



中国管理科学
Chinese Journal of Management Science
ISSN 1003-207X, CN 11-2835/G3

《中国管理科学》网络首发论文

题目：融合大语言模型与动态图神经网络的产学研知识流动测度方法研究
作者：黄璐，任航，刘明熹，王雨路，梅睿寒
DOI：10.16381/j.cnki.issn1003-207x.2025.1785
网络首发日期：2026-03-17
引用格式：黄璐，任航，刘明熹，王雨路，梅睿寒. 融合大语言模型与动态图神经网络的产学研知识流动测度方法研究[J/OL]. 中国管理科学. <https://doi.org/10.16381/j.cnki.issn1003-207x.2025.1785>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

DOI:10.16381/j.cnki.issn1003-207x.2025.1785

融合大语言模型与动态图神经网络的 产学研知识流动测度方法研究

黄璐^{1,2}, 任航³, 刘明熹⁴, 王雨路⁵, 梅睿寒¹

(1. 北京理工大学经济学院, 北京 100081; 2. 数字经济与政策智能工业和信息化部重点实验室, 北京 100081; 3. 北京理工大学(珠海), 广东 珠海 519088; 4. 中国科学院科技战略咨询研究院, 北京 100190; 5. 苏州大学政治与公共管理学院, 江苏 苏州 215123)

摘要: 全面了解和有效度量产学研知识流动及其动态变化, 对于提升国家创新体系整体效能具有重要意义。然而, 现有测度方法多基于单一数据源构建静态单层网络, 难以全面刻画产学研多元异质主体间的复杂关系及其时序演化特征, 影响测度结果的准确性。为此, 本文提出一套基于大语言模型和动态图神经网络的产学研知识流动水平测度方法。首先, 基于论文、专利、新闻报道、地理位置等多源异构数据, 利用 SciBERT 模型、Qwen 大语言模型、层次聚类等方法构建包含机构(学研机构和企业)、成果(学研机构成果和企业成果)、科技人才和知识主题四类节点, 以及引用、合作、隶属、语义关联四类连边, 并嵌入时间维度信息的产学研知识流动动态异质网络; 之后, 结合 Metapath2vec 算法和 DyHNet 模型, 测度产学研知识流动水平。本文选取 2007-2024 年人工智能领域进行实证研究, 以验证方法有效性。本研究能为国家、区域和行业层面的产学研协同创新提供重要的量化决策参考。

关键词: 产学研知识流动; 大语言模型; 动态图神经网络; 动态异质网络

中图分类号: G353.12

文献标识码: A

1 引言

当前, 以人工智能(AI)为代表的新一轮科技革命和产业变革加速演进, 我国正处于推进产学研协同创新、构建现代化产业体系、推进中国式现代化的关键阶段^[1]。党的二十大报告明确提出要“加强企业主导的产学研深度融合, 强化企业科技创新主体地位, 推动创新链产业链资金链人才链深度融合”; 十四届全国人大三次会议进一步强调“加强企业主导的产学研深度融合”“推动科技创新和产业创新融合发展”。推动产学研协同创新、促进科技创新与产业创新深度融合, 是我国深入实施创新驱动发展战略的内在要求, 也是建设现代化产业体系、巩固壮大实体经济根基的迫切需要。

知识流动是以实现知识价值创造为目的, 在来源方和接收方之间进行知识生产、转移和使用的过程^[2]。在产学研协同创新体系中, 知识流动是指高校、科研院所、企业

之间通过知识创造、知识转移和应用, 实现知识共享和资源优势互补, 进而提高创新效率^[1,3]。学研机构与企业之间的知识流动已被广泛视为驱动产学研协同创新的核心机制, 是连接知识供给与需求、推动科技成果转化与再创新的重要纽带^[4]。然而, 在中国情境下, 由于学研机构与企业在知识基础、技术路径和创新逻辑等方面存在显著差异^[5], 跨组织的双向知识流动长期面临障碍, 这种流动受阻进一步引发知识供需错配、科技成果转化乏力、系统性“创新迟滞”等问题^[6], 制约了创新链各环节的有效衔接与协同联动。在此背景下, 对产学研知识流动水平进行全面、准确的测度与分析, 不仅有助于量化不同主体间知识互动的结构特征及其演化规律, 为精准识别产学研协同中的瓶颈问题、优化创新资源配置与合作模式提供数据支撑与理论依据。这对于推动我国实现高水平科技自立自强、促进创新链与产业链深度融合具有重要的现实意义。

基金项目: 国家自然科学基金面上项目(72274013); 北京市社会科学基金重点项目(23JCB021); 青海省重点研发与转化计划科技援青合作专项(2025-QY-202); 中国科学院学部咨询评议项目(2023-ZY04-B-030); 中央高校基本科研业务费专项(100072024CX13021)

通讯作者简介: 刘明熹(1987—), 男(汉族), 山东泰安人, 中国科学院科技战略咨询研究院副研究员、中国科学院特聘(骨干)研究员, 研究方向: 科技和产业政策、四链融合、人工智能、科技人才, E-mail: liumingxi@casisd.cn.

当前,学术界围绕知识流动测度开展了大量前沿研究,一方面,部分学者基于专利、论文和著作等知识载体的引用关系刻画主体间的显性知识流动轨迹^[7];另一方面,通过合著关系构建合作网络,或借助调查问卷、专家访谈等方式对隐性知识流动进行刻画^[8]。然而,现有研究多基于论文或专利等单一数据源构建单层网络^[9],主要聚焦于同类主体间(如高校之间或企业之间)的知识流动水平测度,而较少以产学研多元异质主体间的知识流动为核心展开系统探究。此外,这些网络大多为静态结构,难以全面、动态刻画产学研多元异质主体间知识流动的时间特征。因此,传统的知识流动网络分析方法难以满足对产学研多元异质主体间知识流动进行全面、动态、精准测度的需要。

以 GPT-5、Qwen、DeepSeek 为代表的大语言模型推动了科技文本分析与实体识别的快速发展。相较于传统实体抽取方法,大语言模型具备更强的自然语言理解、多轮推理和上下文生成能力,能够有效从非结构化文本中识别实体及复杂关系^[10],从而提升产学研合作关系识别的准确率和覆盖率。与此同时,基于动态图神经网络的动态图表示学习能够有效整合网络结构特征与语义信息,并刻画节点关系的时间演进过程,生成更具代表性的节点向量表示^[11],从而更有效地揭示产学研知识流动的动态变化规律并提升测度准确性。基于此,本文提出一套基于人工智能大模型和动态图神经网络的产学研知识流动水平测度方法。首先,基于论文和专利数据,使用 SciBERT 文本表示学习模型、Leiden 社区发现算法和 PageRank 算法识别主题;同时,微调 Qwen 大语言模型从多源新闻文本中识别学研机构和企业进而获取产学研合作关系;然后,利用 SciBERT 文本表示学习模型和层次聚类方法识别产学研科技人才,进而构建包含机构(学研机构和企业)、成果(学研机构成果和企业成果)、科技人才和知识主题四类节点,引用、合作、隶属、语义关联四类连边并融合时间信息的产学研知识流动网络;最后,使用 Metapath2vec 算法和 DyHNet 模型对产学研知识流动网络开展动态图表示学习,生成学研机构和企业节点的嵌入向量,并据此测度产学研知识流动水平,并利用人工智能领域验证该测度方法有效性。

2 相关研究

2.1 知识流动

知识流动指以实现知识价值创造为目的,在来源方和接收方之间进行知识生产、转移和使用的过程^[12]。在产学研协同创新体系中,知识流动主要指科学和技术知识在产学研主体之间的扩散与转移,此类知识主要包括在科技创新活动中形成与应用的科学原理、发明或技术方案等^[13]。其中根据知识是否能够清晰表达和有效传递,可以将知识流动分为显性和隐性知识流动^[8,14],前者依托于论文、专利、著作等形式;后者则依赖于合作、面对面交流和人员流动等方式^[14-15]。

近年来,随着全球科技竞争的加剧,知识流动的战略意义愈发凸显,众多学者对其深层影响与实现路径开展了广泛讨论。从流动范围来看,现有研究主要涵盖组织内部部门间的知识流动、不同技术领域间的知识流动、特定国家内部不同地区(如省、市)间的知识流动以及全球范围的知识流动^[16-17],鲜有研究关注产学研不同创新主体间的知识流动;从流动主体来看,当前研究主要探究以论文、专利引用等形式化、结构化的信息转移为代表的显性知识流动,而对基于合作关系、人员流动或组织互动所体现的隐性知识流动关注不足。

从实证研究视角来看,如何测度知识流动水平已成为科技创新管理领域的重要热点问题。近年来,学者们引入科学计量和社会网络分析方法,通过论文的引用、合著关系及专利的引用、共同发明与转让关系来表征知识流动,并从网络角度揭示其结构特征与演化规律^[16-17]。其中,引用关系能够反映显性知识的传播与继承过程,因此论文与专利的引用常被用于表征主体间的显性知识流动^[7]。相较之下,隐性知识流动由于难以直接观测,其测度较为困难,现有研究大多集中于理论分析,少量研究通过调查问卷或专家访谈对隐性知识流动情况进行描述^[8]。此外,也有学者尝试基于合著关系构建合作网络量化测度隐性知识流动水平^[18]。

总的来看,现有知识流动水平测度的研究主要聚焦于单一类型创新主体间(如高校之间或企业之间)的知识流动,且主要基于论文或专利数据构建单层静态网络^[9],仅能反映静态截面特征,难以全面、动态刻画产学研多元异质主体间知识流动的动态演化特征。因此,对产学研多元主体间知识流动水平进行精准测度仍具有较大挑战。

