

# “问题—方法”关联视角下领域知识创新网络演化机制研究\*

——以信息资源管理学科群为例

■ 杨金庆<sup>1</sup> 庞业佳<sup>1</sup> 刘智锋<sup>2</sup> 李鹏程<sup>3</sup> 李洁<sup>4</sup>

<sup>1</sup> 华中师范大学信息管理学院 武汉 430079

<sup>2</sup> 北京大学信息管理系 北京 100871

<sup>3</sup> 湖北工业大学经济与管理学院 武汉 430068

<sup>4</sup> 苏州大学社会学院 苏州 215000

**摘要:** [目的/意义] 从“问题—方法”关联视角,探究组合创新模式下的领域知识创新网络的演化机制,试图揭示科学知识创新的潜在影响因素,以便辅助科技政策制定,及时开辟创新发展新赛道。[方法/过程] 首先选取信息资源管理学科群,识别领域“问题”与“方法”知识单元,然后以“问题—方法”关联为组合单元,构建领域知识创新网络,最后运用以关系为中心的统计建模方法—指数随机图模型对关系形成及关系模式进行建模,分析网络节点属性特征、网络结构特征等多种内生性和外生性因素对领域知识创新网络形成与演化的影响。[结果/结论] 研究发现,期刊分散广度较大的“问题”与“方法”类知识单元形成连边的可能性相对较大,学科领域内部的“问题”与“方法”间形成连边的倾向相对较大,刊载在影响力相对较低的期刊上的“问题”与“方法”相对较易形成连边,下载量较高或被引频次相对较低的“问题”与“方法”知识单元之间形成连边可能性较大。

**关键词:** 知识网络 演化机制 科技创新 组合创新

**分类号:** G251

**DOI:** 10.13266/j.issn.0252-3116.2024.10.009

**引用本文:** 杨金庆,庞业佳,刘智锋,等. “问题—方法”关联视角下领域知识创新网络演化机制研究——以信息资源管理学科群为例[J]. 图书情报工作, 2024, 68(10): 97-108. (Citation: Yang Jinqing, Pang Yejia, Liu Zhifeng, et al. Research on the Evolution Mechanism of Domain Knowledge Innovation Network from the Perspective of “Issue-Method” Correlation: Taking the Discipline Group of Information Resources Management as an Example[J]. Library and Information Service, 2024, 68(10): 97-108.)

## 1 引言 /Introduction

科学知识创新是推动经济社会发展的重要动力,是谋求新一轮科技革命主动权的主要途径之一。及时把握创新方向不仅有助于抢占科技创新高地,而且可以辅助科技政策制定,及时开辟创新发展新赛道。科技文献作为科学知识创新的载体形式,可理解为由不同类型的知识单元按照一定逻辑和功能组合而成。知识组合创新是通过对现有知识进行非典型组合创新产生新想法的过程<sup>[1]</sup>。表示“问题”和“方法”的科学知识单元是科技文献构成的核心元素,科学知识

创新也可视为对“问题—方法”的配对组合,为科学问题发现新的解决途径<sup>[2]</sup>。

“问题—方法”科学知识组合模式能够客观地抽象出科技文献的核心内容及创新之处。“问题—方法”对的形成并不是随机的,是针对特定科学问题而构造的解决方案。例如,根据不同化学物质的基本特性可组合反应产生新产品。类比化学反应式,“问题—方法”对可以相应地视为科学知识创新生产的“知识反应式”。“问题—方法”组合配对是科学知识创新表现形式的最小组合创新单元。特定领域内不同的科学知识“问题—方法”组合创新单元连接在一起便形

\* 本文系国家重点研发子课题“颠覆性技术识别理论、方法与专家预判系统”(项目编号:2019YFA0707200)、国家自然科学基金项目“基于可解释机器学习的科学知识角色转变预测研究”(项目编号:72304108)、江苏省社会科学青年基金项目“疫情常态化背景下图书馆数字资源认知推荐研究”(项目编号:21TQC001)研究成果之一。

作者简介:杨金庆,特任副教授,博士, E-mail: jinq\_yang@163.com;庞业佳,本科生;刘智锋,博士研究生;李鹏程,讲师,博士;李洁,讲师,博士。

收稿日期:2023-09-26 修回日期:2024-01-16 本文起止页码:97-108

成了领域知识创新网络。

领域知识创新网络的形成与演化内嵌了科学知识创新的潜在规律和内在机制。基于此,笔者立足于“问题—方法”关联的领域知识创新网络,将科学知识组合创新影响因素探索问题转变为“问题—方法”关联知识网络的连边形成机制分析问题,分析领域知识创新网络形成与演化的潜在特征与规律,进而揭示科学知识组合创新的内在机制,研究成果一定程度上可为科技创新方向监测提供理论支撑。

## 2 相关研究 / Related work

### 2.1 “问题”与“方法”识别方法研究

学术论文是知识创新的重要载体,而知识创新一般体现在研究问题与方法的创新。因此,从学术论文中识别出研究问题和方法是探索知识创新及其演化机制的基础。问题和方法词的自动识别方法主要分为3类:①基于规则的抽取方法。该方法主要依据方法词和问题词在学术文本中分布的规律,制定相应的规则进行抽取。程齐凯等<sup>[3]</sup>探索了学术论文标题中常见的组合形式,并制定特定标注规则从标题中抽取问题和方法组合。②基于机器学习的识别方法。该方法主要通过构建分类预测模型,识别问题类和方法类词。K. Heffernan等<sup>[4]</sup>认为科学研究是提出问题和解决问题的活动,将研究问题和解决方案的识别视作二分类问题,并采用SVM (Support Vector Machine) 等算法进行分类,在问题和解决方案的识别上分别取得82.3%和79.7%的准确率。③基于深度语义理解的识别方法。陆伟等<sup>[5]</sup>提出融合BERT (Bidirectional Encoder Representations from Transformers) 和LSTM (Long Short-Term Memory) 的深度学习模型,以捕捉关键词的语义和语境信息以及识别关键词是否为问题词或方法词。也有学者融合关键词的多维特征,引入注意力机制和神经网络识别问题和方法类关键词<sup>[6-7]</sup>。此外,学者开始尝试采用以ChatGPT为代表的大模型助力研究方法等学术论文实体的自动识别研究<sup>[8]</sup>。由此可见,“问题”与“方法”识别方法具有完备的实现方案和相对可靠的识别效果。

### 2.2 科学知识网络演化规律揭示研究

知识网络是从知识结构层面揭示科学知识创新演化规律的基本载体。科学知识网络较为常见的是由共现关系、引用关系等连接而成的知识引用网络和知识共现网络。丁玉飞等<sup>[9]</sup>以词共现关系为连边构建了知识共现网络,通过分析网络形成过程中萌芽期、成

长期、成熟期以及衰退期的演化规律,进而揭示知识演化的内在机制;操玉杰等<sup>[10]</sup>从知识引入和知识产出两个方面分析知识网络结构变化,揭示学科交叉领域科学知识处于不同阶段的演化特征与规律;S. Behrouzi等<sup>[11]</sup>利用受控词表构建计算机领域的动态科学知识网络,通过时间性和顺序性知识结构揭示知识演化规律;A. Mina等<sup>[12]</sup>通过对大型引文网络的分析,揭示医学知识从出现、成长到转化过程中的演化规律。随着研究的深入,学者们开始从多层知识网络视角揭示科学知识演化规律。Z. Liang等<sup>[13]</sup>以引用和共现关系为连边,构建“文献—概念”多层知识网络,揭示知识演化过程中知识创造、使用和演化3个阶段的特征与规律;X. Zhang等<sup>[14]</sup>利用关键词间的多层关系,沿着知识产生、成长、过时、转移和共生5个发展阶段,发掘多层知识网络的演化规律。由此可见,目前学者们在以知识网络为载体揭示科学知识演化规律的过程中,较少从知识网络节点类型角度开展上述研究。

