

# 基于知识图谱的高校科技成果统计数据深度挖掘方法研究

荣誉

长春师范大学，吉林长春，中国

**【摘要】**知识图谱是一种新兴的数据处理方法，采用图谱的方式对知识点进行网状联系，挖掘知识点之间的内在关系，发现传统方法不能认清的关键内涵，为知识的发展和事务趋势预测提供新的视角。高校科研科技成果中论文种类繁杂、水平不一、与科研能力的匹配与评价存在断层，传统的表格式统计方法难以发现高校科研人员与成果之间与世界发展热点之间与高校总体科研能力之间的内在关系。本文以东北某省师范类高校 2024 年科技成果统计数据为模板采用知识图谱的分析处理路径对其成果内容、作者、机构、水平等构件实体结构化语义网络，深度统计分析和挖掘各关键点的内涵，为科研管理和趋势预测提供一种数据分析方法，提高科研数据的应用有效性。

**【关键词】**知识图谱；高校科技成果；统计分析；深度挖掘

**【基金项目】**本论文资助项目：吉林省教育厅科学技术研究项目（JJKH20241019KJ），基于知识图谱的高校科技统计应用研究

## 1. 研究背景与意义

高校科技成果是科技创新的重要载体，传统统计方法往往局限于论文专利成果转化等的数量、作者、引用次数等表面指标，难以揭示数据之间的深层关联（如合作网络、主题演化、跨学科交叉等）。知识图谱技术通过构建实体（如作者、机构、关键词）及其关系的结构化语义网络，可支持更复杂的统计分析和深度挖掘，为科研管理、趋势预测、资源优化提供数据驱动支持。在高校科研管理数字化转型背景下，传统科技成果统计方法（如表格分类统计、基础计量分析）虽能反映成果数量与类型分布，却难以揭示学科交叉、研究主题演化、团队协作网络等深层关联特征[1]。知识图谱技术通过实体识别、关系抽取与图结构建模，可将碎片化科研数据转化为语义关联网络，为突破传统统计分析维度局限提供了新范式[2-5]。然而，现有研究多聚焦于宏观领域知识图谱构建[6-8]，针对高校科技成果数据的细粒度知识建模与动态挖掘仍存在三方面不足：其一，科研主体（学者、机构）与成果（论文、专利）的多模态属性融合不足，导致学术影响力评估片面化；其二，学科关键词与科研主题的时空演化路径分析缺乏可视化追踪工具；其三，科研合作网络的隐性关联识别与预测能力较弱，制约了管理决策的预见性。

本文提出一种面向高校科技成果的“数据-知识-决策”三层深度挖掘框架：首先，

基于 Neo4j 图数据库构建包含学者、机构、成果、主题等实体的全域科研知识图谱，实现异构数据的语义对齐；其次，采用 GNN（图神经网络）算法挖掘跨学科研究热点与团队合作强度，并通过时序图谱量化科研主题演进速率；最后，结合青海民族大学等案例[9-10]，验证该方法在科研资源优化配置、人才梯队建设等管理场景的应用效能。本研究为高校科研治理的智能化转型提供了方法论支持，亦拓展了知识图谱技术在垂直领域的的数据价值转化路径。

## 2. 基于知识图谱的科研论文数据建模及数据挖掘

### 2.1 知识图谱构建流程

数据来源：论文数据库（如 Web of Science、CNKI）、机构数据库、作者信息库、专利数据等。外部数据补充（如机构官网、学术社交平台）。本体设计：核心实体：论文、作者、机构、期刊、关键词、研究领域、基金项目等。关系定义：论文-作者（发表、合作）、论文-关键词（主题）、作者-机构（隶属）、论文-基金（资助）等。高阶关系：跨学科关联、技术演化路径等。数据抽取与融合：使用自然语言处理（NLP）技术抽取实体和关系（如 BERT-CRF 模型）。解决同名消歧（如区分同名作者）、数据冲突（如机构名称标准化）。

