0.046***	0.050***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
现有技术类别数量	0.003***	0.003***	0.004***	0.004***	0.005***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
分配给组织	0.075***	0.115***	0.072***	-0.011***	-0.085***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
团队发明	0.076***	0.086***	0.094***	0.086***	0.084***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Ln（发明人经验）	0.010***	-0.000	-0.023***	-0.039***	-0.047***
	[0.000]	[0.880]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
非专利文献	0.009***	0.007***	0.017***	0.019***	0.017***

	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)
权利要求数量	0.018***	0.018***	0.021***	0.023***	0.024***
	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)
家庭规模	-0.169***	-0.101***	-0.203***	-0.268***	-0.308***
	[0.000] (0.001)	[0.000] (0.001)	[0.000] (0.001)	[0.000] (0.002)	[0.000] (0.002)
年份固定效应	YES	YES	YES	YES	YES
常数	0.049	0.099	0.036	0.206***	0.439***
	[0.367] (0.054)	[0.068] (0.054)	[0.525] (0.057)	[0.001] (0.061)	[0.000] (0.060)
离散度模型			采用相同的自变量，结果见附录 ***E***		
观测值	4992915	2010031	2010031	2010031	2010031

注：*p* 值为 e 位于方括号内。稳健标准误位于圆括号内。所有检验均为双尾检验。

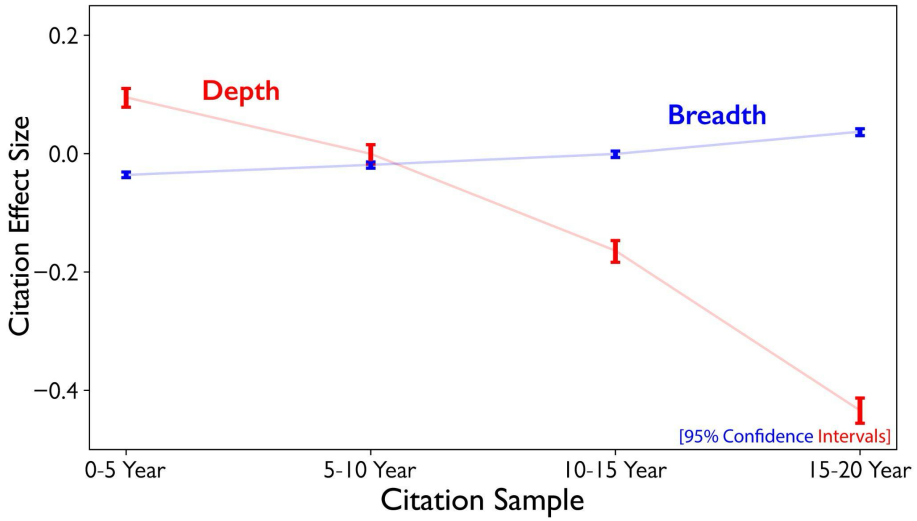


图4 深度和广度效应对短期和长期专利影响力的影响

5. 探索嵌入空间中的知识流动机制

我们的分析表明，搜索深度与短期内更高的前向引用相关，而广度的影响则相反；随着时间推移，这些影响的方向会发生逆转。为什么会这样？由于可以在庞加莱嵌入空间中追踪引用增长的路径和模式，我们的嵌入模型为探索其潜在机制提供了一个窗口。

我们首先基于嵌入空间构建认知锁定和桥接机制的代理度量。图5左侧展示了集体发明过程的概念示意图。焦点发明（深蓝色圆点）出现后，后续专利可能沿着相同的知识脉络（浅蓝色圆点）产生，也可能来自不同的知识领域（黄色圆点）。我们使用焦点专利的θ与后续专利的θ之差的绝对值（记为$|\Delta\theta|$）来度量领域间的水平扩散距离。如图5右侧对这一距离的概念分布所示，后续作品通常倾向于聚集在焦点发明周围，并以递减的概率向更远的区域扩散，这表明存在锁定过程。

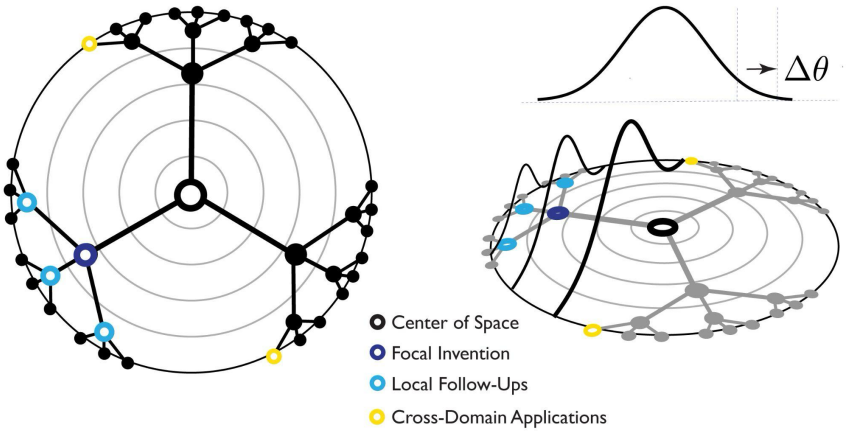


图5 发表后扩散的概念图

如第4.2节所述，该框架使我们能够为这些中介路径构建四个代理指标：(1) 认知锁定、(2) 认知桥接、(3) 网络锁定和(4) 结构桥接。基于这些中介变量，我们开展中介分析。

表4的结果显示，短期内，深度通过认知锁定和网络锁定均与更高的本地影响相关（认知锁定为β=1.214***，网络锁定为β=0.034***）；而广度则与垂直和水平方向上更广的扩散范围相关（认知桥接为β=0.171***，结构桥接为β=0.026***）。深度也有助于促进更多结构桥接，但其效应大小小于广度（β=0.019 < 0.026）。

表4 基于广义结构方程模型的影响力累积中介分析（五年）

	认知锁定	网络锁定	认知桥接	结构桥接	未来五年引用次数
认知锁定					0.166*** [0.000] (0.001)
认知桥接					0.181*** [0.000] (0.004)
结构桥接					0.201*** [0.000] (0.009)
网络锁定					0.070*** [0.000] (0.001)
深度	1.214*** [0.000] (0.020)	0.034*** [0.000] (0.004)	-0.250*** [0.000] (0.002)	0.019*** [0.000] (0.001)	-0.089*** [0.000] (0.007)
广度	-0.473*** [0.000] (0.002)	-0.041*** [0.000] (0.001)	0.171*** [0.000] (0.000)	0.026*** [0.000] (0.000)	0.047*** [0.000] (0.001)
水平位置（均值θ)	0.006*** [0.000] (0.002)	-0.001** [0.007] (0.000)	0.003*** [0.000] (0.000)	0.001*** [0.000] (0.000)	0.017*** [0.000] (0.001)
自然对数（组件熟悉度）	0.134***	0.025***	0.003***	0.000***	0.001^{+}

	[0.000] (0.001)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)	[0.070] (0.001)
自然对数（组合熟悉度）	-0.057*** [0.000] (0.001)	0.001*** [0.000] (0.000)	-0.029*** [0.000] (0.000)	-0.006*** [0.000] (0.000)	0.024*** [0.000] (0.000)
技术类别数量	0.028*** [0.000] (0.001)	-0.009*** [0.000] (0.000)	-0.005*** [0.000] (0.000)	-0.004*** [0.000] (0.000)	0.013*** [0.000] (0.000)
现有技术类别数量	0.001*** [0.000] (0.000)	0.000*** [0.000] (0.000)	-0.000*** [0.000] (0.000)	-0.000*** [0.000] (0.000)	0.002*** [0.000] (0.000)
分配给组织	0.165*** [0.000] (0.008)	0.292*** [0.000] (0.005)	-0.017*** [0.000] (0.001)	0.001*** [0.000] (0.000)	-0.033*** [0.000] (0.005)
团队发明	0.053*** [0.000] (0.005)	0.076*** [0.000] (0.001)	-0.002*** [0.000] (0.000)	0.000 [0.994] (0.000)	0.006** [0.005] (0.002)
自然对数（发明人经验）	0.001 [0.422] (0.001)	-0.044*** [0.000] (0.000)	-0.002*** [0.000] (0.000)	-0.000*** [0.001] (0.000)	0.002** [0.008] (0.001)
非专利文献	0.014*** [0.000] (0.000)	-0.010*** [0.000] (0.000)	-0.002*** [0.000] (0.000)	-0.000*** [0.000] (0.000)	0.007*** [0.000] (0.000)
权利要求数量	0.007*** [0.000] (0.000)	-0.001*** [0.000] (0.000)	-0.000*** [0.000] (0.000)	0.000 [0.170] (0.000)	0.008*** [0.000] (0.000)
家庭规模	-0.102*** [0.000] (0.004)	0.028*** [0.000] (0.001)	0.000 [0.224] (0.000)	0.001*** [0.000] (0.000)	-0.110*** [0.000] (0.001)
年份固定效应	YES	YES	YES	YES	YES
常数	1.634 [0.000] (0.130)	1.163 [0.000] (0.008)	0.608 [0.000] (0.025)	0.152 [0.000] (0.008)	1.091 [0.000] (0.028)
变量与协方差	控制了四个中介变量在成对层面的协变量				
N	2178535				

注：*P*值置于方括号内。稳健标准误置于圆括号内。所有检验均为双尾检验。

我们进一步考察这些效应路径如何随时间演变：针对四个因变量，分别复制同一模型，考察专利授权后0–5年、5–10年、10–15年和15–20年内获得的引文。为确保可比性，我们将样本限定为2002年以前授权的专利，以便所有时间窗口的回归均基于相同的数据集。

图6报告了通过四条路径估计的间接效应（即，深度/广度对中介效应，以及中介对引文效应）。这些结果带来了若干新的见解。

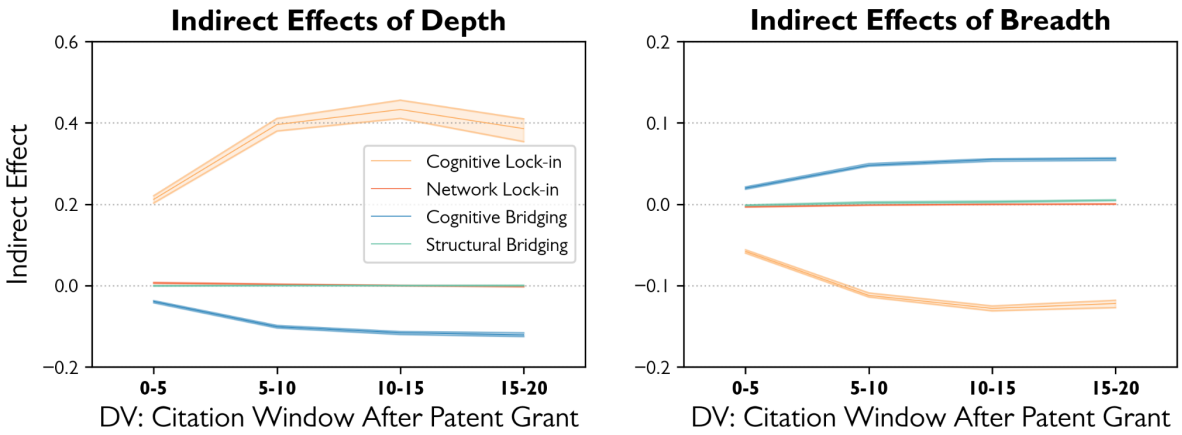


图6 间接　　效应：通过四种拟议机制（CI：1000次自助法重复抽样）

首先，在我们考察的四种机制中，认知机制——具体而言，认知锁定和认知桥接——似乎能够解释大部分观察到的效应。尽管结构桥接也与广度在后期年份带来的影响力收益呈现适度关联，但与认知路径相比，其影响仍然有限。其次，深度专利所获得的相当一部分引文仍来自本地社群（这体现在通过认知锁定产生的间接效应始终为正），而广度专利则始终与认知桥接呈正相关。值得注意的是，无论是深度还是广度，这两种间接效应之间的差距都会随时间扩大，这表明这两类发明获得的引文不仅在数量上有所不同，在性质上也存在差异。第三，深度对引文呈负向直接效应，且这一效应随时间增强（从 -0.086 增至 - 0.537）；而广度则呈现正向且不断增强的直接效应（从 0.030 增至 0.106）。因此，深度和广度很可能与专利质量的某些方面（例如新颖性）部分相关，而这些方面只有在专利公开数年后才会变得更加明显。

最后，在模型中控制四个中介变量后，深度和广度的系数均改变了符号，这表明这些中介变量共同抑制了搜索策略对技术影响的效应。这一发现表明，我们在此识别的路径在很大程度上解释了深度和广度早期的引文模式。与此同时，直接效应（控制机制后）可能反映了这些关联中在考虑社会与信息背景后仍然存在的部分；在这种情况下，评价主要由发明的内在特征所塑造。值得注意的是，这些系数的方向与长期影响模型（例如 15–20 年模型）中的系数一致；在这些模型中，锁定机制似乎明显较弱。

这些结果表明，深度和广度的影响随时间变化，部分可由三个因素的多重效应解释：后期局部回报增长放缓、广泛受众的重要性日益提升，以及直接效应不断增强。

6. 深度和广度效应对集体发明的影响

从企业和社会的角度看，发明并非彼此独立的产品，而是共同编织成一张集体网络，影响未来的创新（Shi 等人，2015）。为评估专利深度和广度如何影响后续作品中的技术搜索，我们开展了另外两项分析，以进一步考察集体情境如何增强或抑制深度和广度效应。

首先，我们探讨深度、广度与后续研究知识结构之间的关系。我们首先计算 5 年内后续专利的平均深度和广度。然后，我们采用普通最小二乘模型，估计深度、广度与后续深度和广度特征之间的关联。如图7左侧面板所示

并见附录表 E7-E8，我们的结果揭示了两种模式：（1）深度会带来更多深度，正如广度会带来更多广度。焦点深度每增加一个单位，后续深度便增加 0.414 个单位（模型1）；焦点广度每增加一个单位，后续广度便增加 0.505 个单位（模型2）。在同一组织内，这两种关联都更为显著（表 E7 中为 $\beta_{depth\_to\_depth} = 0.515 > 0.414$，表 E8 中为 $\beta_{breadth\_to\_breadth} = 0.595 > 0.505$），这可能是因为知识交流的障碍有所减弱。（2）表 E8 中的模型（针对后续广度）显示，焦点专利的深度与广度之间存在正向交互作用。这表明深度与广度相互促进、彼此补充。这种交互作用可能意味着，“深度”发明作为未来与其他技术重组时的可靠组成部分，具有更高的稳健性（模型3 中为 $\beta_{depth\ x\ 广度} = 0.404$，模型 4 中为 $\beta_{depth\ x\ 广度} = 0.411$ ）。这种调节效应在组织内部更大，这表明，无须跨越组织边界，发明深度便能更有效地为焦点发明融入未来工作做好准备。

