

# 国内知识图谱应用研究综述\*

■ 胡泽文 孙建军 武夷山

[摘 要] 以知识图谱在情报学领域中的应用为例,首先简要介绍情报学知识图谱研究的必要性、现状、发展趋势以及知识图谱绘制工具及方法流程,然后对在国内 248 篇知识图谱研究文献进行内容分析,全面梳理知识图谱在情报学领域及其子领域的应用概况,并简要介绍知识图谱在其他学科或领域中的应用。

[关键词] 情报学 知识图谱 应用 研究综述 工具与方法 可视化

[分类号] G350

## 1 情报学知识图谱研究的必要性、现状和发展趋势

### 1.1 情报学知识图谱研究的必要性

情报学发展至今,已有 60 多年的历程。在这 60 年的发展历程中,情报学不断与其他学科交叉融合,产生了很多分支学科领域,同时也引进了很多新的方法和理论。谈及情报学领域比较盛行的方法和理论,不可不谈可视化技术与方法在情报学领域中的广泛应用。可视化技术与方法因其能够绘制学科知识图谱,动态清晰、直观形象地全面解读学科的国内外发展趋势、研究进展、热点前沿、学科知识结构及其动态演化关系,备受情报学研究人员的喜爱。国内已有很多学者将科学知识图谱的理论与方法应用于情报学领域,产出了一批研究成果。尽管关于情报学领域知识图谱的成果已较丰富,但尚缺乏对其研究现状和概况进行全面系统性梳理的文献。

### 1.2 情报学知识图谱研究的现状

采用检索式“KY = 知识图谱\* 可视化图谱”(KY 为关键词),于 2012 年 8 月 24 日检索中国学术文献网络出版总库,并按照情报类期刊名称对文献进行识别,发现目前国内学者已发表 180 余篇情报学知识图谱方面的研究文献,占全部知识图谱研究文献(248 篇)的 74%,其中大部分研究文献发表在表 1 所示的情报类期刊上。

国内情报学知识图谱研究文献最早起源于 2006 年大连理工大学侯海燕和刘则渊分别在《情报杂志》

表 1 情报学知识图谱研究文献的期刊分布

刊名	图书情报工作	情报杂志	情报科学	现代情报	情报理论与实践	中国图书馆学报	情报资料工作	图书情报知识
数量(篇)	19	16	15	14	8	4	4	4

和《科学学研究》上发表的两篇文章<sup>[1-2]</sup>。侯海燕在她的文章中,借助信息可视化、科学知识图谱等新的信息技术手段,界定出 10 位国际上最权威的科学计量学家,分析了他们对科学计量学的突出贡献、所代表的分支学科领域、相互关系及其最为关注的研究课题;而刘则渊在文章中运用引文分析(citation analysis)、多维尺度分析(multidimensional scaling, MDS)、因子分析(主成分分析法, factor analysis, PCA)、聚类分析(cluster analysis)和共词分析(co-word analysis)等方法,对 1995 - 2004 年间 SCI 数据库中 6 种期刊发表的 4 800 篇论文的 126 244 条引文数据进行可视化分析,绘制科学学知识图谱,深入描绘了该学科的主流研究领域及其演变情况。此后,兰州大学管理学院赵勇和沙勇忠在 2008 年发表的一篇文章中,对 1996 - 2008 年期间 SSCI 数据库中 24 种情报学核心期刊发表的 9 820 篇论文的 62 003 位被引作者进行共引分析、聚类分析、多维尺度分析和因子分析,绘制了国际情报学主流学术群体及其代表人物的知识图谱,并发现:目前国际情报学研究热点主要分为计量学研究和信息检索两大部分,科学计量学、信息计量学和网络计量学等新兴分支学科研究活跃,传统的文献计量定律研究逐步淡出计量学研究的主流范围<sup>[3]</sup>。中山大学资讯管理系杨利军利用 Web of Science 的文献数据,借助信息可视化工具

\* 本文系江苏省普通高校研究生科研创新计划项目“科技产出影响因素分析与预测研究”(项目编号: CXZZ12\_0075)研究成果之一。

[作者简介] 胡泽文,中国科学技术信息研究所、南京大学信息管理学院博士研究生, E-mail: huzenwen915@163.com; 孙建军,南京大学信息管理学院教授,博士生导师; 武夷山,中国科学技术信息研究所研究员,副所长。

收稿日期: 2012 - 10 - 15 修回日期: 2012 - 11 - 24 本文起止页码: 131 - 137 84 本文责任编辑: 易飞

CiteSpace II 软件绘制知识图谱,分析了国外竞争情报研究的重要学科领域、知识基础、研究热点和前沿,预测了未来竞争情报研究的趋势<sup>[4]</sup>。嘉兴学院廖胜姣基于汤森路透公司开发的工具 TDA (德文特分析家),绘制了情报学 1991 - 2007 年间的研究前沿及其演变情况的知识图谱,并结合该图谱,预测了未来情报学的发展方向<sup>[5]</sup>。江苏大学科技信息研究所杨国立等学者结合 CiteSpace II 和 GoogleEarthSCI 两款可视化工具,对 SCI 和 SSCI 收录期刊 *Scientometrics* 于 1990 - 2009 年期间发表的 2 097 篇论文进行了可视化分析,并绘制相关知识图谱,直观形象地展示出国际科学计量学领域研究力量的分布和合作情况<sup>[6]</sup>。黑龙江大学马海群和吕红利用 CiteSpace II 软件对中文社会科学引文索引数据库 (CSSCI) 中的 8 290 篇情报学文献数据进行了作者共现分析、机构共现分析、引文共被引的聚类分析和时间演化分析、关键词共现分析,分别绘制了中国情报学作者科研合作网络、机构共现网络、引文共被引聚类图谱和时间演化图谱、关键词共现图谱等,清晰地展示出中国情报学的研究主体 (包括通过作者合作网络识别出的核心作者群及其之间的合作关系、通过机构共现网络识别出的核心机构)、研究基础 (包括通过引文共被引聚类网络识别出的早期经典文献、高被引文献和高中心性文献)、研究热点 (主要通过高频关键词共现聚类图谱来识别) 和研究前沿 (主要通过研究前沿文献和前沿关键词共现图谱的时间演化趋势图来识别)<sup>[7]</sup>。