2.2 大语言模型

大语言模型是基于深度神经网络架构的大规模参数模型,通过大规模无监督预训

练、监督微调以及基于人类反馈的强化学习实现多阶段训练^[19-23]。其中，具有代表性的模型包括 GPT^[19-20]、LLaMA^[21]、DeepSeek^[22]和 Qwen^[23]等。相较于传统预训练模型，大语言模型在自然语言理解、多轮推理和上下文生成能力方面具有显著优势^[10]，使其在需要高度语义推理的领域中表现尤为突出，例如实体识别和信息抽取。具体而言，大语言模型能够通过推理隐含的技术语境、高效提取更具价值的文本表征，从而提升对复杂化或高度领域化文本进行信息抽取质量^[24]。它们能够从海量文本中高效捕捉有价值的信息，尤其适用于处理结构复杂、语境变化丰富的非结构化新闻文本数据，实现产学研协同创新情境中复杂的机构主体的精准识别。

因此，本文采用大语言模型对新闻文本进行实体识别，即从新闻文本中抽取高校、科研院所与企业等创新主体，旨在提升产学研合作关系识别的完整性与准确性，构建更高质量的产学研知识流动网络，提高产学研知识流动水平测度的准确性。

2.3 动态图表示学习

图表示学习是一种将图结构数据映射为低维向量表示的方法，旨在尽可能保留网络的结构与语义信息，已广泛应用于推荐系统、创新管理和社会网络分析等领域。现有图表示学习方法大体可分为三类：一是基于矩阵分解的方法（如 GraRep 和 HOPE），通过分解邻接矩阵来刻画图结构，但在可扩展性方面存在局限^[25]；二是基于随机游走的方法（如 DeepWalk 和 Node2Vec），通过模拟随机遍历路径来学习图的局部与全局模式^[26]；三是基于神经网络的方法（如图卷积网络），能够学习更复杂的结构表示^[27-29]。

上述图表示学习方法大多基于静态图结构，难以有效反映现实世界中节点与关系随时间演化的动态特征。为此，学者们提出了动态图表示学习方法，通过对随时间变化的图结构进行建模，实现对动态图数据的低维向量化表示，并捕捉节点属性、结构及其演化过程^[11]。其中，动态图神经网络能够在刻画节点与边动态变化的同时建模跨时间切片的长期依赖和演化路径，从而实现动态图更精细化和准确的表征。作为代表性模型之一，DyHNet 能够在不同时间切片中学习节点嵌入，并保留网络的异质性与时间演化特征^[30]。因此，本文引入 DyHNet 模型，基于多个年度时间切片开展动态图表示学习，

以捕捉产学研知识流动网络中节点表示的时间演化特征，生成融合了结构语义信息与时间演化特征的低维向量表示，从而提升产学研知识流动水平测度的准确性。

3 产学研知识流动测度方法体系构建

本文将构建产学研知识流动动态异质网络，对产学研不同创新主体间的知识流动水平进行测度。基于协同创新理论和知识溢出理论，知识的流动伴随产学研协同创新全过程，科学和技术知识通过在高校、科研机构和企业等创新主体间的互动实现持续交换与扩散。因此，主体间多元关系成为构建知识流动网络的结构依据。

进一步地，根据知识管理理论，产学研不同创新主体间知识的跨组织流动可分为显性知识流动和隐性知识流动，不同类型的知识依托不同载体实现流动。显性知识是指能够被明确表达、记录和传播的知识，其流动可通过论文、专利等成果的引用关系清晰追踪^[8]；隐性知识则以实践经验、直觉、个人技能以及主观认知等形态存在，其流动往往表现为实践互动、能力迁移与认知对齐的过程^[14]。基于此，本文以论文、专利等成果的引用关系表征显性知识流动，并综合采用产学研机构与企业的合作关系、科技人才跨组织流动以及研究主题耦合三类关系对隐性知识流动进行表征。

尽管显性与隐性知识的流动机制有所区别，但在真实创新活动中二者的流动过程往往交织发生并可相互转化^[31]：①显性知识在组织吸收与实践应用过程中常被内化为隐性知识；②隐性知识也可通过形式化、标准化和文献化过程实现显性化；③显性知识与隐性知识在产学研协同创新中通常并行发挥作用，难以在实践中被清晰分离。因此，本文并未将显性与隐性知识区分测度，而是将前述不同关系类型共同纳入同一网络框架，以整体呈现产学研主体间知识流动的结构特征与演化过程，进而测度产学研知识流动水平。

基于此，本文构建包含机构（学研机构和企业）、成果（论文和专利）、科技人才与主题四类节点，引用、合作、隶属与语义关联四类连边和时间关系的产学研知识流动网络以进行产学研知识流动水平测度与分析，研究的整体框架如图 1 所示。

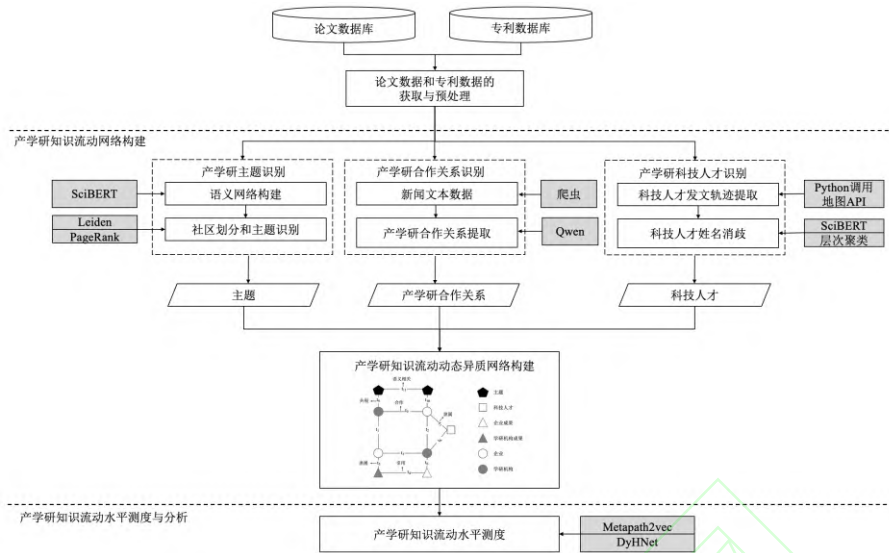


图 1 产学研知识流动测度方法框架

3.1 产学研知识流动网络构建

3.1.1 基于文本表示学习和网络分析的主题识别

主题是一个相对宽泛的概念，涵盖技术、方法和科学概念，并由一系列术语构成^[32]。与传统知识实体相比^[5]，主题能够提供更多的语义信息，更能反映产学研创新主体的知识结构和认知框架。主题耦合指不同创新主体在科研产出中呈现的研究主题关联^[32]。当不同创新主体的科研产出集中于相同或相近的研究主题时，即体现出较强的主题耦合，反映出他们在问题界定、技术路线与认知框架上的深层趋同。这种趋同本质上源于主体间隐性知识的吸收和内化。因此，主题耦合可被视为隐性知识流动与扩散的重要表征。本文主要基于 SciBERT 模型和 Leiden 社区发现算法对产学研主题实体进行识别。

首先，基于 Scopus 论文数据库和 Patsnap 专利数据库，围绕特定研究领域下载论文和专利数据，并利用文本挖掘软件 ITGInsight 和自然语言处理技术^[33]进行数据处理和清洗，获得论文和专利主题词集合以及对应的学研机构和企业集合。随后，采用 SciBERT 模型对论文与专利的主题词集合进行语义表示学习，提取其语义向量表征，并分别基于论文主题词间和专利主题词间的语义相似度构建科学主题词语义网络和技术主题词语义网络。其中，网络节点为论文（专利）主题词，连边代表论文（专利）主题词间的语义关联，连边权重为论文（专利）主题词间的语义相似度。SciBERT 作为基于 BERT 的科技文献文本预训练模型，能较为准确地捕捉文本语义表征^[35]。

在此基础上，通过 Leiden 社区发现算法识别科学社区（主题）和技术社区（主题），并进一步利用 PageRank 算法对各社区（主题）中的主题词进行重要性排序，将数值最高的主题词作为主题名称，为后续知识流动网络的构建提供数据基础。其中，Leiden 算法是一种基于模块度的社区发现方法，能够在快速运行的同时保证社区内部的紧密连接，更好地捕捉网络中的社区结构^[36]。

3.1.2 基于大语言模型的产学研合作关系识别

在产学研协同创新体系中，不同创新主体间的合作既包括通过共同发表论文或申请专利的合著关系，也包括战略签约、科研共建、项目合作等更广泛的互动形式^[37]。这些合作往往伴随资源共享、技术吸收和能力迁移，是隐性知识流动的关键途径。其中，合著成果可以通过论文或专利数据直接识别；但对于战略性、平台性或项目级合作，常规数据库中往往缺乏对应字段、难以通过结构化方式加以刻画，更多以新闻报道等形式对外披露^[37]。因此，新闻报道成为获取此类合作信息的重要来源，能够表征不易通过合著数据捕捉的合作关系。本文利用 Qwen 大语言模型对新闻文本进行实体识别，从中获取产学研不同创新主体之间的合作关系。

首先，围绕“产学研合作”相关关键词^[37]，通过 Python 程序对多家主流新闻网站进行抓取与匹配，获取原始新闻文本数据并完成清洗与预处理，形成新闻文本数据集。其次，采用基于 LoRA 方法微调后的 Qwen-2.5-7B 模型，对全量新闻文本进行实体识别，进而构建了包含学研机构名称、企业名称、

