

文章编号:0559-9350(2019)11-1400-17

## 基于文献计量学的国内水生态环境研究知识图谱构建与应用

毛文山<sup>1, 2, 3, 4</sup>, 赵红莉<sup>4</sup>, 蒋云钟<sup>4</sup>, 段浩<sup>4</sup>, 郝震<sup>4, 5</sup>

(1. 兰州交通大学 测绘与地理信息学院, 甘肃 兰州 730070;

2. 地理国情监测技术应用国家地方联合工程研究中心, 甘肃 兰州 730070;

3. 甘肃省地理国情监测工程实验室, 甘肃 兰州 730070; 4. 中国水利水电科学研究院 水资源研究所, 北京 100038;

5. 大连理工大学 建设与工程学部, 辽宁 大连 116024)

**摘要:**近年来,水生态环境问题已愈发引起学者和社会的广泛关注。本文选取1992—2017年CNKI数据库中文核心期刊中收录的水生态环境相关文献作为样本数据。采用文献计量学中定性分析与定量研究的信息分析方法,基于知识图谱的结构性和时间性指标,以CiteSpace、SPSS信息分析软件为支撑工具,对数据进行主题词共词矩阵分析、关键词共现聚类分析和名词短语突变检测分析,绘制了展现国内水生态环境研究领域知识结构、知识演化的各类知识图谱。研究表明:①本文构建的“文献时序分析&学科分析(宏观)→主题词共词矩阵分析(微观、直观)→关键词共现聚类分析(复杂、整体)&典型聚类分析(局部)→名词短语突变检测分析(特殊)”行业文献数据分析方法是合理有效的;②国内水生态环境研究进展符合“Shneider四阶段理论”,研究主题演进路径明确,以“水环境保护研究→生态需水研究→面向生态环境的应用研究”为主,研究热点具有丰富的进程特征,各时段前沿研究脉络清晰,体现为:“流域生态水文规律研究→水生态环境管理技术研究→流域生态环境改善、恢复研究”;③通过主题词共词矩阵分析识别中文文献研究主题,弥补了中文文献数据因参考文献缺失无法通过文献共被引分析进行主题识别的不足。

**关键词:**水生态环境;知识图谱;社会网络分析;多维尺度分析;CiteSpace

**中图分类号:** P95

**文献标识码:** A

**doi:** 10.13243/j.cnki.slxb.20190422

## 1 研究背景

随着经济社会的高速发展,水生态环境问题成为21世纪全世界面临的新的重大挑战,党的十八大明确提出推进水生态文明建设,加强污染治理,持续改善生态环境质量。《水利改革发展“十三五”规划》开展全国105个水生态文明城市建设试点工作,以加快推进水生态文明建设。水生态环境成为政府和学术界关注的焦点<sup>[1]</sup>。对水生态环境领域内的研究主题、研究热点、研究前沿进行梳理,以知识图谱的方式展现,可为水生态文明领域的研究与建设提供参考。

以往关于科学知识图谱的分析大都建立在对大量文献量化分析的基础上,但在数据采集和分析内容的选择上存在较强的主观性<sup>[2]</sup>。自2006年以来,在科学计量学、数据可视化、知识图谱构建与应用等一系列领域发展背景下,以CiteSpace、VOSviewer等为主流的信息可视化软件,实现某研究领域内热点主题、知识演进等的可视化,为学者后续研究提供了知识储备和创新转向,在众多研究领域得到了广泛应用。目前利用科学知识图谱进行计量分析的领域主要有金融证券、生物医药、图书情报等<sup>[3]</sup>,水生态环境领域研究起步较晚且数量少。Xiong等<sup>[4]</sup>以定量研究代替定性分析的方法对文献进行关键词分析,总结出不同时期的水政策影响下,水文化变迁的阶段性特点、水利开发阶段

收稿日期:2019-06-14

基金项目:中国工程科技知识中心建设项目-水利专业知识服务系统(CKCEST-2019-1-6);兰州交通大学优秀平台支持项目(201806)

作者简介:毛文山(1993-),硕士生,主要从事知识图谱、数据挖掘研究。E-mail: 1098748344@qq.com

及水文极端事件的协同性特征；胡秀芳等<sup>[5]</sup>以定性和定量分析相结合的方法绘制出生态安全研究知识图谱，梳理出研究主题、经历发展阶段、影响生态安全研究的阶段和未来研究方向；许振亮等<sup>[6]</sup>以关键词共现网络的视角，采用社会网络分析方法，揭示了国际生态城市研究前沿的知识结构变化；林卓等<sup>[7]</sup>通过关键词共现分析出生态资产/价值研究领域的热点与发展趋势，明确了生态资产/价值的研究界限，生态资产变化的内在机制是生态资产研究未来趋势的切入点；伍新木等<sup>[8]</sup>利用 CiteSpace 中探测词频突发增长的突现词功能，绘制出国内水资源管理研究领域前沿知识图谱，分析出各研究阶段及未来一段时间的关注热点。这些研究成果对水生态环境研究知识图谱的构建与应用提供了案例参考，但仍需做进一步调整：①生成表征不同研究内容的知识图谱，需选取合适的分析对象；②中文文献研究主题的识别需选取合适的分析对象和方法，避免以关键词共现聚类分析或英文文献共被引分析方式代替。

国内关于水生态环境研究知识图谱分析较为少见，以及不同分析方法在分析对象选择上存在一定的局限性。本文采用定性分析和定量研究相结合的信息分析方法，针对不同研究视角的知识单元，选择合适的分析方法，选取主题词、关键词和名词短语作为分析对象，选取共词矩阵分析、共现聚类分析和突变检测分析作为相对应的分析方法，对水生态环境领域内的权威文献数据进行分析，通过提取聚类标签和附加时间标签，绘制出研究主题、研究热点和研究前沿三类科学知识图谱，选取相对应的指标用于分析和评价，客观揭示主题漂移规律、热点聚类分布和前沿研究趋势，为该领域继续深入相关研究的科研人员提供知识参考和新的研究思路。

2 数据来源与知识图谱构建方法

**2.1 文献数据来源** 本文选择中国知识基础设施工程(简称“CNKI”)权威数据库作为数据来源，以 CNKI 检索采集的文献数据作为水生态环境研究知识图谱构建的实验数据。考虑到国内数据库存在检索方式和数据格式上的差异性，本文采用六元组检索模型在 CNKI 上进行数据采集检索，六元组模型包含检索式、检索字段、检索模式、时间跨度、文献类型、数据来源六方面的检索设定。利用六元组模型构建本文的检索策略(表 1)。

国内水生态环境研究是在水环境和水生态领域探索的基础上发展而来，包括：水质评价、生态水文、生态需水、流域生态、淡水生态等研究内容，由此确定中文文献数据的检索式为：TS<sub>1</sub>="水环境保护 or 水生态保护与修复 or 生态需水 or 流域生态 or 水生态环境管理 or 流域生态环境改善 or 生态水文"，TS<sub>2</sub>="水污染 or 水资源保护 or 海绵城市 or 河流生态 or 河口近岸海域生态 or 湖沼湿地生态 or 坡面生态 or 水土保持"，根据 TS<sub>1</sub> 对 CNKI 进行主题(Topic)检索，可保证查全率，根据 TS<sub>2</sub> 对 CNKI 进行“篇名(Title) or 关键词(Keywords) or 摘要( Abstract) ”检索，可保证查准率，对检索条目进行检查，去除报告、新闻、会议通知等不相关条目。

**2.2 知识图谱构建方法** 知识图谱(Knowledge Graph)以特定知识领域为对象，显示知识发展进程与结构关系的一种图形，具有“图”和“谱”双重性质与特征：既是可视化的知识图形，又是序列化的知识谱系，本质上是具有属性的实体通过关系链接，由“点-边”组成的大规模网络图，其中，节点表示概念/实体，边表示概念/实体间的语义关系<sup>[9]</sup>。本文知识图谱是水生态环境知识工程的形式化表达，以知识网络形态展现该领域的类别、实体、属性或关系关联，显示该研究领域中各知识单元或知识群之间网络、结构、交叉、演化或衍生等诸多复杂的关系，实现事实型知识和过程型知识的陈述，形成对水生态环境知识的专业性认识。核心是建立水生态环境知识库，然后通过引文分析、聚类分析生成特定的知识语义网，最终具

表1 中文文献检索六元组模型

标识	名称	CNKI 检索说明
Se	检索式	“TS <sub>1</sub> ”U“TS <sub>2</sub> ”
Sf	检索字段	主题、篇名+关键词+摘要
Sm	检索模式	精确
Ts	时间跨度	1992—2017
Dt	文献类型	论文、会议论文、综述
Ds	数据来源	EI 期刊、核心期刊、CSCD、CSSCI

备对主题词、关键词和名词短语等文献关键要素进行大规模实时关联和特征分析的能力<sup>[10]</sup>。其中，“节点”表示水生态环境类别、主题及对应的属性值，“边”表示的关系包括：“类别-类别”、“类别-主题”、“主题-主题”、“主题-属性”和“属性-属性”。

水生态环境知识图谱的构建分为数据模式层定义与数据层构建两部分。数据模式层定义研究领域的知识类别、主题(实体)、属性或关系关联等，本文在对数据模式层的定义上，采用CNKI标引的文献字段规则来识别文献中的类别和主题(实体)、关系和属性等<sup>[11]</sup>，数据模式层对文献字段识别规则的定义如表2所示；数据层建设包括：知识获取、知识融合、知识存储、知识推理、知识更新5个阶段(图1)。

