

基于引证关系的作者与主题多重共现网络构建

刘爱琴 吴瑞瑞

(山西大学经济与管理学院, 太原 030006)

摘要: 利用可视化工具描述知识资源及其载体, 深层次地挖掘知识内容及其结构关系, 表征知识之间的关联是有效提高信息检索效率和准确率, 实现数据库服务模式创新的重要途径。本文从数据库检索因子优化入手, 以CNKI中国引文数据库的数据作为依据, 基于引证关系, 对知识发现领域研究群体进行知识图谱构建和聚类分析, 一方面展示了群体族群关系及整体知识架构, 挖掘出知识发现领域作者群体、主题及引文之间的超网络模型; 另一方面验证了将二次凝练因子作为基础特征项进行应用的效果, 提高了信息检索的查全率和查准率, 弥补信息检索的缺陷。

关键词: 知识图谱; 知识关联; 多重共现; 引证关系

中图分类号: G353.1

DOI: 10.3772/j.issn.1673-2286.2019.10.004

论文共现现象在科学社会体系中普遍存在, 通过共现分析可以从多角度、多方位挖掘论文中的隐含信息以揭示其内在关联^[1]。随着可视化技术的不断完善, 共现现象的表述逐步走向形式化, 知识图谱成为形式化表述的一种重要方式。利用可视化工具描述知识资源及其载体, 深层次地挖掘知识内容及其结构关系, 表征知识之间的关联能有效提高信息检索效率和准确率, 也是实现数据库服务模式创新的重要途径^[2]。因此, 基于引证关系, 本文对知识发现领域研究群体进行知识图谱构建和聚类分析, 一方面展示了群体族群关系及整体知识架构, 挖掘出知识发现领域作者与主题之间的超网络模型; 另一方面提高了信息检索的查全率和查准率, 弥补信息检索的缺陷。

1 文献综述

知识图谱最早出现在管理学领域, 随后向信息技术领域和图情领域靠拢, 慢慢渗透到各行各业^[3]。国外对知识图谱的研究起步相对较早, 在文献数量、质

量对比上占有一定优势。O'Donnell等^[4]认为, 知识图谱是一种节点链接, 通过一系列的链接与其他概念相连, 以知识图谱作为认知加工的支架有助于人们在认知过程中快速产生有效响应。Van Eck等^[5]全面概述了VOSviewer以一种易于解释的方式显示大型书目的强大功能, 并通过构建和显示数千种主要期刊的共引图来证明VOSviewer程序构造的合理性。Boyack等^[6]通过直接引用、书目耦合、共引文分析, 以及基于引文与文内耦合的多重方法对生物医学文献的聚类情况进行研究, 比较不同方法的聚类精确度。Porter等^[7]通过运用新的跨学科指标和科学制图可视化方法研究不同领域的跨学科程度, 发现跨学科指数呈现适度增长, 但只略微增加了与远距离认知领域的关系, 主要分布仍然集中在邻近学科; 同时表明, 叠加学科知识图谱能够为未来跨学科研究提供通用标准。

2005年, 由陈悦等^[8]共同署名的《悄然兴起的科学知识图谱》在《科学学研究》上发表, 多数研究者认为其标志着国内知识图谱领域研究的开始; 文章认为科学知识图谱是科学计量学表达形式转换的产物, 在揭