合作年份及合作频次的产学研合作关系数据集。其中, Qwen 是由阿里云发布的开源大语言模型, 具备较强的语义理解、文本生成和长文本推理能力^[23], 已被广泛用作各类下游任务。鉴于其开源、部署灵活, 以及在大规模中文文本处理上的良好性能, 本文将其作为产学研机构实体识别的主要模型。

同时, 基于论文与专利数据, 参考既有研究方法对作者所属机构类型进行识别, 并将机构名称中同时包含学研机构与企业的成果界定为产学研合著成果。进而构建涵盖学研机构名称、企业名称、合著年份与合著频次等关键信息的产学研合著关系数据集。

3.1.3 基于多特征融合消歧的科技人才识别

论文第一作者和专利第一发明人往往承担知识创造与成果产出的核心角色, 可视作推动创新过程的关键科技人才, 其不同创新主体间的跨组织流动被视为隐性知识扩散和能力迁移的重要路径^[15]。因此, 本文聚焦论文第一作者和专利第一发明人, 利用 SciBERT 模型和层次聚类方法 (HAC) 对科技人才进行识别、姓名消歧与对齐, 构建规范化的科技人才集合。

首先, 本文通过 Python 调用百度地图和高德地图 API, 结合天眼查批量查询及人工清洗, 获取论文第一作者和专利第一发明人所属机构的省份信息。在此基础上, 分别处理论文与专利中的核心字段信息: ①对于论文, 提取第一作者、第一作者所属机构、第一作者所属机构类型、第一作者所属机构省份、标题、摘要、关键词、合作者、参考文献、年份。②对于专利数据, 提取专利第一发明人、第一申请 (专利权) 人 (第一发明人所属机构)、IPC 分类号、标题、摘要、合作者、年份、省份信息。需要说明的是, Patsnap 专利数据库通常不提供全部发明人与专利权人的逐一对应关系, 但“第一发明人—第一专利权人”字段清晰可匹配, 可据此准确识别第一专利权人 (第一发明人所属机构)。最终, 得到科技人才数据及其对应的创新成果 (论文和专利) 数据。

然后, 考虑到论文第一作者与专利第一发明人均可能存在重名现象, 本文参考已有研究^[38]构建多特征融合的姓名消歧框架。由于 Scopus 数据库已经对论文作者姓名进行了消歧处理, 不同的作者被赋予了不同的作者 ID, 故不需要单独进行论文作者消歧操作。对于专利数据, 基于标题、摘要、IPC 分类号、省域、合作者五类属性, 为同名发明人构建特征向量。其中, 标题和摘要通过

SciBERT 模型获取语义向量来表示; 对于 IPC 分类号、省域和合作者则直接使用原始的文本形式来表示。鉴于后续网络构建需关注科技人才的机构变更, 因此未将机构作为消歧属性, 以避免在消歧阶段固化其机构信息。随后, 基于特征向量计算同名发明人之间的相似度, 并采用层次聚类算法将相似度较高的记录归为同一实体, 从而实现同发明人的识别与划分。其中, 层次聚类是一种常见的聚类方法, 能够根据数据的层次结构递归地将数据划分为不同的簇, 在处理多维属性、结构复杂的记录归并任务中具有较强的稳定性与适用性^[39]。

在完成专利内部的发明人姓名消歧后, 进一步对论文作者和专利发明人姓名进行对齐。具体而言, 将论文作者与消歧后的专利发明人及其对应的成果记录进行整合, 并基于专利和论文数据的标题、摘要、省域和合作者, 为每位论文作者 (专利发明人) 构建特征向量。随后, 采用与上文类似的方法计算同名的论文作者与专利发明人间的特征相似度, 并对论文作者与专利发明人进行对齐。最终得到科技人才数据, 为后续产学研知识流动网络的构建提供数据基础。

3.1.4 基于多元实体融合的产学研知识流动网络构建

本文融合论文、专利、新闻、地理位置等多源异构数据, 并结合前文识别得到的部分实体与关系, 构建包含机构 (学研机构和企业)、成果 (论文和专利)、科技人才与主题四类节点, 引用、合作、隶属与语义关联四类连边和时间关系的产学研知识流动网络 (如图 2 所示)。在此过程中, 引入时间戳将实体之关系进行时序标记, 使网络能够刻画不同时间节点下的结构变化和演化趋势。构建的产学研知识流动网络为:

$$G_{KF} = G(V, R, W, T) \quad (1)$$

其中, $V \{U, C, S, T, P, Z\}$ 是节点类型的集合, U 、 C 、 S 、 T 、 P 和 Z 分别表示学研机构集合、企业集合、学研机构的成果 (包括论文和专利) 集合、企业的成果 (包括论文和专利) 集合、科技人才集合、主题集合。 $R \{E_1, E_2, E_3, E_4, E_5, E_6\}$ 是关系类型的集合, E_1 表示学研机构与企业之间的合作关系集合, 具体表现为机构间的合著关系与其他合作关系的加权整合, E_2 为学研机构成果与企业成果之间的引用关系集合, E_3 为成果与机构间的隶属关系集合, E_4 为科技人才与机构间的隶属关系集合, E_5 为主题间的语义相似

关系集合, E_6 为机构与主题间的关联关系集合。 $W\{W_1, W_2, W_3, W_4, W_5, W_6\}$ 是权重集合, $W_1, W_2, W_3, W_4, W_5, W_6$ 分别为 $E_1, E_2, E_3, E_4, E_5, E_6$ 的权重, 具体计算公式为:

$$W_1(U_i, C_j) = \alpha * coauthor(U_i, C_j) + \beta * cooperation(U_i, C_j) \quad (2)$$

$$W_2(S_i, T_j) = Citation(S_i, T_j) \quad (3)$$

$$W_3(S_i, U_j) = Cooccurrence(S_i, U_j) \quad (4)$$

$$W_4(P_i, U_j) = I[Cooccurrence(P_i, U_j)] \quad (5)$$

$$W_5(Z_i, Z_j) = Similarity(Z_i, Z_j) \quad (6)$$

$$W_6(Z_i, U_j) = \frac{1}{CZ_i} \sum_{m=1}^{CZ_i} Cooccurrence(S_m, U_j) \quad (7)$$

其中, $W_1(U_i, C_j)$ 是学研机构 U_i 和企业 C_j 之间的权重, $Coauthor(U_i, C_j)$ 表示学研机构 U_i 和企业 C_j 之间的合著次数, $Cooperation(U_i, C_j)$ 表示学研机构 U_i 和企业 C_j 的合作次数, α 和 β 分别是熵权法计算得到的学研机构 U_i 和企业 C_j 之间合著关系与合作关系的权重, 其中, $\alpha = 0.204, \beta = 0.796$; $W_2(S_i, T_j)$ 是学研机构成果和企业成果之间的连边权重, $Citation(S_i, T_j)$ 表示学研机构成果 S_i 和企业成果 T_j 之间的引用次数; $W_3(S_i, U_j)$ 是机构和成果(学研机构成果和学研机构、企业成果与企业)之间的连边权重, $Cooccurrence(S_i, U_j)$ 表示成果 S_i 和机构 U_j 的共现次数; $W_4(P_i, U_j)$ 是科技人才 P_i 和机构 U_j (学研机构或企业) 之间的连边权重, $I[\bullet]$ 是指示函数, 若科技人才 P_i 与机构 U_j 存在共现关系, 则令 $W_4(P_i, U_j) = 1$, 反之 $W_4(P_i, U_j) = 0$; $W_5(Z_i, Z_j)$ 是主题 Z_i 和主题 Z_j 之间的权重, $Similarity(Z_i, Z_j)$ 表示主题 Z_i 和主题 Z_j 之间的语义相似度; $W_6(Z_i, U_j)$ 是主题 Z_i 和机构 U_j (学研机构或企业) 之间的权重, CZ_i 是主题 Z_i 内包含的论文和专利的数量, $Cooccurrence(S_m, U_j)$ 表示主题 Z_i 中成果 S_m 和机构 U_j 的共现次数, 该权重反映主题 Z_i 与机构 U_j 的平均关联强度, 即主题内成果与机构共现次数的归一化结果。 T 表示时间戳集合, 对应网络在不同时间切片下的状态。 本文将时间信息嵌入异质网络的节点连边中, 以刻画关系随时间演进的动态变化。 通过这一网络结构, 能够全面刻画产学研之间知识流动的整体格局及其时序演化特征。

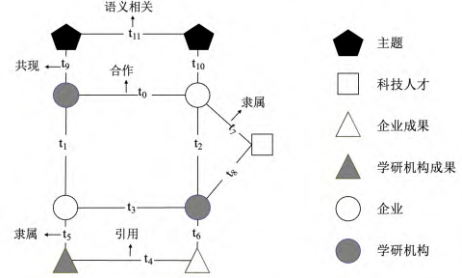


图2 产学研知识流动网络示意图

3.2 产学研知识流动水平测度

在实际的产学研协同创新情境中, 学研机构与企业间的多类型关系往往同时发生。其中, 部分关系具有明确的知识传递方向, 如成果引用以及科技人才跨组织流动; 部分关系则呈现双向互动特征, 如机构合作以及研究主题耦合。这些关系共同推动知识在不同创新主体间的扩散与转移, 使知识在网络内得以持续循环。鉴于产学研知识流动由多类型关系共同作用形成, 且方向特征并不一致, 本文在测度知识流动水平时不单独区分流动方向, 而是基于整体网络结构的关联强度刻画产学研创新主体间的知识流动水平。