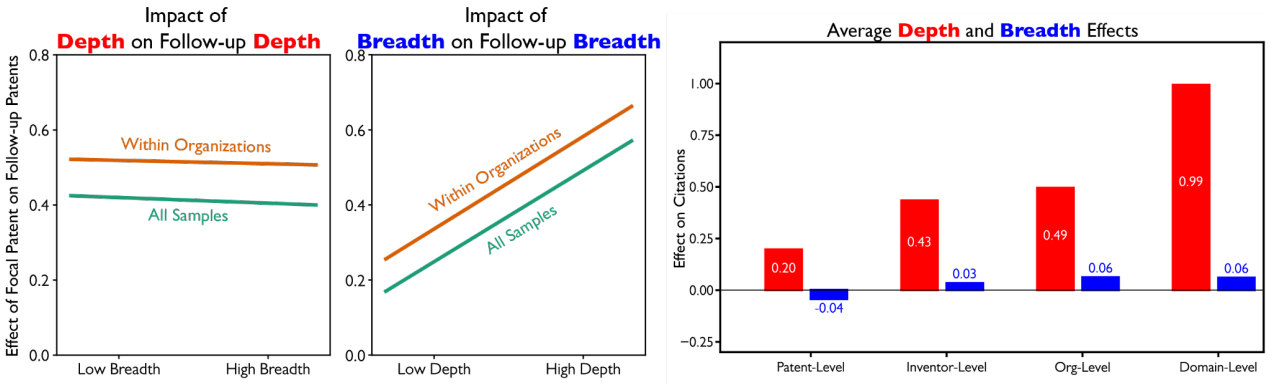


图7 深度与广度的组织效应和集体效应

为进一步考察这种集体效应，我们在不同分析层级上开展分析，以重复验证研究结果。这些层级涵盖技术发展中从小到大的创新发生地——从发明人到组织（即原始专利受让人），再到技术领域（定义为 5 位数 CPC 类别）。如果集体思维通过注意力发挥作用——引导发明人将注意力集中于相关组件，并加快新颖要素整合的进程——那么，随着分析单位从专利层面依次扩大到发明人、组织，最终到领域层级，我们应该会看到搜索深度与广度对引文的影响不断增强，因为更大的单位拥有更多的剩余能力来促进集体智慧。图7右侧面板展示了这一模式。从专利层面模型到领域层级模型，我们观察到深度（$\beta_{patent\_depth} = 0.200$, $\beta_{inventor\_depth} = 0.433$, $\beta_{organization\_depth} = 0.494$, $\beta_{domain\_depth} = 0.992$，所有结果的 $p<10^{-128}$）和广度（$\beta_{patent\_breadth} = -0.035$, $\beta_{inventor\_breadth} = 0.033$, $\beta_{organization\_breadth} = 0.061$, $\beta_{domain\_breadth} = 0.059$，所有结果的 $p<10^{-186}$）的影响均随层级升高而明显增强，这支持了我们关于集体注意力作用机制的假设。

7. 讨论与结论

7.1 通过计算分析重新审视深度与广度之争 s

技术搜索的深度与广度如何影响创新的影响力？数十年来，这个问题一直备受争论。尽管这一问题在理论和实践层面都十分重要，但以往研究对二者的影响得出了相互矛盾的结论。Kaplan 和 Vakili（2015a年）将这场争论概括为“基础性”观点与“张力”观点之争。为推进这一研究方向，我们采用计算方法对技术景观建模，并开发了精细刻画重组结构和扩散路径的度量方法。这一方法使我们能够从集体智慧的视角考察深度与广度的影响，并对以往难以观察的机制进行实证检验。

我们首先运用基于双曲几何的 Poincaré 嵌入方法，在庞加莱圆盘中构建技术类别的知识树。在这一表示中，纵向维度r表示从通用技术组件到特定应用的连续谱，即技术组合的深度；而横向维度θ 表示跨领域组合的广度或多样性。基于这一知识图谱，我们构建了重组深度与广度的度量方法。不同于将这两个因素视为权衡取舍的传统方法，我们依据嵌入空间中的几何定义，分别以.....的第90百分位数和.....的第90百分位数来度量深度与广度。前者衡量发明人或公司在所属领域内纳入并改进现有技术基础的程度，后者则衡量发明人或公司以出乎意料的方式跨领域组合组件的程度。这种方法更符合现实观察：许多组合要么兼具深度与广度，要么两者皆不具备。以整个知识系统的结构为参照框架，我们的模型能够准确捕捉技术功能，清晰区分深度与广度这两个维度，支持跨产业比较，并从集体智慧的角度考量其意义。双曲表示为理解不同搜索策略的结果提供了一个简单、清晰且有说服力的框架，同时也开辟了新的理论研究空间。

Δr $\Delta \theta$

我们的发现为组合式创新的机制提供了新的见解，并具有重要的理论意义。首先，我们观察到一个清晰且一致的模式：深度与短期影响正相关，与长期影响负相关。广度则呈现相反的作用。这表明，基础观与张力观并不矛盾，而是分别揭示了现实的不同侧面。基础性观点认为局部

搜索能获得更多局部认可，这一观点是正确的。张力观点认为，局部搜索会导致认知“锁定”，并限制创造性冲突，这一观点也是正确的。不过，这两种结果并非同时发生。专利授权后会立即获得局部认可，但负面约束则会随着时间推移逐渐显现并加剧。鉴于不同技术领域的演进速度差异很大——在快速发展的人工智能领域，5 年可能已算“长期”，而在传统化学领域则可能仍属“短期”——以往针对特定领域的研究即使使用相同的前向引用指标，也可能只揭示了问题的一个侧面，因而得出彼此不同且受到混杂因素影响的结论。因此，我们的研究强调了时间在组织学习中的作用，为关于深度与广度的长期争论提供了解答。

其次，我们对后续专利特征的模型揭示了利用与探索逐步展开的相互作用，凸显了二者的协同关系，尤其是在能够战略性地配置工程资源、以既有进展为基础继续发展的公司中。在我们通过庞加莱圆盘描绘的技术知识树中，当知识沿纵向移动时，利用会迅速发生。相比之下，探索则间歇性地发生，因为发明者会建立跨越庞加莱纬度的联系。尽管利用的即时优势会随着时间推移而减弱其技术影响力，但它能为组件做好准备，使其顺利整合到横向创新中。这种“隐形影响”质疑了深度重组仅与利用相关联的传统观点，并表明创新型企业有必要维持兼具探索与利用的组合，使二者能够随着时间推移相互促进。

我们承认本研究存在若干局限。首先，尽管我们的模型为追踪引用增长提供了数学基础，但现有数据无法让我们充分区分所有锁定机制和桥接机制。此外，我们使用的中介变量是

对所研究社会过程的不完美代理指标。未来研究应利用更细粒度的数据，更深入地探究这些机制，并揭示更精确的模式。其次，我们对深度与广度的度量依赖于对技术类别的准确划分。尽管遗漏的情况很少，但一旦发生，可能会引入测量误差。第三，我们的嵌入仅依赖引文网络，而非文本信息。引文是知识传承的不精确代理指标，因为它往往受到法律和战略考量的影响。虽然我们的方法显著降低了计算成本、增强了普适性，且与文本分析方法总体一致（见附录 A3），但可能牺牲了粒度。最后，我们的嵌入空间采用3年时间窗口，能够捕捉技术演化的部分动态，但可能会遗漏快速变化。例如，一项新兴的整合型技术若迅速发展为平台，其位置可能在不到三年的时间里发生变化。这类快速发展可能无法在我们当前的模型中得到充分体现。未来研究可以拓展我们的分析、扩大应用案例的范围，并进一步提高算法准确性。

7.2 对研发管理的启示

我们的研究也具有广泛的实践意义，尤其是对研发管理而言。首先，通过揭示搜索中深度与广度的时间收益，我们的研究为发明者和组织根据自身优先事项做出明智的研发决策提供了指导。例如，小型企业和个人发明者往往缺乏长期研发投资所需的资源，又需要立即获得市场回报，因此可能会认为深度搜索是更安全、更可靠的策略。对于这些创新者而言，深度搜索能够带来更快的市场反馈，并为技术迭代提供更清晰的路径。

其次，我们的研究还提出了另一种潜在策略，尤其适用于大型企业和市场在位者：随着时间推移，交替进行深度搜索与广度搜索。通过深度搜索开发出的稳健模块能够支持并增强未来的探索，使深度与广度相互促进。这种做法体现了双元性原则（O'Reilly and Tushman, 2008; Raisch and Birkinshaw, 2008）在不同时间尺度上的应用。由于这一策略受到研发预算、组织规模以及技术领域成熟度的制约（McCutchen and Swamidass, 1996; Sørensen and Stuart, 2000; Tsai and Wang, 2005），拥有充足资源且容错能力较强的大型企业更有条件实施该策略。辉瑞是一个典型案例（图8），其专利申请轨迹展现出清晰的随时间推移不断放大与缩小的模式。

总体而言，我们的工作支持个人发明者和组织制定有效的研发决策，并提升其在市场中的竞争力。

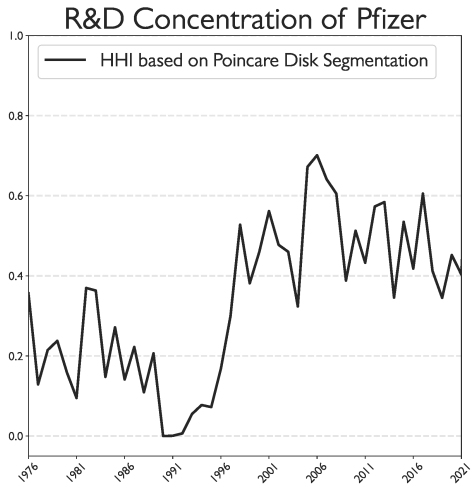


图8 辉瑞随时间变化的技术集中度（按庞加莱分段计算的赫芬达尔-赫希曼指数）

参考文献

Aceves, P., Evans, J., 2024年。《调动概念空间：词嵌入模型如何为组织科学中的测量与理论提供信息》。Organization Science 35, 788–814。Aharonson, B.S., Schilling, M.A., 2016年。《绘制技术景观：测量技术距离、技术足迹与技术演化》。《研究政策》45, 81–96。Alcácer, J., Gittelman, M., 2006年。《专利引用作为知识流的度量：审查员引用的影响》。《经济学与统计学评论》88, 774–779。亚里士多德, 1999年。《形而上学》，新版。Penguin Classics，伦敦。Arthur, W.B., Polak, W., 2006年。《简单计算机模型中的技术演化》。《复杂性》11, 23–31。Augustine, S., 2003年。《上帝之城》。Penguin UK。Benson, C.L., Magee, C.L., 2015年。《专利检索方法扩展所带来的技术结构影响》。《科学计量学》102, 1965–1985。柏林，2013年。《刺猬与狐狸：论托尔斯泰的历史观——第二版》。普林斯顿大学出版社。Bernstein, J.I., Nadiri, M.I., 1989年。《研发与行业内溢出：动态二元性的实证应用》。《经济研究评论》56, 249–267。Bikard, M., 2020年。《创意双生：将同时发现作为研究工具》。《战略管理杂志》41, 1528–1543。Bloom, N., Schankerman, M., Van Reenen, J., 2013年。《识别技术溢出与产品市场竞争》。《计量经济学》81, 1347–1393。博，W.F., Evaristo, R., Ouderkirk, A., 2014年。《平衡创新所需专业知识的广度与深度：3M的案例》。《研究政策》43, 349–366。Brahm, F., Parmigiani, A., Tarzijan, J., 2021年。《企业能否兼具广度与深度？探索企业横向范围与纵向范围之间的相互依赖关系》。《管理学期刊》47, 1219–1254。Brian Arthur, W., 2009年。《技术的本质：技术是什么以及它如何演化》。Simon and Schuster。Carlile, P.R., 2004年。《转移、翻译与转化：跨边界管理知识的整合框架》。《组织科学》15, 555–568。Chamberlain, B.P., Clough, J., Deisenroth, M.P., 2017年。《双曲空间中的图的神经嵌入》。Chami, I., Ying, R., Ré, C., Leskovec, J., 2019年。《双曲图卷积神经网络》。《神经信息处理系统进展》32, 4869–4880。Chang, S.-B., Lai, K.-K., Chang, S.-M., 2009年。《探索商业方法的技术扩散与分类：使用专利引用网络》。《技术预测与社会变迁》76, 107–117。查托帕迪亚，S., 伯科维茨，J., 2020年。《一扇门关闭，另一扇门开启……但对某些人而言如此：来自TRIPS协定后印度制药业的证据》。《战略管理杂志》41, 988–1022。崔，J., Menon, A., Tabakovic, H., 2021年。《利用机器学习重新审视多元化与绩效关系》。《战略管理杂志》42, 1632–1661。乔杜里，P., Allen, R.T., Endres, M.G., 2021年。《用于管理学研究模式发现的机器学习》。《战略管理杂志》42, 30–57。乔杜里，P., Wang, D., Carlson, N.A., Khanna, T., 2019年。《面部与文本分析的机器学习方法：发现首席执行官的口头沟通风格》。《战略管理杂志》40, 1705–1732。Christensen, C.M., Bower, J.L., 1996年。《客户权力、战略投资与领先企业的失败》。《战略管理杂志》17, 197–218。Corrocher, N., Malerba, F., Montobbio, F., 2007年。《信息与通信技术领域创新活动的熊彼特式模式》。《研究政策》36, 418–432。Crafts, N., 2021年。《作为通用技术的人工智能：历史视角》。《牛津经济政策评论》37, 521–536。Csikszentmihalyi, M., 2009年。《创造力：心流与发现心理学》。Harper Collins。Dhanaraj, C., Parkhe, A., 2006年。《协调创新网络》。《管理学会评论》。Dougherty, D., 1992年。《大型企业成功产品创新的解释性障碍》。《组织科学》3, 179–202。Dyson, F.J., 2015年。《鸟与蛙：Freeman Dyson 1990–2014年精选论文》。World Scientific Publishing Company。Escolar, E.G., Hiraoka, Y., Igami, M., Ozcan, Y., 2023年。《绘制企业在技术空间中的位置：专利统计的拓扑分析》。《研究政策》52, 104821。

Ethiraj, S.K., Levinthal, D., Roy, R.R., 2008. 模块化的双重作用：创新与模仿。《管理科学》。Feldman, M.P., Yoon, J.W., 2012. 通用目的技术的实证检验：对 Cohen–Boyer 重组DNA技术的考察。《产业与企业变革》21, 249–275. Fellbaum, C., 2005. WordNet与WordNet词网。载于：Barber, A.（编），《语言与语言学百科全书》。爱思唯尔，第2–665页。Fleming, L., 2001. 技术搜索中的重组不确定性。《管理科学》47, 117–132. Fleming, L., Sorenson, O., 2004. 科学作为技术搜索的地图。《战略管理杂志》25, 909–928. Fleming, L., Waguespack, D.M., 2007. 开放式创新社群中的经纪、跨边界与领导力。《组织科学》18, 165–180. Fontana, R., Nuvolari, A., Verspagen, B., 2008. 将技术轨迹映射为专利引文网络：数据通信标准的应用。《创新与新技术经济学》18. Ghemawat, P., 1991. 承诺：战略的动态。《哈佛商业评论》。Ghosh, A., Ranganathan, R., Rosenkopf, L., 2016. 情境与模型选择对战略联盟形成决定因素的影响：分阶段重复研究的证据。《战略管理杂志》37, 2204–2221. Granovetter, M.S., 1973. 弱关系的力量。《美国社会学杂志》78, 1360–1380. Grover, A., Leskovec, J., 2016. node2vec：网络的可扩展特征学习，KDD ’16。计算机协会。Guillard, C., Martin, R., Mohnen, P., Thomas, C., Verhoeven, D., n.d. 面向创新的高效产业政策：站在隐形巨人的肩膀上（第CEPDP1813号）。经济绩效研究中心，伦敦政治经济学院。Hall, B.H., Jaffe, A.B., Trajtenberg, M., 2001. NBER专利引文数据文件：经验、见解与方法工具。Hamilton, W.L., Leskovec, J., Jurafsky, D., 2016. 历时词嵌入揭示语义变化的统计规律。arXiv [cs.CL]。Hamilton, W.L., Ying, R., Leskovec, J., 2017. 大型图上的归纳式表示学习，NIPS’17。Curran Associates Inc. Hargadon, A.B., 2002. 知识经纪：连接学习与创新。《组织行为研究》24, 41–85. Hargadon, A., Sutton, R.I., 1997. 产品开发企业中的技术经纪与创新。《行政科学季刊》42, 716. Harrison, J.S., Josefy, M.A., Kalm, M., Krause, R., n.d. 运用监督式机器学习扩展人工编码数据：董事会领导力情境中的一种方法与数据集。《战略管理杂志》。Henderson, R.M., Clark, K.B., 1990. 架构创新：现有产品技术的重构与成熟企业的失败。《行政科学季刊》35, 9–30. Huenteler, J., Ossenbrink, J., Schmidt, T.S., Hoffmann, V.H., 2016. 产品设计层级如何塑造技术知识的演进——来自风力发电领域专利引文网络的证据。《研究政策》45, 11951217–。Joulin, A., Grave, E., Bojanowski, P., Mikolov, T., 2017. 高效文本分类技巧集。发表于EACL 2017，计算语言学协会。Kang, D., Danziger, R.S., Rehman, J., Evans, J.A., 2024. 科学知识的有限扩散预示着崩溃。《自然·人类行为》1–9. Kaplan, S., Vakili, K., 2015a. 突破性创新中重组的双刃剑。《战略管理杂志》36, 1435–1457. Kaplan, S., Vakili, K., 2015b. 突破性创新中重组的双刃剑。《战略管理杂志》36, 1435–1457. Kellogg, K.C., Orlikowski, W.J., Yates, J., 2006. 交易区中的生活：后科层制组织中的跨边界协调结构。《组织科学》17, 22–44. Ke, Q., Ferrara, E., Radicchi, F., Flammini, A., 2015. 定义并识别科学睡美人。《美国国家科学院院刊》112, 7426–7431. Kim, K., Guler, I., Karim, S., n.d. 谁会被重新部署？发明者特征与资源重新部署决策。《战略管理杂志》n/a。Kovács, B., Carnabuci, G., Wezel, F.C., 2021. 类别、注意力与发明的影响。《战略管理杂志》42, 992–1023. Kozlowski, A.C., Taddy, M., Evans, J.A., 2019. 文化的几何结构：通过词嵌入分析阶级的意义。《美国社会学评论》84, 905–949. Kumar, P., Liu, X., Zaheer, A., 2022 43, 14331468. 企业的联盟网络有多重要？《战略管理杂志》–。