示科学知识内涵、结构及其活动规律的进程中起到了简化作用。侯海燕等^[9]通过对科学计量学、应用数学及计算机科学等相关学科进行可视化研究,交叉整合各学科代表性成果并绘制图谱,剖析科研热点,同时预测演进趋势。陈悦等^[10]给出知识图发展历程简介,并将传统知识图谱与现代知识图谱的类型和实现原理进行对比,表明其作为知识管理工具的有效性。秦长江等^[11]主要讲述构建知识图谱的理论及方法技术,并结合具体应用疏通知识图谱的发展脉络。赵蓉英等^[12]用Cite Space II围绕图书馆主题展开讨论,并对引文数据和主题词数据进行分析,梳理了该学科的代表性文献和领军人物。庞宏燊等^[13]选取竞争情报学科的2个多重共现网络研究主题发展趋势的分析模型,解析竞争情报在各个因素方向的发展态势与变化,并对技术过程提出改进意见。郭秋萍等^[14]构建基于作者-关键词-引文3个子网的多重共现超网络,并以图书馆、情报与文献学学科领域的知识服务为主题进行实证分析,揭示科技文献网络不同节点之间的关联关系,为研究文献之间的隐性关联关系提供新的研究方法。郭红梅等^[15]将一系列具有语义信息的术语进行叠加,检测多重术语关系在识别文本核心主题方面的效用,结果表明3种关系的叠加使文本主题更为凸显,克服了单独考虑一种关系时造成的信息缺失。Lee等^[16]重点分析主题-方法二重共现关系,考察传播学中的主题变化趋势。Leydesdorff^[17]把异质网络思想进行扩展,构建了作者-关键词-期刊三重共现网络,将不同类型节点及其交叉关系展现在同一网络图谱中,

更加真实地反应研究网络。

综上所述,当前学者主要运用可视化工具对某一领域的基础元素进行一层架构,并分析其具体应用和发展趋势,没有进行整合提升。本文以CNKI中国引文数据库的数据作为依据,首先通过作者间的引证关系形成同被引网络并将知识图谱进行升级应用,运用社会网络分析法对该群体进行凝聚子群分析;随后基于作者群体、主题与引文的关联关系构建多重共现超网络模型,通过逐层凝练整合,构建三者之间更精准的关联体系框架,有效提高信息检索效率和准确率,实现数据库服务模式创新的重要途径。

2 基于引证关系的作者与主题多重共现网络构建

2.1 群体可视化知识图谱构建与分析

在科学文献体系结构中,引证文献是论文的基本属性,也是文献之间得以连接的内在枢纽。在文献中,引证文献以各种形式的标注呈现,用以揭示文献之间的相互引用关系^[18]。著者同被引是由被引文献延伸而来,本文借助被引文献构建著者同被引网络,具体方法及数据处理如下。

(1)以知识发现为检索主题,统计CNKI中国引文数据库中收录的文献。以发表数3篇、被引总数150次作为筛选条件,选出47位高被引作者作为研究对象,见表1。

表1 知识发现领域47位高被引作者

作者姓名	总发文/篇	总被引/次	作者姓名	总发文/篇	总被引/次	作者姓名	总发文/篇	总被引/次	作者姓名	总发文/篇	总被引/次
李德仁	12	1 881	黄亚楼	5	620	潘云鹤	7	349	王清毅	7	259
李德毅	18	1 664	石纯一	3	596	郑诚	8	349	温有奎	11	246
陆玉昌	11	1 474	焦李成	10	580	周傲英	3	341	殷国富	10	244
蔡庆生	43	1 220	唐常杰	15	577	刘君强	9	331	陈恩红	5	238
王新洲	6	1 009	史开泉	8	466	高文	5	322	朱东华	6	237
王树良	3	949	林士敏	6	426	谢榕	3	300	吴朝晖	5	185
张文修	20	822	程继华	7	419	吉根林	3	286	毕强	11	173
胡可云	4	761	史文中	3	410	卢正鼎	6	279	李增智	10	166
刘宗田	13	725	邸凯昌	5	373	张晓林	4	276	张天庆	3	160
施鹏飞	15	708	刘明吉	4	361	周成虎	7	274	吕安民	4	152
欧阳为民	18	691	王秀峰	3	356	蒋永光	5	273	商琳	5	150
胡学钢	22	624	朱绍文	3	354	左劼	3	271			

(2) 以CNKI中国引文数据库知识发现学科为检索范畴, 搜索47位作者彼此之间的共被引频次, 不计同名作者的相关文献, 统计整理得到我国知识发现科学学术群体的同被引矩阵(TOP10), 见表2。

表2 知识发现科学学术群体的同被引矩阵(TOP10)