总体而言,情报学知识图谱研究就是基于引文分析、同被引分析、共现分析、聚类分析、词频分析、社会网络分析和多维尺度分析等可视化分析方法<sup>[8]</sup>,利用可视化分析工具如 CiteSpace II、TDA、Ucinet、VOSviewer、Bibexcel、SPSS、Histcite 和 WordsmithTool 等,对本学科的国内外文献数据进行分析,绘制科研合作网络、共现网络、引文网络和共被引网络等各类科学知识图谱,动态清晰、直观形象地全面解读学科或领域的国内外发展趋势、研究进展、热点与前沿;研究主体 (核心作者、期刊、机构和国家等) 及其之间的互引和合作关系;研究基础 (核心文献、高被引文献和经典文献等) 及其之间的动态演化关系 (反映学科知识结构);学科或领域的引文历史及文献之间的互引关系;等等。

### 1.3 情报学知识图谱研究的发展趋势

对 2006 - 2011 年间国内学者发表的 110 余篇相关文献进行时间分布分析,如图 1 所示:

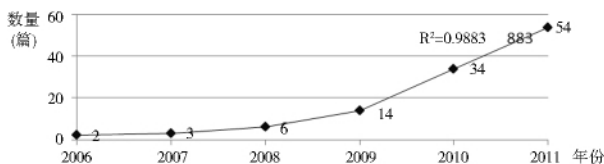


图 1 情报学知识图谱研究文献的年份分布

从图 1 可以看出,情报学知识图谱研究文献以指数级的速度增长,拟合优度  $R^2$  达到 0.988 3。其中,2009 - 2011 年呈现出直线上涨趋势。

## 2 知识图谱的绘制工具及方法流程

### 2.1 知识图谱绘制工具的概况、功能及优缺点

目前比较知名的知识图谱绘制工具主要有 CiteSpace II、Ucinet、VOSviewer、Bibexcel、SPSS 和 Histcite 等,其中 CiteSpace II 是目前最常用的工具。笔者对这几种工具的概况、功能及优缺点进行了分析<sup>[9-17]</sup>,如表 2、表 3 所示:

表 2 知识图谱绘制工具的概况及功能

工具名称	概况	功能描述
CiteSpace II	由美国德雷赛尔 (Drexel) 大学教授、大连理工大学长江学者陈超美开发	作者、机构、国家、术语 (来自文献标题和摘要) 和关键词 (来自文献主题词和标引词) 的共现分析;引文、作者和期刊的共被引分析;文献耦合分析;爆发词或爆发文献探测
Ucinet	由加州大学欧文 (Irvine) 分校 Linton Freeman 编写,是一款功能强大的社会网络分析软件	共现网络分析、中心性分析、子群分析、角色分析、多元回归分析、因子分析、聚类分析和多维尺度分析等
VOSviewer	荷兰莱顿大学 CWTS 资助开发的科学知识图谱绘制工具	作者或期刊的共引关系图,关键词共现关系图
Bibexcel	瑞典科学计量学学者皮尔逊 (Person) 开发的一款科学计量分析软件	引文分析、共引分析、引文耦合分析和聚类分析
SPSS	美国 SPSS 公司 20 世纪 80 年代开发的大型统计学软件包,一种多元统计分析软件	统计分析、距离相关分析、因子分析、多维尺度分析和聚类分析
Histcite	由 SCI 之父、引文分析方法的创立者——尤金·加菲尔德 (E. Garfield) 发明	histcite = history of cite,即引文历史,或称引文图谱分析软件,该软件能够用图示的方式展示某一领域不同文献之间的互引关系,可以帮助绘制出一个领域的发展历史,定位该领域的重要文献以及最新的重要文献

### 2.2 知识图谱绘制的方法流程

目前研究人员基本上是使用成熟的知识图谱绘制工具对收集的海量数据进行处理和可视化分析。知识图谱绘制的方法流程主要包括数据源选择、数据采集、数据处理、数据导入、参数设置、知识图谱绘制和分析几个步骤,见图 2。