Lamoreaux, N.R., Sokoloff, K.L., 1999. 美国1840—1920年的发明活动与技术市场。工作论文系列。Leonard-Barton, D., 1993. 核心能力与核心刚性：管理新产品开发中的悖论。《长期规划》26, 154. Levinthal, D.A., March, J.G., 1993. 学习短视。《战略管理杂志》14, 95–112. Linzhuo, L., Lingfei, W., James, E., 2020. 社会中心化与语义坍缩：网络与文本的双曲嵌入。《诗学》。Lobo, J., Strumsky, D., 2019. 发明新颖性的来源：两套专利分类体系，同一个故事。《科学计量学》120, 19–37. Lodh, S., Battaggion, M.R., 2015. 创新中的技术知识广度与深度：并购在生物技术领域中的作用。《产业与企业变革》24, 383–415. Magee, C.L., Basnet, S., Funk, J.L., Benson, C.L., 2016. 技术绩效的定量实证趋势。《技术预测与社会变迁》104, 237–246. Mannucci, P.V., Yong, K., 2018. 知识深度与知识广度对个人职业生涯中创造力的差异化影响。《管理学会期刊》61, 1741–1763. March, J.G., 1991. 组织学习中的探索与利用。《组织科学》2, 71–87. March, J.G., Simon, H.A., 1993.《组织》，第2版。Wiley-Blackwell，剑桥，马萨诸塞州，美国。McCutchen, W.W., Swamidass, P.M., 1996. 探索大型生物技术研究企业的战略：基于小型与大型企业比较的预测。《技术预测与社会变迁》51, 173–184. Mikolov, T., Chen, K., Corrado, G., Dean, J., 2013. 向量空间中词表示的高效估计。arXiv [cs.CL]。Milojević, S., 2015. 量化科学的认知范围。《信息计量学杂志》9, 962–973. Moorthy, S., Polley, D.E., 2010. 技术知识的广度与深度：绩效影响。《知识管理杂志》14, 359–377. Nakamura, H., Suzuki, S., Sakata, I., Kajikawa, Y., 2015. 知识组合建模：不同技术领域之间知识相似性的测量。《技术预测与社会变迁》94, 187–201. Navis, C., Glynn, M.A., 2010. 新市场类别如何涌现：卫星广播领域合法性、身份与创业的时间动态，1990—2005年。《行政科学季刊》55, 439–471. Nelson, R.R., Winter, S.G., 1990. 经济变迁演化理论。哈佛大学出版社，伦敦，英格兰。Nickel, M., Kiela, D., 2017. 用于学习层次表示的庞加莱嵌入。Curran Associates, Inc. O’Reilly, C.A., Tushman, M.L., 2008. 作为动态能力的双元性：化解创新者的窘境。《组织行为研究》28, 185–206. Paruchuri, S., Awate, S., 2017. 组织知识网络与局部搜索：组织内部发明者网络的作用。《战略管理杂志》38, 657–675. Perozzi, B., Al-Rfou, R., Skiena, S., 2014. DeepWalk：社会表示的在线学习，KDD ’14。计算机协会。Petralia, S., 2020. 用专利数据映射通用技术。《研究政策》49, 104013. Raisch, S., Birkinshaw, J., 2008. 组织双元性：前因、结果与调节变量。《管理学期刊》34, 375–409. Risha, Z., Lin, Y., Leahey, E., Wu, L., 2023. 替代文艺复兴人：团队是否大于其组成部分之和？arXiv [cs.SI]。Rogers, E.M., Singhal, A., Quinlan, M.M., 2014. 创新扩散。载于：《传播理论与研究综合方法》。Routledge，第432–448页。Rosenkopf, L., Nerkar, A., 2001. 超越局部搜索：光盘行业中的跨边界、探索与影响。《战略管理杂志》22, 287–306. Sampat, B.N., Ziedonis, A.A., 2005. 专利引文与专利的经济价值。载于：Moed, H.F., Glänzel, W., Schmoch, U.（编），《定量科学与技术研究手册：科学技术系统研究中的公开与专利统计》。施普林格荷兰，多德雷赫特，第277–298页。Santos, F.M., Eisenhardt, K.M., 2009. 市场构建与边界塑造：新兴领域中的企业家权力。《管理学会期刊》52, 643–671. Sgourev, S.V., Aadland, E., Formilan, G., 2023. 审美空间中的关系：颜色如何助力市场定位。《行政科学季刊》68, 146–185. Shapiro, C., 2000. 驾驭专利丛林：交叉许可、专利池与标准制定。《创新政策与经济》1, 119–150。

Shi, F., Foster, J.G., Evans, J.A., 2015年。编织科学的结构：科学演进结构的动态网络模型。《社会网络》43，73–85。Shrestha, Y.R., He, V.F., Puranam, P., von Krogh, G., 2021年。用于构建理论的算法支持归纳：我们如何利用预测模型进行理论化？《组织科学》32，856–880。Silva, F.N., Tandon, A., Amancio, D.R., Flammini, A., Menczer, F., Milojević, S., Fortunato, S., 2020年。新近性可以预测作者引文演变中的爆发。《定量科学研究》1，1298–1308。Simon, H.A., 1991年。《复杂性的架构》。载于：Klir, G.J.（编），《系统科学的多个方面》。施普林格美国公司，马萨诸塞州波士顿，第457–476页。Sørensen, J.B., Stuart, T.E., 2000年。老化、过时与组织创新。《行政科学季刊》45，81–112。Stark, D., 2011年。《失调感：经济生活中的价值论述》。普林斯顿大学出版社，新泽西州普林斯顿。Sternberg, R.J., Lubart, T.I., 1995年。《挑战大众：在顺从文化中培育创造力》。Sterzi, V., 2013年。专利质量与所有权：英国高校教职人员专利申请分析。《研究政策》42，564–576。Stettner, U., Lavie, D., 2014年。审视双元性：通过内部组织、联盟与收购实现探索与利用。《战略管理杂志》35，1903–1929。Stuart, T.E., Podolny, J.M., 2007年。局部搜索与技术能力的演进。《战略管理杂志》17，21–38。Taylor, A., Greve, H.R., 2006年。超人还是神奇四侠？创新团队中的知识组合与经验。《管理学会期刊》49，723–740。Teece, D.J., 1986年。从技术创新中获利：对整合、合作、许可和公共政策的启示。《研究政策》15，285–305。Teodoridis, F., Bikard, M., Vakili, K., 2019年。知识前沿的创造力：专业化在快节奏与慢节奏领域中的影响*。《行政科学季刊》64，894–927。Teodoridis, F., Lu, J., Furman, J.L., 2022年。绘制知识空间：运用无辅助机器学习工具。工作论文系列。Terjesen, S., Patel, P.C., 2017年。寻找流程创新：搜索深度、搜索广度与行业环境的作用。《管理学期刊》43，1421–1446。Ter Wal, A.L.J., Alexy, O., Block, J., Sandner, P.G., 2016年。兼得两全：开放专业化联合网络与封闭多元化联合网络对新创企业成功的益处。《行政科学季刊》61，393–432。Trajtenberg, M., Henderson, R., Jaffe, A., 1997年。大学专利与企业专利：观察发明基础性的窗口。《创新与新技术经济学》5，19–50。Tripsas, M., Gavetti, G., 2000年。能力、认知与惯性：数字成像领域的证据。《战略管理杂志》21，1147–1161。Triulzi, G., Alstott, J., Magee, C.L., 2020年。通过挖掘专利数据估算技术性能提升率。《技术预测与社会变迁》158，120100。Tsai, K.H., Wang, J.C., 2005年。研发绩效是否会随企业规模扩大而下降？——从弹性角度重新考察。《研究政策》34，966–976。Tushman, M.L., Scanlan, T.J., 1981年。边界跨越者：他们在信息传递中的作用及其前因。《管理学会期刊》24，289–305。Uzzi, B., 1997年。企业间网络中的社会结构与竞争：嵌入性的悖论。《行政科学季刊》42，35。Uzzi, B., Mukherjee, S., Stringer, M., J ones, B., 2013年。非典型组合与科学影响力。《科学》342，468–472。Van de Ven, A.H., 1986年。创新管理中的核心问题。《管理科学》32，590–607。Verhoeven, D., Bakker, J., Veugelers, R., 2016a年。利用基于专利的指标衡量技术新颖性。《研究政策》。Verhoeven, D., Bakker, J., Veugelers, R., 2016b年。利用基于专利的指标衡量技术新颖性。《研究政策》。Verhoeven, D., Bakker, J., Veugelers, R., 2016c年。利用基于专利的指标衡量技术新颖性。《研究政策》。Vicinanza, P., Goldberg, A., Srivastava, S.B., 2023年。先见性创意深度学习模型表明，这些创意源自边缘。《PNAS Nexus》2，gac275。Wang, D., Song, C., Barabási, A.-L., 2013年。量化长期科学影响力。《科学》342，127–132。Weingart, S.B., n.d. 从树状结构到网络结构：通过可视化重塑知识。Weitzman, M.L., 1998年。重组增长*。《经济学季刊》113，331–360。Williamson, O., 1985年。E.（1985）：《资本主义的经济制度》。纽约。

Xiao, T., Makhija, M., Karim, S., 2022年。创新的知识重组视角：综述与新研究方向。《管理学期刊》48，1724–1777。Zheng, L., Cao, L., Ren, J., Li, X., Yin, X., Chen, J., 2022年。风险投资公司如何选择联合投资伙伴：制度不确定性与投资偏好的调节效应。《管理与组织评论》18，463–490。Zhou, K.Z., Li, C.B., 2012年。知识如何影响根本性创新：知识基础、市场知识获取与内部知识共享。《战略管理杂志》33，1090–1102。Zuckerman, E.W., 1999年。《分类命令：证券分析师与非法性折价》。《美国社会学杂志》104，1398–1438。

补充材料：

技术层级重组中的深度与广度

附录A：庞加莱嵌入空间的验证

附录A验证了庞加莱嵌入空间能够有效表示技术树。为此，我们（1）对14个庞加莱圆盘进行可视化，并通过计算各技术类别在不同时期的 r 和 theta 的相关性，评估其跨时期稳定性；（2）将庞加莱嵌入空间中的度量与知识扩散网络联系起来，并检验二者之间的相关性；（3）证明我们的度量比传统方法更准确地描绘了发明者的搜索过程。

*A1. 14*个嵌入空间及其跨时期稳定性

在本研究中，我们以三年为间隔（首个时期为1976—1979年，最后一个时期为2015—2017年），构建了涵盖1976年至2017年的14个双曲嵌入空间。每个空间都刻画了对应时期技术树的结构。图A1展示了这14个庞加莱圆盘。我们采用普罗克拉斯提斯过程（Hamilton、Leskovec 和 Jurafsky，2016）对它们进行对齐，以提升可读性。庞加莱圆盘直观地呈现了不同时期之间的连续性，凸显了技术演化的稳定性。

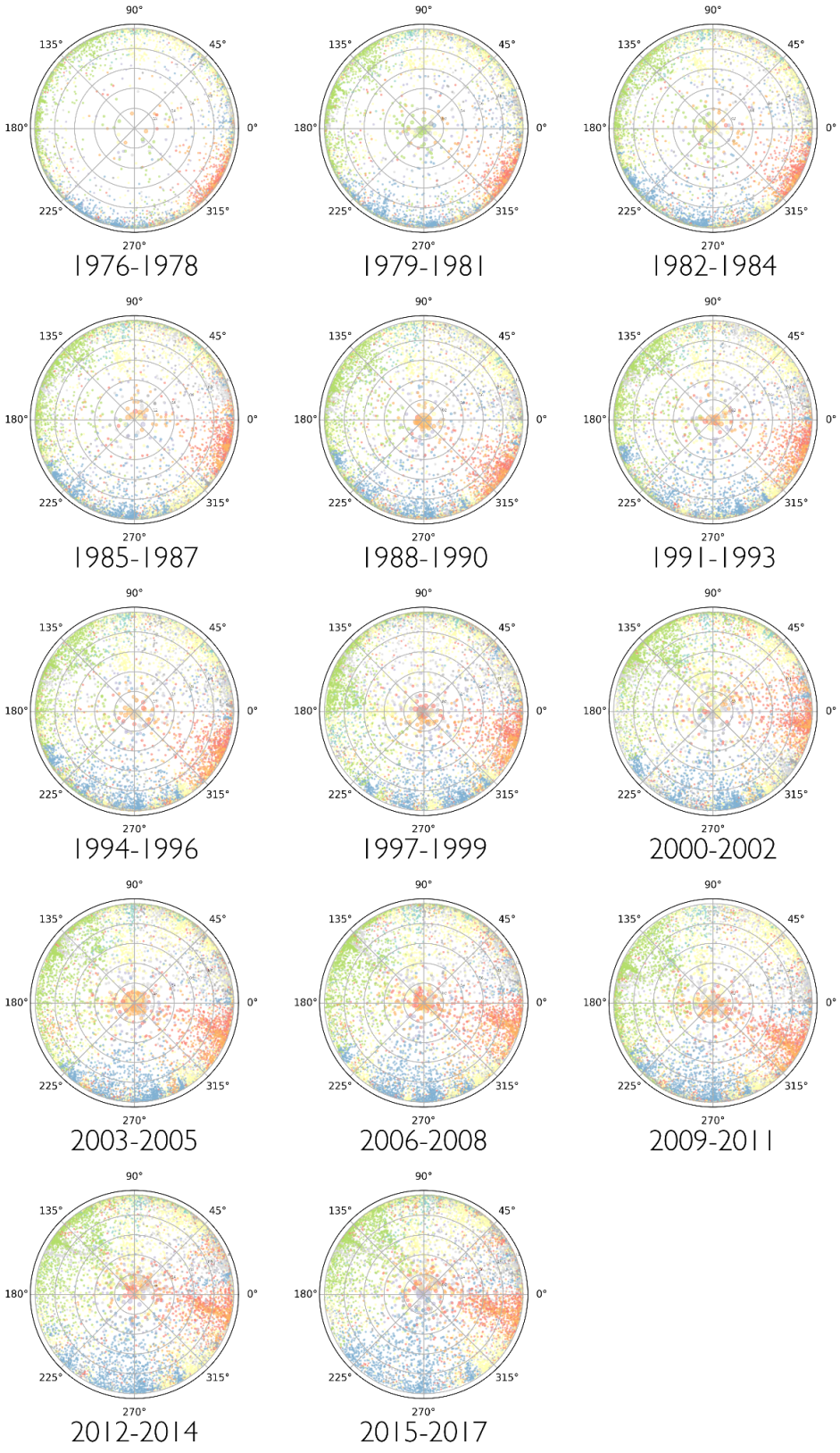


图 A1. 各时期的 14 个嵌入空间

我们假设，随着时间推移，稳定的技术轨迹会使各时期的径向（r）和角度（θ）值保持一致，表明其一般性和水平关系处于相似层级。此外，我们预期相邻时期之间的相关系数更高。图 A2 中的热力图展示了这些相关性。结果表明，嵌入空间具有稳定性，r 值的相关系数均超过 0.6。删除低频类别后，θ 值的相关系数大多超过 0.5。值得注意的是，相邻时期之间的相关系数有所提高，进一步支持了我们的初始假设。