图表示学习能够在低维空间中保留网络的结构特征和语义信息, 为度量主体间关系强度及演化规律提供了有效工具。为此, 本文引入 Metapath2vec 算法和 DyHNet 模型, 对产学研知识流动网络中的多元关系和方向信息进行表示学习, 生成学研机构和企业的嵌入向量, 最终通过计算向量间的相似度来衡量产学研知识流动水平。

首先, 本文使用 Metapath2vec 算法生成网络中节点的预嵌入向量, 将节点映射到低维的向量空间以获得节点的稠密向量表示。Metapath2vec 是一种基于结构游走的异质图表示学习方法, 通过在设定的元路径模式下采样节点序列并进行训练, 从而学习节点向量表示^[40]。为刻画产学研主体间多维度、多层级的知识交互过程, 本文设计了六条具有代表性的元路径, 具体如附表 1 所示。

接着, 引入基于动态图神经网络的 DyHNet 模型, 基于多个年度时间切片开展动态图表示学习, 生成各节点的低维向量表示。DyHNet^[30]综合考虑局部节点语义、全局网络语义、网络拓扑结构、时间序列信息, 以及动态网络快照间的时间关联信息, 能够捕捉节点交互关系随时间的演化过程, 从而有效提取跨时间维度的异质网络特征。

最终生成的节点嵌入向量综合了学研机构和企业知识流动网络中的结构位置、关联关系及其时序特征, 能够反映创新主体

在知识产生、吸收与应用过程中的参与程度。因此，嵌入向量相似度可用于衡量产学研主体间的知识流动水平，两个创新主体嵌入向量的相似度越高，表明其知识互动与关联程度越高，知识流动水平越高。省域 i 在 t 年的产学研知识流动水平 KF_{it} 的计算方法为：

$$KF_{it} = \frac{1}{|U_{it}| |C_{it}|} \sum_{u \in U_{it}} \sum_{e \in E_{it}} \cos(u, e) \quad (8)$$

其中， U_{it} 和 C_{it} 分别为省域 i 在 t 年的学研机构节点集合与企业节点集合， u 和 e 分别为对应节点的嵌入向量； $\cos(u, e)$ 为用两向量夹角余弦表示的相似度。最终将所有学研机构与企业节点嵌入向量间的两两相似度取平均，用以反映该省份产学研主体在嵌入空间中的整体对齐程度，以此度量该地区的产学研知识流动水平。

4 实证研究

随着算法能力提升、数据规模增长和算力持续增强，人工智能正成为引领科技创新、助推产业结构升级的重要力量。在此背景下，各国不断加强人工智能战略布局，我国也相继出台多项政策推动人工智能发展与应用。2022 年，我国科技部、工业和信息化部等六部门发布《关于加快场景创新以人工智能高水平应用促进经济高质量发展的指导意见》，部署推动企业、学研机构等创新主体协同促进人工智能场景创新，着力解决人工智能应用和产业化难题；2024 年，中央经济工作会议明确提出要开展“人工智能+”行动；2025 年 8 月，国务院印发《关于深入实施“人工智能+”行动的意见》，进一步凸显了人工智能领域的重要性。然而，在实践中，人工智能领域在产学研协同创新质量和效率上仍有待提升。因此，深入剖析人工智能领域产学研知识流动，对于优化创新资源配置、推动高水平产学研协同发展具有重要意义。基于此，本文选择人工智能领域开展实证分析。

4.1 数据收集与预处理

本文分别从 Scopus 论文数据库与 PatSnap 专利数据库中获取人工智能领域的论文和专利数据。其中，论文数据检索策略主要参考 Liu 等^[41]的研究，构建初始论文数据库，提取这些论文中的高频关键词形成扩展关键词，进而建立完整的检索框架；专利数据检索策略是参考 Jiang 等^[42]的研究，从人工智能领域的机器学习、计算机视觉、自然语言处理、专家系统和机器人五方面构建专利检索框架。最终获得 2007-2024 年

305120 条论文数据和 230923 条专利数据。然后，参考 Chen 等^[32]的研究，利用 ITGInsight 文本挖掘软件分别遴选词频及长度至少为 2、术语度最大的 6000 个论文主题词和专利主题词作为主题词集合。

4.2 人工智能领域的产学研知识流动网络构建

首先，通过 SciBERT 模型对论文与专利主题词进行语义表示学习，构建论文语义网络和专利语义网络。随后，利用 Leiden 社区发现算法进行主题划分，计算主题中各主题词的 PageRank 值，将数值最高的主题词作为主题名称，经整合与筛选获得 56 个主题。

接着，本文围绕人工智能领域“产学研合作”“产学合作”等指定关键词构建查询规则，以人民日报、新华网、中国新闻网、人民网等中央级主流新闻媒体，以及上海证券报、证券时报、21 世纪经济报道、新京报等专业财经与综合新闻媒体发布的新闻报道为数据来源，经检索、采集与整理，共获得 6907 条新闻文本数据。随后，使用 Qwen 模型识别学研机构和企业实体，最终获得人工智能领域 2354 对合作实体，即 2354 条产学研合作关系数据。在学研机构和企业合著关系识别中，共识别出 22024 个产学研合著成果，即 22024 条产学研合著关系数据。

然后，基于 SciBERT 模型和层次聚类方法，对人工智能领域科技人才进行多特征融合的识别、消歧与对齐。具体而言，本文共识别论文第一作者 156102 人，专利第一发明人 114556 人，经对齐后，最终得到 201366 位在中国人工智能领域以第一作者发表论文或以第一发明人申请专利的科技人才。

最后，按照 3.1.4 小节中的步骤，构建产学研知识流动网络。网络整体和各年度切片的结构特征如附表 2 和附表 3 所示。其中，数目表示各类节点或连边数量；最大边权重和最小边权重共同反映了关系权重变化范围；边权加权平均用于刻画节点之间关系的整体强度水平；边权标准差衡量权重分布的离散程度，反映不同关系间差异的大小；平均度值表示网络中节点平均拥有的连接数量；最大度值表示网络中连接数最多的节点的度值；密度表示实际连边占所有可能连边的比例，用以衡量网络的整体紧密程度。

为了更加直观地展示产学研知识流动网络，本文基于 2007-2024 年各年度的网络数据，按年度进行可视化。由于网络中的节点和连边数量过多，仅展示连边权重排名在前 100 的边，结果如附图 1 所示。

4.3 人工智能领域的产学研知识流动水平测度

基于构建的产学研知识流动网络，本文在各年度时间切片上采用 Metapath2vec 算法生成节点的预嵌入向量，并将其作为输入，利用 DyHNet 模型对我国 31 个省（自治区、直辖市）的年度产学研知识流动水平进行测度。同时，为了刻画我国人工智能领域的整体产学研协同创新趋势，本文对国家整体层面的年度知识流动水平进行了汇总计算，从宏观与区域两个维度反映知识流动的时序演化特征。2007-2024 年间的中国人工智能领域整体知识流动趋势如图 3 所示。

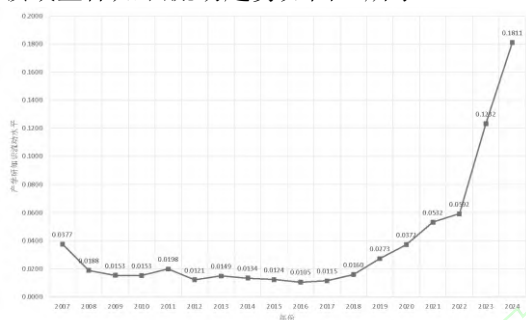


图 3 中国人工智能领域知识在产学研间的整体流动趋势

由上图可知，自 2007 年人工智能蓬勃发展以来，我国人工智能领域知识的流动规模呈现逐年扩大、指数增长的趋势，大致可

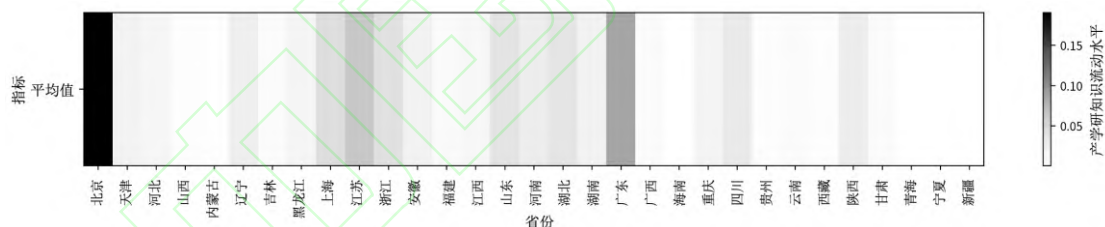


图 4 中国人工智能领域各省份产学研知识流动水平热力图

由上图可知，我国人工智能领域产学研知识流动的区域分布不均衡。东部沿海地区，尤其是长三角、珠三角等科技资源与产业基础雄厚的区域，成为产学研合作最为活跃的核心区域；而西北等中西部地区则整体流动水平偏低，处于网络边缘位置，体现出区域发展不平衡问题。进一步地，从省域维度来看，全国产学研知识流动水平排名前五的省份依次为北京、广东、江苏、上海和浙江，均为人口密集、数字经济基础扎实、创新资源富集和科研产业合作活跃的地区。这表明我国人工智能领域的产学研知识流动高度集中于部分经济发达、创新能力强的核心区域并形成了明显的空间集聚格局。这种集聚虽有助于优势区域的高效协同，但也进一步加剧了地区间在创新要素获取、科研成果转