表 3 知识图谱绘制工具的优缺点

工具名称	优点	缺点
CiteSpace II	使用简单,可直接导入 Web of Science 和 arXiv 数据库中的数据,进行可视化分析,并自带格式转换程序,可将 CNKI、CSSCI 和 Derwent 等数据库导出的数据进行格式转换后,进行可视化分析;功能强大,可以绘制各类知识图谱;提供聚类、时间序列和时区三种图谱布局;可视化效果好,易于解读,通过图谱中各节点之间连线的粗细和各节点圆圈的大小、年轮的厚度、颜色差异等,可以了解该节点的共现频次、中心度、在不同年份的被引情况和热点动向等	内存消耗大;各节点字体大小一样,不能随着节点频次或中心度的增大而增大;如果数据量较大的话,软件通常会进入无反应状态
Ucinet	使用简单;处理数据量大;内存消耗较小;视图简洁清晰,易于解读;绘图功能强大,集成了两大可视化软件: NetDraw 和 Pajek;能够处理转换后的中文数据	只识别矩阵数据,因此需花费很多时间和精力处理原始数据
VOSviewer	与 CiteSpace II 类似,在理论算法和图谱显示效果上有所不同,相对来说, CiteSpace II 处理的数据源种类更多,绘制的图谱种类更全面,不过在数据量处理上,不如 VOSviewer	需输入规定格式数据,无法处理中文数据
Bibexcel	功能强大,内在消耗低	需对数据进行处理,构建特定格式数据,比较复杂,且不能处理中文数据;可视化效果差
SPSS	易学易用、功能全面;支持丰富的数据源,具备强大数据访问和管理能力;数据统计分析功能强大,主要用于社会学研究和调查统计;软件的多维尺度分析和聚类分析功能可用于构建清晰的科学知识图谱	仅识别数值型数据,需对数据做复杂处理;利用多维尺度分析和聚类分析功能绘制科学知识图谱的过程较复杂;不具有数据清理功能;生成的多维尺度图比较简单,动态性较弱,图中信息量也比较少

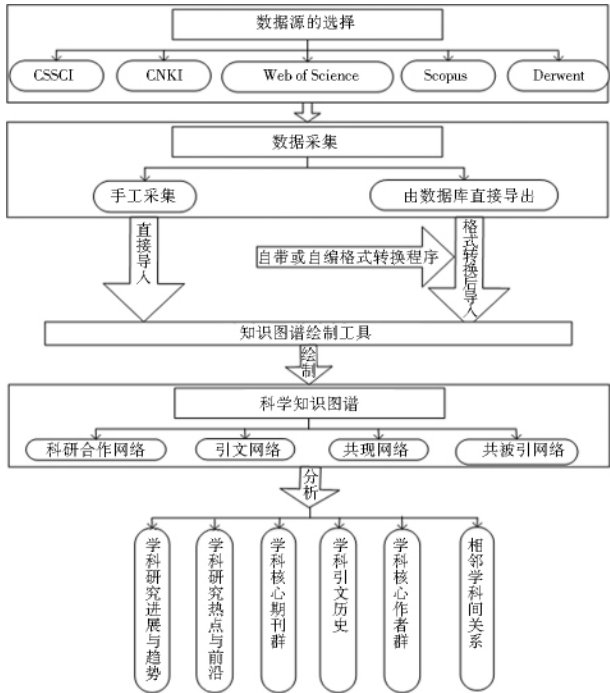


图 2 知识图谱绘制的方法流程

2.2.1 数据源选择 知识图谱的数据源有很多,主要分为学术和专利两个方面。学术方面的数据源有国内的中国社会科学引文索引( CSSCI)、中国知网( CNKI)和《中文科技期刊数据库》以及国外的 Web of Science、Scopus、arXiv、ProQuest 博硕士论文全文数据库、ISI Medline、Ei Engineering Village 2 等。专利数据源有 Derwent 专利数据库、各国官方专利数据库、国际性专利数据库等。笔者认为选择数据源时要考虑下面几个原则:①数据源尽量可靠、权威和覆盖面广;②尽量选择数据字段规范、有数据定制和导出功能的数据源;③尽量选择知识图谱绘制工具能够直接识别的数据源(可节省数据处理的时间)。

2.2.2 数据采集 对于一些有数据导出功能的数据源,可以直接利用导出功能导出所需主题的数据;如果没有导出功能,需要手工收集或向数据库生产商购买所需主题和字段的数据,或者利用数据库接口,编制数据采集程序进行采集,不过很少有数据库支持这种功能。

2.2.3 数据处理 不同的知识图谱绘制工具对数据的格式要求有所不同,因此需要对采集到的数据进行处理。有些知识图谱绘制工具可以直接识别和处理一些数据源的数据,如 CiteSpace II 和 Histcite 可以直接处理 Web of Science 数据库中的数据,因此从这个数据库中下载的数据无需处理,不过从其他数据源获取的数据需要利用软件自带的格式转换程序或自编格式转换程序,转换成该工具要求的数据格式。有些知识绘制工具只识别特定格式的数据,如 Ucinet 仅识别矩阵数据、SPSS 只能处理和分析数值型数据,因此这时需要利用自编的格式转换程序,将从数据库中采集的数据转换成矩阵数据或数值型数据后,进行可视化分析。

2.2.4 数据导入、参数设置和知识图谱绘制 将处理后的数据导入知识图谱绘制工具中,设置相关参数,绘制科研合作网络、共现网络、引文网络和共被引网络等各类科学知识图谱。

2.2.5 知识图谱分析 基于绘制的各类知识图谱进行全面解释和分析:①学科的国内外发展趋势、研究进展、热点与前沿动态;②学科知识结构及其之间的动态演化关系;核心作者、机构、期刊、国家及它们之间的合作关系;③学科引文历史及文献之间的互引关系;等等。