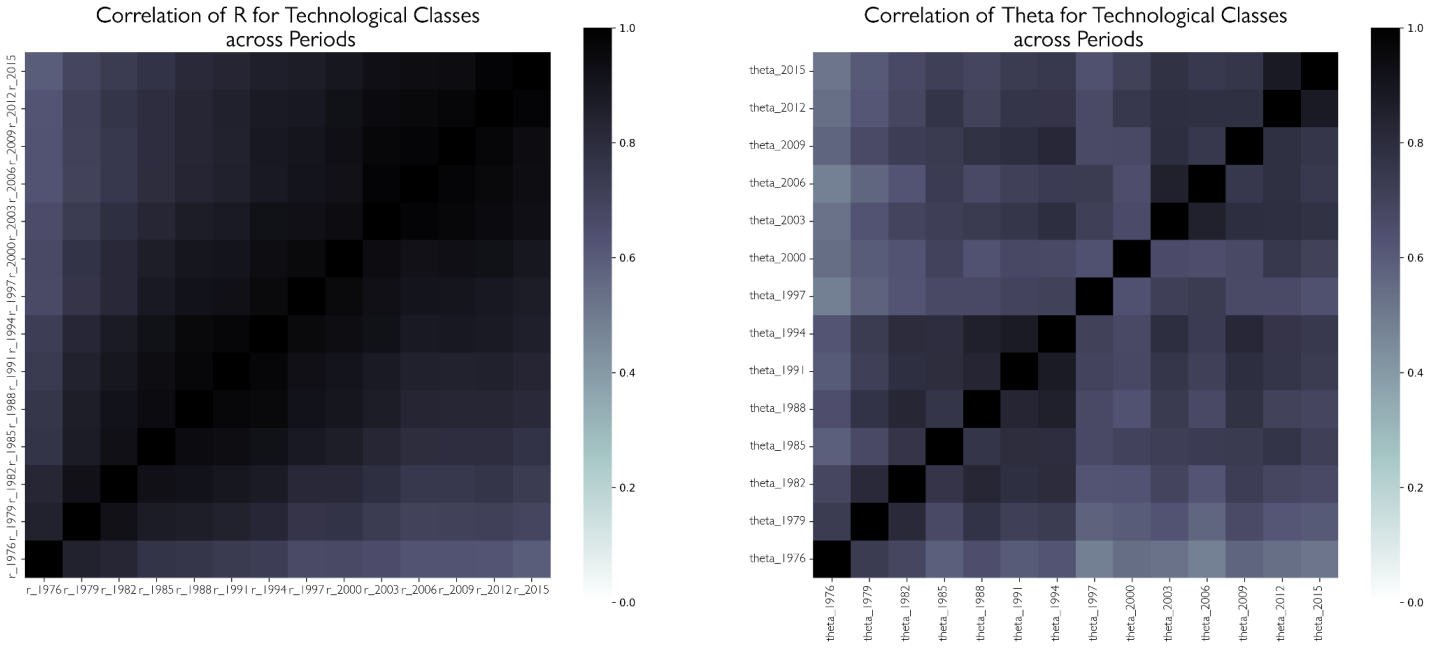


图 A2. 技术类别径向 (r) 和角向 (θ) 值的跨时期相关性

针对技术类别

A2. 基于庞加莱的度量与网络度量之间的相关性

我们使用嵌入方法，简化表示引文网络中的信息流。将网络嵌入庞加莱圆盘后，我们假设中心性较高的节点会更靠近圆盘中心，在极坐标中表现为更小的径向距离*r*。先前的研究表明，这些位于中心的组件在推动技术性能提升率方面发挥着重要作用（Triulzi 等人，2020）。测地路径较长的节点预计彼此距离更远，表现为更大的角距离。

在表 A1 中，我们考察网络中心性度量与径向距离（r）之间的相关性。出度中心性与径向距离之间的相关系数大多大于 0.6，而特征向量中心性与径向距离之间的相关系数大多超过 0.8。尽管接近中心性与径向距离之间的相关系数略低，但仍达到 0.4 的基准值。

表 A1. 中心性度量与 *r* 在庞加莱圆盘上的相关性

	出度 中心性与 r 的关系	特征向量 中心性与 r 的关系	接近中心性 与 r 对比
1976—1978	0.597 [0.000]	0.590 [0.000]	0.355 [0.000]
1979—1981 年	0.646 [0.000]	0.709 [0.000]	0.402 [0.000]
*1982—1984*年	0.677 [0.000]	0.721 [0.000]	0.433 [0.000]
*1985—1987*年	0.732 [0.000]	0.807 [0.000]	0.468 [0.000]
*1988—1990*年	0.761 [0.000]	0.827 [0.000]	0.475 [0.000]
*1991—1993*年	0.773 [0.000]	0.817 [0.000]	0.497 [0.000]
*1994—1996*年	0.802 [0.000]	0.841 [0.000]	0.482 [0.000]

*1997—1999*年	0.816 [0.000]	0.841 [0.000]	0.517 [0.000]
*2000—2002*年	0.834 [0.000]	0.859 [0.000]	0.508 [0.000]
*2003—2005*年	0.849 [0.000]	0.875 [0.000]	0.533 [0.000]
*2006—2008*年	0.864 [0.000]	0.871 [0.000]	0.542 [0.000]
*2009—2011*年	0.871 [0.000]	0.865 [0.000]	0.524 [0.000]
*2012—2014*年	0.883 [0.000]	0.857 [0.000]	0.480 [0.000]
*2015—2017*年	0.886 [0.000]	0.859 [0.000]	0.473 [0.000]

注：相关系数基于皮尔逊相关计算。*P*值列于方括号内，所有检验均为双尾检验。

在表A2中，我们考察相应节点对的测地路径长度与角距离（$|\Delta\theta|$）之间的相关性。几乎每个技术类别都与其他类别存在一定程度的互连。虽然有些节点对之间仅由一条或两条连接相连，但这些偶然的连接可能会改变测地距离的整体分布。为解决这一潜在问题，我们采用了各种连接权重阈值。这种方法有助于保留引文网络的核心结构。随后，我们将这些网络中的测地路径与庞加莱圆盘得出的角距离进行相关性分析。

表A2. 庞加莱圆盘上的测地距离与 $\Delta\theta$ 之间的相关性

	连接频率阈值			
	前10%	前5%	前0.5%	前0.1%
1976—1978	0.176 [0.000]	0.204 [0.000]	0.442 [0.000]	0.668 [0.000]
*1979—1981*年	0.172 [0.000]	0.151 [0.000]	0.212 [0.000]	0.737 [0.000]
*1982—1984*年	0.150 [0.000]	0.206 [0.000]	0.227 [0.000]	0.505 [0.000]
*1985—1987*年	0.153 [0.000]	0.173 [0.000]	0.229 [0.000]	0.390 [0.000]

*1988—1990*年	0.133 [0.000]	0.136 [0.000]	0.255 [0.000]	0.388 [0.000]
*1991—1993*年	0.142 [0.000]	0.161 [0.000]	0.233 [0.000]	0.254 [0.000]
*1994—1996*年	0.154 [0.000]	0.159 [0.000]	0.240 [0.000]	0.247 [0.000]
*1997—1999*年	0.147 [0.000]	0.151 [0.000]	0.268 [0.000]	0.261 [0.000]
*2000—2002*年	0.147 [0.000]	0.151 [0.000]	0.250 [0.000]	0.364 [0.000]
*2003—2005*年	0.138 [0.000]	0.145 [0.000]	0.197 [0.000]	0.264 [0.000]
*2006—2008*年	0.129 [0.000]	0.150 [0.000]	0.207 [0.000]	0.252 [0.000]
*2009—2011*年	0.122 [0.000]	0.136 [0.000]	0.133 [0.000]	0.240 [0.000]
*2012—2014*年	0.136 [0.000]	0.182 [0.000]	0.215 [0.000]	0.324 [0.000]
*2015—2017*年	0.141 [0.000]	0.166 [0.000]	0.262 [0.000]	0.297 [0.000]

注：相关系数基于皮尔逊相关。*P*值以方括号表示，所有检验均为双尾检验。

表A2显示，在大多数时期，测地距离与角距离之间存在正相关。随着连接权重阈值提高，引文网络中界定得更加精确的“核心”逐渐凸显。这个核心保留了最重要的距离信息，而我们的嵌入空间应能有效捕捉这些信息。在构建最严格的“核心”时，仅纳入排名前0.1%的连接，此时测地距离与角距离之间的相关性可超过0.5。

A3. 使用*Word2Vec*的语义嵌入验证基于庞加莱的度量

我们的度量并未直接捕捉特定的语义距离，而且可能受到技术领域跨学科性质的混淆。为解决这一潜在混淆并验证庞加莱距离度量，我们进行了三项额外的稳健性检验。

我们首先使用语义嵌入（Word2Vec）计算每项专利和技术类别（所有相关专利的平均值）在欧几里得空间中的位置，然后

将结果与庞加莱嵌入的结果进行比较（2015—2017年切片）。在类别对层级——仅限于包含至少2,000项专利的类别——我们观察到，Word2Vec与庞加莱嵌入所得距离之间的斯皮尔曼相关系数约为0.57。鉴于Word2Vec并未考虑时间动态，这一相关性之强令人瞩目。这一结果表明，尽管我们的嵌入不包含显式的语义信息，但它通过利用源自引文网络的知识流，有效捕捉了技术距离。基于这一证据，我们认为$\Delta\theta$ 可作为语义距离的合理代理变量。

其次，为应对跨学科性可能产生的混淆效应，我们基于 Word2Vec 模型，比较了每个 CPC类别的类别间距离与类别内距离的分布。由于我们的方法不包含语义信息，因此主要捕捉类别间距离，而忽略了类别内变化。该比较结果见下方图 A3。如图所示，类别间距离显著大于类别内距离，t统计量为 88.946（$p < 0.001$），表明存在统计显著差异。此外，每个类别内的专利数量与类别内距离之间的相关性较低（约为 0.11），这表明规模较大的领域并不一定具有更广的认知跨度。因此，尽管我们的度量可能无法捕捉技术领域内部的跨学科性所引入的认知跨度的细粒度变化，但我们认为，这一局限并不会从根本上影响总体距离尺度，因而也不会改变我们的主要结论。

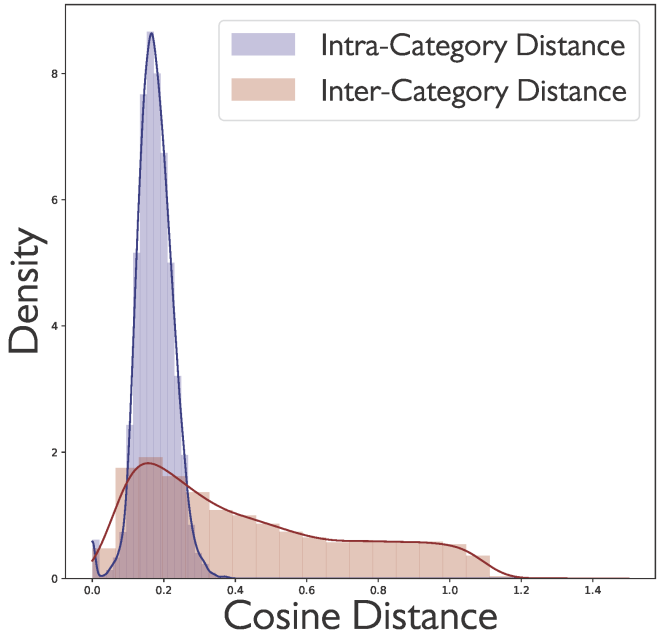


图 A3. 类别内与类别间距离分布的比较

附录B. 主要模型和机制分析的补充模型

B1. 基于不同样本纳入标准的补充模型

正文中，我们选取了两组样本用于构建模型：模型1使用所有2017年之前授权的专利，模型2-5使用所有2002年之前授权的专利。较早的专利为构建引文变量提供了更长的时间跨度，使我们能够估计不同时期深度与广度的效应。为检验模型的稳健性，我们还使用了另外两个子样本：2012年之前授权的专利，可计算0-5年和5-10年引用次数；以及2007年之前的专利，可计算0-5年、5-10年和10-15年引用次数。随后，我们在这两个子样本上重新运行了模型。结果见表B1。

表B1. 基于1976-2007年和1976-2012年子样本的重组深度与广度对专利短期和长期影响的效应

	专利模型 1976年-2007年		**专利模型 1976年-2012年**		
	模型1：预测0-5年份 引文	**模型2 预测5-10年引文**	**模型3 预测0-5年引文**	**模型 4 预测5-10年引文**	**模型5 预测10-15年引文**
主要模型					
深度	0.146	0.041	0.121	0.034	-0.104
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.005)	(0.006)	(0.006)	(0.006)	(0.008)
广度	-0.055	-0.036	-0.054	-0.041	-0.018
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
水平位置 (平均 theta)	0.031	0.025	0.043	0.035	0.029
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
自然对数（组件熟悉度）	0.031	0.037	0.031	0.039	0.037
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
自然对数（组合	0.030	0.022	0.033	0.024	0.019