表2 数据模式层定义的文献字段识别规则

标识	说明	标识	说明	标识	说明
PT	出版物类型	AB	文献摘要	OPY	出版时间
JAU	文献作者	C1	文献来源机构	VL	卷
TI	文献标题	CR	参考文献	PG	文献页数
SO	文献来源期刊/出版物名称	NR	引用的参考文献数	UT	入藏号
DE	文献关键词	OTC	文献被引次数	ER	记录结束

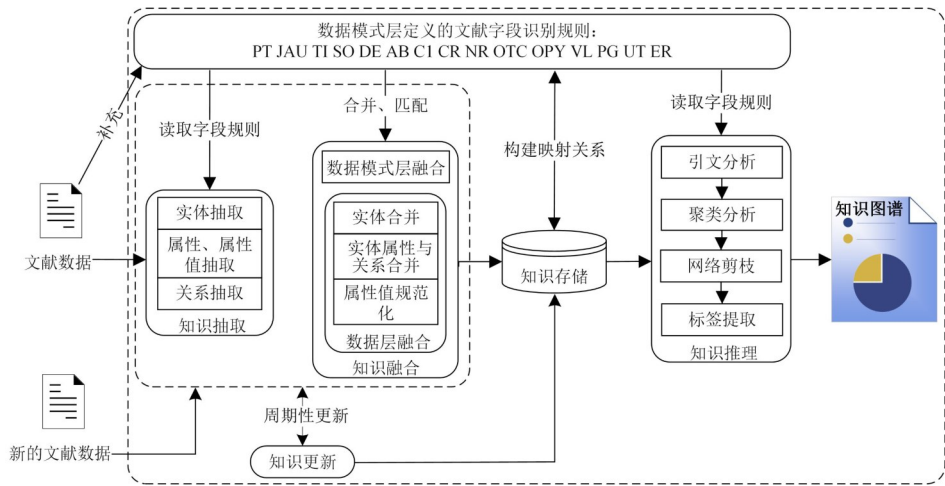


图1 水生态环境研究知识图谱数据层构建流程

(1)知识获取。从采集的文献中抽取知识，包括实体、属性和属性值、关系，首先，实体抽取根据数据模式层定义的文献字段识别规则识别命名实体(分布在标题、关键词、摘要中的主题词、关键词和名词短语)<sup>[12]</sup>；其次，属性和属性值抽取是为每个研究类别构造属性列表，并为研究类别中的实体附加属性值(各命名实体的出现时间、分布时段、出现频次等)<sup>[13]</sup>；最后，关系抽取是施引文献和被引文献之间引文关系和相似性关系的抽取<sup>[14]</sup>，引文关系包括共现关系和共引关系(图2)，相似性关系包括字符相似、属性相似和结构相似。

(2)知识融合。知识融合包括数据模式层、数据层的融合<sup>[15]</sup>，数据模式层融合是根据数据模式层定义的文献字段识别规则建立数据映射时，通过设置合并规则来确保数据的统一；数据层融合包括实体合并、实体属性与关系的合并、属性值的规范化<sup>[16]</sup>，其中，实体合并是将具有歧义的实体指称项链接到文献字段识别规则中的符号项，实现实体消歧，实体属性与关系的合并是更新具有时态特性的属性，将再次出现的实体并入该实体首次出现的时段属性中。

(3)知识存储。融合后的知识通过建立图数据库中概念和文献字段识别规则的映射关系实现存储和查询，实体查询及重要性排序是通过计算实体属性与关系的相关度来实现<sup>[17]</sup>。



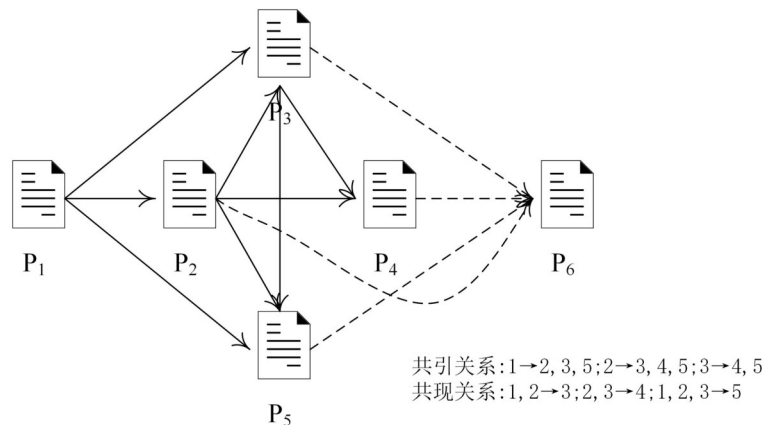


图2 引文关系网络

(4)知识推理。首先，基于文献字段识别规则进行引文分析和相似性分析，再对分析结果进行聚类分析，采用面向引文网络链接关系的聚类，避免由节点属性聚类引起的局部最优缺陷，形成关系网络，提取隐含的知识和关系<sup>[18]</sup>；其次，对聚类结果进行网络剪枝，寻径网络算法(Pathfinder)对每一时段的网络结构和整体网络结构进行剪枝，获得最优网络结构，最后，使用LLR(Log-Likelihood Ratio)算法提取优化后的网络结构标签，得到的聚类标签会有效减少重复项<sup>[3]</sup>。

(5)知识更新。随时间变化，水生态环境知识会发生不断变化，对新发表的相关文献再次进行知识抽取和知识融合，完成领域中知识类别、主题(实体)、属性或关系关联的周期性更新，最后将其存入水生态环境知识库。

**2.3 研究方法和分析指标选取** 本文采用定性分析与定量研究相结合的信息分析方法，对经六元组模型检索整理后的文献数据进行知识图谱构建和分析。选取发文量、学科、主题词、关键词和名词短语作为研究对象，进行时序分布、学科分布、研究主题、研究热点、研究前沿等五个方面的特征分析，形成对水生态环境研究领域发展变化的初步认知：①时序分析。对发文量按时序进行统计分析，根据不同时段发文量的变化趋势，分析研究重心的转移特征，探究1992—2017年不同时段内水生态环境研究的发展变化及其影响因素；②学科分析。对水生态环境研究的学科分布进行统计分析，揭示该领域权威文献的研究层次、学科分布现状和研究主题在学科中的受关注程度；③研究主题分析。借助共被引形成的主题相似性思想，建立主题词两两共被引矩阵，聚类分析生成的主题词谱系图反映主题之间的亲疏关系，多维尺度分析生成的多维尺度图反映主题知识结构特征和核心主题内容，进一步揭示不同主题词相互影响下的时段分布特征，加上时间标签，探究不同时段以主题词为主要表征方式的主题漂移特点<sup>[19-20]</sup>；④研究热点分析。关键词共现聚类分析生成研究热点图谱，通过对整体网络、聚类子网络的横向和纵向分析，探究不同时段内由各研究热点聚类生成的知识群组；⑤研究前沿分析。研究领域内的最新演变趋势和相互影响，通过Kleinberg突变检测算法的研究前沿分析，探究不同时段下的研究前沿及未来研究方向。本文提出面向水生态环境研究领域的知识分析框架(图3)，考虑领域专业知识的所有表征方式，主要通过词间关系的数据挖掘进行知识发现，为将来更深入的信息挖掘提供一些借鉴。

从发文量探究国内水生态环境研究的时序分布特征及同国际间的区别，从学科分类分布揭示国内水生态环境研究的科学领域结构，完成该研究领域的宏观分析。其次，从微观视角出发，主题词共词矩阵分析确定该研究领域的主题知识结构，通过多维尺度分析，进一步揭示研究主题之间的亲疏关系和研究主题所处的发展阶段；关键词共现聚类分析则从复杂、整体上按聚类的时段分布和聚类之间的关联程度，展现研究领域的热点知识结构，每个时段局部的典型聚类分析，得到其研究热点。最后，研究前沿分析是从特殊视角出发，以时区视图的方式反映各个子领域发展演进的时间跨度和下阶段的研究方向。

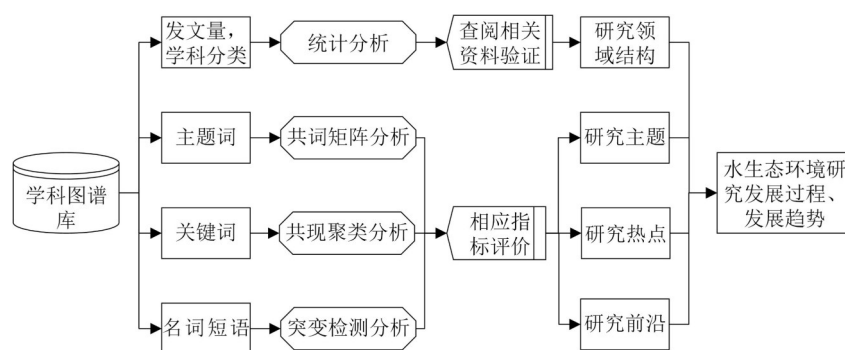


图3 水生态环境知识量化分析流程

表3 知识图谱分析指标说明

分析内容	分析对象	对象分析指标选取	图谱评价指标选取
研究主题	主题词	被引词频	载荷系数、贡献率、方差累计贡献率(因子分析) 相关系数、熵值、基于类的 $F$ 值(R型聚类分析) 拟合度指标:Stress、RSQ(多维尺度分析)
研究热点	关键词	共现词频、出现时间、分布时段	整体网络结构指标:节点数、连线数、密度 聚类指标: $Q$ 值、 $S$ 值 单节点指标:突现度、中心性、结构洞
研究前沿	名词短语	词频、频次变化率、出现时间、分布时段	信息熵、中心性、突现度