### 2.3 科学知识网络连边形成机制研究

知识创新是一个复杂的、隐含规律与秩序的演化过程。学者们已从择优连接机制、传递性机制和同质性机制3个层面探究了引文网络和共词网络连边形成机制。杨冠灿等<sup>[15]</sup>根据引文网络的连边、度分布、传递闭合等网络结构属性以及专利的地域、领域、学科等节点属性试图揭示专利引文网络连边形成机制;段庆锋等<sup>[16]</sup>利用网络系统理论将专利扩散因素抽象为网络结构嵌入的内生性因素和节点属性的外生性因素,以揭示专利价值、专利保护、多维同配和网络嵌入对专利引文网络形成机制的影响;操玉杰等<sup>[17]</sup>运用网络嵌入性理论,从网络结构因素(内生变量)和关键词属性因素(外生变量)两个方面揭示了领域共词网络连边形成的影响因素及其作用机理,并发现择优连接机制和传递性机制对节点间连边形成具有显著的正向作用;宫雪等<sup>[18]</sup>进一步发现共词网络中新节点的引入更偏向与学科基础词汇和学科前沿词汇连接。共词网络中的择优连接机制在对学科基础词汇的具体研究中经常被具象为节点度数中心度<sup>[19]</sup>和中介中心数<sup>[20]</sup>。综上所述,引文网络和共词网络形成与演化过程中表现出了明显的“择优连接”等特征机制。

## 3 数据来源与预处理 / Data sources and preprocessing

### 3.1 数据来源

笔者选取信息资源管理学科作为实证分析对象,

主要考虑到：① 选择信息资源管理学科便于对分类识别出的“问题—方法”词进行人工控制筛选；② 该领域交叉性较强，实验结果具有代表性，可为科学创新发展给予指导性建议。笔者依据 2021—2022 年《中文社会科学索引（CSSCI）》来源期刊目录（见表 1），选定了 20 本核心期刊，并从中国知网（CNKI）数据库中获取科技文献数据，去除其中的征文、贺词等非学术性文本后，共 94 812 篇论文。

3.2 数据预处理

本文获取的科技文献数据主要以 XML 形式存储（见图 1），从 XML 格式文献中抽取标题、摘要、关键词、期刊名称等科技文献元数据，然后利用标题字段从数据库中检索每篇文献的下载量和被引用频次。作者选取的关键词一定程度上包含了论文的研究问题和所采用的研究方法。问题类和方法类关键词识别是构建“问题—方法”关联知识网络的基础。因此，

数据预处理的核心内容包含两个主要部分：① 识别“问题—方法”关键词；② 抽取关键词间的共现关系。

表 1 2021—2022 年信息资源管理领域《中文社会科学索引（CSSCI）》来源期刊

Table 1 Journals in the information resources management in Chinese Social Science Index (CSSCI), 2021-2022

子学科类别	期刊名称	子学科类别	期刊名称
情报学	情报学报	图书馆学	中国图书馆学报
	图书情报工作		大学图书馆学报
	情报理论与实践		国家图书馆学刊
	情报杂志		图书馆建设
	情报资料工作		图书馆论坛
	数据分析与知识发现		图书馆学研究
	图书情报知识		图书馆杂志
	情报科学	档案学	档案学通讯
	图书与情报		档案学研究
	现代情报		
	信息资源管理学报		

```
<metadata><category>G251</category><keywords>
  <keyword>公共图书馆</keyword>
  <keyword>转型与定位</keyword>
  <keyword>办馆模式</keyword>
  <keyword>信息化</keyword>
  <keyword>数字图书馆</keyword></keywords><keywords xml:lang="en">
  <keyword>Public library. Transformation and orientation. Management pattern. Info
</keywords><received date="2002-04-01"/></critdates></prolog>
<body>
  <p id="1">所谓图书馆的“转型”，是指图书馆办馆模式的转变，主要表现为图书馆的服务理念更新，|
  <title><b>1 面对“信息化”的探索</b></title>
  <p id="3">自党的十一届三中全会以来，我国实行改革开放政策。市场经济在神州大地日益活跃
  <section id="14" level="1">
  <title><b>2 面对“数字图书馆”的探索</b></title>
  <p id="15">1978年，美国著名图书馆学家兰开斯特（F.W.Lancaster）出版《走向无纸信息社会》
</body>
```

图 1 科技文献关键词存储形式

Figure 1 Storage of keywords in scientific and technological literature

3.2.1 “问题—方法”关键词识别与处理

“问题—方法”关键词识别是学术文本挖掘领域被广泛关注的研究问题。学者们将科学研究定义为提出问题和解决问题的过程，据此关键词词汇语义功能可大致分为问题类、方法类<sup>[4,21]</sup>。笔者所在团队前期已对学术文本词汇功识别方法开展了系列研究，结合现有的技术方案和任务形式，识别方法可分为两大类：基于预设类别的分类标注和基于限定内容的文本生成。笔者采用基于有监督学习的神经网络方法，对科技文献中问题类和方法类关键词进行识别。具体而言，笔者选取当前分类准确率较高的 BERT+LSTM 关键词词汇功能标注方法<sup>[22]</sup>，获取了多本计算机及图书情报领域期刊进行数据标注和模型训练，研究方法识别 F1 值为 0.89，研究问题识别 F1 值为 0.79，训练后的模型可用于本文对图书情报领域 20 本核心期刊论文中的关键词进行分类。考虑到分类结果仍存在误差以及异形同义，笔者对识别出的问题方法和方法词集合进行相应处理，发现

问题类关键词占比（平均占比 76.7%）远高于方法类关键词（平均占比 23.3%），笔者首先人工校对方法类关键词，将方法关键词集合中的异形同义词进行合并，如表 2 所示：

表 2 人工校对方法类关键词结果示例

Table 2 Method-type keywords by manual proofreading

方法类关键词归并表述	方法类关键词原始表述
本体	知识本体
本体	情境本体
本体	微本体
本体	语义本体
神经网络	神经网络模型
神经网络	神经网络
协同过滤	协同过滤算法
协同过滤	协同过滤
CiteSpace	Citespace
CiteSpace	Citespace II
CiteSpace	CiteSpace
CiteSpace	CiteSpace 软件
.....	.....