熟悉度)	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
技术类别数量	0.027	0.036	0.029	0.037	0.043
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
现有技术类别数量	0.003	0.004	0.003	0.004	0.005
类别	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
分配给组织	0.088	0.049	0.104	0.057	-0.025
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
团队发明	0.078	0.094	0.079	0.092	0.085
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
Ln（发明人经验）	0.006	-0.019	0.005	-0.019	-0.039
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
非专利文献	0.009	0.010	0.008	0.013	0.013
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
权利要求数量	0.018	0.021	0.018	0.021	0.023
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
家庭规模	-0.139	-0.253	-0.118	-0.226	-0.286
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
年份固定效应	YES	YES	YES	YES	YES
常数	0.098	0.036	0.090	0.008	0.178**
	[0.069]	[0.532]	[0.095]	[0.889]	[0.003]
	(0.054)	(0.057)	(0.054)	(0.057)	(0.061)
色散模型					
深度	-0.011	-0.074	0.085	0.002	-0.101
	[0.145]	[0.000]	[0.000]	[0.840]	[0.000]
	(0.008)	(0.007)	(0.010)	(0.008)	(0.008)
广度	-0.023	-0.034	-0.032	-0.031	-0.039
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
水平位置 (平均 theta)	0.013	0.002	0.010	0.006	-0.002
	[0.000]	[0.002]	[0.000]	[0.000]	[0.001]
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
自然对数（组件熟悉度）	0.031	0.042	0.022	0.033	0.049
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
自然对数（组合熟悉度）	-0.016	-0.004	-0.021	-0.009	0.003
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
技术类别数量	0.002	0.002	-0.000	0.000	-0.001
	[0.000]	[0.000]	[0.228]	[0.480]	[0.002]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
现有技术	-0.001	-0.000	-0.001	-0.001	-0.000
类别	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
分配给组织	0.067	0.120	0.065	0.114	0.147
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.004)	(0.003)	(0.004)	(0.003)	(0.003)
团队发明	0.031	0.011	0.032	0.021	0.003
	[0.000]	[0.000]	[0.000]	[0.000]	[0.104]
	(0.002)	(0.002)	(0.003)	(0.002)	(0.002)
Ln（发明人经验）	0.011	0.033	0.012	0.029	0.039
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
非专利文献	0.012	0.005	0.018	0.009	0.004
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
权利要求数量	-0.007	-0.007	-0.007	-0.006	-0.007
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
家庭规模	-0.025	0.024	-0.030	-0.000	0.049
	[0.000]	[0.000]	[0.000]	[0.991]	[0.000]
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
年份固定效应	YES	YES	YES	YES	YES
常数	-0.039	-0.021	-0.045	-0.012	0.075
	[0.673]	[0.819]	[0.626]	[0.893]	[0.417]
	(0.092)	(0.092)	(0.092)	(0.092)	(0.093)
观测值	3711809	3711809	2774652	2774652	2774652

注：*P*值置于方括号内。稳健标准误置于圆括号内。所有检验均为双尾检验。

表B1呈现出与正文表3类似的模式。短期内，深度的影响为正，而广度的影响为负；长期内，这些影响则相反。

B2. 同一组织内部的深度与广度效应

我们仅使用组织内引用复现了表3的结果。这些结果见表B2。

表B2. 重组深度与广度对专利短期和长期影响的效应（组织内部）

	采用所有专利的模型：1976—2017年	采用早期专利的模型：1976—2002年			
	模型1：预测0-5年 引文	**模型2 预测0-5年引文**	**模型3 预测5-10年引文**	**模型 4 预测10-15年引文**	**模型5 预测15-20年引文**
主要模型					
深度	0.057	0.031	0.080	0.060	-0.167
	[0.000]	[0.012]	[0.000]	[0.014]	[0.000]
	(0.008)	(0.012)	(0.017)	(0.024)	(0.030)
广度	-0.006	-0.016	0.014	0.050	0.076
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.001)	(0.002)	(0.003)	（0.003）
水平位置（平均 theta）	0.012	0.017	0.028	0.024	0.034
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
自然对数（组件熟悉度）	0.021	0.012	0.031	0.046	0.042
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
自然对数（组合熟悉度）	0.021	0.018	0.022	0.025	0.024
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
技术类别数量	0.016	0.017	0.022	0.023	0.022
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
现有技术类别数量类别	0.003	0.002	0.002	0.002	0.002
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
分配给组织	0.000	0.000	0.000	0.000	0.000

	[.] (.)	[.] (.)	[.] (.)	[.] (.)	[.] (.)
团队发明	0.070	0.088	0.089	0.086	0.149
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.002)	(0.003)	(0.004)	(0.006)	(0.008)
Ln（发明人经验）	0.077	0.073	0.075	0.068	0.056
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
非专利文献	0.018	0.011	0.016	0.014	0.007
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)
权利要求数量	0.011	0.013	0.017	0.015	0.011
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
家庭规模	-0.100	-0.063	-0.190	-0.264	-0.332
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.002)	(0.002)	(0.003)	(0.004)	(0.005)
年份固定效应	YES	YES	YES	YES	YES
常数	-0.138	-0.065	-0.152	0.292	0.238
	[0.096]	[0.442]	[0.232]	[0.189]	[0.217]
	(0.083)	(0.084)	(0.127)	(0.222)	(0.193)
色散模型					
深度	0.029	0.093	0.068	0.071	-0.070
	[0.042]	[0.002]	[0.018]	[0.036]	[0.084]
	(0.014)	(0.030)	(0.029)	(0.034)	(0.040)
广度	0.030	0.026	0.071	0.089	0.093
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.002)	(0.003)	(0.003)	(0.004)	(0.005)
水平位置（平均 theta）	0.026	0.016	0.031	0.026	0.048
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.002)	(0.002)	(0.003)	(0.004)
自然对数（组件熟悉度）	0.048	0.015	0.016	0.038	0.080
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
自然对数（组合熟悉度）	0.017	0.009	0.008	0.015	0.022
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
技术类别数量	0.005	0.016	0.017	0.016	0.018
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)

现有技术	0.002	0.001	-0.000	0.000	0.000
类别	[0.000]	[0.000]	[0.575]	[0.892]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
分配给组织	0.000	0.000	0.000	0.000	0.000
	[.]	[.]	[.]	[.]	[.]
	(.)	(.)	(.)	(.)	(.)
团队发明	0.060	0.043	0.036	0.033	0.184
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.004)	(0.008)	(0.008)	(0.009)	(0.012)
Ln（发明人经验）	0.042	0.060	0.055	0.042	0.054
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.002)	(0.002)	(0.003)	(0.003)
非专利文献	-0.001	0.010	0.000	-0.006	-0.001
	[0.000]	[0.000]	[0.530]	[0.000]	[0.223]
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
权利要求数量	0.008	0.011	0.008	0.006	0.007
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
家庭规模	-0.294	-0.115	-0.154	-0.169	-0.300
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.004)	(0.005)	(0.005)	(0.007)	(0.009)
年份固定效应	YES	YES	YES	YES	YES
常数	-1.297	-1.185	-0.529	0.023	-2.621
	[0.000]	[0.000]	[0.023]	[0.930]	[0.027]
	(0.259)	(0.259)	(0.232)	(0.261)	(1.188)
观测值	1440070	613180	369316	204775	101397

注：*P*值置于方括号内。稳健标准误置于圆括号内。所有检验均为双尾检验。

表B2呈现的模式与正文表3中的模式相似。短期内，深度具有正向影响，而广度具有负向影响；长期来看，这些影响则相反。表中还显示，这种反转在组织内部发生得更快。在组织外部，广度的正向影响要到授权日期15年后才会出现；而在组织内部，焦点专利授权后最早5年就可能发生这种转变。这可能是因为组织边界内的信息扩散遇到的障碍较少，从而更快实现集体发明。

B3. 使用泊松拟最大似然估计量的稳健性检验

我们使用泊松拟最大似然估计量复现表3。结果见表B3。

表B3. 重组深度与广度对专利短期和长期影响的作用（基于泊松准极大似然估计）

	包含所有专利的模型：1976—2017年	包含早期专利的模型：1976—2002年			
	模型1：预测0-5年份引文	**模型2 预测0-5年引文**	**模型3 预测5-10年引文**	**模型 4 预测10-15年引文**	**模型5 预测15-20年引文**
主要模型					
深度	0.278	0.122	0.117	-0.053	-0.395
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.008)	(0.007)	(0.009)	(0.011)	(0.015)
广度	-0.007	-0.040	-0.034	-0.014	0.054
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)
水平位置 (平均 theta)	0.020	0.065	0.059	0.038	0.050
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
自然对数（组件熟悉度）	0.033	0.026	0.033	0.035	0.023
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.001)	(0.001)	(0.001)	（0.001）
自然对数（组合熟悉度）	0.032	0.035	0.026	0.025	0.026
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.004)	(0.004)	(0.004)	(0.001)	(0.001)
技术类别数量	0.027	0.031	0.038	0.042	0.045
	[0.003]	[0.003]	[0.003]	[0.004]	[0.001]
	(0.000)	(0.000)	(0.000)	(0.001)	(0.001)
现有技术的类别数量	0.003	0.003	0.003	0.004	0.004
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
分配给组织	0.080	0.130	0.088	-0.017	-0.120
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]

	(0.004)	(0.003)	(0.004)	(0.005)	(0.006)
团队发明	0.074	0.085	0.094	0.094	0.107
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)
Ln（发明人经验）	0.014	-0.007	-0.031	-0.050	-0.052
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
非专利文献	0.013	0.005	0.012	0.013	0.012
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
权利要求数量	0.018	0.018	0.022	0.025	0.025
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
家庭规模	-0.220	-0.124	-0.248	-0.336	-0.384
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.002)	(0.002)	(0.002)	(0.003)	(0.004)
年份固定效应	YES	YES	YES	YES	YES
常数	-0.034	0.075	-0.070	0.104	0.416
	[0.775]	[0.274]	[0.511]	[0.476]	[0.042]
	(0.118)	(0.068)	(0.106)	(0.146)	(0.204)
观测值	4992915	2010031	2010031	2010031	2010031

注：*P*值位于方括号内。稳健标准误位于圆括号内。所有检验均为双尾检验。

B4. R 使用深度和广度度量的替代阈值进行稳健性检验

在主要模型中，我们使用 r 和 \theta 的第90百分位数来构建深度和广度度量。这个特定阈值可能影响了模型结果，进而影响我们的主要结论。为评估研究结果的稳健性，我们以第95百分位数和最大值作为阈值，构建了替代性的深度和广度度量，并采用这些新设定重新估计主要模型。表B4所报告的结果与主要模型的结果一致。

表B4 . 使用替代阈值的深度和广度指标主要模型 s

	所有专利的模型: 1976—2017年		包含早期专利的模型: 1976—2002年		
	模型1: 预测0—5年引用	**模型2 预测0-5年引文**	**模型3 预测5-10年引文**	**模型 4 预测 10-15年份 引文**	**模型5 预测 15-20年份 引文**
使用深度和广度指标的模型（第 ***95*** 百分位数）					
深度	0.197***	0.084***	-0.001	-0.160***	-0.412***
	[0.000]	[0.000]	[0.914]	[0.000]	[0.000]
	(0.005)	(0.007)	(0.007)	(0.009)	(0.010)
广度	-0.045***	-0.037***	-0.019***	0.000	0.040***
	[0.000]	[0.000]	[0.000]	[0.780]	[0.000]
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
控制变量（与正文中的表3相同）	YES	YES	YES	YES	YES
使用第***100***百分位数深度和广度指标的模型					
深度	0.182***	0.062***	-0.006	-0.150***	-0.372***
	[0.000]	[0.000]	[0.378]	[0.000]	[0.000]
	(0.005)	(0.007)	(0.007)	(0.009)	(0.010)
广度	-0.049***	-0.039***	-0.022***	-0.004***	0.037***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
控制变量（与正文中的表3相同）	YES	YES	YES	YES	YES
观测值	4992915	2010031	2010031	2010031	2010031

注：*p*值 ar 位于方括号内。稳健标准误位于圆括号内。所有检验均为双尾检验。

B5. 深度与广度效应的因果推断

在本节中，我们进行了两项额外分析，以解决内生性问题。首先，我们采用了美欧双胞胎专利设计；其次，我们使用工具变量法，以加强对结果的因果解释。

B5.1 美欧双胞胎专利检验

在这项检验中，我们识别出美国与欧洲专利对，以控制专利质量和内容。按照先前研究的做法，我们仅保留同时在美国专利商标局和欧洲专利局提交的有效双胞胎专利，并排除标题不同或存在一对多／多对一匹配的专利。最终样本由84,916对专利组成。

我们使用欧洲专利局的专利，而非在其他国家专利局提交的专利，作为匹配样本，主要有两个原因。首先是语言问题。例如，许多在德国专利商标局提交的专利仅以德语撰写，只有少量内容提供英文版本。这降低了不同样本之间的可比性，也给匹配和文本分析带来了挑战。其次是样本量问题。与匹配后的美国专利商标局—德国专利商标局数据集相比，匹配后的美国专利商标局—欧洲专利局数据集所提供的样本量要大得多。

对于美国专利商标局授予的专利，引文既可能由申请人添加，也可能由审查员添加；而欧洲专利局授予的专利仅包含审查员添加的引文。为确保成对样本具有可比性，我们对美国专利商标局和欧洲专利局的样本均只统计审查员添加的引文。与首次分析不同，在此我们根据美国专利分类系统计算美国专利的深度与广度，并根据国际专利分类体系计算欧洲专利的深度与广度，以有意保持两组配对样本采用不同的操作化方式。

欧洲类别被嵌入各个3年时间窗口内的独立庞加莱圆盘中。我们将样本限定为2001年至2012年间授权的专利，因为2001年是美国专利商标局首次提供审查员添加的引文数据的年份，而2013年则是美国专利商标局将标准化分类体系从美国专利分类改为CPC的年份，这会改变比较基准。

一项发明在美国专利商标局的引文数量可表示为：而其在欧洲专利局的专利孪生体的影响可表示为：

$$log(Citations_{i,US}) = \beta_1 \times Depth_{i,US} + \beta_2 \times Breadth_{i,US} + \beta_3 \times Controls_{i,US} + Quality_i + \epsilon_{i,US}$$

$$log(Citations_{i,EU}) = \gamma_1 \times Depth_{i,EU} + \gamma_2 \times Breadth_{i,EU} + \gamma_3 \times Controls_{i,EU} + Quality_i + \epsilon_{i,EU}$$

我们将第二个方程代入第一个方程。得到的方程形式如下：

$$log(Citations_{i,US}) = \alpha_1 \times Depth_{i,US} + \alpha_2 \times Breadth_{i,US} + \alpha_3 \times Controls_{i,US}$$
$$+ \alpha_4 \times log(Citations_{i,EU}) - \alpha_5 \times Depth_{i,EU} - \alpha_6 \times Breadth_{i,EU} - \alpha_7 \times Controls_{i,EU}$$
$$+ \epsilon_{i,US} - \epsilon_{i,EU}$$

我们在成对样本上估计该模型，以便在控制专利质量和其他个体层面异质性的同时，检验深度和广度效应的稳健性。通过将欧洲引文方程代入原始模型，我们严格控制了专利质量，从而估计独立于发明过程的深度和广度效应。

我们在表B5中报告了该模型的结果。如果我们的论点具有稳健性，我们便采用Kovács及其同事（2021年）设定的标准，并预期：(1) 深度和广度效应与表3中先前模型的结果相同，保持相同的正负号；(2) 美国和欧洲引文计数在每个目标时期均呈正相关；(3) 在美国

引文模型中，欧洲自变量（例如，同一项专利的欧洲深度和广度）的符号将与美国变量相反，这表明它们体现的是相同的效应，只是因我们的模型设定而方向相反（见第4.2节）。表B5所报告的结果符合这些预期，并表明在严格控制底层发明的质量后，正文表3中的模型依然稳健。

表B5. 专利孪生测试：控制欧洲专利局度量后的美国专利商标局0–5年期前向引用

	模型1	模型2	模型3	模型 4
	模型：泊松准极大似然估计		模型：负二项	
美国专利商标局深度	0.750***	0.624***	0.548***	0.571***
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.038)	(0.040)	(0.031)	(0.036)
美国专利商标局广度	-0.053***	-0.050***	-0.020***	-0.043***
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.007)	(0.007)	(0.005)	(0.006)
欧洲五年期前向引文		0.079***		0.103***
		[0.000]		[0.000]
		(0.003)		(0.003)
欧洲深度		-0.075**		-0.063*
		[0.006]		[0.010]
		(0.027)		(0.025)
欧洲广度		0.012		0.008
		[0.065]		[0.166]
		(0.006)		(0.006)
美国/欧盟专利控制变量	YES	YES	YES	YES
美国/欧盟年份固定效应	YES	YES	YES	YES
负二项离散度模型			YES	YES
N	84,956	84,956	84,956	84,956
对数似然	-137492.823	-131271.191	-118030.06	-116809.3