本文选取文献量、学科分类作为水生态环境研究统计分析的指标。以研究主题、研究热点、研究前沿分析为例,进行水生态环境研究知识图谱的应用研究。选取主题词、关键词和名词短语作为知识图谱分析的具体对象,探索研究主题、研究热点和研究前沿。相关分析评价指标选取说明(表3):

(1)研究主题。主题词是文献核心内容的高度概括,反映某研究领域的研究方向。确定受控的、被统一标引的主题词作为共词矩阵分析的基本单元,其在检索平台规范化处理之后在检索性能方面优于关键词<sup>[5,20]</sup>。选取主题词的被引词频作为研究主题的分析指标,被引词频记录一组主题词被同一篇或同一组文献引证的次数,以测度之间的依赖关系。研究主题知识图谱的评价指标选取,主题词R型因子分析的载荷系数反映研究主题的显著程度,因子得分(贡献率、方差累计贡献率 $\geq 60\%$ )确定研究主题分类。相关系数度量主题词之间的相似性,各研究主题分支聚类时,合并两个相关系数最大的类。熵值和基于类的 $F$ 值用于评价、验证单个或整个聚类结果<sup>[21]</sup>。拟合度量值Stress确定维度数,值越小,说明模型适合度越高,效度估计值RSQ越大,说明分布在战略坐标中各主题词间的距离与实际输入距离(被引词频)越合适,Stress和RSQ确定水生态环境研究领域内的核心研究主题和研究主题整体网络结构。

其中,主题词提取通过源文献共被引检索手段,对1992—2017年国内水生态环境文献数据中的主题词归纳整理,获得961个,利用《水利水电科技主题词表》、《水利水电工程技术术语》(SL 26—2012)等中出现的科技主题词,对语料库中的文献进行分词,分词结果数据清洗后,借助TF-IDF计算方法从文献中提取反映不同时段知识群组研究特色的高频主题词,时段高频主题词的提取结果为:

$$TF-IDF(X) \quad (1)$$

其中:

$$IDF(x) = \log \frac{N}{N(x)} \quad (2a)$$

$$IDF(X) = \log \frac{N+1}{N(x)+1} + 1 \quad (2b)$$

$$TF-IDF(X) = TF(x) \cdot IDF(X) \quad (3)$$

式中： $TF(x)$ 为主题词在某时段中的词频； $IDF(x)$ 为主题词在全时段中的词频， $IDF(X)$ 为平滑处理后主题词在全时段中的词频； $N(x)$ 为原始主题词表中某主题词的时段词频， $N$ 为整个时段的主题词总数。

为防止新主题词( $N(x)=0$ )的出现，常用的 $IDF$ 通过式(2b)进行平滑处理。该矩阵为邻近矩阵，对角线值取该主题词与其它主题词共被引频次最高值+1表示，为测度多个研究主题之间的相似性，且保证共被引矩阵中0存在条件下的稳定性，原矩阵系统聚类方法为：组间连接法，将共被引矩阵转化为Pearson's相关系数矩阵，并进行标准化处理。假设主题词的聚类结果为 $C=\{C_1, C_2, \dots, C_m\}(1 \leq i < m)$ ， $m$ 是样本个数，人工判定主题结构为 $P=\{P_1, P_2, \dots, P_n\}(1 \leq j \leq n)$ ， $n$ 是数据分析指标，则可计算 $C$ 中每一个聚类 $C_i$ 的熵值：

$$Entropy = \frac{1}{\log(n)} \sum_{j=1}^n \frac{|C_i \cap P_j|}{|C_i|} \log \left( \frac{|C_i|}{|C_i \cap P_j|} \right) \quad (4)$$

对每一个聚类 $C_i$ 存在： $F(C_i) = \max_{j=1, 2, \dots, n} \{F(P_j, C_i)\}$ ，表示在人工判定主题范围内循环一次后聚类 $C_i$ 的最大值，计算最终 $F$ 值：

$$F = \frac{\sum_{i=1}^m (|C_i| * F(|C_i|))}{\sum_{i=1}^m |C_i|} \quad (5)$$

(2)研究热点。关键词是文章的核心议题，关键词共现(Keyword Co-appearance Analysis, KCA)网络揭示某一领域当前及过去产生的热点研究<sup>[22]</sup>。选取关键词的共现词频、出现年份和分布时段作为研究热点的分析指标，共现词频记录一组关键词在同组文献中出现的次数，以测度之间的耦合关系。研究热点知识图谱的评价指标包括：①图谱整体网络结构的评价指标：节点数、连线数和密度，用来描述网络链接程度；②图谱聚类指标：聚类模块值(Modularity,  $Q$ 值)， $Q > 0.3$ 认为图谱聚类结构明显，聚类平均轮廓值(Silhouette,  $S$ 值)， $S > 0.5$ 认为聚类合理， $S > 0.7$ 认为聚类具有较高的可信度；③图谱节点指标：结构洞、中心性和突现度，突现度是一定时间段内关键词的频次变化率较高，预示着研究热点的转变<sup>[23]</sup>，借助社会网络分析(Social Network Analysis, SNA)中不同个体构成整体网络的结构及相互关系概念，构建的“结构洞+中心性”分析体系可研究知识实体的关系链接及结构<sup>[24]</sup>。

其中，结构洞描述与个体关键词相关联的其他关键词之间不存在非冗余关系，借助网络结构中某节点与其它节点的联系紧密程度而体现该主体在整体网络中的丰富性，本文结构洞分析以限制度指标为准(式(6))。中心性用来测度高频关键词是否处于核心地位，以体现节点在整体网络中的地位重要性，中心性分析则参照中间中心度指标(式(7))。中间中心度的数值越大，结构洞值也最小，表明关键词在共现网络中既处于“核心”位置，又具有信息沟通的“桥接”作用<sup>[25-26]</sup>。

$$C_i = \sum_j \left( P_{ij} + \sum_q P_{iq} P_{qj} \right)^2 \quad (6)$$

$$C_B(n_i) = \sum_j \sum_k g_{jk}(i) / g_{jk}, j \neq k \neq i, j < k \quad (7)$$

式中： $i$ 、 $j$ 、 $q$ 为节点， $P$ 为节点间连接关系的权重比例， $C$ 为所求节点的限制度指标， $g$ 为节点之间存在的捷径数， $C(n)$ 为所求节点的中介中心度指标。

(3)研究前沿。名词短语反映某研究领域未来研究热点问题的新动向，为研究者提供该学科领域的最新演化动态<sup>[27]</sup>。选取名词短语的词频、频次变化率、出现年份和分布时段作为研究前沿的分



析指标, 频次变化率较高的名词短语可反映当前时期领域中的关键研究。研究前沿知识图谱的评价指标包括: 信息熵、中心性和突现度, 信息熵是对研究前沿中知识变化不确定性的一种度量, 根据整个时期累积的词汇量进行回顾性计算, 预示着领域前沿研究总体格局改变的必要性, 若分析指标的信息熵越小, 表明指标值变异程度越大, 指标所提供的信息量越多, 相应权重越大, 在综合评价中该分析指标所起作用越大。

### 3 水生态环境研究的结果分析

#### 3.1 水生态环境研究统计分析

3.1.1 文献量时序分布特征 为比较国内水生态环境研究发展同国际间的差距, 本文另统计了中文社会科学引文索引数据库(简称“CSSCI”)和WOS(Web of Science)中水生态环境研究主题的发文量。图4是体现文献量时序分布特点和变化规律的时序分布图。从水生态环境研究领域文献发布的年代来看, 国内进展符合“Shneider 四阶段理论”<sup>[28]</sup>, WOS与CNKI、CSSCI在发文量及变化趋势上存在明显差异, 前者处于稳步提升阶段, CNKI波动较大, 在2014年停止增长, 之后处于波动状态; CSSCI则一直处于持平阶段, 表明在水生态环境研究领域中国内与国际研究进展存在较大差距, 国内该研究领域地位的提升未来需更深入的研究积累, 其中的科技类研究发展进程丰富, 而人文社科类研究达到饱和状态。通过指数曲线回归模型对累计百分比进行偏差分析,  $y = 0.0116e^{0.1856x} (R^2 = 0.9538)$  表明曲线与数据点拟合效果较好, 表明该领域总体发文量呈现指数式增长规律, 说明水生态环境研究逐渐受到国内众多学者的关注和重视。