经统计发现,与方法类关键词相比,问题类关键词较多,方法类关键词共有 538 个,问题类关键词共

获取 1 014 个。笔者通过关键词被使用词频排序,对问题类和方法类关键词进行可视化分析,如图 2 所示:

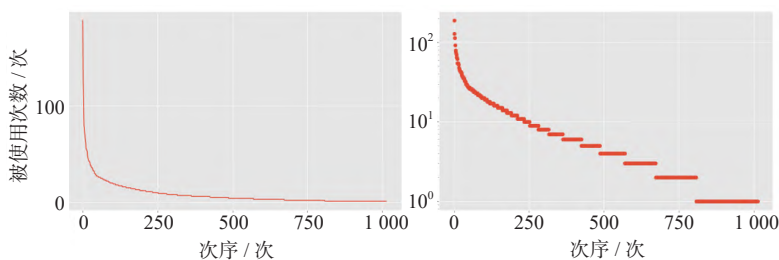


图 2 问题类关键词被使用次数分布

Figure 2 Usage frequency for issue-type keywords

由图 2 可知,问题类与方法关键词被使用次数排序呈幂律分布,笔者选取位次排序前部的

问题类和方法类关键词进行内容可视化,如图 3 所示:

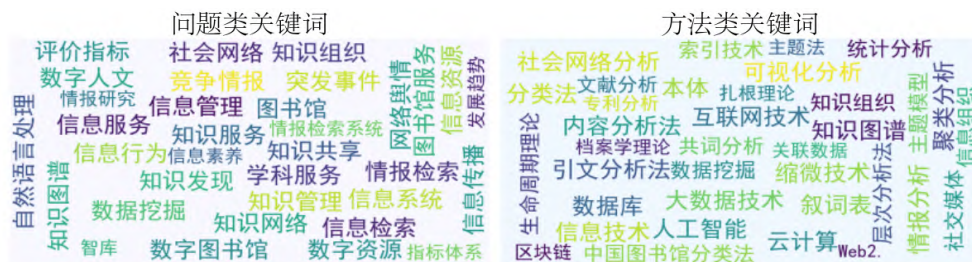


图 3 问题类与方法类关键词可视化

Figure 3 Issue- and method-type keyword visualization

### 3.2.2 “问题—方法”关联知识创新网络构建

“问题—方法”组合一定程度上可以表示科技文献承载的主要研究内容。以关键词表示知识单元,学科领域内所有科技文献中的“问题—方法”组合关联在一起就形成了“问题—方法”关联的领域知识创新网络。笔者将抽取的 8 046 对“问题—方法”组合成领域知识创新网络。由于各关键词通用性不同,方法类关键词被归类后语义较为宽泛,其连边较多,与其关联的节点间的权重相对较高。如图 2 所示,问题类和方法类知识单元在知识创新网络中分布不均衡,为了解决这一问题,笔者采用一种规范化方法对“问题—方法”知识单元关联关系权重进行处理,具体计算方法<sup>[23]</sup>如公式(1)所示:

$$W_{ij} = \frac{c_{ij}}{s_i \times s_j} \quad \text{公式(1)}$$

其中, $i$ 表示问题类知识节点, $j$ 表示方法类知识节点, $W_{ij}$ 表示规范化后的权重, $c_{ij}$ 表示“问题—方法”关联频次, $s_i$ 表示问题类知识节点截止当前被使用频次, $s_j$ 表示方法类知识节点截止当前被使用频次。

由图 2 可知,问题类和方法类知识节点符合幂律分布,对连边权重规范化计算之后,赋予连边权重的“问题—方法”知识节点对相互连接形成了规范化后

的“问题—方法”关联知识创新网络。“问题—方法”关联网络的基本情况统计如表 3 所示:

表 3 “问题—方法”关联知识创新网络基本信息统计  
Table 3 Statistics on the basic information of the “issue-method” knowledge innovation network

网络结构特征	数值
节点数	1 520
连边数	8 406
网络密度	0.006 6
网络平均度数	10.01
网络平均中介中心度数	0.001 6

## 4 领域知识创新网络特征分析 /Feature analysis of domain knowledge innovation networks

### 4.1 “问题—方法”类网络结构演化分析

领域知识创新网络是由问题类和关系类知识节点关联起来构成的复杂网络,网络的演化不仅受领域知识自身属性过程的影响,而且网络自组织过程也是不可忽视的因素<sup>[24]</sup>。领域知识自身属性过程强调网络中领域知识单元个体的类型、学科属性、影响力等存在差异,对关系网络形成具有影响。网络自组织过程是指网络在学科领域情境下自发的促使“问题”与“方法”关联关系的形成,长期累积形成了有序的

网络构型<sup>[25]</sup>。由此,笔者通过对“问题—方法”类网络结构信息进行历时分析,进而揭示网络演化的特征与规律。根据网络嵌入理论,网络中的三角形结构、2-路径、星型结构等是网络中的基本构型<sup>[24]</sup>(见图4)。

“问题—方法”关联的知识创新网络是由问题和方法类关键词关联关系构成,故不存在闭合三角形构型。但是可以从“问题—方法”对组合情况来揭示领域知识组合创新的趋势。

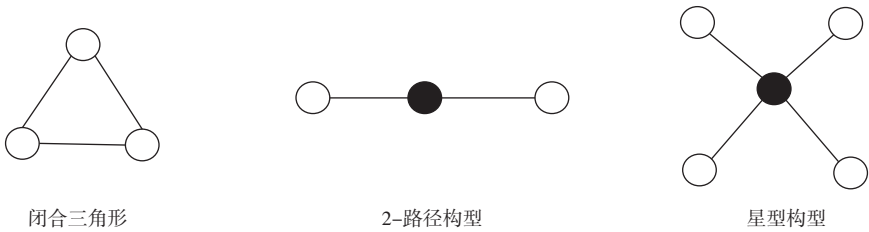


图 4 知识网络中常见的基本构型

Figure 4 Common basic configurations in knowledge networks

笔者采用网络指标中介中心性和点度中心度,计算 2-路径构型和星型结构以及“问题—方法”组合

对等网络结构指标来分析知识创新网络的结构演化情况,如图5所示:

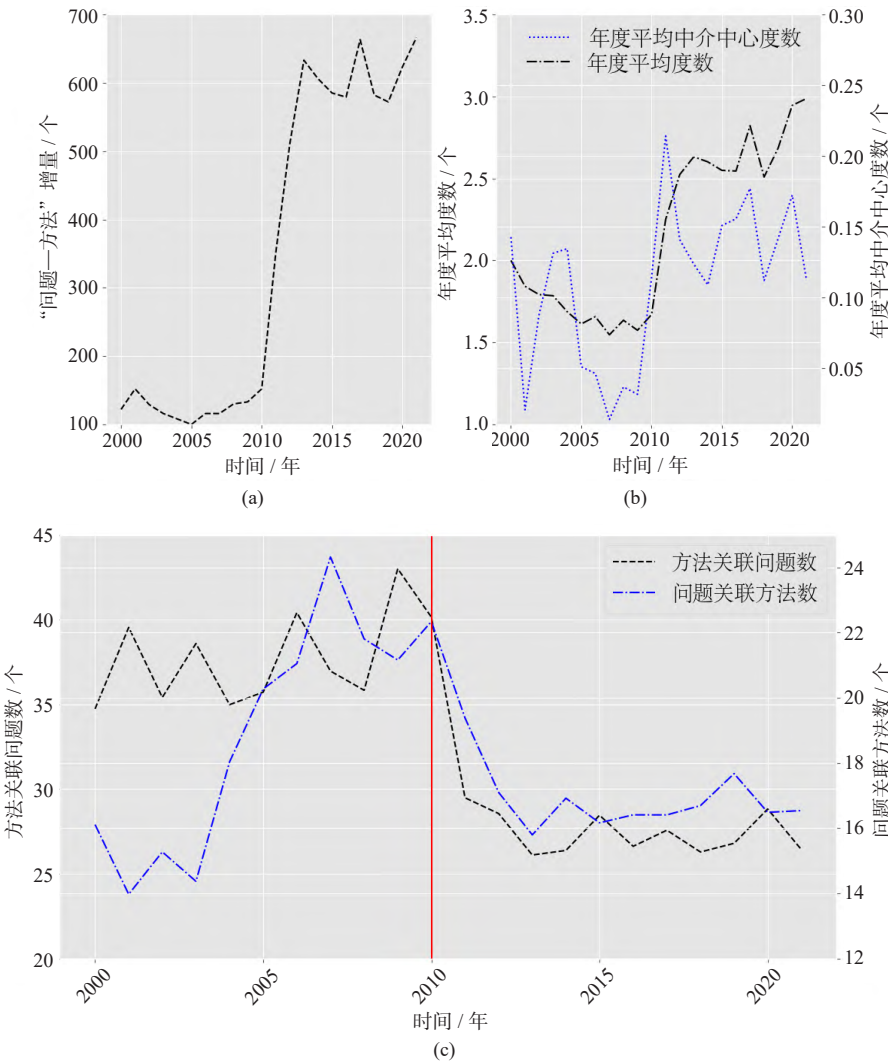


图 5 “问题—方法”关联知识创新网络结构演化分析

Figure 5 Structure evolution of the “issue-method” correlation knowledge and innovation network

从图5(a)中可知,“问题—方法”关联 2010 年之后出现突增现象,2013 年以后表现出稳定波动

趋势。这意味着从 2010 到 2013 年涌现了大量“问题—方法”关键词对,侧面反映出在此期间“图情档”

领域学术研究发展迅速,领域研究范式和内容逐步稳定。图5(b)中的年度平均度数时序分析显示,从2010到2013年知识创新网络的平均度数也出现快速增长,然后逐渐稳定。不难发现,年平均度数整体演化趋势与“问题—方法”增量趋势基本一致,说明星型结构特征能够反应知识网络所表征的学科演化。可是,年度平均中介中心度历时演化整体趋势波动较大,难以很好地揭示知识创新网络演化特征与规律。为了突出“问题—方法”关联组合的演化规律,笔者计算了各年度解决研究问题的平均方法数(即“问题”关联“方法”平均数)和研究方法用于解决的研究问题的平均数(即“方法”关联“问题”平均数)。如图5(c)所示,可以观察到两

条曲线呈现出上下交错的趋势,2000—2010年间“问题”关联“方法”平均数整体低于“方法”关联“问题”平均数且前者相对较为稳定,后者呈快速上升趋势,反映出这一时期不断尝试新的研究方法解决现有问题,而2010年以后两者均呈现出下降后趋于稳定,可推测“图情档”领域研究方法的专用性在增强。为了更好地理解领域知识创新网络结构演化过程,笔者选取4个年份,分别可视化以“信息检索”问题类知识单元为中心的星型构型结构,整体上表现出下降趋势,与图5(c)中的数值计算结果基本一致(见图6)。其中,以“1”结尾表示方法类知识单元,以“2”结尾表示问题类知识单元。

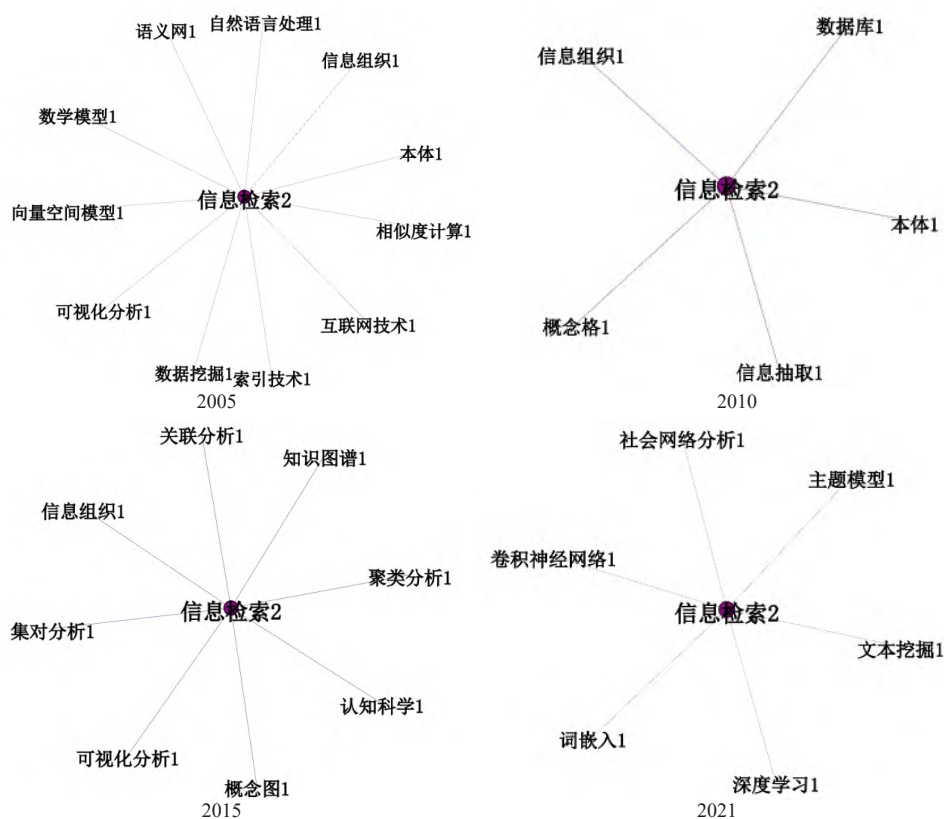


图6 “信息检索”问题关联方法的星型结构演化

Figure 6 Evolution of the star configuration of the linking method for the “information retrieval” issue

#### 4.2 “问题—方法”类网络节点属性分析

领域知识创新网络由“问题—方法”两类知识单元节点构成,领域知识自身属性也是揭示网络形成与演化的重要影响因素<sup>[25]</sup>。知识扩散过程中涉及的外在因素一定程度上对领域知识创新网络演化产生影响。以期刊为载体形式,领域知识在期刊中的分布变化一定程度上可以反映出知识创新网络的形成特征与演化规律。刊载领域知识的期刊数量、领域内学科

种类以及期刊影响因子可分别用于衡量领域知识分散广度、交叉扩散程度和期刊影响力。领域期刊是科学知识传播和扩散的正式载体形式,常被用于测度研究主题或者科学知识的发展潜力和重要性程度<sup>[26]</sup>。研究发现科技文献后期影响力与所发表期刊影响因子之间成正相关关系<sup>[27]</sup>。笔者从期刊维度计算了领域知识创新网络中问题类和方法类知识节点属性值的分布情况,如图7所示:

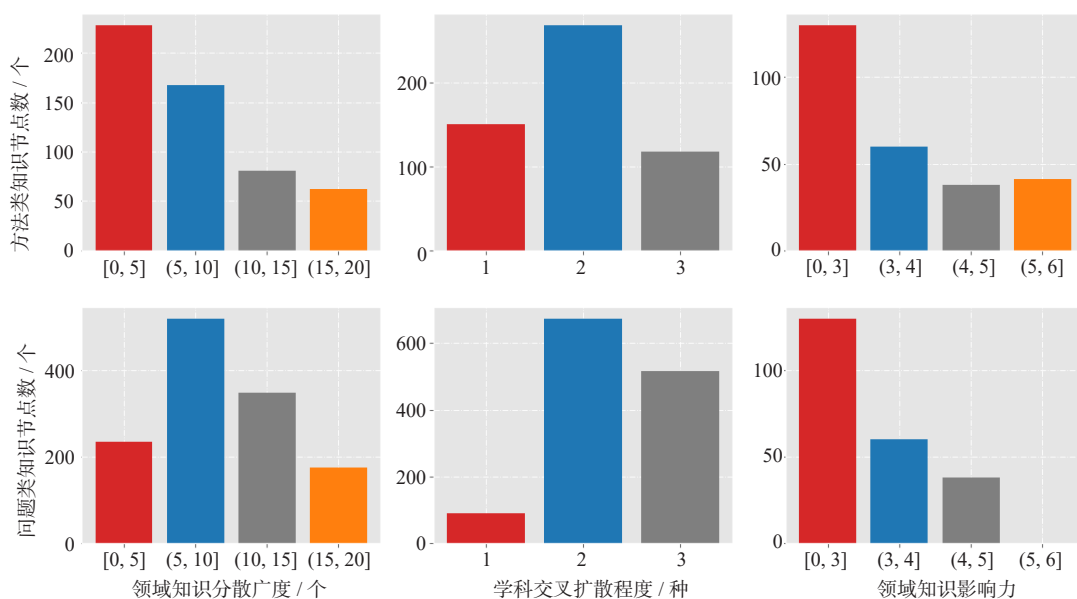


图 7 “问题—方法”类知识节点期刊属性值分布

Figure 7 Distribution of journal attribute values for the “issue-method” type of knowledge nodes

以科技文献为载体形式，领域知识在科技文献之间传播扩散的过程中，科研人员对文献的引用和下载等行为将影响知识创新网络的形成与演化。问题类和方法类知识单元除了以期刊影响因子来度量其影响

力外，还可以以论文被引量和下载量来量化其影响力。根据论文的引用和下载频次，以科技文献为纽带计算问题类和方法类知识单元的平均被引和下载频次，分布如图 8 所示：

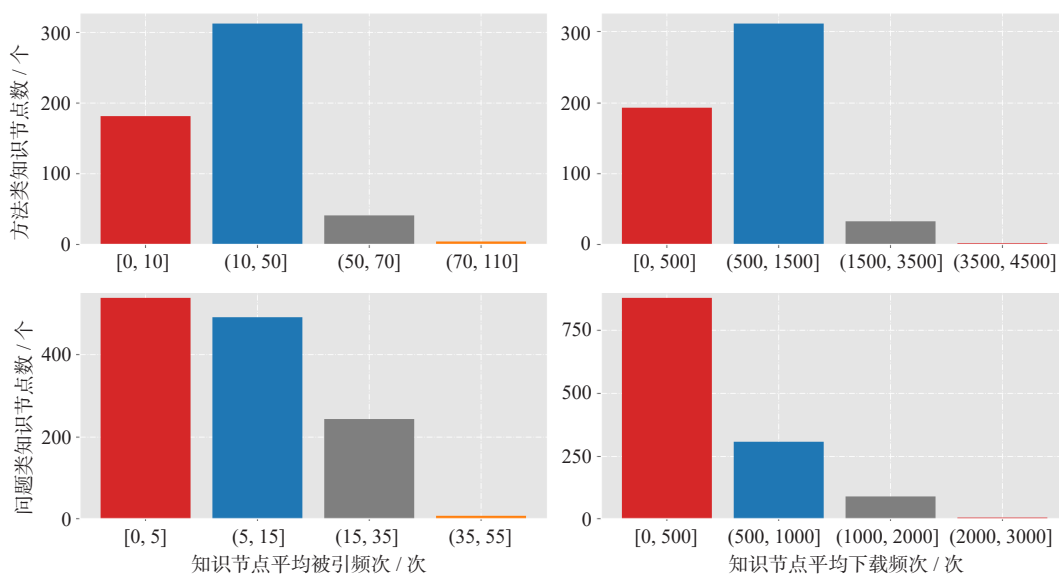


图 8 “问题—方法”类知识节点被引和下载频次属性值分布

Figure 8 Distribution of citation and download frequency attribute values for “issue-method” type of knowledge nodes

由图 8 可知，“问题—方法”类网络节点被引和下载频次属性值分布不均衡，其中方法类知识单元被引和下载频次呈现先升后降的“凸”分布，问题类知识单元被引和下载频次呈现逐级下降的“幂”分布。方法类知识单元多聚集在中度被引层次，问题类知识单元多聚集在低度被引层次。由此可见，问题类和方

法类知识单元的被引频次与下载频次分布不均衡且存在差异，可能对创新知识网络形成与演化产生影响。为了进一步揭示被引和下载频次不平衡分布对创新知识网络形成与演化的影响机制，笔者采用文献[28]将不均衡分布分别划分为高被引/下载、中被引/下载和低被引/下载 3 种类别，其分布如图 9 所示：



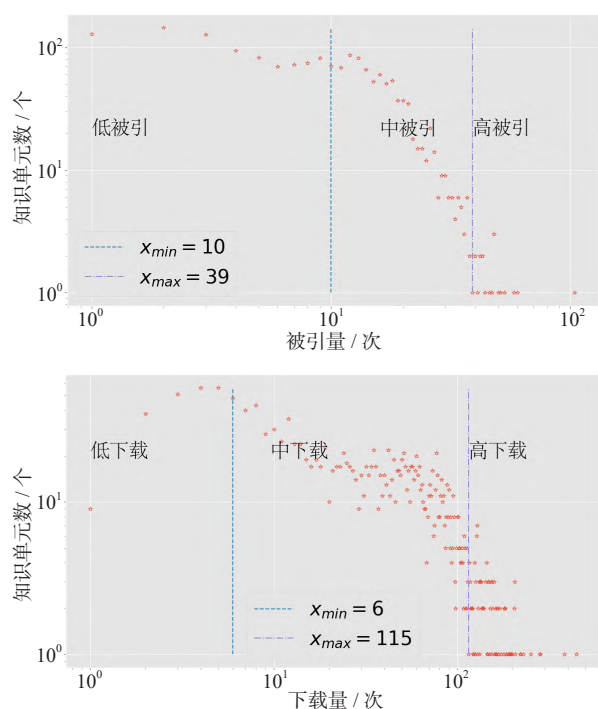


图9 领域知识被引和下载频次不均衡分类  
Figure 9 Uneven distribution of citation and download frequency of domain knowledge

由图9可知,创新知识网络中节点被引频次和下载频次可划分为高、中、低3类,从分布数量来看,主要集中于中和低两类,其中下载频次通过除以10进行了量级处理。

## 5 领域知识创新网络连边形成影响因素分析 / Analysis of factors influencing edge formation of domain knowledge innovation network

### 5.1 指数随机图模型构建

领域知识创新网络是由“问题—方法”对连接而成的二分网络,仅“问题”和“方法”类节点之间存在连边。“问题”与“方法”之间连边的形成一定程度上反应了科学知识的创新。领域知识创新网络如同其他网络一样,其形成与演化受领域知识自身属性过程和网络自组织过程的影响。

指数随机图模型(exponential random graph model, ERGM)<sup>[29]</sup>是一种能够从网络自组织构型和科学知识节点属性解释网络形成的内在影响因素的网络分析方法,可以检验出对网络形成起到促进作用的格局,这将有助于理解领域知识创新网络的形成与演化。从数学原理上看,指数随机图是一种不同于标准回归的方法,而是类似于逻辑回归的一种预测网络中任意两节点间形成连边概率的统计方法,其一般数学