分为以下三个阶段：（1）在 2007-2018 年间，产学研知识流动水平较低且波动较大。这主要由于人工智能尚处于早期发展阶段，学术研究更集中于自然语言处理、图像识别等特定领域的基础研究与核心算法探索，企业应用场景有限，导致技术转化不足、产业链尚未完善、人才需求不高，加之政策支持和协同机制仍在探索中，科技人才流动性较低，限制了知识在学研机构与企业之间的传播；

（2）随着国家新一轮科技创新政策和人工智能战略的深入实施、产学研协同制度改革进程的加快，产学研不同创新主体间的知识流动水平在 2018-2022 年间显著增加，标志着知识流动进入了一个新的活跃期；（3）从 2022 年开始，整体知识流动水平迎来了井喷式增长，并在 2024 年达到峰值（0.1811）。这反映出人工智能技术及相关产业的高速发展，核心技术突破与商业化应用加速推进，催生出更强的产学研协同需求。同时，国家层面密集出台促进科技成果转化和创新生态构建的政策，为高校、科研机构与企业之间的协同提供制度保障和资源支持，有效推动了知识要素的跨主体高效流动。

此外，为深入探究产学研知识流动的区域空间分布，本文基于 2007-2024 年各省份产学研知识流动水平的平均值绘制热力图，如图 4 所示。

化能力和产业协同水平等方面的差异，影响国家创新体系整体效能的提升。

为进一步刻画区域产学研不同创新主体间知识流动的演化趋势，本文基于各省份各年度的知识流动水平，对近五年排名前五的省份进行统计与比较，如图 5 所示。

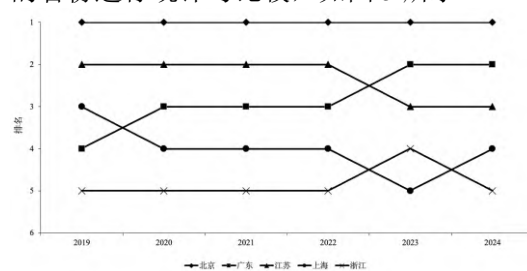


图 5 2019-2024 年中国人工智能领域各省份产学研知识流动水平排名演化（Top 5）

可以看到，北京、广东、江苏、上海和浙江的人工智能领域产学研知识流动水平随时间变化始终稳居前列，表明其在创新资源供给、产业技术吸纳与成果转化等方面具备持续性优势，其人工智能领域的产学研协同体系已形成较为成熟稳定的运行格局。其中，北京的表现尤为突出，在近五年内一直保持全国领先。从地方政府政策布局来看，北京已形成从顶层设计到生态赋能的政策路径，展现出较强的战略主动性与政策导向性。例如，北京市政府率先出台《北京市加快人工智能赋能科学研究高质量发展行动计划》《北京具身智能科技创新与产业培育行动计划》等政策，形成“1+20+N”政策框架。在这一系列政策的持续推动下，北京逐步构建起“基础研究—技术转化—产业落地”

的完整创新链条，为人工智能领域知识在产学研不同创新主体间的高效传导与顺畅流动提供了重要支撑。

为识别产学研知识流动网络中的关键创新主体，本研究参考 Zhang 等^[33]的研究，计算每个学研机构与企业节点的度中心性、接近中心性与中介中心性。这三项指标分别反映了节点的局部影响力、全局位置优势与网络控制能力，综合使用能够更全面地刻画节点在网络中的重要性。为保持指标的一致性，本文采用 min-max 归一化方法对所有指标进行标准化处理，并对 3 项指标使用熵权法得到每个节点的综合得分。最终，分别统计了得分排名前 5 位的学研机构和企业，如表 1 和表 2 所示。

表 1 Top-5 学研机构及指标值

序号	学研机构	所属省份	度中心性	接近中心性	中介中心性	综合得分
1	清华大学	北京	0.513	0.316	0.206	0.326
2	浙江大学	浙江	0.558	0.325	0.099	0.279
3	上海交通大学	上海	0.583	0.310	0.054	0.262
4	北京大学	北京	0.418	0.311	0.114	0.234
5	中国科学院	北京	0.376	0.325	0.034	0.170

表 2 Top-5 企业及指标值

序号	企业	所属省份	度中心性	接近中心性	中介中心性	综合得分
1	华为技术有限公司	广东	0.713	0.308	0.034	0.300
2	北京百度网讯科技有限公司	北京	0.581	0.279	0.081	0.277
3	腾讯科技（深圳）有限公司	广东	0.521	0.304	0.042	0.230
4	阿里云计算有限公司	浙江	0.335	0.307	0.015	0.142
5	国家电网有限公司	北京	0.117	0.345	0.048	0.078

表 1 显示，清华大学、浙江大学、上海交通大学、北京大学和中国科学院的综合得分较高，表明这些高水平研究型大学和国立科研机构在人工智能领域具有较强的知识辐射能力和协同创新驱动作用。表 2 显示，华为、百度、腾讯、阿里云和国家电网的得分居于前列，说明这些头部企业在产学研知识流动体系中发挥着重要的技术应用推动与产业协同枢纽作用。

进一步观察整体得分水平可以发现，学研机构的得分整体高于企业，反映高校和科研机构在知识跨组织流动的过程中占据更突出的主导地位。同时，从机构所属省份来看，北京和广东包含的关键创新主体数量最多，呈现较强的区域集聚效应，这与两地长期处于产学研知识流动领先地位的表现相一致。上述结果可为相关主管部门制定人工智能领域创新发展战略提供决策参考。例如，党的二十届三中全会通过的《中共中央关于进一步全面深化改革、推进中国式现代化的

决定》对“构建支持全面创新体制机制”作出系统部署，特别要求“优化国家科研机构、高水平研究型大学、科技领军企业定位和布局”。高水平研究型大学、科研机构和科技领军企业作为国家创新体系中的关键主体和国家战略科技力量的重要组成部分，在加强基础研究、培养创新人才、促进科研成果转化和区域创新集群发展等方面充分发挥基础性支撑作用。这也侧面印证了本文关键节点识别结果的可靠性。

4.4 方法验证

作为对产学研知识流动测度体系的探索性构建，本文所提出方法尚难以通过直接的指标验证其整体效果。为此，本文分别对“实体识别”“姓名消歧”与“图表示学习”三个关键技术环节的有效性进行检验。

(1) 实体识别方法验证

为定量验证本文训练得到的基于 Qwen 模型的实体识别方法的性能优势，本文从识

别性能和输出稳定性两个维度进行验证。

针对实体识别性能的评估,本文首先依托人工标注方式构建测试集。该测试集包含500条新闻文本,以及其对应的学研机构、企业两类实体标注的参考结果,为后续实体识别性能的量化分析提供基准数据。基于测试集获取各模型的实体识别输出结果,分别从各模型输出文本与测试集参考标注中提取“学研机构”与“企业”两类实体集合,并以测试集参考结果为标准,通过计算模型输出实体与标准实体的重合度,得到本文采

用的Qwen-2.5-7B模型与其他主流大语言模型(LLaMA^[21]、DeepSeek^[22]、GPT-5^[19]和GPT-4o^[20])在实体识别任务上的精确率(Precision)、召回率(Recall)与F1分数(F1 Score)。特别的,为保证不同模型输出在评价维度上的一致性,若模型未输出任何合作主体识别结果,则将其空输出统一记为“无”以消除因输出格式差异造成的评估偏差。针对输出稳定性,本文进一步对本文方法与其他主流大语言模型进行稳定性测试,结果如表3所示。

表3 大模型实体识别性能和稳定性评估结果

模型	Precision	Recall	F1 Score	平均离散程度 (归一化编辑距离)	离散程度 标准差
本文方法	0.732	0.737	0.730	0.052	0.081
LLaMa-3.2-8B	0.356	0.396	0.367	0.066	0.064
DeepSeek-R1-7B	0.372	0.384	0.374	0.605	0.097
GPT-5	0.377	0.363	0.361	0.097	0.279
GPT-4o	0.471	0.502	0.483	0.195	0.292

由此可见,本文所提方法的综合性能表现最优。在识别性能维度,本文方法在Precision、Recall和F1 Score三项核心指标上均显著优于其他对比模型,表明其在实体识别任务中具有更优的识别准确性与目标实体覆盖能力。在输出稳定性维度,本文方法结果的平均离散程度最低且标准差较小,表明其在不同的实体识别任务中,能输出一致性较强、可重复性较高的识别结果。

(2) 科技人才姓名消歧方法验证

为定量验证本文提出的基于SciBERT模型和层次聚类法的多特征融合消歧方法在同名专家消歧任务中的有效性,本文以Scopus数据库作为外部参照基准构建验证样本子集。Scopus数据库为每位论文作者分配了唯一的作者ID,可作为具有高可靠性的真实作者身份标识。具体而言,本文从初始样本中选取20个出现频率较高的作者姓名,以Scopus提供的作者ID为真实标签,构建包含1447条作者记录的验证样本集。在此基础上,选取字符串匹配、TF-IDF+HAC、Word2Vec+HAC三种方法作为对比方法,在Precision、Recall与F1 Score三项指标上与本文所提方法进行比较,结果如表4所示。

表4 姓名消歧实验结果

模型	Precision	Recall	F1 Score
本文方法	0.932	0.998	0.953
字符串匹配	0.982	0.406	0.480
TF-IDF+HAC	1	0.434	0.515
Word2Vec+HAC	0.931	0.532	0.631