3 知识图谱在情报学及其子领域中的应用

3.1 展示发展趋势与研究进展

武汉大学邱均平在 2009 年对 1978 - 2007 年间情



报学领域高频关键词进行了分阶段统计分析,绘制了情报学高频关键词在各阶段的数量分布图,以此界定出改革开放以来情报学的三个发展阶段:复苏与发展时期(1978-1991年)、发展转折期(1992-1998年)和快速发展时期(1999-2007年)<sup>[18]</sup>。2011年,邱均平利用CiteSpace II中的爆发词探测算法,通过SCI和SSCI收录的13种高影响力外文源刊在2006-2010年间刊载的3 881条论文的题目、摘要和索引词,探测出2 171个反映图书情报学研究前沿的突变专业术语,并基于这些术语数据绘制出近5年图书情报学研究前沿及其随时间演化的时间序列图谱。基于图谱所示的发展趋势和研究前沿主题术语在不同时间段的变化,他将2006-2010年的情报学发展划分为三个发展阶段:繁荣期(2006年)、稳定期(2007-2009年)和新一轮繁荣的孕育期(2010年),并根据三个发展阶段的高频术语及其随时间的变化情况,描述了三个时期的研究进展、热点与前沿<sup>[19]</sup>。武汉大学学者赵蓉英和许丽敏利用CiteSpace II软件,对1995-2010年间Web of Science数据库中包含文献计量学(Bibliometric)主题词的1 459篇文献数据及其引文数据进行了引文共被引分析,绘制出文献计量学领域代表人物与代表作品共被引的时间序列图谱。基于图谱的时间演化趋势,作者将文献计量学划分为三个发展阶段:文献计量学的萌芽阶段(1917-1933年);文献计量学的奠定时期(1934-1960年)和文献计量学的全面发展与分化时期(1961年至今),并根据各个发展阶段中代表人物、文献及文献引文的数据对各个发展阶段的研究进展和概况进行梳理和分析<sup>[20]</sup>。

3.2 识别情报学研究热点与前沿

3.2.1 识别情报学领域研究热点 识别研究热点的方法有很多,如词频统计分析法(以高频关键词反映研究热点)、词共现法(以共现频繁的词聚类反映研究热点主题)和共被引聚类方法(以比较大的聚类反映研究热点主题)。其中共被引聚类方法是常用的方法。

美国德雷塞尔大学副教授、大连理工大学长江学者陈超美对国际情报学领域的12本权威期刊在1996-2008年期间发表的10 853篇文章的129 060篇引文进行了引文共被引聚类分析,并分别使用TF\* IDF和Log 似然率从施引文献标题术语集中选择Top 权重术语标引聚类,绘制了引文共被引聚类图谱<sup>[21]</sup>。

基于引文共被引聚类图谱,陈超美识别出情报学领域规模最大的5个聚类,聚类成员数量及两种权重方法选出的高权重聚类主题词,如表4所示:

表4 5个聚类的成员数量及选出的聚类主题术语

聚类成员数量(个)	TF* IDF 选择出的 Top 主题术语	Log 似然率选择出的 Top 主题术语
150	(80.82) interactive information retrieval (72.92) information retrieval (51.5) user information problem	interactive information retrieval (294.13) user information problem (167.06) various aspect (167.06)
69	(131.97) academic Web (103.62) Web site (54.77) exploratory hyperlink	academic Web (174.79) exploratory hyperlink (152.94) linguistic consideration (152.94)
46	(42.16) information retrieval (23.47) probabilistic model (22.53) query expansion	information retrieval (104.03) probabilistic model (81.54) using heterogeneous thesauri (67.95)
44	(14.07) citation behavior (14.07) citing literature (11.74) citation theory	citation behavior (56.66) citing literature (56.66) citation theory (43.92)
29	(83.69) h index	h index (212.76)

从表4可以看出1996-2008年间国际情报学研究热点有:①交互式信息检索(interactive information retrieval);②用户信息问题(user information problem);③网络信息计量学(academic Web、exploratory hyperlink和Web site三个高权重主题术语反映出这个研究热点);④信息检索(information retrieval)、概率模型(probabilistic model)和查询扩展(query expansion);⑤引用行为(citation behavior)、引用文化(citing literature)和引用理论(citation theory)研究;⑥H指数(h index)。

马海群和吕红基于CSSCI数据库中2001-2010年发表的8 290篇中国情报学文献的关键词数据,绘制高频关键词共现聚类图谱,基于图谱识别出2001-2010年间中国情报学领域的11个研究热点领域,分别是竞争情报、信息检索、情报学理论研究、知识管理、搜索引擎、信息服务、文献计量学、本体、数据挖掘与数据库、信息资源管理、信息管理与信息组织。与陈超美识别出的国际情报学研究热点相比,国内学者对信息检索及相关方面的研究比较多,而对引用行为和H指数等方面的关注比较少。