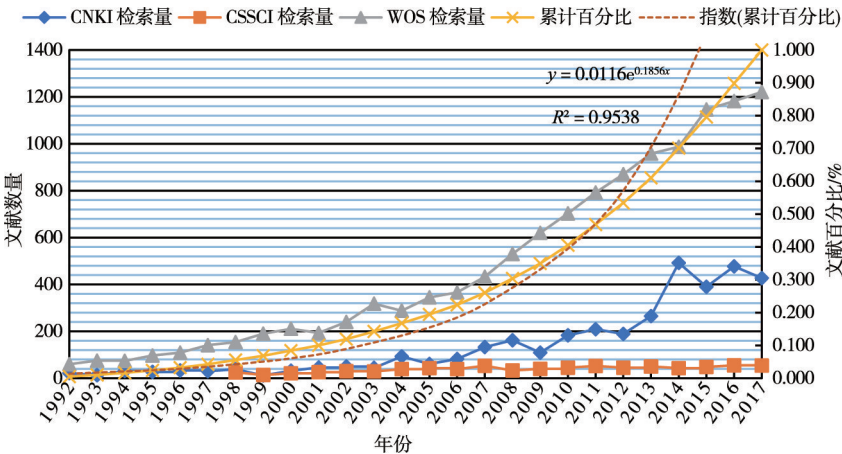


图4 水生态环境研究领域中外文献时序分布

国内水生态环境研究由1970年代末水资源研究扩展至1980年代初水环境保护研究演变而来, 依据图2该研究领域发展历程大致分为3个阶段: ①1992—2004年是水生态环境研究的萌芽起步阶段, 国内学者在传统水资源研究的基础上, 逐步开展水环境保护方面的研究, 主要是全国城市饮用水水源地安全保障规划提出了调查评价的理论方法和技术体系, 解决了水源地保护中存在的监控及标识等若干关键技术问题<sup>[29]</sup>。该阶段文献发表较少但研究范围不断扩大, 其理论体系不断完善, 相关研究逐步开展, 已引起相关领域研究学者的关注; ②2004—2012年处于曲折增长阶段, 是我国生态需水研究、生态水文研究、典型地区流域水循环过程中水文水资源的生态效应研究的高峰期。为缓解我国干旱西北地区水资源供需矛盾, 开展了以生态需水计算方法为主的生态需水及生态水文理论及技术研究<sup>[30]</sup>。1980年代至1990年代初水土保持的减水减沙效应研究为水土保持的水资源和水环境效应研究(2005—2007)提供了理论支撑, 形成了面向水生态与环境多任务需求的水土保持规划理论及技术体系, 自此国内生态需水研究及探索流域生态水文规律为主的基础研究已基本成型<sup>[31-32]</sup>。2009年CNKI趋势线出现波动, 国内水生态环境研究以支撑流域水资源调配与管理的生态环境需求为前

提, 转向流域生态环境相关的计算及工程技术研究; ③2012—2017年该领域呈现迅速繁荣态势, 该阶段主要面向水生态环境保护 and 修复的理论及技术应用研究。开展了针对不同地貌类型区域的植被恢复、修复技术及产生的水文效应研究<sup>[33]</sup>; 关注点转向河流生态健康评价理论、水生态环境保护与修复等方面, 促进了生态水利的提出(2012年)和发展, 面向最严格水资源管理制度下的水资源论证技术得以完善<sup>[34]</sup>。2015年《水污染防治行动计划》发布, 旨在解决我国面临的水生态环境问题, 也是CNKI趋势线从2014年开始呈波浪式变化的主要原因, 说明当前水生态环境研究需扩展研究思路, 提高创新力度。党的十八大提出水生态文明建设, 水生态环境研究领域从支撑水生态环境管理为目标的应用技术研究转向促进流域生态环境改善为前提的工程实践研究, 国内水生态环境研究进入全新阶段。研究领域呈现出以生态水利相关理论与实践研究为基础, 以国内水生态环境问题和国家关于水生态环境恢复与建设的战略方针为导向的特征<sup>[35-36]</sup>。

3.1.2 学科分类分布特征 对国内水生态环境研究文献统计分类, 得到其学科分类分布图(图5)。左侧是其施引文献所在的学科分布, 主要学科: 环境科学与资源利用(28.13%)、生态(15.30%)、水利水电工程(13.59%)、城乡规划与市政(12.08%)、资源科学(9.87%)、农业经济(7.01%)、建筑科学与工程(4.37%)、农业资源与环境(2.90%)等, 表明水生态环境研究的领域应用。右侧是对应被引文献所在的学科分布, 其引用的主要学科具体为环境科学、水资源研究、自然资源学、生态学、生物学(水生物学)等, 表明水生态环境研究的研究基础, 水生态环境是多学科共同关注的研究。当前水生态环境研究的期刊来源中共30个研究子学科左右, 主要集中于地理环境(9.39%)、流域生态(8.18%)、水环境保护(7.15%)、生态补偿(6.83%)、生态环境(6.23%)、水资源(5.89%)、水生态文明建设(3.92%)、可持续性发展(3.65%)、水生态(3.37%)、生态补偿机制(2.48%)等, 同时, 该领域研究具有理论基础研究向应用技术和工程实践转型的特征, 呈现多学科交叉。

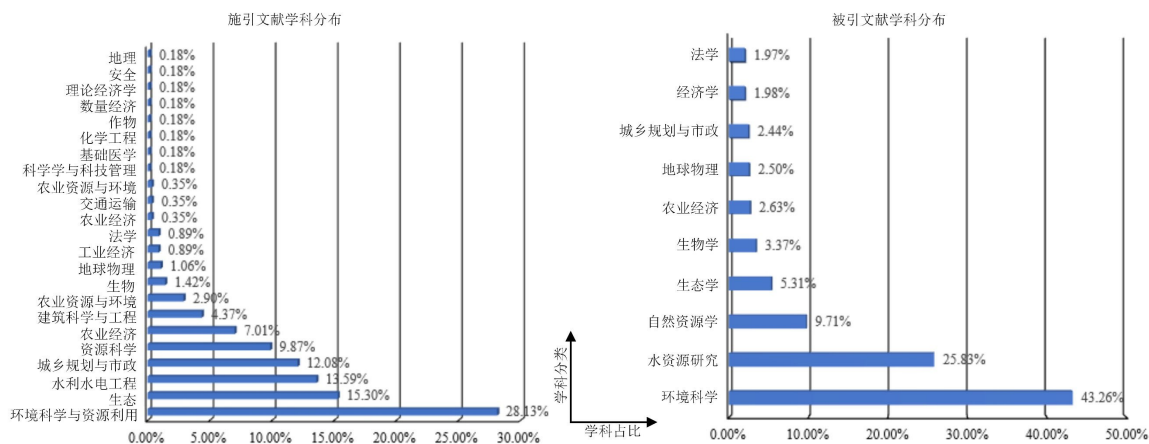


图5 水生态环境研究文献学科分类分布

3.2 基于主题词共词矩阵分析的研究主题识别 通过主题词提取方法获得49个时段高频主题词作为分析对象(表4), 通过这些主题词之间的被引词频构建49×49共被引矩阵, 即因子模型, 因子分析确定研究分支数后, 转化为Pearson's相关系数矩阵, 作为系统聚类分析、多维尺度分析的基础, 综上, 绘制水生态环境研究领域研究主题的知识图谱。

因子分析是在少损失信息或尽可能不损失信息的情况下, 将多个变量指标降维为少数几个因子的多元统计分析方法。主题词共词矩阵(49×49)下的R型因子分析, 提取到12个因子, 每个主题载荷系数的大小表示该主题在各个因子中的载荷程度, 值越高表示主题研究在水生态环境领域的显著程度越强。15个主题词在单个因子中具有较高的负载值(得分≥±0.60以上), 特别是“指标体系”(因子1)和“生态需水”(因子3)各自连接的分支5和分支3研究在该领域有显著体现(得分≥±0.70以上), 12个主题词在组间连接下的多个研究分支中具有显著体现(得分≥±0.40以上), 如流域治理、生态补偿机制、水利水电工程等。其中, 累积贡献率约78.067%的总方差存在12个潜在因子, 前7个主因子累计



表4 1992—2017年国内水生态环境研究时段高频主题词

主题词	频次	分布时段	主题词	频次	分布时段	主题词	频次	分布时段
流域治理	622	1998—2017	水资源	455	2005—2017	生态修复	185	2001—2017
流域综合治理	381	2003—2017	水环境	237	2008—2017	水生态安全	28	2008—2017
综合治理	433	1998—2017	水生态系统	165	2001—2017	生态补偿机制	205	2007—2017
小流域治理	247	2006—2017	水污染	273	1997—2017	水生态修复	43	2001—2017
小流域	398	2002—2017	流域水污染	241	2002—2017	石羊河流域	246	2003—2017
太湖流域	278	2005—2017	生态环境建设	168	1997—2017	流域水环境	180	2007—2017
小流域综合治理	221	2002—2017	黑河流域	206	2001—2017	最严格水资源管理制度	54	2011—2017
长江流域	338	2001—2017	水污染防治	222	2001—2017	水生态文明城市	128	2012—2017
淮河流域	230	2000—2017	水利工程	57	1997—2017	生态补偿	485	2008—2017
流域生态	760	2008—2017	水利管理	64	2002—2017	水生态文明建设	232	2013—2017
生态环境	602	2004—2017	水资源管理	50	2001—2017	水生态文明	190	2014—2017
地理环境	970	2004—2017	水利水电工程	43	2004—2017	海绵城市	42	2011—2017
水生态	277	1998—2017	可持续性发展	323	2006—2017	水生态功能区划	212	2007—2017
水环境保护	713	2001—2017	指标体系	62	2005—2017	水生态环境	201	1992—2017
环境质量	206	2006—2017	需水量	45	2001—2017	水利风景区	39	2006—2017
水资源污染	234	2001—2017	水功能区	55	2001—2017			
海河流域	206	2001—2017	流域水资源	231	1997—2017			

贡献率为62.900%且取值均大于1，表示其所代表的主题是水生态环境研究领域的主要研究方向，贡献率依次为：18.888%、10.975%、8.180%、7.780%、6.621%、5.935%、4.521%，说明主题词共被引矩阵降维至12，能表达整体信息的78.067%。借助因子分析结果，结合聚类分析中的R型聚类方法进行共引聚类分析，得到主题词聚类分析谱系图(图6)，纵轴为各研究主题词，反映研究内容亲疏关系的主题词相互连接，谱系图更揭示了水生态环境领域中各研究分支之间的内在关联及归属层次，每个研究分支由不同亲属关系的研究主题聚合而成。计算Pearson’s相关系数的聚类评价指标的平均得分，熵值为0.3917，基于类的F值为0.7145，熵值较小，基于类的F较大，说明聚类结果较为合理；同时，与因子分析结果相比有较高的一致性，图中虚线为各主题词聚类最优解。综合因子得分 $\geq \pm 0.40$ 的12个潜在因子，根据聚类结果可解释水生态环境领域中存在12个研究分支：分支1，水资源保护科学的理论方法及技术研究；分支2，面向流域(区域)的水功能区划研究；分支3，不同典型区域下的生态需水计算模型及生态水文研究；分支4，水生态保护和修复的理论及技术研究；分支5，评价指标体系研究；分支6，流域生态水文规律及作用机理研究；分支7，水生态环境管理理论及技术研究；分支8，流域生态环境污染治理及修复研究；分支9，生态环境保护与修复研究；分支10，水利水电工程和生态水利研究；分支11，水生态文明城市 and 海绵城市研究；分支12，水资源管理理论及技术研究。