注：*P*值置于方括号内。稳健标准误置于圆括号内。所有检验均为双尾检验。

B5.2 工具变量法

接下来，我们采用工具变量检验，进一步排除混杂效应并处理潜在的内生性问题。为此，我们选取一个工具变量：焦点专利申请之前现有技术类别编码的重新分类。

当专利管理部门调整 CPC 分类系统时，就会发生重新分类。具体而言，当焦点专利引用其他专利，且这些被引专利的 CPC 类别发生更新时，现有技术类别就会重新分类。该变量满足有效工具变量的两个关键标准：首先，它与自变量（深度与广度）相关，因为它会影响焦点专利发明人收集信息和探索技术空间的方式。其次，它不会直接影响结果变量（焦点专利的引文），因为它主要反映的是一种外部的系统层面冲击。到焦点专利申请时，其后向引用和现有技术类别的重新分类都已发生，因此这些调整不太可能影响其前向引用（即焦点专利获得的引用）。

对于每项焦点专利，我们度量其现有技术类别中受美国专利商标局调整的类别所占比例。由于 CPC 分类体系变更的数据（收集自
h
t
t
p
s://
www.uspto.gov/web/patents/classification/cpc/html/cpc-notices-of-changes.html）仅从2013年开始，我们使用2013年至2017年间授权的专利来估计工具变量模型。由于这些样本无法提供完整的10 年窗口期来观察前向引用（或更长时期），我们将结果变量限定为0至5年的前向引用。尽管如此，这项分析有助于增强我们的因果推断，并减轻潜在混杂因素的影响。

基于上述工具变量，我们采用带有稳健标准误的两阶段最小二乘法（2SLS）估计模型。由于我们只有一个工具变量，因此分别使用该工具变量检验深度和广度。对于深度模型，Kleibergen-Paap rk Wald F 统计量为 5703.794（$p < 0.001$），Cragg-Donald Wald F 统计量为 6059.688（$p <0.001$）。对于广度模型，Kleibergen-Paap rk Wald F 统计量为 296.505（$p < 0.001$），Cragg-Donald Wald F 统计量为 298.091（$p < 0.001$）。所有 F 统计量都远高于

两种情况下最高的 Stock-Yogo 阈值，这表明我们的工具变量满足相关性条件。

原始模型和两阶段最小二乘估计的结果均报告于表 B6。结果显示，工具变量估计结果与原始模型结果总体一致，表明在考虑潜在内生性后，搜索深度和广度对前向引用的影响依然稳健，并证实我们关于搜索策略对专利影响力的影响的主要结论并非由遗漏变量偏误或反向因果关系所致。

表 B6. 深度和广度效应的两阶段最小二乘分析（数据：2013年至2017年的专利，因变量：未来五年引用次数）

	原始模型	**原始模型**	**两阶段最小二乘：深度**		**两阶段最小二乘：广度**	
	负面 负二项	OLS	**第一阶段： 因变量= 深度**	**第二阶段： 因变量=引文**	**第一阶段： 因变量=广度**	**第二阶段： 因变量=引文**
工具变量： 对先前分类的重新界定 艺术类别			-0.018***		0.027***	
			[0.000]		[0.000]	
			(0.000)		(0.002)	
深度 / **预测深度**	0.546***	5.120***		**17.282***	2.408***	24.395***
	[0.000]	[0.000]		**[0.000]**	[0.000]	[0.000]
	(0.016)	(0.149)		**(1.662)**	(0.006)	(2.808)
广度 / **预测广度**	-0.021***	-0.078***	0.051***	-0.704***		**-8.104***
	[0.000]	[0.000]	[0.000]	[0.000]		**[0.000]**
	(0.002)	(0.022)	(0.000)	(0.085)		**(1.152)**
水平位置（均值 平均 theta）	0.022***	0.056***	-0.001***	0.064***	-0.022***	-0.124***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.002)	(0.013)	(0.000)	(0.006)	(0.000)	(0.028)
自然对数（组件熟悉度）	0.029***	0.066***	-0.003***	0.106***	-0.100***	-0.736***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.008)	(0.000)	(0.007)	(0.000)	(0.115)
自然对数（组合 熟悉度）	0.020***	0.117***	-0.003***	0.150***	-0.017***	-0.019
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.336]
	(0.001)	(0.005)	(0.000)	(0.008)	(0.000)	(0.020)
技术类别数量	0.029***	0.123***	-0.001***	0.135***	-0.001***	0.111***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.004)	(0.000)	(0.005)	(0.000)	(0.005)
现有技术类别数量	0.004***	0.016***	0.000***	0.011***	0.001***	0.027***

	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.002)
分配给组织	-0.079***	-0.155*	-0.007*	-0.071	-0.111*	-1.048***
	[0.000]	[0.016]	[0.000]	[0.061]	[0.000]	[0.000]
	(0.007)	(0.065)	(0.000)	(0.038)	(0.003)	(0.136)
团队发明	0.060***	0.117***	-0.002***	0.140***	-0.007***	0.057
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.068]
	(0.004)	(0.033)	(0.000)	(0.028)	(0.001)	(0.031)
Ln（发明人经验）	0.019***	0.121***	-0.000***	0.125***	-0.003***	0.090***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.008)	(0.000)	(0.009)	(0.000)	(0.009)
非专利文献	0.011***	0.094***	0.000***	0.091***	-0.004***	0.063***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.002)	(0.000)	(0.003)	(0.000)	(0.005)
权利要求数量	0.022***	0.051***	-0.000**	0.052***	-0.001**	0.046***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.002)	(0.005)	(0.002)	(0.000)	(0.002)
家庭规模	-0.423***	-1.070***	-0.005***	-1.012***	-0.003*	-1.091***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.012]	[0.000]
	(0.003)	(0.027)	(0.000)	(0.018)	(0.001)	(0.019)
年份固定效应	YES	YES	YES	YES	YES	YES
常数	-0.282***	-1.719***	0.325***	-5.639***	3.119***	23.371***
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.019)	(0.161)	(0.001)	(0.543)	(0.007)	(3.594)
观测值	1281106	1281106	1281106	1281106	1281106	1281106

注：*p*值　　位于方括号中。稳健标准误位于圆括号中。所有检验均为双尾检验　　.

附录 C. 比较庞加莱嵌入与CPC分类系统在捕捉技术类别间距离方面的表现

在技术创新领域的文献中，知识及其功能之间的区别早已得到充分确立（Benson and Magee，2015；Magee 等人，2016）。本文认为，传统分类系统无法充分捕捉技术类别之间的实际关联，因为它以知识相似性而非功能关系为基础。为提供支持这一观点的经验证据，我们实施了一项统计检验。

对于嵌入的14个时期，我们首先识别专利中出现过的所有技术类别。假设符合条件的类别数量为N，则可得到N(N-1)/2个类别对。对于每一对，我们通过点互信息计算其共同出现情况：

$$\text{PMI(i,j)}=log2(\frac{P(i,j)}{P(i)*P(j)})$$

点互信息（PMI）得分越高，表示一对类别越倾向于在专利中共同使用，因而说明它们之间的距离越小。本研究将PMI作为基准，与我们的度量进行比较。能够准确概括这一距离的度量，更有助于界定搜索范围。

在这方面，我们评估了两种候选方案：我们提出的双曲嵌入和传统的CPC分类系统。

- 对于双曲嵌入，我们将两点之间的距离计算如下：

$$d(u,v) = arcosh(1 + 2\frac{||u-v||^2}{(1-||u||^2)(1-||v||^2)})$$

按照（Nickle 和 Kiela，2017年）的定义。我们将两点之间的相似性计算为：

$similarity(u, v) = \frac{1}{1-.(*,,)}$

遵循 Gensim 软件包中关于庞加莱嵌入的源代码。

❖ 对于传统的层级式 CPC 分类，我们采用 Trajtenberg 等（1997）的方法。在合作专利分类（“CPC”）系统中，每项专利都会被分配一个或多个四级分类。例如，分类 A01B33 在最高层级属于“A”，在第二层级属于“A01”，在第三层级属于“A01B”，在最低层级属于“A01B33”。对于每一对技术类别，如果它们属于同一个三位数类别（例如，“A01B33”和“A01B35”），则将技术距离 $d_{i,\ j}$设为 0.25；如果它们属于同一个两位数类别（例如，“A01B33”和“A01F12”），则设为 0.5；如果它们属于不同的一位数类别或没有现有技术（例如，“A01B33”和“B01B1”），则设为 1。

技术相似性简单地设为 1-$d_{i,\ j}$。

在确立经验基准以及成对技术相关性的两种度量后，我们进一步计算这三种度量之间的相关性。为此，鉴于基于 CPC 层级结构的距离本质上是序数型的，我们采用斯皮尔曼秩相关系数。该分析结果见表 C1。

研究结果表明，在所有时期和所有纳入阈值下，基于我们的庞加莱嵌入算法的相似性度量，在捕捉实际技术相关性方面均优于传统的 CPC 分类法。值得注意的是，阈值的提高会导致性能差异进一步扩大。这凸显了我们的方法在准确表示知识流动“骨架”方面的优势：当数据输入中包含更多信息时，该方法能够实现更好的拟合。

表 C1. 斯皮尔曼相关系数 基准共同出现与两种相似性度量之间的

频率阈值 / **技术类别**	**0**		**100**		**200**		**500**		**1000**	
度量	CPC Tree	庞加莱	CPC Tree	庞加莱	CPC Tree	庞加莱	CPC Tree	庞加莱	CPC Tree	庞加莱
1976—1978	0.144	0.224	0.183	0.293	0.233	0.395	0.276	0.529	0.399	0.643
1979—1981 年	0.143	0.236	0.180	0.308	0.233	0.413	0.279	0.556	0.482	0.711
*1982—1984*年	0.138	0.232	0.175	0.307	0.217	0.416	0.274	0.532	0.385	0.678
*1985—1987*年	0.133	0.225	0.165	0.291	0.215	0.386	0.273	0.482	0.261	0.513
*1988—1990*年	0.136	0.227	0.164	0.283	0.207	0.371	0.249	0.476	0.233	0.488
*1991—1993*年	0.137	0.232	0.170	0.292	0.209	0.375	0.261	0.494	0.266	0.511
*1994—1996*年	0.143	0.238	0.169	0.295	0.212	0.379	0.264	0.494	0.278	0.552
*1997—1999*年	0.141	0.234	0.168	0.285	0.203	0.357	0.254	0.482	0.305	0.569
*2000—2002*年	0.141	0.235	0.167	0.282	0.194	0.345	0.243	0.456	0.290	0.543
*2003—2005*年	0.144	0.244	0.172	0.293	0.201	0.360	0.248	0.456	0.307	0.534
*2006—2008*年	0.149	0.251	0.175	0.301	0.207	0.366	0.267	0.471	0.320	0.546
*2009—2011*年	0.150	0.258	0.174	0.304	0.202	0.367	0.252	0.459	0.290	0.538
*2012—2014*年	0.148	0.261	0.169	0.301	0.196	0.359	0.249	0.453	0.293	0.520
*2015—2017*年	0.150	0.280	0.169	0.318	0.195	0.371	0.225	0.452	0.275	0.529

附录 D. 指标构建的技术细节

D1. 重组广度传统指标的详细说明

在第 3.3.4 节中，我们将基于庞加莱圆盘的度量与以往研究中最广泛使用的其他度量进行比较，包括技术多样性、技术距离、原始组合指标和组合熟悉度。下文将详细说明每项度量。

首先，我们使用现有技术的技术类别来计算基于赫芬达尔指数的技术多样性和技术距离。

❖ 技术多样性：该度量最初由（Hall 等，2001）提出。它考察一个引用了 n 个技术类别的焦点专利 i，其中每个类别的频次为 $S_{i,\ j}$，j = 1，2，3...n。焦点专利 i 的赫芬达尔多样性计算如下：1- $\sum_1^n \quad S_{i,\ j}$。为修正现有技术较少的专利可能存在的向下偏差，我们引入一个乘数进行调整。因此，修正后的技术多样性度量定义如下：

$$\text{Technological diversity} = \frac{number\ of\ prior\ art}{number\ of\ prior\ art-1}(1- \sum_1^n \quad S_{ij})$$

❖ 技术距离：技术距离：依照 Trajtenberg 等，1997 的方法，我们采用美国专利商标局分类系统中知识领域的层级组织，而不自行构建知识层级表示。在合作专利分类（“CPC”）系统中，每项专利会被分配一个或多个四级分类。例如，分类

A01B33 在最高层级属于“A”，在第二层级属于“A01”，在第三层级属于“A01B”，在最低层级属于“A01B33”。我们关注的是现有技术类别之间的距离。对于现有技术中的每一对技术类别，我们将技术距离$d_{i,\ j}$设为：若现有技术类别 i 和 j 属于相同的四位技术类别（例如，均为“A01B33”），则为 0；若它们属于相同的三位数类别（例如，“A01B33”和“A01B35”），则为 0.25；若它们属于相同的两位数类别（例如，“A01B33”和“A01F12”），则为 0.5；若它们属于不同的一位数类别或没有现有技术（例如，“A01B33”和“B01B1”），则为 1。得到所有距离后，我们计算其平均值。需要指出的是，我们对 Trajtenberg 等人提出的原始度量进行了调整：（1）关注技术类别之间的距离，而非单项专利之间的距离；（2）采用 CPC 分类系统，该系统的技术类别有四个层级，而非三个层级。因此，技术距离定义为：技术多样性 = $\sum^{n^2}_{i,j} d_{i,\ j}$，其中 n 为现有技术类别总数。

第二步，我们考察另外两种经典度量，这些度量依据焦点专利的技术类别，而非现有技术类别。这些度量用于评估组合的熟悉度或新颖性。由于现有技术类别数量庞大，现有技术中的技术组合通常较不熟悉，也更具原创性。这种倾向可能导致沿连续谱出现明显的过度新颖性偏差。为解决这一问题，我们遵循以往研究，改用焦点专利的技术类别。

● 第一种度量是原始组合指标，相对简单，取决于该组合是否首次出现在我们的数据集中。如果该组合完全新颖，则原创性虚拟变量赋值为 1；反之，如果该组合此前已出现过，则该虚拟变量赋值为 0。

● 第二项度量——组合熟悉度，采用自（Fleming，2001）。专利 i 的组合熟悉度 =

$$\sum_{\square\square\square\ \square\square\square\square\square\square\ \square\ \square\square\square\square\square\square\square\ \square\square\square\square\square\square\ \square} 1\{\text{专利 k 使用与 i 相同的子类组合}\} \times e^{-\left(\frac{application\ date\ of\ i - application\ date\ of\ k}{time\ constant\ of\ knowledge\ loss}\right)}$$

第二项度量背后的假设是，过去使用频率越高的组合就越熟悉。该度量还考虑到，随着集体记忆逐渐淡化，组合的熟悉度可能会随时间下降。根据 Fleming（2001），我们将知识流失的时间常数设为五年。

值得注意的是，尽管我们的组成部分熟悉度变量所生成的分布与 Fleming（2001）以及 Kaplan 和 Vakili（2015年）所观察到的分布高度相似，但我们的组合熟悉度和累计组合使用量呈现出显著更高的均值和方差。这一差异可归因于我们的样本量大于上述两项研究。由于组件数量随样本量线性增加，而组合数量则呈二次增长，因此每种组合在历史数据中达到更高熟悉度值的可能性也随之增加。