可以看出,四种方法在Precision上均表现较好,但部分传统方法的Recall偏低,限制了整体识别能力。相比之下,本文方法在Recall和F1Score上均显著优于其他方法,说明其精确率虽并非最佳,但在科技人才姓名消歧任务中具有更稳定、综合更优的识别效果。

(3) 图表示学习模型验证

为定量验证本文训练的DyHNet模型在动态异质网络中的节点表示学习效果,本文在产学研知识流动网络上基于AUC、Accuracy、Precision、Recall与F1 Score五项指标,将本文方法与四种传统图表示学习方法(DeepWalk^[26]、GCN^[27]、HAN^[28]、HetGNN^[29])进行对比,结果如表5所示。

表5 图表示学习模型性能对比

模型	AUC	Accuracy	Precision	Recall	F1 Score
本文方法	0.740	0.665	0.668	0.658	0.657
DeepWalk	0.500	0.503	0.502	0.503	0.502
GCN	0.547	0.526	0.524	0.526	0.524
HAN	0.565	0.537	0.767	0.537	0.435
HetGNN	0.538	0.524	0.520	0.524	0.521

可以看出,本文方法在五项指标上显著优于其他四种传统的图表示学习方法,充分体现出其在动态异质网络中对节点表示学习的优越性能。

5 结语

5.1 研究结论

本文提出了一套基于大语言模型和动态图神经网络的产学研知识流动水平测度方法。通过整合论文、专利、新闻、地理位置等多源异构数据,采用 SciBERT 模型、Leiden 社区发现算法和 PageRank 算法识别主题,微调 Qwen 大语言模型从多源新闻文本中识别科研机构和企业并获取其合作关系,结合 SciBERT 模型和层次聚类方法识别产学研科技人才,构建包含机构、成果、科技人才和知识主题四类节点,引用、合作、隶属、语义关联四类连边并融合时间信息的产学研知识流动动态异质网络。在此基础上,利用 Metapath2vec 算法和 DyHNet 模型获取科研机构和企业节点的嵌入向量并测度产学研知识流动水平。研究发现:(1)在时间演化趋势上,2007-2024 年间我国人工智能领域的产学研知识流动水平持续提升,呈现由缓慢增长、稳步活跃到快速跃升的演化趋势;(2)在空间维度上,我国人工智能领域产学研知识流动呈现明显的区域不均衡特征,核心流动区域主要集中于东部沿海省份。北京、广东、江苏、上海、浙江等地区的产学研知识流动水平随时间变化始终稳居前列,反映出其在创新资源供给、产业技术吸纳与成果转化等方面具备持续性优势,产学研协同体系已较为成熟稳定;(3)在关键创新主体方面,清华大学、浙江大学、上海交通大学等高水平科研机构以及华为、百度、腾讯等科技领军企业在人工智能领域产学研知识流动网络中居于核心位置,呈现科研机构主导性与区域集聚并存的结构特征。

5.2 理论贡献

本研究的理论贡献如下:

(1)拓展了产学研知识流动的研究视角。现有关于知识流动的研究多聚焦于单一类型创新主体(如高校或企业)之间的知识流动,且主要基于论文或专利数据构建单层静态网络,难以全面刻画产学研多元异质主体间知识流动的动态特征。本文综合论文、专利、新闻和地理位置等多源异构数据,构建了包含机构(科研机构和企业)、成果(论文和专利)、科技人才与主题四类节点,引

用、合作、隶属、语义关联四种连边和时间信息的动态异质网络,从而系统揭示产学研知识流动的结构特征与时序演化过程。

(2)融合多源数据和大语言模型,为产学研知识流动网络的构建提供了关键支撑。现有研究主要依托专利、论文和著作等知识载体,基于引用、合著等行为刻画创新主体间的知识流动。本文在此基础上,创新性地融合论文、专利、新闻和地理位置等多源异构数据,利用 Qwen 大语言模型从新闻文本中批量抽取科研机构和企业实体,识别产学研合作关系,并引入基于 Python 的大规模文件数据清洗技术和 API 驱动的位置信息调取方法,深入挖掘科技人才在学术界与产业界之间的流动轨迹,从而拓展了知识流动的刻画维度,提升了测度的全面性与解释力。

(3)基于动态异质图表示学习,为产学研知识流动量化分析提供了有益补充。本文综合使用 Metapath2vec 算法和 DyHNet 模型,深入挖掘产学研知识流动网络中的多类型关系及时序特征,并基于得到的机构节点嵌入向量计算产学研知识流动水平,该方法还可以用到产学研合作伙伴推荐等研究中。

(4)丰富了知识流动水平测度和产学研协同创新等领域的方法研究。本文提出的基于大语言模型和动态图神经网络的产学研知识流动水平测度方法,能够为精准、客观把握产学研合作水平,破解核心技术“卡脖子”困境,形成高质量高效率的产学研协同创新体系提供重要的理论支撑。

5.3 实践启示

本研究能为政府部门系统性引导企业、高校与科研院所开展高质量产学研协同创新提供关键决策支持,为推动科技创新和产业创新深度融合提供量化依据。

一是强化协同创新平台建设,提升知识流动空间载体。建议国家在重大科技基础设施、产业技术创新平台的建设中,增加面向机构协同的“产学研交互平台”,促进高校、科研院所与企业之间的知识交流与合作深度。例如,政府可以加大对高校和科研院所的科研基础设施投入,建设一批高水平的实验室和技术创新中心,并鼓励企业参与共建共享,为企业与高校、科研院所之间的合作提供便利条件。同时,政府可以通过举办产学研对接活动、搭建线上交流平台等方式,促进各方之间的信息交流与合作。

二是健全隐性知识转化激励机制。建议设立面向机构合作成果转化的绩效评估与激励体系。通过成果转化积分、专利共享收

益、合作项目优先立项等方式，提升机构对隐性知识转化的积极性与参与度。例如，政府可以对高校、科研院所与企业之间的合作项目进行绩效评估，根据评估结果给予相应的奖励和支持。对于在成果转化过程中做出突出贡献的个人和团队，给予相应的荣誉和物质奖励，激发其积极性和创造力。

三是构建国家层面的知识流动记录系统。建议由科技部牵头，搭建全国统一的“产学研知识流动与协同创新管理平台”，实现对高校、科研院所、企业之间的知识共享、技术转移、合作成果的数字化备案与追踪记录，进而强化知识流动的透明度与价值体现，增强跨机构合作的可持续性。该平台可以整合各方的资源和信息，提供一站式的产学研合作服务，包括项目发布、合作对接、成果管理等功能。同时，通过对知识流动的记录和分析，可以为政府制定相关政策提供数据支持，为企业和高校、科研院所的合作提供参考依据，促进产学研协同创新的健康发展。

四是优化区域协同机制，促进知识要素在全国范围内更高效流动。政府应加大对中西部和东北地区等知识流动水平较低地区的支持力度，通过政策引导、资金扶持、人才引进等方式，促进当地高校、科研院所与企业之间的合作与交流，提升知识流动水平和创新能力。加强区域间协同合作，打破区域壁垒，促进知识、技术、人才等创新要素在全国范围内的自由流动和优化配置，实现区域协调发展，提升国家创新体系整体效能。

5.4 研究局限与展望

本研究也存在一些不足。首先，当前测度方法主要基于图表示学习生成的节点嵌入向量相似度，未来可进一步引入更细粒度的刻画维度：一方面，细化知识类型，区分显性与隐性知识流动的贡献差异；另一方面，刻画知识流动方向，以揭示主体间影响关系及其动态演化。其次，本文聚焦省域内部知识流动特征，未涉及省域间的联动机制，后续将探索跨省份知识扩散路径（如京津冀、长三角、粤港澳、成渝地区等）。此外，实证主要聚焦人工智能领域，结论在其他战略性新兴产业领域（如新能源、新材料、航空航天、低空经济）的普适性有待验证，未来将结合不同领域的特征拓展多领域实证分析；本文对产学研知识流动测度方法中三个关键技术环节的有效性进行了检验，未来可进一步结合专家评分法、案例分析法等方法对整体测度结果进行进一步检验。