### 2.2 知识存储与表示

存储方式：图数据库（如 Neo4j、

JanusGraph) 存储三元组 (实体-关系-实体)。

语义增强: 引入外部知识库 (如 DBpedia、领域本体) 扩展语义关联。

### 2.3 统计分析与可视化

基础统计: 论文数量趋势、作者/机构贡献度、高被引论文分布、期刊影响力排名等。高阶统计: 合作网络分析: 利用图算法 (如 PageRank、社区检测) 识别核心作者/机构。主题演化分析: 基于关键词共现网络, 结合时间维度追踪研究热点变迁。跨学科关联: 通过领域关键词与外部知识库的链接, 量化交叉学科强度。

### 2.4 动态演化建模

时间序列分析: 构建动态知识图谱, 分析科研趋势的阶段特征 (如突发检测算法)。技术演化路径: 结合论文引用关系与关键词共现, 预测未来研究方向 (如链路预测)。

### 2.5 影响力评估与预测

多维度影响力模型: 结合引用次数、期刊影响因子、作者合作网络中心性等指标, 构建综合评价体系。预测模型: 基于图神经网络 (GNN) 预测论文未来被引趋势或作者潜在合作机会。

### 2.6 知识推理与发现

隐含关系挖掘: 通过规则推理 (如若两篇论文共享多个关键词且作者合作, 则主题相关性高) 发现潜在关联。知识补全: 使用 TransE、RotatE 等知识表示学习模型补全缺失关系 (如未被标注的合作关系)。

### 3. 基于知识图谱的科技成果分析应用场景及案例

基于知识图谱工具的高校科技成果深度挖掘可将结果应用于多个方面, 包括科研管理决策, 能够识别高产团队/机构, 优化科研资源配置、发现跨学科合作潜力, 推动交叉领域研究。学术趋势预测, 能够预测新兴技术领域 (如 AI+医疗), 指导科研立项。个性化推荐, 能够为学者推荐潜在合作者或相关论文。科研绩效评估, 能够多维评估学者/机构的学术影响力, 支持职称评审与基金分配。本节以东北某省 2024 年东北某省师范类高校某统计数据为基础, 结合知识图谱分析方法, 对该数据进行深度加工, 获得相应的分析结果。表 1 为东北某省份省属师范类高校 2024 年科技成果统计数据, 以此数据为案例进行基于知识图谱的科技成果数据分析。

表 1. 东北某省省属师范类高校 2024 年科技成果统计表

学科门类	编号	发表科技论文 (篇)		论文检索系统			科技专著				大专院校教科书		编著	
		合计	国外刊物	SCIE	EI	CPCI-S	部	千字	国 (境) 外出版		部	千字	部	千字
									部	千字				
甲	乙	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11	L12	L13
合计	01	956	451	366	22	9	7	1537	1	120	3	1100	2	1736
自然科学	02	584	228	210	9	8	6	1417	0	0	0	0	2	1736
工程与技术	03	224	36	54	1	1	2	270	1	120	3	1100	0	0
医药科学	04	79	55	25	1	0	0	0	0	0	0	0	0	0
农业科学	05	64	22	20	2	0	0	0	0	0	0	0	0	0

基于知识图谱的数据分析方法如下:

步骤一: 确定节点和边

节点: 将学科门类 (甲、乙列)、发表科技论文相关指标 (L1-L4 列)、科技专著相关指标 (L5-L8 列)、大专院校教科书相关指标 (L9-L10 列) 以及编著相关指标 (L11-L13 列) 作为节点。

边: 根据数据之间的逻辑关系确定边。例如, 学科门类与发表科技论文数量、论文检索情况、科技专著、大专院校教科书以及编著之间存在关联, 这些关联可以用边来表

示。

步骤二: 构建知识图谱

以学科门类为中心节点, 分别与其他相关指标节点建立连接。例如, “自然科学节点”与“发表科技论文 (584 篇)” SCIE (228 篇) EI (210 篇) 等节点相连, 表明自然科学领域在这些方面的成果。同样, “工程与技术” “医药科学” “农业科学” 等学科门类也与各自对应的指标节点建立连接。