表示形式如下所示:

$$P_r(Y_{ij} = y_{ij}) = \frac{1}{K} \exp \left\{ \sum_A \eta_A g_A(y) \right\} \quad \text{公式(2)}$$

在公式(2)中,假设观测网络由N节点及其连边组成,每一对节点描述为 $\langle i, j \rangle$ ,  $Y_{ij}$ 代表节点连接的随机变量,  $Y_{ij}=1$ 表示两节点间存在连边,否则表示不存在连边。 $y_{ij}$ 是 $Y_{ij}$ 集合中的一个观测值, $y$ 是 $Y$ 的观测矩阵, $Y$ 即可有向也可无向。 $A$ 表示构型, $\eta_A$ 是 $A$ 的参数, $g_A(y)$ 是 $A$ 的统计量, $K$ 对应概率分布的归一化因子,确保随机网络的概率是在0到1之间。

(1) 领域知识自身属性过程。领域知识自身属性过程中的主效应和同质效应是用于测度节点属性对于“问题”——“方法”连接形成的影响。领域期刊是科学知识传播和扩散的主要载体形式,期刊影响力一定程度上反应了其刊载知识的重要性程度<sup>[26]</sup>。研究发现科技文献后期影响力与所发期刊影响因子之间存在正相关关系<sup>[27]</sup>。结合4.2节的分析,从期刊载体粒度,刊载领域知识的期刊数量、领域内学科种类以及期刊影响因子可用于测量领域知识分散广度、交叉扩散程度和期刊影响力。从科技文献载体粒度,被引频次和下载频次可用于测度领域知识节点属性的主效应,被引频次与下载频次类别(高/中/低)可用于揭示相同属性类别对“问题”——“方法”连边形成的影响。

(2) 网络自组织过程。网络自组织过程中的网络内嵌构型对“问题—方法”连接的形成产生影响。从整体结构视角,领域知识创新网络是一个无向二部图网络结构,其最小组成是由4环构成。领域知识的网络中心度和中介中心性对网络连边形成产生影响<sup>[27]</sup>。几何加权度(geometrically weighted degree distribution, GWD)用于考察观测网络中不断递减的度分布特征,其是以每个中心度值所对应的频数乘以一个加权参数,然后求和得到。中介中心性用于测度领域知识节点是否处于其他节点间的最短路径上,起到连接不同知识模块的作用,不过本文构建的二部图领域知识创新网络中处于中介位置上的知识节点较少,仅占84.08%,因此中介中心性缺乏统计效应。

笔者从领域知识自身属性过程和网络自组织过程两个方面探究领域知识创新网络连边形成的影响因素,根据指数随机图模型(ERGM)原理,可用于计量的网络统计项如表4所示:



表 4 影响领域知识创新网络连边形成的变量总结  
Table 4 Summary on variables influencing the edge formation of domain knowledge innovation network

类型	变量名称	作用机制	ERGM 函数
网络结构变量	网络节点度分布	“问题”“方法”对之间的度分布对网络连边形成产生影响	gwdegree(0.1,T)
领域知识属性变量	领域知识分散广度	领域知识分散广度对网络连边形成产生影响	nodecov()
	学科交叉扩散程度	领域知识的学科交叉扩散程度对网络连边形成产生影响	nodecov()
	领域知识期刊影响力	领域知识期刊影响力对网络连边形成产生影响	nodecov()
	领域知识被下载频次	领域知识被下载频次对网络连边形成产生影响	nodecov()
	领域知识被引用频次	领域知识被引用频次对网络连边形成产生影响	nodecov()
	被下载强度等级	被下载强度等级相同的领域知识之间更有可能连接在一起	nodematch()
	被引用强度等级	被引用强度等级相同的领域知识之间更有可能连接在一起	nodematch()

5.2 模型拟合及参数分析

指数随机图模型的拟合优度一般通过统计摘要表中的 *AIC* (*Akaike Information Criterion*) 和 *BIC* (*Bayesian Information Criterion*) 值进行评估, *AIC* 和 *BIC* 的值越低, 说明仿真网络的拟合优度越好。此外, 统计摘要表中的 *p* 值也能够对观测网络与仿真网络的测度结果是否具有相同分布进行评价。

根据表 4 中的变量, 笔者运用 *R* 语言 *statnet* 包中的 *ergm(formula = g ~ edge)* 命令构建领域知识创新网络的指数随机图模型。具体过程为: 以零模型为基线, 先后加入网络结构变量和领域知识属性变量, 观察 *AIC* 和 *BIC* 值的变化以及 *p* 值的大小, 进而确定领域知识创新网络连边形成的影响因素。笔者通过实验获取领域知识创新网络指数随机图模型的统计摘要表, 如表 5 所示:

表 5 领域知识创新网络指数随机图模型的统计摘要  
Table 5 Summary on the exponential random graph model of the domain knowledge innovation network index

变量	零模型	结构效应模型	主效应模型	综合模型
边	-4.915 1*** (0.000)	-5.407 4*** (0.000)	-8.974*** (0.000)	-9.206*** (0.000)
网络节点度分布		0.153 6*** (0.000)	—	1.599*** (0.000)
领域知识分散广度		—	0.214 9*** (0.000)	0.220 4*** (0.000)
学科交叉扩散程度		—	-0.287 7*** (0.000)	-0.294 1*** (0.000)
领域知识期刊影响力		—	-0.117 5*** (0.000)	-0.104 0*** (0.000)
领域知识下载频次		—	0.000 5*** (0.000)	0.000 5*** (0.000)
领域知识被引频次		—	-0.006 1** (0.002)	-0.006 2** (0.002)
下载强度等级		—	0.037 0 (0.299)	0.036 4 (0.330)
被引强度等级		—	0.898 8*** (0.000)	0.897 6*** (0.000)
<i>AIC</i>	99 508	92 544	83 638	83 553
<i>BIC</i>	99 520	92 568	83 734	83 661

注: \*\*\* 表示  $p<0.001$ , \*\* 表示  $p<0.01$ , \* 表示  $p<0.05$

由表 5 可知, 零模型边数统计项系数小于 0, 表示领域知识创新网络的密度小于 0.5, 属于稀疏网络。零模型的构建是为复杂模型提供拟合优度评价的基准。赤池信息准则 (*AIC*) 和贝叶斯信息标准 (*BIC*) 是两种评价模型拟合效果较好的方法, 值越低拟合优度越好。由此, 综合模型拟合效果好于结构效应模型和主效应模型, 其中主效应模型拟合效果优于结构效应模型。

5.3 领域知识创新网络连边形成影响因素分析

根据 5.2 节参数分析可知, 综合模型拟合效果最好, 最为接近领域知识创新网络的观测网络。网络节点的几何加权重度分布在结构效应模型和综合模型中系数均为正, 表明度数越大的知识节点之间更易形成连接关系。

从期刊载体角度来看, 分散广度越大的“问题”与“方法”类知识节点之间倾向于形成连边, 侧面反映了由于认知局限, 学者们倾向于选择被普遍关注的“问题”与“方法”类知识, 进而寻求创新点。学科交叉扩散程度变量系数在主效应模型和综合效应模型中均为负, 学科内的“问题”与“方法”更易形成连边, 说明了信息资源管理学科群中的图书馆学、情报学和档案学之间一定程度上存在着学科边界。领域知识期刊影响力变量在主效应模型和综合效应模型中均为负, 期刊影响力较低的“问题”与“方法”类知识节点之间更易形成连边, 意味着出现在影响力较低期刊上的“问题”与“方法”更易得到学者们的关注, 该发现与 F. Peset 等<sup>[30]</sup>发现的“出现于低影响力期刊的关键词生存时间较长”相符。