因子分析和聚类分析从微观视角探究了主题词之间的内在关联程度，而多维尺度分析以向心度和密度为参数绘制的主题知识图谱从宏观角度探究了主题集群间的相似性<sup>[37]</sup>，图7中各主题词(点)的相关位置及距离反映了该领域各主题的集群关系，主题词间距离越近，关系越紧密，反之关系越疏远。其中，Stress值是0.12394，Stress 优劣尺度评价：12.4% $\leq$ 15%，近似程度为满意，值较小，表明观察数据与分析结果拟合较好，说明模型拟合度较高，RSQ值为0.99118，接近于1，表示图中各点构形距离与输入实际距离匹配度较高，二者反映出各主题间的相互影响强度较高，联系强度紧密。坐标横轴为向心度，表示各主题词间相互影响强度，纵轴为密度，表示水生态环境研究领域内部联系强度，I象限内各主题词在流域生态需水及生态水文研究方向上都有体现，流域生态环境、不同修复措施下的水环境效应机理等研究主题处于“次核心”地位，理论成果发展较成熟，且研究主题间存在较高的交叉度；II象限内主题领域内部联系紧密，各主题词表现于整个水生态环境领域的“核心”地位，共同构成了水生态环境管理下的应用技术研究方向，水功能区划、最严格水资源管理制度下

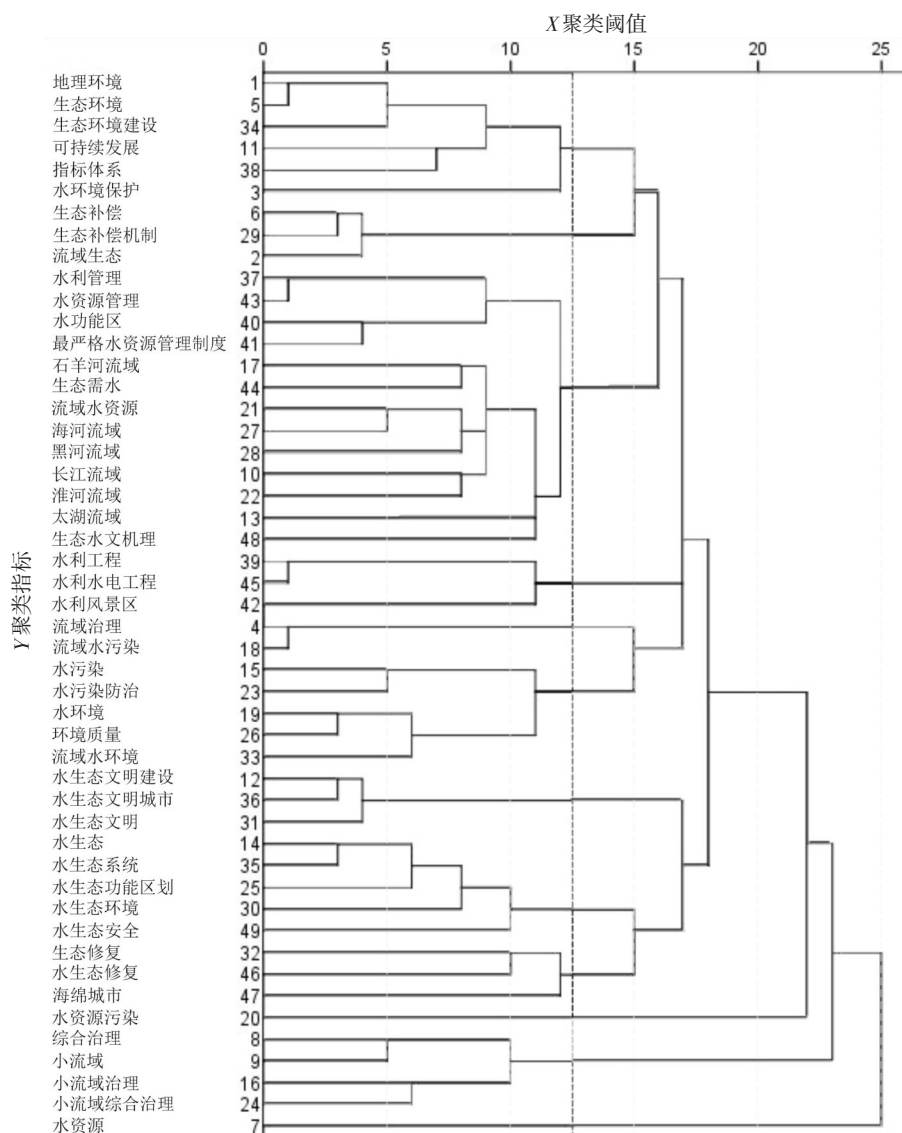


图6 主题词聚类分析谱系图

的水资源论证技术、水利工程环境影响评价、水资源保护技术等研究主题间的关联程度较高，理论及技术成果较为成熟；Ⅲ象限内主题领域结构松散，各主题词定位于生态环境应用研究方向中，水生态系统保护与修复、水资源论证、水生态文明建设等研究主题关联度较低，且发展不成熟，处于研究领域的“边缘”地带，表明上述研究具有较大的发展空间；Ⅳ象限内各主题词聚类到流域生态环境改善研究方向，水土保持、面源污染治理等研究主题下的相关理论研究成果处于“核心”地位，但以此为基础的工程实践技术仍不成熟，需不断创新。

**3.3 基于关键词共现聚类分析的研究热点辨识** 经寻径网络(Pathfinder)算法修剪的水生态环境研究热点知识图谱保留了最重要的节点关联，将致密冗余的原网络简化为表达清晰研究热点聚类结构的最有效网络，其中节点半径、颜色深浅和连线距离、密度反映了关键词的出现频率、被引年份和关键词的直接、间接联系(图8)。共现网络存在428个节点，529条连线，密度为0.0058，说明网络链接程度紧密，各主题词间共现程度较高。 $Q$ 值为0.8499，说明热点研究聚类效果明显， $S$ 值为0.6328，说明研究热点的同质性较高，呈现较集中化的特点。图8中明显的关键节点为：水环境、海绵城市、生态补偿、流域、水资源、生态环境等，即出现的高频热点词；突变强度较大的关键词有13个：生态补偿( $Burst=50.2362$ )、低影响开发( $Burst=43.8269$ )、流域( $Burst=24.944$ )、水生态文明( $Burst=22.473$ )、年径流总量控制率( $Burst=18.4514$ )、湿地( $Burst=18.3414$ )、生态补偿机制( $Burst=$