步骤三: 分析内在联系

学科与论文发表的关系: 从知识图谱中

可以看出,自然科学领域发表的科技论文数量最多,占总论文数的比例较大(584/956≈61.1%),说明该领域的科研活动较为活跃,研究成果丰富。其中,在 SCIE 和 EI 检索系统中的论文数量也相对较多,分别为 228 篇和 210 篇,反映出自然科学领域在国际学术研究中的影响力较大。工程与技术领域的论文发表数量位居第二(224 篇),但在 SCIE 检索中的论文数量相对较少(36 篇),而在 EI 检索中的论文数量(54 篇)相对较为突出,这表明该领域在工程技术应用方面的研究成果在国际上有一定的认可度。医药科学和农业科学领域的论文发表数量较少,分别为 79 篇和 64 篇,但医药科学领域在 SCIE 检索中的论文占比相对较高(55/79≈69.6%),说明该领域对国际前沿研究的关注度较高,注重在国际学术期刊上发表成果。

论文检索与学科的关系: SCIE、EI 和 CPCI-S 等论文检索系统是衡量科研成果国际影响力的重要指标。通过知识图谱可以发现,不同学科在各检索系统中的分布存在差异。自然科学领域在 SCIE 和 EI 检索中的论文数量均较多,说明该领域的研究成果在国际学术界具有较高的认可度和影响力。工程与技术领域在 EI 检索中的论文数量相对较多,这与该领域注重技术应用和工程实践的特点相符,表明其在相关领域的技术创新和应用研究方面得到了国际同行的关注。医药科学领域在 SCIE 检索中的论文占比较高,反映出该领域的研究成果在国际医学研究领域具有一定的地位,可能与国际上对医药科学研究的关注度较高以及该领域研究的国际化程度较高有关。学科与科研成果类型的关系:科技专著、大专院校教科书和编著是科研成果的重要表现形式。从知识图谱中可以看出,自然科学领域在科技专著方面的成果较为突出,有 6 部,1417 千字,且有 2 部编著,1736 千字,说明该领域不仅在科研论文发表方面表现出色,还注重对研究成果的系统总结和理论升华,通过编写科技专著和编著来传播学科知识和研究成果。工程与技术领域在科技专著和大专院校教科书的编写上也有一定成果,分别有 2 部和 3 部,这与该领域需要将研究成果应用于实际工程和教学实践的需求相契合。相比之下,医药科学和农业科学领域在科技专著、大专院校教科书和编著方面的成果较少,可能需要

进一步加强在这些方面的工作,以促进学科知识的传承和推广。学科与成果的联系:知识图谱中,学科门类是核心节点,与发表科技论文、科技专著、大专院校教科书等成果节点相连接。可以看出自然科学领域发表的科技论文数量最多(584 篇),占总论文数的比例较大,且在科技专著方面也有较多成果(6 部,1417 千字)。工程与技术领域虽然论文数量(224 篇)少于自然科学,但在科技专著和大专院校教科书的编写上有一定成果。医药科学和农业科学领域的论文和专著成果相对较少。论文检索与学科的联系: SCIE、EI、CPCI-S 等论文检索系统与各学科相连,反映出不同学科在国际学术影响力方面的差异。自然科学领域在 SCIE 和 EI 检索中的论文数量较多,分别为 228 篇和 210 篇,说明该领域在国际学术研究中的活跃度和认可度较高。工程与技术领域在 EI 检索中的论文有 54 篇,也有一定的国际影响力。医药科学领域在 SCIE 检索中的论文占比较高(55 篇),体现出该领域对国际前沿研究的关注和参与度。成果类型之间的联系:科技论文是科研成果的主要呈现形式之一,而科技专著和大专院校教科书则是对科研成果的进一步总结和系统化阐述。从数据中可以发现,发表科技论文数量较多的学科,在科技专著和大专院校教科书的编写上也有相对较多的成果,说明这些学科在科研积累和知识传播方面较为活跃。