武汉大学学者赵蓉英利用CiteSpace II可视化软件对ISI Web of knowledge收录的“Bibliometric”相关文献题录数据进行高频关键词共现分析,得到关键词共现的时间序列知识图谱。借助知识图谱及其后台运行数据,作者识别出频次大于50的文献计量学热点研究主题,如科学(science)、影响力(impact)、期刊(journals)、指标(indicators)、文献计量分析(bibliometric analysis)、引文分析(citation analysis)、影响因子(impact factor)、出版物(publications)、文献计量指标(bibliometric

indicators) 等。

3.2.2 展示情报学研究前沿 嘉兴学院廖胜姣从 Web of Science 数据库中 25 种情报学领域 SCI 收录期刊的文献数据中, 选出 78 篇高被引文献及 4 000 多条引文。然后利用 TDA 对 78 篇文献制作自动关联图, 绘制出近 20 年情报学研究前沿文献聚类的知识图谱。基于图谱, 作者将 1991 - 2007 年期间情报学研究前沿划分为三个大的聚类: 文献计量学、科学交流与知识管理研究; 信息技术与信息系统研究; 信息检索研究。笔者认为, 这三个主题更应该被视作这段时期的研究热点, 而非前沿。通常研究前沿是近几年刚兴起的具有前瞻性、先导性和探索性的研究领域或主题, 而研究热点是指一定时期内大家关注比较多的、稳定集中的研究领域或主题, 两者有一定交叉。廖胜姣通过 1991 - 2007 年间 78 篇高被引文献的聚类识别出的主题是这段时间研究比较热的主题, 应该作为研究热点来对待。

从高频关键词、前沿术语、文献和引文聚类的时间演化来探寻学科或领域的研究前沿, 不失为一种相对科学合理的方法。赵蓉英利用 CiteSpace II 软件中的爆发词探测 (burst detection) 方法, 探测出 656 个反映文献计量学前沿的爆发词 (术语频次在较短时间内突然增加的术语), 并绘制爆发词随时间演化的学科前沿发展趋势图, 以此识别出文献计量学领域的研究前沿, 如科学文献 (scientific literature)、科学研究 (scientific research)、文献研究 (research-literature)、高被引 (highly cited)、科学信息 (scientific information)、同行评议 (peer review)、引文影响 (citation impact), 等等。中山大学杨利军和魏晓峰基于同样方法识别出反映国外竞争情报前沿领域的突现术语, 如信息技术 (突现年份: 2000)、竞争战略 (突现年份: 2003)、公共信息 (突现年份: 2007)、市场营销情报 (突现年份: 2008)、生存或环境选择 (突现年份: 2009), 等等<sup>[22]</sup>。

### 3.3 识别核心作者群及其之间的互引关系

通过作者共现或共被引分析, 能够识别出一个学科或领域的核心作者群及其之间的合作强度和互引关系。除此之外, 通过了解聚类内作者的研究主题, 能够发现一个学科或领域的知识结构和研究热点。同时, 通过作者共现或共被引聚类的时间演化图谱, 能够发现学科或领域的研究前沿。山西大学马瑞敏和倪超群利用 SPSS 软件对 CSSCI 中 1998 - 2007 年间图书情报学高被引作者做总时间周期和分阶段的共被引聚类分析, 绘制 1998 - 2007、1998 - 2002 和 2003 - 2007 三个

时间阶段的作者共被引聚类图谱, 展示出 10 年间图书情报学的知识结构及其在 1998 - 2002 年和 2003 - 2007 年两个连续时间段的演化情况。通过浏览和分析 1998 - 2007 年期间作者共被引聚类图谱中各个聚类内作者们的研究方向, 发现国内图情学研究主要集中在 11 个方向 “三计学” (指文献计量学、信息计量学和科学计量学)、竞争情报、情报学基本理论与方法、知识产权、知识管理与信息资源管理、信息标引与检索、数字信息资源开发与利用、目录学、图书馆学基本理论和网络信息组织、检索与服务<sup>[23]</sup>。关于图情学知识结构在两个时间段的演化情况, 在此不予细述。

马瑞敏和倪超群绘制的知识图谱的不足之处是不能够展示作者之间的共被引强度。南京大学鲍杨和朱庆华基于 1998 - 2007 年间 CSSCI 收录的所有情报学领域文献数据, 利用 Pajek 可视化工具绘制了一个能够反映作者之间共被引强度的图谱。基于该图谱, 作者发现严怡民的两部著作《现代情报学理论》和《情报学概论》的共被引次数最多, 达到 29 次。其次是马费成和严怡民的两部著作《情报学的进展与深化》和《现代情报学理论》(共被引频次 28 次)。另外, 作者也通过分析图谱结构, 发现作者共被引的两种模式: ①以某一重要学者文章为核心向外辐射的共被引模式, 如梁战平 2003 年发表的论文《情报学若干问题辨析》所引出的共被引模式; ②由某一重要学者文章引出高度聚团的一群节点文章, 如马费成 1996 年的论文《情报学的进展与深化》引出一个较大的共被引聚类。

同时, 鲍杨和朱庆华基于作者共现关系绘制了一个作者合著网络 (也称作者共现网络), 发现: 马费成和陈锐、张玉峰与金燕、毕强和刘甲学以及包昌火与谢新洲之间的合作强度比较大。值得注意的是, 这些合作大多是导师与学生之间的合作, 像包昌火与谢新洲这种教授与教授之间的合作较少。另外他们发现: 不同学者的研究风格不同; 不同的学术 “小团体” 往往是由于研究领域的不同造成的<sup>[24]</sup>。

江苏大学科技信息研究所杨国立、张垒基于科学计量学权威杂志 *Scientometrics* 在 1990 - 2009 年期间的文献数据, 利用 CiteSpace II 绘制了国际科学计量学的作者合著网络, 从中发现国际科学计量学领域 4 个较大的合作团体: ①以格伦采尔 - 舒伯特 - 鲁索 - 埃格赫为中心的第一个合作网络; ②以舒伯特 - 布劳温 - 格伦采尔为中心的第二个合作网络; ③以格伦采尔 - 舒伯特 - 埃格赫为中心的第三个合作网络; ④以维尔森 - 克雷其默 - 梁立明为中心的第四个合作网络<sup>[25]</sup>。