参考文献

- [1] 陈晓红, 许冠英, 徐雪松, 等. 新质生产力视域下管理科学变革: 内涵特征、现实挑战与发展路径[J]. 中国管理科学, 2025, 33(01): 14-21. Chen X H, Xu G Y, Xu X S, et al. The transformation of management science from the perspective of new quality productive forces: Connotation, practical challenges, and development path[J]. Chinese Journal of Management Science, 2025, 33(01): 14-21.
- [2] Szulanski G. Exploring internal stickiness: Impediments to the transfer of best practice within the firm[J]. Strategic Management Journal, 1996, 17(S2): 27-43.
- [3] 薛澜, 姜李丹, 黄颖, 等. 资源异质性、知识流动与产学研协同创新——以人工智能产业为例[J]. 科学学研究, 2019, 37(12): 2241-2251. Xue L, Jiang L D, Huang Y, Liang Z. Resource heterogeneity, knowledge flow and industry-university-research collaborative innovation——A case study of artificial intelligence industry[J]. Studies in Science of Science, 2019, 37(12): 2241-2251.
- [4] Steinmo M, Rasmussen E. How firms collaborate with public research organizations: The evolution of proximity dimensions in successful innovation projects[J]. Journal of Business Research, 2016, 69(03): 1250-1259.
- [5] Zhao H, Zhao Y, Zhang C. Exploring novelty differences between industry and academia: A knowledge entity-centric perspective[C]// Proceedings of the 20th International Conference on Scientometrics and Informetrics (ISSI 2025), Yerevan, Armenia, 2025.
- [6] 任广乾, 景曼, 焦豪, 等. “链主”国企能引领链上民企创新能力提升吗? ——基于产业链溢出效应的视角[J/OL]. 中国管理科学, 1-16. Ren G Q, Jing M, Jiao H, et al. Can state-owned "chain leader" enterprises lead the innovation capability improvement of private enterprises on the chain? A perspective based on industrial chain spillover effects[J/OL]. Chinese Journal of Management Science, 1-16.
- [7] 李思志, 李佳骏, 李艳红. 管理科学与工程领域的创新轨迹研究——基于 TOP 期刊的文献计量和文本挖掘视角[J]. 中国管理科学, 2014, 22(S1): 56-62. Li S Z, Li J J, Li Y H. Research on the innovation trajectory in the field of management science and engineering: A perspective based on bibliometrics and text mining of TOP journals[J]. Chinese Journal of Management Science, 2014, 22(S1): 56-62.
- [8] Bai J, Jin W, Zhou S. Proximity and knowledge spillovers: Evidence from the introduction of new airline routes[J]. Management Science, 2024, 70(11): 7464-7485.
- [9] 梁爽, 刘小平, 柴文越. 主题—引文融合视角下重要主题发现及知识流动路径研究[J]. 数据分析与知识发现, 2024, 8(02): 99-113. Liang S, Liu X P, Chai W Y. Identifying important topics and knowledge flow paths with topic-citation fusion[J]. Data Analysis and Knowledge Discovery, 2024, 8(02): 99-113.

- [10] Luo M, Xu X, Liu Y, et al. In-context learning with retrieved demonstrations for language models: A survey [J]. arXiv preprint arXiv:240111624, 2024.
- [11] Kazemi S M, Goel R, Jain K, et al. Representation learning for dynamic graphs: A survey[J]. Journal of Machine Learning Research, 2020, 21(70): 1-73.
- [12] Teece D J. Technology transfer by multinational firms: The resource cost of transferring technological know-how[J]. The Economic Journal, 1977, 87(346): 242-261.
- [13] 涂振洲, 顾新. 基于知识流动的产学研协同创新过程研究[J]. 科学学研究, 2013, 31(09): 1381-1390.
Tu Z Z, Gu X. Study on process of industry - university - research institute collaborative innovation based on knowledge flow[J]. Studies in Science of Science, 2013, 31(09): 1381-1390.
- [14] Maurseth P, Svensson R. The importance of tacit knowledge: Dynamic inventor activity in the commercialization phase[J]. Research Policy, 2020, 49(07): 104012.
- [15] 张生太, 朱宏森. 人员流动对组织间隐性知识共享影响研究[J]. 管理科学学报, 2016, 19(7): 78-84.
Zhang S T, Zhu H M. The effect of employee mobility on tacit knowledge sharing among organizations[J]. Journal of Management Sciences in China, 2016, 19(7): 78-84.
- [16] Liu W, Tao Y, Bi K. Capturing information on global knowledge flows from patent transfers: An empirical study using USPTO patents[J]. Research Policy, 2022, 51(05): 104509.
- [17] Cai Z, Ma D, Zhou R, et al. Unraveling the impact of patent transfers on regional innovation: Empirical insights through the lens of entity relationships[J]. Technological Forecasting and Social Change, 2024, 208: 123666.
- [18] Hsu D H, Hsu P H, Zhou K, et al. Industry-university collaboration and commercializing Chinese corporate innovation[J]. Management Science, 2025, 71(6): 5351-5375.
- [19] Leon M. GPT-5 and open-weight large language models: Advances in reasoning, transparency, and control[J]. Information Systems, 2025: 102620.
- [20] Islam R, Moushi O M. Gpt-4o: The cutting-edge advancement in multimodal llm[C]//Intelligent Computing-Proceedings of the Computing Conference. Cham: Springer Nature Switzerland, 2025: 47-60.
- [21] Grattafiori A, Dubey A, Jauhri A, et al. The llama 3 herd of models[J]. arXiv preprint arXiv:2407.21783, 2024.
- [22] Liu A, Feng B, Xue B, et al. Deepseek-v3 technical report[J]. arXiv preprint arXiv:2412.19437, 2024.
- [23] Bai J, Bai S, Chu Y, et al. Qwen technical report[J]. arXiv preprint arXiv:2309.16609, 2023.
- [24] Hou W, Zhao W, Liu X, et al. Knowledge-enriched prompt for low-resource named entity recognition[J]. ACM Transactions on Asian and Low-Resource Language Information Processing, 2024, 23(05): 1-15.
- [25] Sun Z, Wang L, Sun J. A multi-scale graph embedding method via multiple corpora[J]. Neurocomputing, 2023, 540: 126192.
- [26] Perozzi B, Al-Rfou R, Skiena S. Deep Walk: Online learning of social representations[C]//Proc of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014: 701-710.
- [27] Kipf T, Welling M. Semi-supervised classification with graph convolutional networks[J]. arXiv preprint arXiv:1609.02907, 2016.
- [28] Wang X, Ji H, Shi C, et al. Heterogeneous graph attention network[C]//The World Wide Web Conference. 2019: 2022-2032.
- [29] Zhang C, Song D, Huang C, et al. Heterogeneous graph neural network[C]//Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019: 793-803.
- [30] Nguyen H, Rad R, Zarrinkalam F, et al. DyHNet: Learning dynamic heterogeneous network representations[J]. Information Sciences, 2023, 646: 119371.
- [31] Nonaka I, Takeuchi H. The knowledge-creating company[J]. Harvard business review, 2007, 85(7/8): 162.
- [32] Chen X, Ye P, Huang L, et al. Exploring science-technology linkages: A deep learning empowered solution[J]. Information Processing & Management, 2023, 60(02): 103255.
- [33] Zhang Y, Wu M, Miao W, et al. Bi-layer network analytics: A methodology for characterizing emerging general-purpose technologies[J]. Journal of Informetrics, 2021, 15(4): 101202.
- [34] 许元杰, 李建平, 吴登生. 基于期刊论文主题耦合的学科结构识别方法研究[J]. 图书情报工作, 2024, 68(21): 3-15.
Xu Y J, Li J P, Wu D S. Research on disciplinary structure identification method based on topic coupling of journal papers[J]. Library and Information Service, 2024, 68(21): 3-15.
- [35] Dinh T N, Pham P, Nguyen G L, et al. Enhancing local citation recommendation with recurrent highway networks and SciBERT - based embedding[J]. Expert Systems with Applications, 2024, 243: 122911.
- [36] Traag V, Waltman L, Van Eck N. From Louvain to Leiden: guaranteeing well-connected communities[J]. Scientific reports, 2019, 9(01): 5233.
- [37] 苏涛永, 王柯, 孟丽. 产学研协同对绿色技术创新的影响——专业相关度、行业联盟和政府参与的调节效应[J]. 科学学与科学技术管理, 2025, 46(06): 46-62.
Su T Y, Wang K, Meng L. The Impact of industry-university-research collaboration on green technology innovation: The moderating effects of major relevance, industry alliances and government participation[J]. Science of Science and Management of S. & T., 2025, 46(06): 46-62.
- [38] 曾健荣, 张仰森, 王思远, 等. 基于多特征融合的同名专家消歧方法研究[J]. 北京大学学报(自然科学版), 2020, 56(4): 607-613.
Zeng J R, Zhang Y S, Wang S Y, et al. Research on name disambiguation method for

- homonymous experts based on multi-feature fusion[J]. *Acta Scientiarum Naturalium Universitatis Pekinensis*, 2020, 56(4): 607-613.
- [39] Ran X, Xi Y, Lu Y, et al. Comprehensive survey on hierarchical clustering algorithms and the recent developments[J]. *Artificial Intelligence Review*, 2023, 56(8): 8219-8264.
- [40] Dong Y, Chawla N, Swami A. Metapath2vec: Scalable representation learning for heterogeneous networks[C]//*Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2017: 135-144.
- [41] Liu N, Shapira P, Yue X. Tracking developments in artificial intelligence research: constructing and applying a new search strategy[J]. *Scientometrics*, 2021, 126(04): 3153-3192.
- [42] Jiang L, Chen J, Bao Y, et al. Exploring the patterns of international technology diffusion in AI from the perspective of patent citations[J]. *Scientometrics*, 2022, 127: 5307-5323.