#### 4. 科研管理与建议

学科发展规划方面:加强优势学科建设:自然科学是优势学科,应继续加大对该领域的投入,鼓励科研人员开展前沿性、创新性的研究,争取在更多高影响力的国际学术期刊上发表论文,进一步提升其国际影响力。同时,注重将科研成果转化为科技专著和教材,推动学科知识的传承和发展。扶持薄弱学科发展:对于医药科学和农业科学等成果相对较少的学科,制定针对性的扶持政策。例如,设立专项科研基金,鼓励跨学科合作研究,为年轻科研人员提供更多的培训和交流机会,提升这些学科的科研实力和成果产出。

科研人员激励方面:建立多元化激励机制:除了对发表高影响力论文的科研人员给予奖励外,对于编写高质量科技专著和大专院校教科书的人员也应进行表彰和奖励。激励科研人员在不同的成果形式上进行全面

发展，为学科建设和知识传播做出更大的贡献。注重国际合作与交流：对于在国际学术期刊上发表论文较多的学科和科研人员，提供更多的国际合作与交流机会，如参加国际学术会议、访问国外知名科研机构等。同时，鼓励其他学科的科研人员积极参与国际合作，提高整体的国际学术水平。

科研成果管理方面：加强成果质量监控：建立严格的科研成果审核机制，确保发表的科技论文、科技专著和大专院校教科书等成果的质量。对于多作者的成果，明确各作者的贡献和责任，避免出现学术不端行为。促进成果转化与应用：鼓励科研人员将科研成果应用到实际生产和教学中，推动科技成果的转化。例如，对于工程与技术领域的科研成果，加强与企业的合作，促进技术的产业化应用；对于自然科学和农业科学领域的成果，通过技术推广等方式，提高其社会效益和经济效益。

## 5. 总结与展望

本文采用知识图谱的方法对东北某省2024年科技成果总体数据进行了分析，得到了传统分析方法难以发现的细节，提出了建设性的管理意见，具有较高的推广应用价值。基于知识图谱的深度挖掘方法突破了传统统计的局限性，通过结构化语义网络和复杂图分析技术，能够全面揭示科研数据的隐藏规律。未来研究方向包括开发低代码平台，降低知识图谱构建门槛，结合文本、图表、代码等多元数据丰富知识表示，避免算法偏见，确保评估模型的透明性与公平性。通过该方法，高校可显著提升科研数据的管理水平与决策科学性，助力科技创新与学科

发展。

## 参考文献

- [1] 侯剑华, 李旭彦. 基础科学研究前沿格局的知识图谱解析[M]. 北京: 科学出版社, 2019.
- [2] 王雅娇, 刘育兵, 等. 基于CNKI数据的知识图谱分析[J]. 企业科技与发展, 2018, (9): 148-151.
- [3] 徐伊儿, 李志红. 基于知识图谱的大数据治理研究热点与趋势分析[J]. 情报科学, 2022, 40(3): 45-52.
- [4] 柳卓萍. 近十年知网收录青海民族大学科研论文计量分析[J]. 企业科技与发展, 2018, (9): 148-151.
- [5] 李会, 胡笑梅. 基于文献计量的数据素养领域研究脉络分析[J]. 图书情报工作, 2020, 64(12): 88-95.
- [6] 邓三鸿, 苏新宁, 徐瑶. 基于中文数据库的知识图谱绘制方法及应用: 以创新研究论文的分析为例. 现代图书情报技术, 2011, 27(1): 55-62.
- [7] 刘春辉, 陈宏, 王慧. 知识图谱在科研选题中的应用价值探讨. 现代医药卫生, 2017, 33(22): 3534-3536.
- [8] 李忠云, 李岩, 李太平. 如何评价大学高层次人才——基于CiteSpace可视化分析. 高等农业教育, 2017, (4): 15-20.
- [9] 刘叶玲, 孙继湖, 陈宝国. 基于知识图谱的效率可视化研究. 系统工程理论与实践, 2018, 38(4): 1026-1037.
- [10] 陈悦, 陈超美, 刘则渊, 胡志刚, 王贤文. 知识图谱的作用. 科学通报, 2015, 60(10): 910-918.