### 3.4 描述情报学的核心期刊群及其之间的互引关系

中国科学院赵勇利用 TDA 和 SPSS 软件对 Web of Science 数据库中被引频次大于 200 的 35 种图情核心期刊进行了期刊共被引聚类分析和多维尺度分析。作者基于多维尺度视图中各期刊之间的距离,将国际图书情报学研究划分为 7 个核心期刊群区域,这 7 个区域及其代表的研究主题为:①网络计量研究、信息(引文)计量研究;②信息检索及用户检索行为理论研究;③电子信息系统与服务及用户行为研究;④科学计量学理论、科研指标与评价研究;⑤数字图书馆、网络资源组织研究;⑥用户服务研究;⑦知识组织、资源开放获取研究<sup>[26]</sup>。

### 3.5 识别核心国家、机构及其之间的合作情况

杨国立利、张垒用 CiteSpace II 对 *Scientometrics* 杂志中文文献的发表机构数据进行共现分析,绘制了机构合作网络。通过合作网络,不仅可以看出国际科学计量学领域的核心研究机构及其之间的合作关系,还可以看出国际科学计量学研究存在四大机构合作网络。这四大机构合作网络包括:①比利时天主教鲁汶大学(Katholieke Universiteit Leuven) - 匈牙利科学院(Hungarian Academy of Sciences, HAS) - 荷兰阿姆斯特丹大学(University of Amsterdam);②印度国立科技与发展研究所(National Institute of Science, Technology and Development Studies) - 比利时林堡大学(Limburgs Universitair Centrum) - 比利时安特卫普大学(University of Antwerp);③荷兰莱顿大学(Leiden University) - 西班牙格拉那达大学(University of Granada);④西班牙国家科学研究委员会(Consejo Superior de Investigaciones Científicas) - 英国伍尔弗汉普顿大学(University of Wolverhampton)<sup>[25]</sup>。

## 4 知识图谱在其他学科或领域中的应用

除情报学领域外,也有一些学者将知识图谱应用于其他科学或领域。中国医科大学董立平和郭继军从 Web of Science 中套录人胚胎干细胞研究的书目信息,利用 Histcite 软件对其中被引频次最高的前 30 篇文献进行引文分析,绘制了引文编年图。从该图可以看出人胚胎干细胞研究的引文历史,直接观察到主要文献之间的相互引用关系,清晰地看到人胚胎干细胞研究的发展历程和变化趋势。另外,通过研读重要文献节点的内容,可以看出人胚胎干细胞研究的重点内容和方向<sup>[27-28]</sup>。

大连理工大学侯海燕和刘则渊等学者借助科学知

识图谱的可视化技术手段,对 6 个科学学国际权威学术期刊于 1995 - 2004 年间发表的 4 800 篇论文的 126 244 条引文数据进行多维尺度与聚类分析,分别绘制出 1995 - 1996、1997 - 1998、1999 - 2000、2001 - 2002、2003 - 2004 年间的国际科学学研究热点演进趋势知识图谱,展示出 1995 - 2004 年间各个时间段的国际科学学研究的热点领域及其发展态势。基于 2003 - 2004 年时间段的研究热点演进趋势知识图谱,作者识别出 2003 - 2004 年期间的国际科学学研究的热点领域,包括:科学政策与管理、战略管理、科学经济学、科学合作、信息检索、信息科学、科学知识图谱与可视化;科学行动理论(科学知识社会学)和国家创新系统;网络计量学、知识管理、科学知识图谱和信息计量学理论<sup>[29]</sup>。

大连理工大学 21 世纪发展研究中心许振亮和陈悦等人对国际创新管理的 10 个代表性期刊进行关键词共现分析,识别出当前国际创新管理的三个主流研究前沿领域:企业与产业创新系统、创新政策与信息管、市场创新与用户创新<sup>[30]</sup>。陈悦还与王续琨等人以 10 种管理学代表性期刊所载论文的引文数据为分析对象,绘制引文共被引和关键词共现知识图谱,揭示出管理学依次形成的三大主导理论——组织行为理论、组织结构理论和战略管理理论,并确认了当前的学科理论前沿<sup>[31]</sup>。天津师范大学体育科学学院王琪和大连理工大学网络—信息—科学—经济计量实验室胡志刚合作利用 CiteSpace II 对科学引文索引(SCI)和社会科学引文索引(SSCI)收录奥林匹克运动研究论文的关键词数据进行高频关键词共现分析,总结出当前国际奥林匹克运动研究的 4 个主要特征,识别出奥林匹克运动的九大研究前沿<sup>[32]</sup>。哈尔滨工程大学经济管理学院王吉武和黄鲁成等人通过对 LED 照明的历史文献进行多维尺度分析和高频关键词共现分析,绘制相关知识图谱,挖掘出当前 LED 照明的技术成熟度、研究热点和关键技术点等信息<sup>[33]</sup>。军事医学科学院科技部张明华和雷二庆利用 CiteSpace II 对认知神经科学领域中的作者、机构、国别、关键词和被引文献等节点进行科学知识图谱分析,识别出认知神经科学领域中的权威作者、研究实力较强的机构和国家以及认知神经科学领域的研究基础(经典文献和关键文献)、研究热点和学科发展趋势等信息<sup>[34]</sup>。江苏大学刘桂锋利用 CiteSpace II 对 Web of Science 数据库中 1990 - 2010 年间太赫兹技术领域研究文献数据进行期刊共被引、作者共被引、文献共被引和关键词共现分析,绘制