Measuring industry-university-research knowledge flow based on large language models and dynamic graph neural networks

HUANG Lu^{1,2}, REN Hang³, LIU Ming-xi⁴, WANG Yu-lu⁵, MEI Rui-han¹

- (1. School of Economics, Beijing Institute of Technology, Beijing 100081, China;
2. Digital Economy and Policy Intelligentization Key Laboratory of Ministry of Industry and Information Technology, Beijing 100081, China;
3. Beijing Institute of Technology, Zhuhai 519088, China;
4. Institutes of Science and Development, Chinese Academy of Sciences, Beijing 100190, China;
5. School of Political Science and Public Administration, Soochow University, Suzhou 215123, China)

Abstract: In the industry-university-research (IUR) collaborative innovation system, knowledge flow refers to the sharing, transfer, and application of knowledge among universities, research institutes, and enterprises. Understanding and effectively measuring of IUR knowledge flow and its dynamic evolution are crucial for improving the overall performance of the national innovation system. However, existing methods mainly rely on single data sources and static single-layer networks, making it difficult to fully capture the complex relationships among IUR actors and their temporal evolution, thereby constraining the comprehensiveness and accuracy of measurement results. For this reason, this study proposes a measurement framework for IUR knowledge flow based on large language models and dynamic graph neural networks. Firstly, based on multi-source heterogeneous data, including papers, patents, news reports, and geographical information, SciBERT, Qwen, and hierarchical clustering methods are used to identify entities and relationships. Then, a dynamic heterogeneous network is constructed to represent the IUR knowledge flow, which comprises four main types of nodes: institutions (universities, research institutes and enterprises), research outputs (papers and patents), scientific and technological talents, and knowledge topics, as well as four types of edges: citation, collaboration, affiliation, and semantic relevance. On this basis, Metapath2vec and DyHNet are utilized to measure the IUR knowledge flow level. An empirical study in the field of artificial intelligence (AI) is conducted to validate the effectiveness of the proposed method. The results show that: (1) the level of IUR knowledge flow in China's AI field exhibits a continuous upward trend from 2007 to 2024, evolving from slow growth to rapid expansion; (2) spatially, knowledge flow shows significant regional imbalance, with core regions concentrated in eastern coastal provinces such as Beijing, Guangdong, Jiangsu, Shanghai, and Zhejiang; and (3) key innovation actors, including leading universities such as Tsinghua University, Zhejiang University, and Shanghai Jiao Tong University, as well as major technology firms such as Huawei, Baidu, and Tencent, occupy central positions in the knowledge flow network. These empirical and validation results demonstrate that the proposed method effectively captures the multi-actor structure, complex relationships, and temporal dynamics of IUR knowledge flow, providing more comprehensive and accurate measurements and offering solid support for identifying knowledge dissemination paths and uncovering core drivers of innovation.

Key words: industry-university-research knowledge flow; large language model; dynamic graph neural network; dynamic heterogeneous network

附录

附表 1 产学研知识流动网络中的代表元路径及其含义

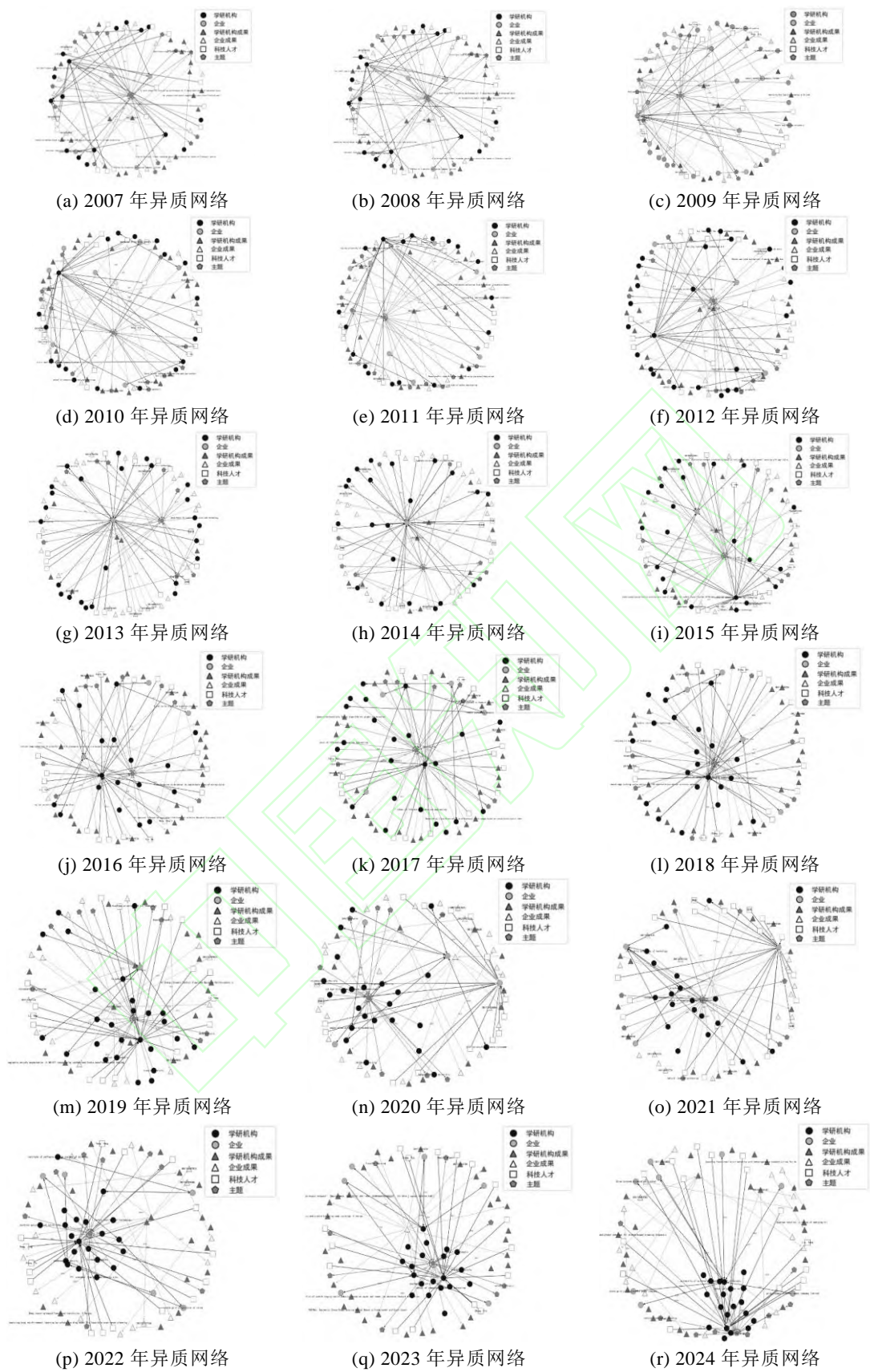
元路径	含义
$\text{学研机构} \xrightarrow{\text{发表}} \text{学研机构的成果} \xrightarrow{\text{被引用}} \text{企业的成果} \xrightarrow{\text{隶属}} \text{企业}$	学研机构的研究成果被企业引用，代表学研机构的学术成果对企业的显性知识输出。
$\text{企业} \xrightarrow{\text{发表}} \text{企业的成果} \xrightarrow{\text{被引用}} \text{学研机构的成果} \xrightarrow{\text{隶属}} \text{学研机构}$	企业发表的成果反过来被学研机构引用，代表企业的产业实践对学研机构的显性知识输出。
$\text{学研机构} \xrightarrow{\text{曾隶属}} \text{科技人才} \xrightarrow{\text{现隶属}} \text{企业}$	科技人才从学研机构流向企业，体现了隐性知识通过人员流动从学研机构流向企业。
$\text{学研机构} \xrightarrow{\text{曾隶属}} \text{科技人才} \xrightarrow{\text{现隶属}} \text{企业}$	科技人才从企业流向学研机构，体现了隐性知识通过人员流动从企业流向学研机构。
$\text{学研机构} \xleftrightarrow{\text{合作}} \text{企业}$	学研机构与企业之间存在直接合作关系，构成隐性知识双向流动的重要通道。
$\text{学研机构} \xleftrightarrow{\text{聚焦}} \text{主题} \xleftrightarrow{\text{语义相似}} \text{主题} \xleftrightarrow{\text{聚焦}} \text{企业}$	学研机构与企业之间存在研究主题上存在的关联，构成隐性知识双向流动的重要通道。

附表 2 2007-2024 年产学研知识流动网络的描述性统计

	类型	数目	最大边权重	最小边权重	边权加权平均	边权标准差
节点	学研机构成果	337544		/		
	企业成果	88730		/		
	企业	28650		/		
	学研机构	25172		/		
	科技人才	201366		/		
	主题	57		/		
边	学研机构成果—企业成果	14140	4	1	1.021	0.145
	企业成果—学研机构成果	28774	10	1	1.262	0.842
	学研机构—企业	10465	11.550	0.462	0.702	0.636
	学研机构成果—学研机构	337544	1	1	1	0
	企业成果—企业	88730	1	1	1	0
	科技人才—学研机构	302946	1	1	1	0
	科技人才—企业	73854	1	1	1	0
	主题—学研机构	13973	146	1	2.543	5.498
	主题—企业	263	3	1	1.118	0.357
主题—主题	147509	1	0.555	0.758	0.066	

附表 3 2007-2024 年产学研知识流动网络年度时间切片的节点与结构特征指标统计

年份	2007	2008	2009	2010	2011	2012	2013	2014	2015
节点数	681519	681519	681519	681519	681519	681519	681519	681519	681519
连边数	153117	156540	158965	158881	157596	158449	159556	162193	166674
平均度值	1.72	1.77	1.77	1.75	1.71	1.72	1.74	1.75	1.78
最大度值	120	146	192	203	172	168	245	312	265
密度	6.6E-07	6.7E-07	6.8E-07	6.8E-07	6.7E-07	6.8E-07	6.8E-07	6.9E-07	7.1E-07
年份	2016	2017	2018	2019	2020	2021	2022	2023	2024
节点数	681519	681519	681519	681519	681519	681519	681519	681519	681519
连边数	174291	185687	208013	237336	255333	274576	263648	260465	250378
平均度值	1.83	1.89	1.96	2.04	2.06	2.06	2.04	2.01	2.03
最大度值	414	477	803	1251	1728	2544	1947	1985	1633
密度	7.4E-07	7.9E-07	8.9E-07	1.1E-06	1.1E-06	1.2E-06	1.1E-06	1.1E-06	1.1E-06



附图 1 2007-2024 年产研学知识流动网络