从文献载体角度来看, 在择优连接机制上, 下载频次和被引频次对“问题”与“方法”类知识节点之间连边形成存在相反的影响。被引频次越低、下载频次越高的“问题”与“方法”之间更倾向于形成连边。在同质性机制上, 下载强度越高的“问题”与“方法”之间更易形成连边, 被引强度变量对“问

题—方法”连边形成影响不显著。由此可见，高下载量对“问题—方法”连边形成产生较为稳定的影响，反映出学者们对下载的领域知识进行吸收后，形成学术研究新思路。

## 6 总结 /Conclusion

由“问题—方法”对构成的领域知识创新网络不是随机网络，而是以关系为中心的二部图复杂网络，受网络节点属性特征、网络全局结构特征等多种内生性和外生性因素的影响。本文运用以关系为中心的统计建模方法（*ERGM*），以节点度分布、领域知识分散广度、学科交叉扩散程度、领域知识期刊影响力、领域知识下载频次和领域知识被引频次等变量为输入参数构建模型，通过计算拟合优度值的变化，分析网络连边形成与演化机制。本文分析发现学科交叉扩散程度变量系数在主效应模型和综合效应模型中均为负，表明学科内的“问题”类与“方法”类知识单元间形成连边的可能性较大。期刊影响力较低的“问题”与“方法”类知识单元间有形成连边的倾向，意味着出现在影响力相对较低期刊上的“问题”与“方法”易得到学者们的关注。被引频次越低或下载频次越高的“问题”与“方法”之间更倾向于形成连边，意味着高下载量对“问题—方法”连边的形成有较为稳定的影响。

本文探究了领域知识网络中“问题”与“方法”类知识单元间连边形成的影响因素，进而从“问题—方法”关联视角揭示知识创新机制。研究结论可用于预测领域潜在创新点，为科研人员选题、科研机构政策制定提供参考。未来，可将“问题”与“方法”进一步细分为新问题、老问题、新方法、老方法，进而以知识网络为依托探究领域知识创新的潜在模式。本文仍存在一些不足之处：仅从知识传播载体视角构建网络知识节点属性特征变量，未来可拓展到作者、机构等其他外在特征；实验数据仅选取了信息资源管理学科，未来可将本文研究方案拓展到其他学科领域，进一步验证实验结果。

### 参考文献 /References:

- [1] UZZI B, MUKHERJEE S, STRINGER M, et al. Atypical combinations and scientific impact[J]. *Science*, 2013, 342(6157): 468-472.
- [2] 钱佳佳, 罗卓然, 陆伟. 基于问题-方法组合的科技论文新颖性度量与创新类型识别[J]. *图书情报工作*, 2021, 65(14): 82-89. (QIAN J J, LUO Z R, LU W. Novelty measurement and innovation type identification of scientific literature based on question-method combination[J]. *Library and information service*, 2021, 65(14): 82-89.)
- [3] 程齐凯, 李信, 陆伟. 领域无关学术文献词汇功能标准化数据集构建及分析[J]. *情报科学*, 2019, 37(7): 41-47. (CHENG Q K, LI X, LU W. Construction and analysis of standard data set for domain-independent term function in academic texts[J]. *Information science*, 2019, 37(7): 41-47.)
- [4] HEFFERNAN K, TEUFEL S. Identifying problems and solutions in scientific text[J]. *Scientometrics*, 2018, 116(2): 1367-1382.
- [5] 陆伟, 李鹏程, 张国标, 等. 学术文本词汇功能识别——基于 BERT 向量化表示的关键词自动分类研究[J]. *情报学报*, 2020, 39(12): 1320-1329. (LU W, LI P C, ZHANG G B, et al. Recognition of lexical functions in academic texts: automatic classification of keywords based on BERT vectorization[J]. *Journal of the China Society for Scientific and Technical Information*, 2020, 39(12): 1320-1329.)
- [6] 张国标, 李鹏程, 陆伟, 等. 多特征融合的关键词语义功能识别研究[J]. *图书情报工作*, 2021, 65(9): 89-96. (ZHANG G B, LI P C, LU W, et al. Research on keyword semantic function recognition based on multi-feature fusion[J]. *Library and information service*, 2021, 65(9): 89-96.)
- [7] 程齐凯, 李鹏程, 张国标, 等. 学术文本词汇功能识别——基于标题生成策略和注意力机制的问题方法抽取[J]. *情报学报*, 2021, 40(1): 43-52. (CHENG Q K, LI P C, ZHANG G B, et al. Recognition of lexical functions in academic texts: problem method extraction based on title generation strategy and attention mechanism [J]. *Journal of the China Society for Scientific and Technical Information*, 2021, 40(1): 43-52.)
- [8] 张颖怡, 章成志, 周毅, 等. 基于 ChatGPT 的多视角学术论文实体识别: 性能测评与可用性研究[J]. *数据分析与知识发现*, 2023, 7(9): 12-24. (ZHANG Y Y, ZHANG C Z, ZHOU Y, et al. ChatGPT-Based scientific paper entity recognition: performance measurement and availability research[J]. *Data analysis and knowledge discovery*, 2023, 7(9): 12-24.)
- [9] 丁玉飞, 关鹏. 知识进化视角下科学文献传播网络演化与预测研究及应用[J]. *图书情报工作*, 2018, 62(4): 72-80. (DING Y F, GUAN P. Study and application of translation and prediction of the scientific literature communication network from the perspective of knowledge evolution[J]. *Library and information service*, 2018, 62(4): 72-80.)
- [10] 操玉杰, 毛进, 潘荣清, 等. 学科交叉研究的演化阶段特征分