相关知识图谱,揭示出该领域的领军人物、知识基础和研究前沿等信息<sup>[35]</sup>。

## 5 结 语

自信息可视化和科学知识图谱绘制领域权威专家陈超美将可视化工具 CiteSpace 及知识图谱绘制方法引入中国后,国内学者对该主题的研究呈井喷之势。目前,知识图谱除了在情报学领域得到广泛和深入应用外,正快速地向其他学科或领域(如体育科学、教育学、科学学、管理学、军事科学、医学、经济学、物理学和新兴技术领域)进行扩散。在该背景下,本文以国内 248 篇知识图谱研究文献为研究对象,在简要分析情报学知识图谱研究的必要性、现状和发展趋势以及简要介绍知识图谱绘制工具及其方法流程的基础上,详细综述了知识图谱在情报学领域及其子领域、其他学科或领域中的具体应用概况,以便使读者对国内知识图谱的应用情况有一个综合性的了解,为以后的相关研究提供指导。

总体来说,国内知识图谱应用研究主要基于学术数据库,对各学科及其子学科的研究热点与前沿、研究主体及研究基础等进行可视化分析。而利用专利数据,对技术领域的研究热点与前沿进行可视化图谱分析的研究比较少。另外,对科学-技术之间关系的可视化分析文献也非常少。

### 参考文献:

- [1] 侯海燕. 权威科学计量学家对科学的关注及贡献[J]. 情报杂志 2006 25(4): 118-120.
- [2] 刘则渊. 科学学理论体系建构的思考——基于科学计量学的中外科学学进展研究报告[J]. 科学学研究 2006 24(1): 1-11.
- [3] 赵勇, 沙勇忠. 当代情报学研究的知识图谱: 基于 ACA 的分析[J]. 图书馆论坛 2008 28(6): 63-69.
- [4] 杨利军, 魏晓峰. 国外竞争情报研究热点、前沿及趋势的可视化分析[J]. 图书情报工作 2010 54(20): 62-66.
- [5] 廖胜蛟. 基于 TDA 的情报学研究前沿知识图谱的绘制及分析[J]. 情报理论与实践 2009 32(11): 98-101.
- [6] 杨国立, 张垒. 国际科学计量学研究力量分布与合作网络分析[J]. 图书情报研究 2012 5(1): 34-39.
- [7] 梁秀娟. 科学知识图谱研究综述[J]. 图书馆杂志 2009 28(6): 58-62.
- [8] 马海群, 吕红. 基于中文社会科学引文索引的中国情报学知识图谱分析[J]. 情报学报 2012 31(5): 470-478.
- [9] 陈超美, 陈悦. CiteSpace II: 科学文献中新趋势与新动态的识别与可视化[J]. 情报学报 2009(3): 401-421.
- [10] Chen Chaomei. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature[J]. Journal of

the American Society for Information Science and Technology, 2006 57(3): 359-377.

- [11] 胡泽文, 王效岳, 白如江. 国内外文本分类研究计量分析与综述[J]. 图书情报工作 2011 55(6): 78-81.
- [12] 2008 Nips Ucinet & NetDraw Workshop [EB/OL]. [2012-09-22]. [http://www.hks.harvard.edu/netgov/files/NIPS/Halgin\\_NIPS\\_2008.pdf](http://www.hks.harvard.edu/netgov/files/NIPS/Halgin_NIPS_2008.pdf).
- [13] 廖胜蛟. 科学知识图谱绘制工具 VOSviewer 与 Citespace 的比较研究[J]. 科技情报开发与经济 2011 21(7): 137-139.
- [14] VOSviewer manual [EB/OL]. [2012-09-20]. <http://www.vosviewer.com>.
- [15] BibExcel [EB/OL]. [2012-09-23]. <http://www8.umu.se/inforsk/Bibexcel/>.
- [16] 廖胜蛟. 科学知识图谱绘制工具: SPSS 和 TDA 的比较研究[J]. 图书馆学研究 2011(5): 46-49.
- [17] SPSS Manual [EB/OL]. [2012-09-21]. <http://spssmanual.com/index.html>.
- [18] 邱均平, 周春雷, 杨思洛. 改革开放 30 年来我国情报学研究的回顾与展望(三)——情报学的发展阶段及趋势分析[J]. 图书情报研究 2009 2(3): 1-9.
- [19] 邱均平, 温芳芳. 近五年来图书情报学研究热点与前沿的可视化分析——基于 13 种高影响力外文源刊的计量研究[J]. 中国图书馆学报 2011 37(2): 51-60.
- [20] 赵蓉英, 许丽敏. 文献计量学发展演进与研究前沿的知识图谱探析[J]. 中国图书馆学报 2010 36(5): 60-68.
- [21] Chen Chaomei, Ibekwe-SanJuan F, Hou Jianhua. The structure and dynamics of co-citation clusters: A multiple-perspective co-citation analysis [J]. Journal of the American Society for Information Science and Technology 2010 61(7): 1386-1409.
- [22] 杨利军, 魏晓峰. 国外竞争情报研究热点、前沿及趋势的可视化分析[J]. 图书情报工作 2010 54(20): 62-66.
- [23] 马瑞敏, 倪超群. 基于作者同被引分析的我国图书情报学知识结构[J]. 中国图书馆学报 2011 37(6): 17-26.
- [24] 鲍杨, 朱庆华. 近 10 年我国情报学研究领域主要作者和论文的可视化分析——基于社会网络分析方法的探讨[J]. 情报理论与实践 2009 32(4): 9-13.
- [25] 杨国立, 张垒. 国际科学计量学研究力量分布与合作网络分析[J]. 图书情报研究 2012 5(1): 34-39.
- [26] 赵勇. 期刊共引分析及可视化实证研究——以图书情报学研究为例[J]. 图书与情报 2009(3): 89-94.
- [27] 董立平, 郭继军. 利用 Histcite 的人胚胎干细胞引文编年图主要路径分析[J]. 医学信息研究 2010 31(11): 38-49.
- [28] 许丹, 侯跃芳. 国外医学信息学研究领域可视化引文分析[J]. 医学信息研究 2011 32(10): 12-19.
- [29] 侯海燕, 刘则渊, 陈悦, 等. 当代国际科学学研究热点演进趋势知识图谱[J]. 科研管理 2006 27(3): 90-96.
- [30] 许振亮, 陈悦, 葛莉, 等. 基于知识图谱的国际创新管理前沿与热点的分析[J]. 管理评论 2009 21(5): 13-18.