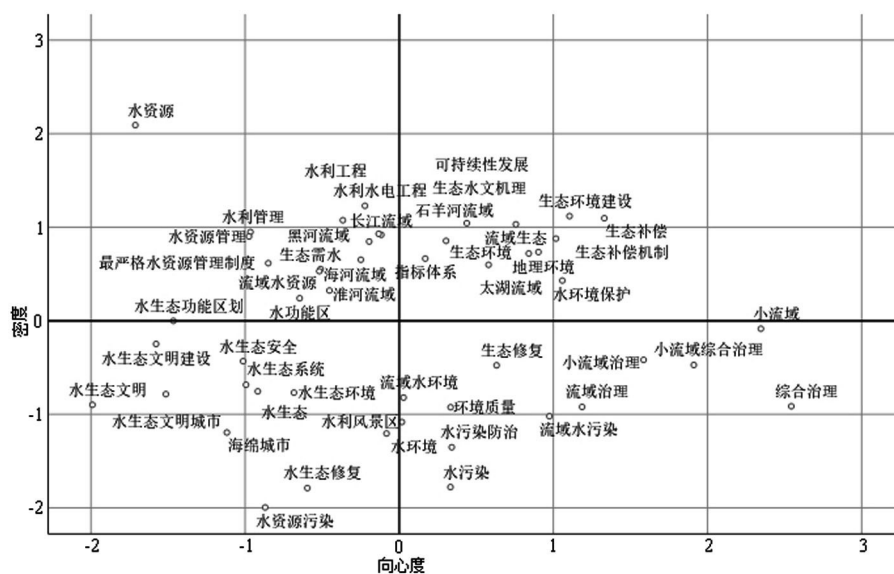


图7 研究主题知识图谱

17.7084)、生态系统服务( $Burst=16.0453$ )、生态需水( $Burst=15.0045$ )、生态补偿标准( $Burst=14.433$ )、生态恢复( $Burst=13.7246$ )、生态水文( $Burst=13.3648$ )、流域生态补偿( $Burst=13.3148$ )，这些关键词出现频次增长较快，且主要分布于2008—2017年，是该领域亟需解决也是学者关注的热点问题。此外，海绵城市、低影响开发、雨洪管理、城市内涝及水生态文明是2013年后出现的新研究热点。经“Circle Packing Of Clusters”方法生成13个子聚类，分别对应图8中13个凝聚子群，凝聚子群在水生态环境领域中体现出一定的生命周期和多元化特征，依次为：(1)水环境知识群组(1992—2017年)，在传统水资源研究基础上拓展，包括水质评价、生态需水等热点；(2)水资源污染及治理知识群组(1996—2013年)，为构建健康河流理论体系奠基，包括小流域、综合治理等热点；(3)生态系统及水土保持措施知识群组(1994—2016年)，促进流域尺度下的水土资源开发研究，包括水土流失、水源地水质等热点；(4)水资源污染治理知识群组(1996—2013年)，是生态环境研究出现的萌芽，包括：岸边污染带、对策等热点；(5)生态环境及生态经济系统知识群组(1993—2013年)，水生态环境研究为工程水利逐步定位，包括质量、可持续发展等热点；(6)流域生态及生态需水知识群组(1997—2014年)，研究成果在水资源和水工程规划中发挥了关键性作用，包括水生态系统、河流生态等热点；(7)生态补偿及指标体系知识群组(1994—2012年)，包括生态承载力、补偿标准等热点；(8)生态补偿机制及生态修复知识群组(1997—2011年)，水生态修复为水生态文明的发展提供了契机，包括土地利用、生态安全等热点；(9)生态系统服务及生态足迹知识群组(1997—2012年)，包括水生态足迹等热点；(10)海绵城市知识群组(2014—2017年)，其是城市水生态环境的范畴，包括低影响开发、年径流总量控制率等热点；(11)水生态文明知识群组(1998—2016年)，为水生态文明试点建设提供了指导，包括指标评价体系等；(12)低影响开发知识群组(2014—2016年)；包括城市内涝、雨水花园等；(13)生态水利工程知识群组(2005—2017年)，流域生态环境改善及恢复下的理论及技术实践，包括湿地、面源污染治理、流域综合规划等。由此看出，国内水生态环境领域具有较好的水环境保护、特定领域下的生态需水、面向生态环境的水生态保护与修复等理论及技术研究基础，以满足水资源管理和调配的生态环境为前提，发展流域生态环境下的计算及工程应用研究、水生态文明建设、海绵城市已成为目前研究热点，因此图谱呈现多中心交叉的枝状分布特点。

通过“结构洞+中心性”分析体系，得到1992—2017年水生态环境领域内按时序凸显的关键词表(表5)，图8中，“水资源”节点 $q$ 是“生态补偿”节点 $i$ 和“流域”节点 $j$ 的共同邻接点， $P_{ij}$ 是“生态补偿”节点 $i$ 的所有邻接点中“流域”节点 $j$ 所占的权重比例， $C_i$ 即为“生态补偿”节点 $i$ 的限制度指标；图8中，“流域”节点 $j$ 和“水资源”节点 $k$ 之间存在捷径数 $g_{jk}$ ，且两者之间存在经过“水环境”节点 $i$



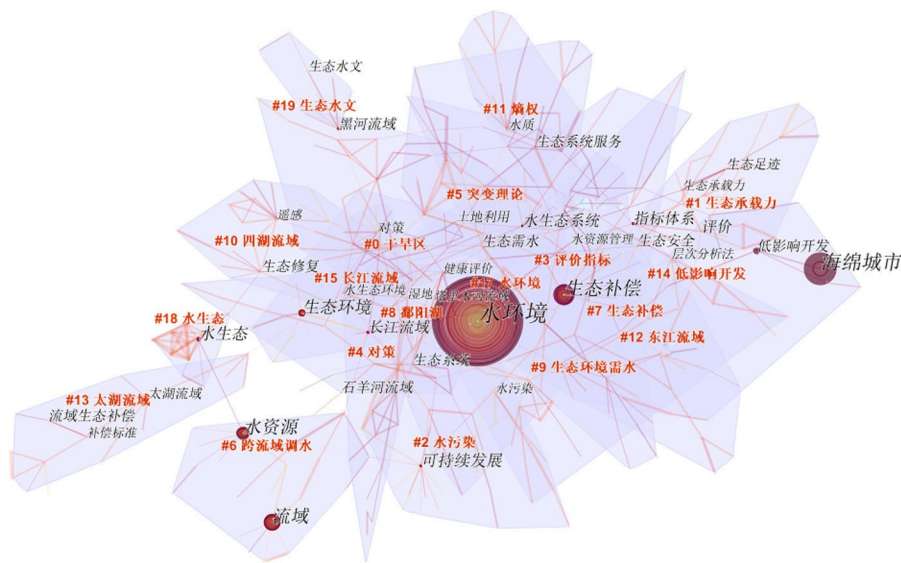


图8 研究热点知识图谱

表5 1992—2017年国内水生态环境领域研究热点关键词（频次≥100）

频次	关键词	中心性	结构洞	出现年份	频次	关键词	中心性	结构洞	出现年份
1410	水环境	0.47	0.526	1992	88	生态系统服务	0.02	0.705	2001
454	海绵城市	0.02	0.478	2015	88	GIS	0.05	0.731	1999
382	生态补偿	0.04	0.496	2001	88	对策	0.10	0.830	1993
353	流域	0.19	0.518	1997	81	流域生态补偿	0.08	0.701	2008
310	水资源	0.12	0.533	1996	80	太湖流域	0.04	0.695	2005
243	生态环境	0.17	0.581	1992	78	土地利用	0.02	0.882	1996
170	可持续发展	0.14	0.560	1997	73	水质	0.05	0.785	2000
146	水生态	0.06	0.548	1998	71	生态足迹	0.02	0.793	2006
132	水生态系统	0.08	0.551	1998	70	生态水文	0.06	0.720	1992
122	健康评价	0.06	0.612	2001	65	水污染	0.08	0.683	1999
121	长江流域	0.04	0.653	1994	59	生态承载力	0.02	0.788	2006
121	黑河流域	0.06	0.661	1998	57	层次分析法	0.02	0.801	2005
116	指标体系	0.03	0.607	1997	54	水资源管理	0.01	0.826	2009
113	石羊河流域	0.03	0.597	1992	53	塔里木河流域	0.01	0.896	1993
111	生态系统	0.06	0.632	1994	51	湿地	0.01	0.922	2014
103	低影响开发	0.01	0.599	2015	50	遥感	0.03	0.835	2001
101	生态需水	0.04	0.601	2001	49	补偿标准	0.02	0.843	2003
95	生态安全	0.04	0.678	2001	47	小流域	0.06	0.763	1996
89	生态修复	0.05	0.693	2002	45	生态文明	0.02	0.894	2004

的捷径数为  $g_{jk}(i)$ ， $g_{jk}(i)/g_{jk}$  为“水环境”节点  $i$  控制此两节点的联接能力， $C_B(n_i)$  即为中介中心度指标。图8中角标明显的节点，对应表5内容，在水生态环境领域中，水环境、海绵城市、生态补偿、水资源、流域、生态环境和水生态是学者关注的研究热点的集中体现。

**3.4 基于 Kleinberg 突变检测算法的研究前沿分析** 通过 Kleinberg 突变检测算法从标题和摘要中探测出按词频时间分布的词频和频次变化率均高的名词短语(Noun Phrases)，将追踪到的名词短语根据主题进行归纳，得到反映研究前沿趋势变化的信息熵图(图9)，对应1999年至2010年上半年，突出显示了连续和陡峭的熵值增加，不确定性的显著增加是新兴宏观性质的体现，即水生态环境领域的研究重心发生变化，预示着新研究内容的注入。图10是研究前沿的样点分布趋势图，样点位置显示在XY平面中，每个样点的感兴趣属性值通过Z轴竖线高度表示，利用全局多项式的内插方法拟合样

点趋势面，探测纵向趋势(X轴)和维度趋势(Y轴)发现，样点分布趋势与研究前沿信息熵的变化趋势相一致，样点分布趋势图中的峰谷基本定格了前沿研究的具体类型，根据首次被引用的时间，2000年时区中的文献最少，1999年、2001年时区次之，表明该领域处于低谷时期，其余时区是该领域发展的较繁荣时期。1992年、1993年时区与1994—1999年间各个时区中的节点的连线关系密集，表明1992年、1993年与1994—1999年间各个时间段的传承关系较强，与此后各个时区的传承关系较弱，2000—2017年间各个时区的传承关系主要来源于1996年、1998年、1999年3个时区，其中1998年时区对各个时区的传承关系最强，且影响至今，说明水生态环境研究前沿的时变特征。

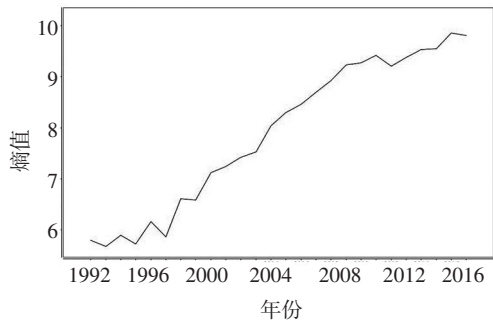


图9 研究前沿信息熵图

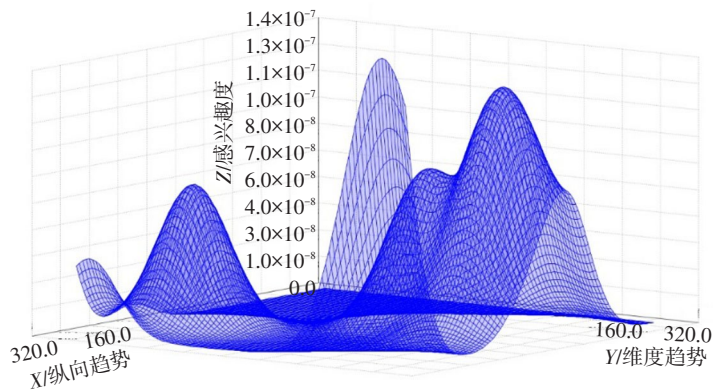


图10 研究前沿样点分布趋势图

对生成时区视图前的整体网络结构进行社会网络分析，自动聚类的网络存在475个节点，1082条连线，密度为0.0096，将名词短语作为网络节点，名词短语之间的共现关系作为节点连线，节点越大越接近中心地位，属于前沿研究，统计具有中心性和突现度的名词短语，如表6所示。