- 析——以医学信息学为例[J]. 数据分析与知识发现, 2019, 3(5): 107-116. (CAO Y J, MAO J, PAN R Q, et al. Analyzing characteristics of interdisciplinary research evolutions: case study of medical informatics[J]. Data analysis and knowledge discovery, 2019, 3(5): 107-116.)
- [11] BEHROUZI S, SARMOOR Z S, HAJSADEGHI K, et al. Predicting scientific research trends based on link prediction in keyword networks[J]. Journal of informetrics, 2020, 14(4): 101079.
- [12] MINA A, RAMLOGAN R, TAMPUBOLON G, et al. Mapping evolutionary trajectories: applications to the growth and transformation of medical knowledge[J]. Research policy, 2007, 36(5): 789-806.
- [13] LIANG Z, LIU F, MAO J, et al. A knowledge representation model for studying knowledge creation, usage, and evolution[C]// International conference on information. Berlin: Springer, Cham, 2021: 97-111.
- [14] ZHANG X, XIE Q, SONG C, et al. Mining the evolutionary process of knowledge through multiple relationships between keywords[J]. Scientometrics, 2022, 127(4): 2023-2053.
- [15] 杨冠灿, 陈亮, 张静, 等. 专利引用关系形成的解释框架: 一个指数随机图模型视角[J]. 图书情报工作, 2019, 63(5): 100-109. (YANG G C, CHENG L, ZHANG J, et al. Framework for explanations of patent citation formation: an exponential random graph model perspective[J]. Library and information service, 2019, 63(5): 100-109.)
- [16] 段庆锋, 马丹丹. 基于指数随机图模型的专利技术扩散机制实证研究[J]. 科技进步与对策, 2018, 35(22): 23-29. (DUAN Q F, MA D D. Empirical research of patent technology diffusion mechanism based on ERGM[J]. Science & technology progress and policy, 2018, 35(22): 23-29.)
- [17] 操玉杰, 李纲, 毛进, 等. 基于 ERGM 的学科交叉领域知识连接机制实证研究[J]. 图书情报工作, 2019, 63(19): 128-135. (CAO Y J, LI G, MAO J, et al. An empirical study on knowledge connection mechanism of interdisciplinary field based on ERGM[J]. Library and information service, 2019, 63(19): 128-135.)
- [18] 宫雪, 崔雷. 基于医学主题词共现网络的链接预测研究[J]. 情报杂志, 2018, 37(1): 66-71, 52. (GONG X, CUI L. Link prediction in *MeSH* terms co-occurring networks[J]. Journal of information, 2018, 37(1): 66-71, 52.)
- [19] 彭陶, 王建冬, 孙慧明. 基于关键词共现网络的我国国情领域近三十年学科发展脉络分析[J]. 大学图书馆学报, 2012, 30(2): 29-34. (PENG T, WANG J D, SUN H M. An analysis on development vein of Chinese library and information science during the past 30 years based on keywords co-occurrence network[J]. Journal of academic libraries, 2012, 30(2): 29-34.)
- [20] 王晓光. 科学知识网络的形成与演化 (II): 共词网络可视化与增长动力学[J]. 情报学报, 2010(2): 314-322. (WANG X G. Formation and evolution of science knowledge network(II): co-word network visualization and growth dynamics[J]. Journal of the China Society for Scientific and Technical Information, 2010(2): 314-322.)
- [21] 程齐凯. 学术文本的词汇功能识别[D]. 武汉: 武汉大学, 2015. (CHENG Q K. Recognition of lexical functions in academic texts [D]. Wuhan: Wuhan University, 2015.)
- [22] 陆伟, 李鹏程, 张国标, 等. 学术文本词汇功能识别——基于 BERT 向量化表示的关键词自动分类研究[J]. 情报学报, 2020, 39(12): 1320-1329. (LU W, LI P C, ZHANG G B, et al. Recognition of lexical functions in academic texts: automatic classification of keywords based on BERT vectorization[J]. Journal of the China Society for Scientific and Technical Information, 2020, 39(12): 1320-1329.)
- [23] ECK N J, WALTMAN L. How to normalize co-occurrence data? An analysis of some well-known similarity measures[J]. Journal of the American Society for Information Science and Technology, 2009, 60(8): 1635-1651.
- [24] HUNTER D R, HANDCOCK M S, BUTTS C T, et al. Ergm: a package to fit, simulate and diagnose exponential-family models for networks[J]. Journal of statistical software, 2008, 24(3): 1-29.
- [25] 杨冠灿, 陈亮, 张静, 等. 专利引用关系形成的解释框架: 一个指数随机图模型视角[J]. 图书情报工作, 2019, 63(5): 100-109. (YANG G C, CHENG L, ZHANG J, et al. Framework for explanations of patent citation formation: an exponential random graph model perspective[J]. Library and information service, 2019, 63(5): 100-109.)
- [26] MOED H F. Measuring contextual citation impact of scientific journals[J]. Journal of informetrics, 2010, 4(3): 265-277.
- [27] DINESH K S. Ranking of arts and humanities journals published in India: a scientometric analysis[J]. Pearl: a journal of library and information science, 2017, 11(2): 155-158.
- [28] HUANG Y, BU Y, DING Y, et al. Partitioning highly, medium and lowly cited publications[J]. Journal of information science, 2021, 47(5): 609-614.
- [29] KRIVITSKY P N, HUNTER D R, MORRIS M, et al. Ergm 4: new features for analyzing exponential-family random graph



- models[J]. Journal of statistical software, 2023, 105(6): 1-44.
- [30] PESET F, GARZÓN-FARINÓS F, GONZÁLEZ L M, et al. Survival analysis of author keywords: an application to the library and information sciences area[J]. Journal of the Association for Information Science and Technology, 2020, 71(4): 462-473.

#### 作者贡献说明 /Author contributions:

杨金庆: 论文选题、数据分析及初稿撰写;  
庞业佳: 数据采集及分析;  
刘智锋: 文献调研及相关研究梳理;  
李鹏程: 数据分析;  
李洁: 论文修改。

### Research on the Evolution Mechanism of Domain Knowledge Innovation Network from the Perspective of “Issue-Method” Correlation: Taking the Discipline Group of Information Resources Management as an Example\*

Yang Jinqing<sup>1</sup> Pang Yeja<sup>1</sup> Liu Zhifeng<sup>2</sup> Li Pengcheng<sup>3</sup> Li Jie<sup>4</sup>

<sup>1</sup>School of Information Management, Central China Normal University, Wuhan 430079

<sup>2</sup>Department of Information Management, Peking University, Beijing 100871

<sup>3</sup>College of Economics and Management, Hubei University of Technology, Wuhan 430068

<sup>4</sup>College of Social Sciences, Soochow University, Suzhou 215000

**Abstract:** [Purpose/Significance] From the correlation of “issue-method”, this study explores the evolution mechanism of domain knowledge innovation networks under the combined innovation mode, reveals the potential influencing factors of scientific knowledge innovation, in order to assist in science and technology policy formulation and timely open up new way of innovation and development. [Method/Process] This paper focused on the discipline group of information resource management, identified the knowledge units of “issues” and “methods” in it, and then constructed a domain knowledge innovation network using “Issues-Method” association as a combination unit. Finally, it constructed the model on the relationship formation and relationship patterns with the relationship centered statistical modeling method, Exponential Random Graph Model, and analyzed the impact of various endogenous and exogenous factors, such as network node attribute and network structure, on the formation and evolution of domain knowledge innovation networks. [Result/Conclusion] It shows that the possibility of “issue-method” correlation is relatively large in journals with large dispersion breadth. Within disciplines, the tendency of “issue-method” correlation is relatively high. “Issues-method” correlation in journals with relatively low impact are more easily to form edges. In literatures with high downloads or relatively low citation frequency, it is more likely that the edge between “issues” and “method” knowledge units is formed.

**Keywords:** knowledge network evolution mechanism technological innovation combinatorial innovation

\*This work is supported by the National Key Research and Development Program of China titled “Theory, Method and Expert Prediction System for Disruptive Technology Identification” (Grant No. 2019YFA0707200), and the National Natural Science Foundation of China titled “Research on Role Change Prediction of Scientific Knowledge Based on Explainable Machine Learning” (Grant No. 72304108), and the Jiangsu Provincial Social Science Foundation for Youths titled “Research on Library Digital Resource Cognition Recommendation under the Normalization of epidemic situation” (Grant No. 21TQC001).

**Author(s):** Yang Jinqing, special appointment associate professor, PhD, E-mail: jinq\_yang@163.com; Pang Yeja, undergraduate; Liu Zhifeng, doctoral candidate; Li Pengcheng, lecturer, PhD; Li Jie, lecturer, PhD.

Received: 2023-09-26 Revised: 2024-01-16 Pages: 97-108