D2. 与重组深度和广度经典度量的比较

接下来，我们将该度量与既有文献中用于衡量深度与广度的经典指标进行相关性分析。应用最广泛的指标包括赫芬达尔多样性（Hall 等，2001；Kaplan 和 Vakili，2015a年）、技术距离（Trajtenberg 等，1997）、组合熟悉度（Fleming，2001）以及用于表示该组合是否为全球首次出现的虚拟变量（Verhoeven 等人，2016年）。这些传统指标的详细信息见附录 C 和附录 D，相关性结果见表 D1。

表 D1. 庞加莱深度和广度与既有度量的相关性。

	赫芬达尔多样性	技术距离	原创组合（虚拟变量）	组合熟悉度
深度（最大值 *r*）	0.376	0.272	0.104	-0.242
广度（最大值 *D*）	0.415	0.410	0.096	-0.160

注：*N=4,992,915*，所有 *P*值*<0.001*。

结果表明，我们对深度与广度的度量与传统操作化指标均呈现不同程度的正相关。先前研究识别出的重组广度中，有相当大一部分源自庞加莱嵌入的水平（跨领域）维度。这表明，发明者组合多个技术类别时，往往会跨越知识层级结构中的多个分支。矛盾的是，组合新颖性与深度的相关性高于与广度的相关性（r_{depth}=0.104 >$r_{breadth}$=0.096）。这说明，大多数新组合都源自单一知识领域内的利用式搜索，证实了 Kaplan 和 Vakili（2015年）发现的规律

D3. 两个示例

为直观地解释我们的度量，我们借助图 D1 中的庞加莱圆盘展示两个重组示例，标出两项专利的位置及其组合的现有技术类别。左图中的专利 US8837824 由微软公司提交，涉及一种为图像选择最佳编码编解码器的方法，可降低通信中的网络带宽消耗。在构成该专利基础的现有技术类别中，H04N/L/W（图像通信、数字信息传输、通信网络）和 G06F/Q（数字数据处理、用于管理和行政用途的信息技术）都与数字信息密切相关。G06F/Q 位于 H04N/L/W 的上游：提高通信（H04N/L/W）效率和质量的方法，也会改善整体数据处理（G06F/Q）。这项专利体现了深而窄的重组，因为所有类别都处于技术树的同一分支，并运用了数字图像分析这一特定领域内广泛使用的技术组件。

相比之下，图 D1 右侧面板展示了专利 US9513224（授权给 Labrador Diagnostics LLC），该专利提供了一种针对生物样本的图像分析和测量方法，使用专门的样本支架和系统，在不同环境中检查样本。该专利（1）旨在改进图像数据处理（G06T）；（2）面向实验室环境，重点关注材料的化学和物理性质（G01N 和 B01L）。这体现了广而浅的重组，其元素横跨技术树的不同分支，出乎意料地将不同候选技术中的通用设计结合起来，而不是借鉴深藏于

特定应用复杂领域中的细致技术实现。由此，图 D1 展示了我们各项度量的依据和灵感。

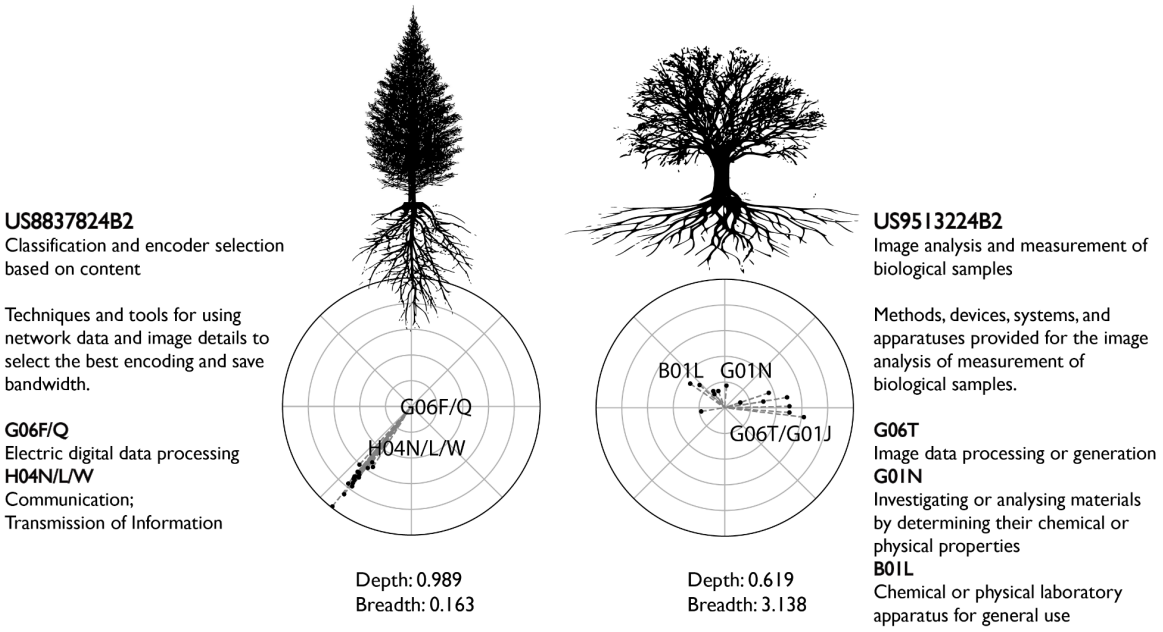


图 D1. 深度与广度重组示例

D4. 机制分析中$\Delta\theta$ 的详细内容

在机制分析中，我们使用 $\Delta\theta$ 来度量焦点专利公开后知识在庞加莱圆盘上的扩散。我们用一个例子（图 D2）来更直观地说明这一点。

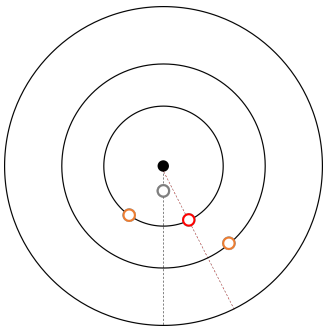


图 D2. 庞加莱圆盘上知识扩散的概念图

假设我们已经构建了一个庞加莱嵌入空间。专利公开时，其在圆盘上的位置用一个空心灰色点表示。专利公开后的三年内，出现了一项引用焦点专利的后续作品，以红点表示。在首次公开后的三至六年间，又出现了两项引用灰色专利的后续作品，我们用橙色表示它们。

首先，我们比较灰色点和红色点。它们之间的距离可分别通过垂直维度上的 Δr 和水平维度上的 $|\Delta\theta|$ 来度量。水平差异度量 $|\Delta\theta|$反映了技术组件对其他领域施加的影响。在机制分析中，我们使用 $|\Delta\theta|$ 的标准差作为离散度的代理变量；在专利层面，我们计算所有后续作品的 $|\Delta\theta|$ 平均值，并将其作为正文表 E8 的因变量。同样，我们使用所有后续作品的 Δr 平均值作为表 E7 的因变量。需要注意的是，Δr 可以为正（表示下游移动），也可以为负（表示上游移动），而 $\Delta\theta$ 没有方向性，因此我们使用其绝对值。因此，后续作品沿 Δr 维度的分布服从正态分布，而沿 $|\Delta\theta|$ 维度的分布则高度偏斜。

附录 E：机制分析模型详情

我们基于广义结构方程模型开展了一系列分析，以揭示效应路径的时间变化。结果报告于表 E1–E5。另有两项分析报告于表 E6 至 E8：（1）深度与广度在发明人、组织和领域层级上对专利影响力的作用；（2）深度与广度对后续专利特征的影响。最后，我们估计了广义负二项回归（均值-离散度模型）。均值模型和离散度模型使用完全相同的一组预测变量。我们在正文中报告均值模型，并在表 E9 中报告离散度模型。

表 E1。影响力累积的中介分析

（0-5年，采用2002年前的专利样本），基于广义结构方程模型

	认知锁定	网络锁定	认知桥接	结构桥接	0-5年 前向引文
认知锁定					0.133
					[0.000]
					(0.001)
认知桥接					0.129
					[0.000]
					(0.004)
结构桥接					-0.056
					[0.000]
					(0.011)
网络锁定					0.078
					[0.000]
					(0.001)
深度	1.592	0.086	-0.305	0.013	-0.086
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.024)	(0.007)	(0.003)	(0.001)	(0.008)
广度	-0.438	-0.041	0.155	0.023	0.030
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.003)	(0.001)	(0.000)	(0.000)	(0.001)
水平位置（平均 theta）	0.030	0.018	-0.005	0.001	0.038

	[0.000] (0.002)	[0.000] (0.001)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.001)
自然对数（组件熟悉度）	0.111	0.035	0.013	0.002	-0.003
	[0.000] (0.002)	[0.000] (0.001)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.001)
自然对数（组合熟悉度）	-0.059	0.006	-0.035	-0.008	0.023
	[0.000] (0.001)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)
技术类别数量	0.032	-0.004	-0.007	-0.005	0.016
	[0.000] (0.001)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)
现有技术类别数量	0.002	0.000	-0.000	-0.000	0.001
	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)
分配给组织	0.178	0.410	-0.016	0.002	-0.034
	[0.000] (0.009)	[0.000] (0.007)	[0.000] (0.001)	[0.000] (0.000)	[0.000] (0.006)
团队发明	0.053	0.094	-0.007	-0.000	0.002
	[0.000] (0.006)	[0.000] (0.002)	[0.000] (0.001)	[0.057] (0.000)	[0.289] (0.002)
Ln（发明人经验）	-0.001	-0.028	-0.002	-0.000	-0.015
	[0.685] (0.002)	[0.000] (0.001)	[0.000] (0.000)	[0.566] (0.000)	[0.000] (0.001)
非专利文献	0.018	-0.005	-0.003	0.000	0.006
	[0.000] (0.001)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)
权利要求数量	0.007	-0.003	-0.000	0.000	0.010
	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)	[0.180] (0.000)	[0.000] (0.000)
家庭规模	-0.065	0.033	-0.003	0.000	-0.079
	[0.000] (0.004)	[0.000] (0.001)	[0.000] (0.001)	[0.301] (0.000)	[0.000] (0.002)
年份固定效应	YES	YES	YES	YES	YES
常数	1.405	1.049	0.618	0.153	0.859
	[0.000] (0.131)	[0.000] (0.013)	[0.000] (0.025)	[0.000] (0.008)	[0.000] (0.014)

Vars 并在成对层面控制协变量的协方差，以及四个中介变量注：*P*值位于方括号内。稳健标准误位于圆括号内。所有检验均为双尾检验。

表 E2. 基于 GSE 的影响力累积中介分析（5至10年，使用2002年之前的专利样本）

M

	认知锁定	网络锁定	认知桥接	结构桥接	未来5至10年引文
认知锁定					0.275
					[0.000]
					(0.001)
认知桥接					0.297
					[0.000]
					(0.004)
结构桥接					0.066
					[0.000]
					(0.011)
网络锁定					0.041
					[0.000]
					(0.002)
深度	1.441	0.075	-0.340	0.006	-0.232
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.021)	(0.005)	(0.003)	(0.001)	(0.008)
广度	-0.406	-0.026	0.162	0.024	0.068
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.003)	(0.001)	(0.000)	(0.000)	(0.001)
水平位置（平均 theta）	0.039	0.004	-0.006	0.000	0.031
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.002)	(0.001)	(0.000)	(0.000)	(0.001)
自然对数（组件熟悉度）	0.112	0.014	0.009	0.000	-0.010
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.002)	(0.001)	(0.000)	(0.000)	(0.001)
自然对数（组合熟悉度）	-0.047	-0.002	-0.034	-0.008	0.026
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
技术类别数量	0.026	-0.003	-0.007	-0.005	0.019
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
现有技术类别数量	0.002	0.000	-0.000	-0.000	0.002
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
分配给组织	0.144	0.255	-0.013	0.003	-0.078
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.008)	(0.006)	(0.001)	(0.000)	(0.006)
团队发明	0.057	0.075	-0.006	0.000	0.003
	[0.000]	[0.000]	[0.000]	[0.627]	[0.202]
	(0.005)	(0.002)	(0.001)	(0.000)	(0.002)
Ln（发明人经验）	-0.009	-0.040	-0.002	0.001	-0.021

	[0.000] (0.002)	[0.000] (0.001)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.001)
非专利文献	0.022	-0.004	-0.003	0.000	0.007
	[0.000] (0.001)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)
权利要求数量	0.008	-0.002	-0.000	0.000	0.011
	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)	[0.041] (0.000)	[0.000] (0.000)
家庭规模	-0.091	-0.011	-0.004	-0.001	-0.132
	[0.000] (0.004)	[0.000] (0.001)	[0.000] (0.001)	[0.005] (0.000)	[0.000] (0.002)
年份固定效应	YES	YES	YES	YES	YES
常数	1.135	1.256	0.692	0.178	0.519
	[0.000] (0.116)	[0.000] (0.127)	[0.000] (0.023)	[0.000] (0.008)	[0.000] (0.102)
变量与协方差	控制四个中介变量成对层面的协变量				

注：*P*值置于方括号内。稳健标准误置于圆括号内。所有检验均为双尾检验。

表E3. 影响力累积的中介分析

基于广义结构方程模型（10至15年，采用2002年之前的专利样本）

	认知锁定	网络锁定	认知桥接	结构桥接	未来10至15年引文
认知锁定					0.335
					[0.000] (0.002)
认知桥接					0.329
					[0.000] (0.005)
结构桥接					0.118
					[0.000] (0.012)
网络锁定					0.010
					[0.000] (0.003)
深度	1.294	0.069	-0.354	0.004	-0.371
	[0.000] (0.026)	[0.000] (0.005)	[0.000] (0.003)	[0.001] (0.001)	[0.000] (0.010)
广度	-0.383	-0.013	0.167	0.025	0.085

	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.003)	(0.001)	(0.000)	(0.000)	(0.002)
水平位置（平均 theta）	0.040	0.001	-0.007	0.000	0.020
	[0.000]	[0.004]	[0.000]	[0.388]	[0.000]
	(0.002)	(0.001)	(0.000)	(0.000)	(0.001)
自然对数（组件熟悉度）	0.114	0.002	0.007	-0.001	-0.013
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.002)	(0.000)	(0.000)	(0.000)	(0.001)
自然对数（组合熟悉度）	-0.046	-0.005	-0.033	-0.007	0.026
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
技术类别数量	0.024	-0.002	-0.007	-0.004	0.019
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
现有技术类别数量	0.002	0.000	-0.000	-0.000	0.002
	[0.000]	[0.910]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
分配给组织	0.097	0.132	-0.009	0.004	-0.127
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.009)	(0.005)	(0.001)	(0.000)	(0.006)
团队发明	0.057	0.057	-0.007	-0.000	-0.004
	[0.000]	[0.000]	[0.000]	[0.622]	[0.164]
	(0.006)	(0.002)	(0.001)	(0.000)	(0.003)
Ln（发明人经验）	-0.009	-0.036	-0.001	0.001	-0.025
	[0.000]	[0.000]	[0.009]	[0.000]	[0.000]
	(0.002)	(0.000)	(0.000)	(0.000)	(0.001)
非专利文献	0.024	-0.003	-0.003	0.000	0.005
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
权利要求数量	0.007	-0.002	-0.000	0.000	0.011
	[0.000]	[0.000]	[0.000]	[0.008]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
家庭规模	-0.112	-0.018	-0.003	-0.001	-0.151
	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]
	(0.005)	(0.001)	(0.001)	(0.000)	(0.002)
年份固定效应	YES	YES	YES	YES	YES
常数	1.763	2.013	0.742	0.196	0.579
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.195)	(0.123)	(0.026)	(0.008)	(0.124)