生成包含表6中名词短语的研究前沿知识图谱(图11)，反映前沿知识演进的时区视图，以时区视图的可视化方式描绘水生态环境研究在时间跨度上的过渡本质、演变趋势和相互影响。参考图9和图10，水生态环境领域研究前沿显著表现为3个研究阶段。1992—1998年，初期研究主要体现为具体案例(饮水安全问题、西北地区生态需水研究等)和发展完善水资源、水工程的理论及技术实践研究，表征该阶段研究前沿的名词短语包括：水环境、生态环境、水土流失、质量评价、水污染、水土保持、生态经济系统等，中心性为0.27的“水环境”是初期各研究分支的根节点，“生态环境”、“水土保持”、“水土流失”、“生态服务”及水库个案研究成为该领域早期受关注程度较高的前沿研究分支，呈现出较高的突现度。1999—2011年，该阶段突现出较多的名词短语，多个研究集群中呈现出中心性和突现度的集中分布，在1992—1998年研究基础上，水生态环境领域衍生出多个新的前沿研究分支，表征该活跃阶段研究前沿的名词短语包括：生态需水、生态用水、生态恢复、生态水文过程、健康风险评价、生态系统健康、生态水文学等，其中面向具体流域的生态系统研究集群、以生态补偿为核心的研究集群和面向生态环境问题的水生态保护与修复、生态系统服务研究集群拓展力度较大，表明该研究领域进入活跃期，各研究分支转向应用研究，趋向具体化。2012—2017年，水生态环境领域前沿研究的瓶颈期，海绵城市、流域尺度下水土资源合理开发及水生态文明建设等经济思想，生态水文学等新兴学科的交叉引进，既注重坡面植被、河流水系、湖沼湿地、河口近岸海域等流域水循环耦合生态系统的生态水文规律基础研究，也偏向于管理和改善水、流域生态环境的工程

表6 1992—2017年国内水生态环境领域研究前沿名词短语

名词短语	中心性	突现度	时段分布	名词短语	中心性	突现度	时段分布
水环境	0.27	124.68	1992—2005	生态系统服务	0.01	7.46	2015—2017
生态环境	0.13	41.17	1996—2007	生态环境需水量	0.12	6.69	2004—2007
低影响开发	0.06	39.59	2015—2017	健康风险评价	0.02	6.54	2006—2010
水土保持	0.09	16.69	1998—2004	水生态功能分区	0.07	6.49	2010—2012
生态补偿	0.07	16.00	2011—2014	流域生态补偿机制	0.04	6.33	2008—2013
水生态文明	0.03	14.41	2014—2017	水生态安全	0.14	6.19	2013—2017
生态需水	0.22	12.85	2002—2008	生态系统健康	0.02	5.99	2006—2010
生态用水	0.01	12.84	2004—2007	生态调度	0.02	5.62	2008—2012
生态恢复	0.22	12.71	2005—2012	景观格局	0.09	5.32	2014—2017
生态系统服务价值	0.06	11.45	2014—2016	生态健康	0.06	5.14	2014—2015
生态补偿机制	0.04	9.65	2011—2015	生态风险评价	0.03	5.10	2012—2015
水土流失	0.02	9.59	1995—2006	生态水文学	0.02	5.01	2009—2012
海绵城市建设	0.01	9.14	2015—2017	水资源保护	0.23	4.92	2013—2017
风景园林	0.01	8.63	2014—2017	水生态承载力	0.07	4.88	2014—2017
生态水文过程	0.07	8.54	2005—2012	生态安全评价	0.11	4.87	2010—2015
生态文明	0.22	8.25	2014—2016	生态服务功能	0.06	4.72	2010—2016
流域生态	0.08	7.71	2008—2014	生态敏感性	0.03	3.78	2014—2017
评价指标体系	0.03	7.51	2015—2017				

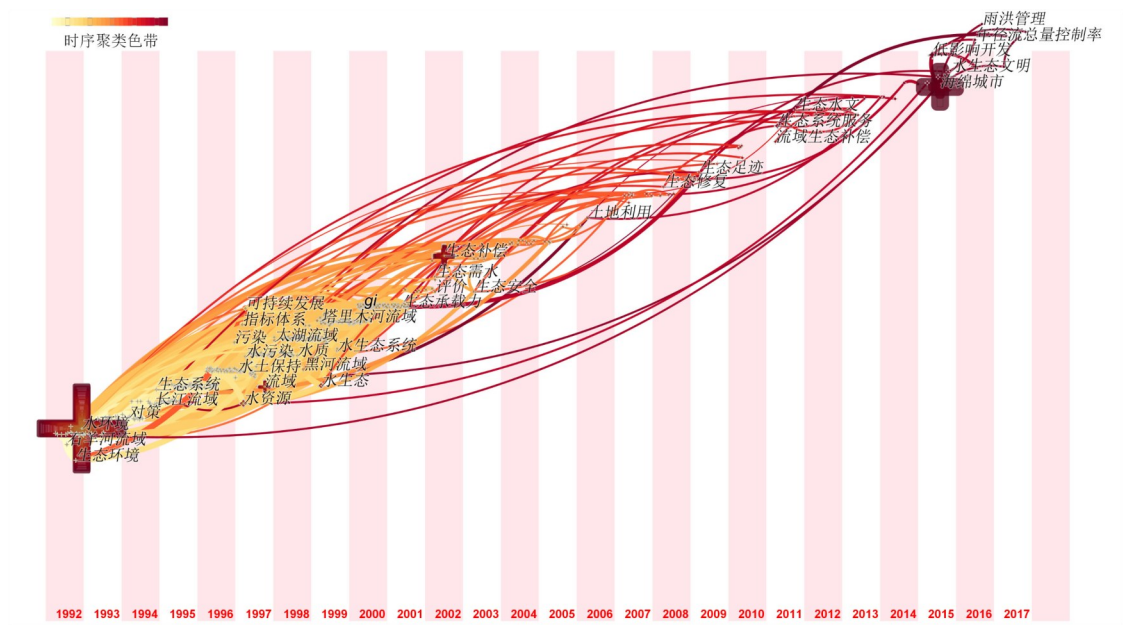


图11 研究前沿知识图谱

应用研究，表征研究前沿的名词短语包括：水生态文明、海绵城市建设、风景园林、生态系统服务、生态环境需水量、景观格局、生态健康、生态系统服务价值、评价指标体系、径流总量控制（年径流总量控制率）、水生态承载力、风景园林、生态安全等，水生态文明、低影响开发、海绵城市建设和面向最严格水资源管理的水资源论证技术具有较高的突现度和一定的中心性，成为当前和未来一段时间内的前沿研究分支。

4 结论

本文以CNKI文献检索引擎为数据来源，通过构建水生态环境研究知识图谱，进行行业知识挖



掘。以 CiteSpace、SPSS 为支撑工具,从知识图谱视角出发,采用定性分析和定量研究相结合的方法,对 1992—2017 年国内水生态环境主题文献数据进行多方位分析。研究表明,国内水生态环境研究发展符合“Shneider 四阶段理论”描述,本文构建的“文献量时序分析&学科分析(宏观)→主题词共词矩阵分析(微观、直观)→关键词共现聚类分析(复杂、整体)&典型聚类分析(局部)→名词短语突变检测分析(特殊)”信息分析方法是合理有效的。

本文以统计分析方式分析了国内水生态环境研究的发展现状,以知识图谱方式分析了国内水生态环境领域的研究主题、研究热点和研究前沿,研究发现:

(1)国内水生态环境研究发展水平同国际间存在一定差距,发展效率较低,主要受国内水生态环境问题和国家关于水生态环境政策方针的影响,近十多年来,该领域学者对水生态环境的研究兴趣增长较显著,引发了较多新的研究创新点。

(2)国内水生态环境研究在多个领域内备受关注,且在环境科学与资源利用、生态、水利水电工程、城乡规划与市政中受关注程度较高,说明水生态环境研究呈现多研究目标、多研究层面、多研究学科交叉融合的复杂特征。

(3)研究主题演进路径明确,分为三个阶段:①1980 年代初,开展针对饮水安全问题的水环境保护研究;②90 年代末,开展面向水资源和水工程规划的生态需水研究;③自 2009 年开始,开展促进生态水利发展的生态环境应用研究。各阶段热点主题丰富明确,发展较集中,研究主题紧跟与国内水生态环境相关的时政热点问题,具有充足的政策导向性。

(4)研究热点知识网络结构联系紧密,时间维度上具有丰富的进程特征,图谱中各阶段有明确突出的研究热点,且集群化程度较高,说明水生态环境研究领域内有较好的研究氛围。

(5)各时段研究前沿脉络清晰,有较强的继承关系且彼此联系紧密,核心演变路径体现为:“流域生态水文规律研究→水生态环境管理技术研究→流域生态环境改善、恢复研究”,以支撑流域水资源调配管理的生态环境需求为前提,促进生态水利、水生态文明建设、流域生态环境相关的计算及工程技术等研究,表示国内水生态环境研究从宏观、共性的理论研究正转向以此为基础的区域、特性的工程技术应用研究。