(下转第 84 页)

Indicators of structural changes and interdisciplinary developments [J]. Journal of the American Society for Information Science and Technology ,2008 ,59( 11) : 1810 – 1818.

[ 8 ] Glänzel W , Schubert A. A new classification scheme of science fields and subfields designed for scientometric evaluation purposes [J]. Scientometrics ,2003 ,56( 3) : 357 – 367.

[ 9 ] 张琳. 基于期刊聚类的科学结构研究[D]. 大连: 大连理工大学 2010.

[10] Zhang Lin , Glänzel W , Liang Liming. Tracing the role of individual journals in a cross-citation network based on different indicators [J]. Scientometrics ,2009 ,81( 3) : 821 – 838.

[11] Zhang Lin , Janssens F , Liang Liming , et al. Journal cross-citation analysis for validation and improvement of journal-based subject classification in bibliometric research [J]. Scientometrics ,2010 ,82( 3) : 687 – 706.

[12] 张琳 , 梁立明 , 刘则渊 , 等. 基于期刊互引聚类与 SOOI 分类体系的科学结构研究 [J]. 科学学研究 ,2012 ,30( 9) : 1287 – 1291.

Analysis of Subject Classification and Interdisciplinary Structure Based on Journal Hybrid Clustering

Zhang Lin<sup>1 2</sup>

<sup>1</sup>Department Management and Economics , North China University of Water Conservancy and Electric Power , Zhengzhou 450011

<sup>2</sup>Center for ECOOM , K. U. Leuven University , Leuven 3000

**[Abstract]** A hybrid clustering algorithm based on fitting similarity of “citation-text” , is used to cluster 8 305 journals in SCI , SSCI and A&HCI in Web of Science , to interpret the internal structure of science. Essential Science Indication ( ESI) classification system is compared with the journal hybrid clustering structure. The “coincidence” of the two schemes is thought having strong “subject correlation” and defined as the “subject kernel” of corresponding disciplines , and the “difference” is “subject peripheral”. The “Core journals” define the core knowledge scope of corresponding , while the “peripheral journals” interpret the structure characteristics such as differentiation and cross fusion of different subjects.

**[Keywords]** journal hybrid clustering interdisciplinary structure of science

( 上接第 137 页)

[31] 陈悦 , 王续琨 , 郑刚. 基于知识图谱的管理学理论前沿分析 [J]. 科学学研究 2007( S1) : 22 – 28.

[32] 王琪 , 胡志刚. 国际奥林匹克运动研究前沿的知识图谱分析 [J]. 西安体育学院学报 2011 28( 4) : 433 – 436.

[33] 王吉武 , 黄鲁成 , 李剑 , 等. 基于文献计量分析的 LED 知识图谱及产业化对策 [J]. 情报学报 2009 28( 3) : 443 – 450.

[34] 张明华 , 雷二庆. 基于知识图谱的认知神经科学前沿与演化研究 [J]. 军事医学科学院院刊 2010 34( 1) : 71 – 75.

[35] 刘桂锋. 基于 Citespace II 的国际太赫兹技术知识图谱研究 [J]. 图书情报研究 2012 ,16( 3) : 51 – 57.

Research Review on Application of Knowledge Mapping in China

Hu Zewen<sup>1 2</sup> Sun Jianjun<sup>2</sup> Wu Yishan<sup>1</sup>

<sup>1</sup>Institute of Scientific & Technical Information of China , Beijing 100038

<sup>2</sup>School of Information Management , Nanjing University , Nanjing 210093

**[Abstract]** This paper takes the application of knowledge mapping in information science as example. Firstly , it makes a brief introduction to the necessity of studying knowledge mapping of information science , its research status and development trend. Then the tools , methods and processes of drawing knowledge mapping are introduced. Finally , based on the analysis of contents of 248 papers about knowledge mapping , it comprehensively reviews the general situation of application of knowledge mapping in information science and its sub – domains , and briefly introduces the application of knowledge mapping in the other disciplines or domains.

**[ Keywords ]** knowledge mapping information science application research and review tools and methods visualization