在成对层面控制协变量的四个中介变量的变量与协方差注：*P*值置于方括号内。稳健标准误置于圆括号内。所有检验均为双尾检验。

表 E4. 基于 GSE 的影响力累积中介分析（15-20 年，采用 2002年之前的专利样本）

M

	认知锁定	网络锁定	认知桥接	结构桥接	15-20 年前向引文
认知锁定					0.376
					[0.000]
					(0.002)
认知桥接					0.330
					[0.000]
					(0.005)
结构桥接					0.217
					[0.000]
					(0.014)
网络锁定					-0.026
					[0.000]
					(0.004)
深度	1.027	0.059	-0.366	0.001	-0.537
	[0.000]	[0.000]	[0.000]	[0.423]	[0.000]
	(0.031)	(0.005)	(0.003)	(0.001)	(0.012)
广度	-0.326	-0.005	0.169	0.024	0.106
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.004)	(0.001)	(0.000)	(0.000)	(0.002)
水平位置（平均 theta）	0.056	-0.001	-0.012	-0.001	0.020
	[0.000]	[0.110]	[0.000]	[0.000]	[0.000]
	(0.002)	(0.000)	(0.000)	(0.000)	(0.001)
自然对数（组件熟悉度）	0.106	-0.004	0.005	-0.001	-0.016
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.002)	(0.000)	(0.000)	(0.000)	(0.001)
自然对数（组合熟悉度）	-0.044	-0.005	-0.033	-0.006	0.027
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
技术类别数量	0.021	-0.001	-0.008	-0.004	0.019
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
现有技术类别数量	0.002	-0.000	-0.000	-0.000	0.002
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
分配给组织	0.081	0.061	-0.008	0.003	-0.146
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.010)	(0.004)	(0.001)	(0.000)	(0.007)
团队发明	0.049	0.039	-0.008	-0.001	-0.000
	[0.000]	[0.000]	[0.000]	[0.029]	[0.991]
	(0.008)	(0.002)	(0.001)	(0.000)	(0.004)
Ln（发明人经验）	-0.003	-0.032	-0.000	0.001	-0.024

	[0.249]	[0.000]	[0.080]	[0.000]	[0.000]
	(0.002)	(0.000)	(0.000)	(0.000)	(0.001)
非专利文献	0.021	-0.002	-0.003	0.000	0.004
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)
权利要求数量	0.006	-0.001	-0.000	0.000	0.009
	[0.000]	[0.000]	[0.000]	[0.067]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
家庭规模	-0.097	-0.007	-0.001	0.000	-0.152
	[0.000]	[0.000]	[0.275]	[0.674]	[0.000]
	(0.006)	(0.001)	(0.001)	(0.000)	(0.003)
年份固定效应	YES	YES	YES	YES	YES
常数	2.424	2.224	0.792	0.206	0.562
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.367)	(0.094)	(0.028)	(0.010)	(0.112)
变量与协方差	针对四个中介变量控制成对层面的协变量				

注：*P*值位于方括号中。稳健标准误位于圆括号中。所有检验均为双尾检验。

表 E5. 直接效应 和四条路径的间接效应（自助法抽样，重复次数 1000）

	直接效应	认知锁定	网络锁定	认知桥接	结构桥接	Full
深度						
所有专利（1976—2017年） 因变量：0–5年引用次数	-0.089*** [-0.103, -0.076]	0.201*** [0.193, 0.209]	0.003*** [0.002, 0.003]	-0.045*** [-0.047, -0.043]	0.004*** [0.003，0.005]	0.073*** [0.058，0.088]
早期专利（1976—2002年） 因变量：0–5年引用次数	-0.086*** [-0.103, -0.070]	0.212*** [0.203, 0.221]	0.007*** [0.006, 0.008]	-0.039*** [-0.042，-0.037]	-0.001** [-0.001,0.000]	0.093*** [0.076, 0.110]
早期专利（1976—2002年） 因变量：5–10年引用次数	-0.232*** [-0.248，-0.215]	0.396*** [0.380，0.411]	0.003*** [0.003, 0.004]	-0.101*** [-0.104, -0.098]	0.000+ [-0.000, 0.001]	0.067*** [0.045, 0.088]
早期专利（1976—2002年） 因变量：10–15年引用次数	-0.371*** [-0.391, -0.350]	0.433*** [0.411, 0.456]	0.001*** [-0.000, 0.001]	-0.116*** [-0.120, -0.113]	0.000+ [-0.000, 0.001]	-0.053*** [-0.081, -0.024]
早期专利（1976—2002年） 因变量：15至20年引用次数	-0.537*** [-0.562, -0.513]	0.386*** [0.354, 0.410]	-0.002*** [-0.002, -0.001]	-0.121*** [-0.125, -0.116]	0.000 [-0.001, 0.001]	-0.274*** [-0.314，-0.233]
广度						
所有专利（1976—2017年） 因变量：0–5年引用次数	0.047*** [0.045, 0.049]	-0.078*** [-0.080，-0.077]	-0.003*** [-0.003, -0.003]	0.031*** [0.030, 0.032]	0.005*** [0.005, 0.006]	0.002 [0.000, 0.004]
早期专利（1976—2002年） 因变量：0–5年引用次数	0.030*** [0.028, 0.033]	-0.058*** [-0.060, -0.057]	-0.003*** [-0.003, -0.003]	0.020*** [0.019, 0.021]	-0.001*** [-0.002, -0.001]	-0.013*** [-0.015, -0.010]
早期专利（1976—2002年） 因变量：5–10年引用次数	0.068*** [0.065，0.071]	-0.112*** [-0.114, -0.109]	-0.001*** [-0.001, -0.001]	0.048*** [0.047, 0.050]	0.002*** [0.001，0.002]	0.005*** [0.002, 0.008]
早期专利（1976—2002年） 因变量：10–15年引用次数	0.085*** [0.082，0.089]	-0.128*** [-0.131, -0.125]	-0.000*** [-0.000，-0.000]	0.055*** [0.053, 0.056]	0.003*** [0.002, 0.004]	0.015*** [0.011, 0.019]
早期专利（1976—2002年） 因变量：15至20年引用次数	0.106*** [0.102，0.109]	-0.122*** [-0.127, -0.118]	0.000*** [0.000, 0.000]	0.056*** [0.054, 0.057]	0.005*** [0.005, 0.006]	0.044*** [0.039, 0.050]

注：*P*值位于方括号内。稳健标准误位于圆括号内。所有检验均为双尾检验。

表 E6. 深　度与广度对发明人、企业和技术领域层级的影响　　el

	专利层面影响（负二项-基准） 模型1	发明人层级含发明人-Year（xtNBREG） 模型2	组织层级 with 组织-Year (xtNBREG) 模型3	领域层级与领域-Year (xtNBREG) 模型 4
深度（最大值）	0.200	0.433	0.494	0.992
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.005)	(0.004)	(0.014)	(0.041)
广度（最大值）	-0.035	0.033	0.061	0.059
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.001)	(0.002)	(0.002)
水平位置（平均 theta）	0.023	0.005	0.001	-0.008
（平均值）	[0.000]	[0.000]	[0.493]	[0.000]
	(0.000)	(0.000)	(0.001)	(0.001)
自然对数（组件熟悉度）	0.031	-0.031	-0.042	-0.017
（平均值）	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.001)	(0.001)
自然对数（组合熟悉度）	0.028	0.011	0.009	-0.025
（平均值）	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.001)	(0.003)
技术类别数量	0.027	0.021	0.021	0.020
(平均值)	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.001)
现有技术类别数量	0.004	0.001	0.001	0.001
(平均值)	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)
分配给组织（平均值）	0.061	0.023	0.000	0.170
	[0.000]	[0.000]	[.]	[0.000]
	(0.002)	(0.003)	(.)	(0.008)
团队发明	0.072	0.083	0.059	0.086
(平均值)	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.002)	(0.004)	(0.006)
Ln（发明人经验）	0.013	-0.047	-0.046	-0.023
(平均值)	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.001)	(0.002)
非专利文献	0.010	-0.000	-0.002	0.000
(平均值)	[0.000]	[0.001]	[0.000]	[0.969]
	(0.000)	(0.000)	(0.000)	(0.000)
权利要求数量	0.019	0.014	0.011	0.021
（平均值）	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)
家庭规模	-0.205	0.115	0.066	-0.122

(平均值)	[0.000] (0.001)	[0.000] (0.001)	[0.000] (0.005)	[0.000] (0.005)
年度专利数量的对数		0.230	0.153	0.090
		[0.000] (0.001)	[0.000] (0.002)	[0.000] (0.002)
发明人/组织/领域固定效应		YES	YES	YES
年份固定效应		YES	YES	YES
职业生涯年份固定效应		YES	YES	YES
常数	0.076	-0.250	0.255	-0.021
	[0.217]	[0.000]	[0.000]	[0.626]
	(0.062)	(0.006)	(0.018)	(0.044)
N	4992915	6068972	681500	334648

注：P值位于方括号内。稳健标准误位于圆括号内。所有检验均为双尾检验。

Ta表 E7. 焦点专利结构对后续技术深度的影响

	模型1	模型2	模型3	模型 4
	预测随访深度	预测后续深度在.....之内 Same 组织	预测随访深度	预测随访其中的深度 Same 组织
深度	0.414	0.515	0.425	0.522
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.001)	(0.001)	(0.001)
广度	-0.018	-0.017	-0.012	-0.013
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.001)
交互作用：深度*广度			-0.008	-0.005
			[0.000]	[0.000]
			(0.000)	(0.001)
水平位置（平均 theta）	-0.002	-0.002	-0.002	-0.002
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)
自然对数（组件熟悉度）	-0.012	-0.012	-0.012	-0.012
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)
自然对数（组合熟悉度）	-0.004	-0.003	-0.004	-0.003
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)
技术类别数量	0.000	0.000	0.000	0.000

	[0.000] (0.000)	[0.095] (0.000)	[0.000] (0.000)	[0.146] (0.000)
现有技术类别数量	-0.000	-0.000	-0.000	-0.000
	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)
分配给组织	-0.004	0.000	-0.004	0.000
	[0.000] (0.000)	[.] (.)	[0.000] (0.000)	[.] (.)
团队发明	-0.002	-0.002	-0.002	-0.002
	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)
Ln（发明人经验）	-0.000	0.001	-0.000	0.001
	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)
非专利文献	0.001	0.001	0.001	0.001
	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)
权利要求数量	0.000	0.000	0.000	0.000
	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)
家庭规模	-0.005	-0.008	-0.005	-0.008
	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)	[0.000] (0.000)
年份固定效应	YES	YES	YES	YES
常数	0.641	0.540	0.631	0.534
	[0.000] (0.005)	[0.000] (0.009)	[0.000] (0.005)	[0.000] (0.009)
调整后的 R 平方	0.489	0.524	0.489	0.524
N	2925660	929483	2925660	929483

注：*P*值置于方括号内。稳健标准误置于圆括号内。所有检验均为双尾检验。

表 E8。焦点专利结构对后续技术广度的影响

	模型1	模型2	模型3	模型 4
	预测随访广度	预测随访内部广度 Same 组织	预测随访广度	预测随访内部广度 Same 组织
深度	-1.271	-1.020	-1.858	-1.615
	[0.000] (0.003)	[0.000] (0.006)	[0.000] (0.005)	[0.000] (0.009)
广度	0.505	0.595	0.168	0.254
	[0.000]	[0.000]	[0.000]	[0.000]

	(0.000)	(0.001)	(0.002)	(0.004)
交互作用：深度*广度			0.404	0.411
			[0.000]	[0.000]
			(0.003)	(0.005)
水平位置（平均 theta）	0.006	0.015	0.007	0.015
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)
自然对数（组件熟悉度）	-0.024	-0.017	-0.024	-0.017
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)
自然对数（组合熟悉度）	-0.023	-0.017	-0.023	-0.017
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)
技术类别数量	0.002	0.000	0.003	0.001
	[0.000]	[0.163]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)
现有技术类别数量	0.000	0.001	0.000	0.001
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)
分配给组织	-0.108	0.000	-0.103	0.000
	[0.000]	[.]	[0.000]	[.]
	(0.001)	(.)	(0.001)	(.)
团队发明	-0.006	-0.009	-0.005	-0.007
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.002)	(0.001)	(0.002)
Ln（发明人经验）	0.002	0.009	0.003	0.010
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)
非专利文献	-0.005	-0.004	-0.005	-0.004
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)
权利要求数量	-0.001	0.001	-0.001	0.001
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)
家庭规模	-0.012	-0.037	-0.011	-0.037
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.001)	(0.001)	(0.001)
年份固定效应	YES	YES	YES	YES
常数	2.030	1.524	2.547	2.052
	[0.000]	[0.000]	[0.000]	[0.000]
	(0.038)	(0.066)	(0.038)	(0.066)
调整后的 R 平方	0.425	0.487	0.429	0.491
N	2925660	929483	2925660	929483

注：*P*值置于方括号中。稳健标准误置于圆括号中。所有检验均为双尾检验。

表 E9. 广义负二项回归的离散度模型（沿用正文表3中的设定）

均值模型	采用相同的自变量，见表*3*				
离散度模型	沿用 模型1	沿用 模型2	后续 模型3	后续 模型 4	后续 模型5
深度	-0.117	0.111	0.086	-0.008	-0.037
	[0.000]	[0.000]	[0.000]	[0.364]	[0.000]
	(0.006)	(0.011)	(0.010)	(0.009)	(0.009)
广度	-0.031	-0.052	-0.049	-0.053	-0.055
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
水平位置（平均 theta）	0.021	0.007	0.014	-0.000	0.014
	[0.000]	[0.000]	[0.000]	[0.946]	[0.000]
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
自然对数（组件熟悉度）	0.036	0.016	0.019	0.037	0.055
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
自然对数（组合熟悉度）	-0.011	-0.022	-0.014	-0.003	0.009
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)
技术类别数量	-0.000	-0.001	-0.001	-0.003	-0.005
	[0.844]	[0.252]	[0.016]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
现有技术数量	-0.000	-0.001	-0.001	-0.000	-0.000
类别	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
分配给组织	0.085	0.054	0.109	0.131	0.174
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.003)	(0.005)	(0.004)	(0.004)	(0.004)
团队发明	0.022	0.033	0.030	0.018	0.021
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.002)	(0.003)	(0.003)	(0.003)	(0.003)
Ln（发明人经验）	0.016	0.012	0.030	0.038	0.050
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
非专利文献	0.008	0.019	0.011	0.006	0.006
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
权利要求数量	-0.008	-0.005	-0.005	-0.005	-0.007
	[0.000]	[0.000]	[0.000]	[0.000]	[0。000]

	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
家庭规模	-0.021	-0.023	-0.003	0.025	0.071
	[0.000]	[0.000]	[0.096]	[0.000]	[0.000]
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
年份固定效应	YES	YES	YES	YES	YES
常数	0.017	-0.003	0.006	0.081	-0.152
	[0.850]	[0.976]	[0.945]	[0.386]	[0.136]
	(0.092)	(0.092)	(0.093)	(0.093)	(0.102)

注：*P*值以方括号标示。稳健标准误以圆括号标示。所有检验均为双尾检验。

现有关于技术演化与重组的大多数文献，主要采用美国专利分类（“美国专利分类”）系统（Fleming and Sorenson，2004；Petralia，2020）或国际专利分类（“IPC”）系统（Corrocher 等人，2007；Fontana 等人，2008）作为分类体系。2013年，美国专利商标局引入合作专利分类（“CPC”）系统，并逐步取代美国专利分类。CPC 分类系统以欧洲专利体系采用的 IPC 为基础，其粒度比美国专利分类更细；美国专利商标局也倾向于在未来使用该系统，以配合欧洲的做法，维护美国的知识产权。实证证据表明，基于美国专利分类和 CPC 分类构建的度量具有高度一致性（Lobo 和 Strumsky，2019）。