目前,以科研论文为媒介的科学知识图谱已成为研究领域知识发现的重要方式,能够对过去的事实型知识和过程型知识进行陈述,但由于缺少语义关系的解析和梳理,所以对文献要素进行大规模实时关联和特征分析的推理能力仍比较薄弱。下一步将以词句、词篇关系的数据挖掘方式探索水生态环境研究领域内的新知识,以提升水利信息化的专业性。

## 参 考 文 献:

- [ 1 ] 王浩. 水生态文明建设的理论基础及若干关键问题[J]. 中国水利, 2016(19): 5-7.
- [ 2 ] 邱江涛,唐常杰,李庆. 面向领域的学术文献检索框架[J]. 软件学报, 2013, 24(4): 798-809.
- [ 3 ] 陈悦,陈超美,刘则渊,等. CiteSpace 知识图谱的方法论功能[J]. 科学学研究, 2015, 33(2): 242-253.
- [ 4 ] 熊永兰,张志强,尉永平,等. 基于科学知识图谱的水文化变迁研究方法探析[J]. 地球科学进展, 2014, 29(1): 92-103.
- [ 5 ] 胡秀芳,赵军,查书平,等. 生态安全研究的主题漂移与趋势分析[J]. 生态学报, 2015, 35(21): 6934-6946.
- [ 6 ] 许振亮,赵宇娜,周文霞. 国际生态城市研究前沿的知识图谱分析——基于共词网络分析视角[J]. 城市发展研究, 2013, 20(2): 115-124.
- [ 7 ] 林卓. 基于知识图谱的生态资产/价值研究可视化分析[J]. 生态学报, 2015, 35(5): 142-154.
- [ 8 ] 伍新木,任俊霖,孙博文,等. 基于文献分析工具的国内水资源管理研究论文的可视化综述[J]. 长江流域资源与环境, 2015, 24(3): 489-497.
- [ 9 ] KROEGER A. The Road to BIBFRAME: The evolution of the idea of bibliographic transition into a post-marc future[J]. Cataloging & Classification Quarterly, 2013, 51(8): 873-890.

- [ 10 ] 蒋秉川, 万刚, 许剑, 等. 多源异构数据的大规模地理知识图谱构建[J]. 测绘学报, 2018, 47(8): 1051-1061.
- [ 11 ] 周春雷, 王伟军, 成江东. CNKI输出文件在文献计量中的应用[J]. 图书情报工作, 2007, 51(7): 124-126.
- [ 12 ] SUNDHEIM B M. Overview of results of the MUC-6 evaluation[C]//Proceedings of the 6th Conference on Message Understanding, Association for Computational Linguistics, 1995.
- [ 13 ] CHEN Z, JI H. Collaborative ranking: a case study on entity linking[C]//Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2011.
- [ 14 ] JIANG X T, WANG Q, LI P, et al. Relation extraction with multi-instance multi-label convolutional neural Networks[C]//COLING, 2016.
- [ 15 ] DONG X L, GABRILOVICH E, HEITZ G, et al. From data fusion to knowledge fusion[J]. Proceedings of the VLDB Endowment, 2014, 7(10): 881-892.
- [ 16 ] SHEN W, WANG J Y, HAN J W. Entity linking with a knowledge base: issues, techniques, and solutions[J]. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(2): 443-460.
- [ 17 ] ZOU L M, TAMER ÖZSU, CHEN L, et al. gStore: a graph-based SPARQL query engine[J]. The VLDB Journal, 2014, 23(4): 565-590.
- [ 18 ] 李涓子, 侯磊. 知识图谱研究综述[J]. 山西大学学报(自然科学版), 2017, 40(3): 454-459.
- [ 19 ] 朱星宇. SPSS多元统计分析方法及应用[M]. 北京: 清华大学出版社, 2011.
- [ 20 ] 傅柱, 王曰芬. 共词分析中术语收集阶段的若干问题研究[J]. 情报学报, 2016, 35(7): 704-713.
- [ 21 ] 隋明爽, 崔雷. 基于SPSS的共现聚类分析参数选择的实例研究[J]. 中华医学图书情报杂志, 2016, 25(1): 52-56.
- [ 22 ] 吴健, 李子运, 王洪梅. 基于关键词共现聚类的深阅读研究热点分析[J]. 图书馆建设, 2016(12): 53-59.
- [ 23 ] KLEINBERG J. Bursty and hierarchical structure in streams[J]. Data Mining and Knowledge Discovery, 2003, 7(4): 373-397.
- [ 24 ] PARK H, YOON J. Assessing coreness and intermediarity of technology sectors using patent co-classification analysis: the case of Korean national R&D[J]. Scientometrics, 2014, 98(2): 853-890.
- [ 25 ] 付举磊, 孙多勇, 肖进, 等. 基于社会网络分析理论的恐怖组织网络研究综述[J]. 系统工程理论与实践, 2013, 33(9): 2177-2186.
- [ 26 ] 秦奇, 吴良, 李飞, 等. 基于社会网络分析的东南亚地缘关系研究[J]. 地理学报, 2018, 73(10): 2014-2030.
- [ 27 ] 陈晓红, 周宏浩. 城市化与生态环境关系研究热点与前沿的图谱分析[J]. 地理科学进展, 2018, 37(9): 1171-1185.
- [ 28 ] SHNEIDER A M. Four stages of a scientific discipline: four types of scientists[J]. Trends in Biochemical Sciences, 2009, 34(5): 217-223.
- [ 29 ] 程军蕊, 王侃, 冯秀丽, 等. 基于GIS的流域水质模拟及可视化应用研究[J]. 水利学报, 2014, 45(11): 1352-1360.
- [ 30 ] 李丽琴, 王志璋, 贺华翔, 等. 基于生态水文阈值调控的内陆干旱区水资源多维均衡配置研究[J]. 水利学报, 2019, 50(3): 377-387.
- [ 31 ] 徐志侠, 陈敏建, 董增川. 基于生态系统分析的河道最小生态需水计算方法研究( I ) [J]. 水利水电技术, 2004, 35(12): 15-18.
- [ 32 ] 徐志侠, 王浩, 陈敏建, 等. 基于生态系统分析的河道最小生态需水计算方法研究( II ) [J]. 水利水电技术, 2005, 36(1): 31-34.
- [ 33 ] 林俊强, 陈凯麒, 曹晓红, 等. 河流生态修复的顶层思考[J]. 水利学报, 2018, 49(4): 483-491.
- [ 34 ] 周娜, 贾仰文, 胡鹏, 等. 松花江流域冰封期水功能区限制纳污控制研究[J]. 水利学报, 2014, 45(5): 557-565.
- [ 35 ] 颜雷, 田庶慧. 水生态环境修复研究综述[J]. 水利科技与经济, 2011, 17(9): 73-75.
- [ 36 ] 高学平, 崔广涛. 生态城市与水利现代化[J]. 水利学报, 2007, 38(11): 1306-1311.
- [ 37 ] LAW J, BAUIN S, COURTIAL J P, et al. Policy and the mapping of scientific change: a co-word analysis of research into environmental acidification[J]. Scientometrics, 1988, 14(3/4): 251-264.

## Construction and application of knowledge graph of domestic water eco-environment based on bibliometrics

MAO Wenshan<sup>1, 2, 3, 4</sup>, ZHAO Hongli<sup>4</sup>, JIANG Yunzhong<sup>4</sup>, DUAN Hao<sup>4</sup>, HAO Zhen<sup>4, 5</sup>

(1. Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, China;

2. National-Local Joint Engineering Research Center of Technologies

and Applications for National Geographic State Monitoring, Lanzhou 730070, China;

3. Gansu Provincial Engineering Laboratory for National Geographic State Monitoring, Lanzhou 730070, China;

4. Department of Water Resources, China Institute of Water Resources and Hydropower Research, Beijing 100038, China;

5. Faculty of Infrastructure Engineering, Dalian University of Technology, Dalian 116024, China)

**Abstract:** In recent years, the problem of water eco-environment has attracted more and more attention from scholars and society. Taking the China National Knowledge Infrastructure (CNKI) as the literature search engine, this paper chooses the papers indexed by Chinese core journals which published articles with the theme of water eco-environment as sample data. Using the information analysis method of qualitative analysis and quantitative research in bibliometrics, based on the structural and temporal indicators of knowledge graph, using CiteSpace, SPSS information analysis software as the supporting tool, terms co-word matrix analysis, keywords co-occurrence cluster analysis and noun phrases burst detection analysis are carried out on the data, and mapping various knowledge graphs such as knowledge structure and knowledge evolution in the field of domestic water eco-environment research. The research shows that: ①“Literature timing analysis & subject analysis (macro)→co-occurrence matrix analysis of terms (microscopic, intuitive)→co-occurrence cluster analysis of keywords (complex, overall)→typical cluster analysis (local)→noun phrases burst detection analysis (special)” industry literature data analysis method is reasonable and effective; ②The progress of domestic water eco-environment research is in line with the “Shneider Four-stage Theory”, and the research theme of this field is clear, with the theme of “water environmental protection research→ ecological water requirement research→ecological environment-oriented application research”, the research hotspots after clustering have rich process characteristics, and the context of frontier research in each period is clear. it is embodied as: “Study on the law of watershed eco-hydrology→water eco-environment management technology research→watershed eco-environment improvement and restoration research”; ③ Identifying the subject of Chinese literature research through the co-word matrix analysis of the terms, which makes up for the insufficiency of Chinese literature data due to the lack of reference literature and the ability to identify the subject through the co-citation analysis of the literature.

**Keywords:** water eco-environment; knowledge graph; social network analysis; multidimensional scaling; CiteSpace

(责任编辑: 杨 虹)