	李德仁	李德毅	陆玉昌	蔡庆生	王新洲	王树良	张文修	胡可云	刘宗田	施鹏飞
李德仁	-	-	-	-	-	-	-	-	-	-
李德毅	162	-	-	-	-	-	-	-	-	-
陆玉昌	12	22	-	-	-	-	-	-	-	-
蔡庆生	18	37	40	-	-	-	-	-	-	-
王新洲	61	56	7	8	-	-	-	-	-	-
王树良	147	118	11	11	61	-	-	-	-	-
张文修	21	47	65	26	10	14	-	-	-	-
胡可云	13	29	93	23	5	7	113	-	-	-
刘宗田	5	8	33	11	2	3	42	53	-	-
施鹏飞	13	25	27	33	7	11	14	17	14	-

(3) 以Ucinet和Netdraw软件作为分析工具, 利用凝聚子群解析著者同被引矩阵并生成同被引网络图谱, 再现整体网络集合中所有成员之间的关系密度, 探究我国知识发现领域作者之间的内在关联, 全部作者凝聚得到8个子群, 最少3人, 最多18人, 见表3。

表3 我国知识发现领域研究群体的凝聚子群

群体编号	人数	作者姓名
群体A	3	李德仁、李德毅、王新洲
群体B	4	史文中、周成虎、王树良、邸凯昌
群体C	3	陆玉昌、张文修、胡可云
群体D	3	蔡庆生、施鹏飞、黄亚楼
群体E	9	胡学钢、焦李成、石纯一、史开泉、林士敏、商琳、刘宗田、王秀峰、刘明吉
群体F	18	唐常杰、程继华、潘云鹤、郑诚、周傲英、刘君强、高文、谢榕、吉根林、卢正鼎、李增智、欧阳为民、吕安民、朱绍文、王清毅、陈恩红、殷国富、张天庆
群体G	4	温有奎、张晓林、朱东华、毕强
群体H	3	吴朝晖、蒋永光、左劼

其中, 群体A中3位作者集中于数据库方面的研究, 李德仁与王新洲侧重空间数据挖掘的理论方法与应用, 李德毅则比较注重数据库与知识发现的应用。从发文情况来看, 多数作品均由任意2人或3人共同署名, 总发文数相差不多, 但被引次数李德仁最多(1 881次), 其次是李德毅(1 664次)、王新洲(1 009次)。从该群体同被引频次来看, 李德仁与李德毅同被引频次最高, 达162次; 李德毅与王新洲、李德仁与王新洲所差无几, 分别为56次和61次。李德毅的同被引频次总和在整个网络中最高(1 106次), 其次是李德仁(954次)、王新洲(423次), 3位学者总被引频次整体排位靠前, 在整个网络中地位十分重要。

群体B中, 4位作者的主要研究方向是空间数据挖掘。其中, 邸凯昌偏重于相关理论与实际应用, 其他3人更倾向于方法、分类等客观性描述。从该群体的同被引频次分析, 王树良与史文中和邸凯昌的同被引频次分别为89次和81次, 史文中和邸凯昌为57次, 王树良和周成虎以及邸凯昌和周成虎均为56次, 史文中和周成虎为44次。从同被引频次总和看, 王树良是557次, 史文中555次, 邸凯昌718次, 周成虎433次。可见, 邸凯昌不论是在该群体还是在整个网络中影响力较大。

群体C中, 3位作者主要关注知识发现领域的人工智能、粗糙集等技术, 且以算法为主。从该群体彼此间的同被引频次来看, 张文修和胡可云最多(113次),

其次是陆玉昌和胡可云(93次)、陆玉昌和张文修(65次)。从被引频次总和看,陆玉昌742次,张文修750次,胡可云695次。由数据可知,该群体成员关系较为亲密,学术地位相当。

群体D中,3位作者研究方向集中在数据挖掘与关联规则。其中,施鹏飞以算法运用为主,蔡庆生与黄亚楼主要描述算法的实现过程。该群体成员同被引次数整体相对较少,蔡庆生与施鹏飞33次,蔡庆生与黄亚楼11次,施鹏飞与黄亚楼7次。从同被引频次总和看,蔡庆生709次,施鹏飞439次,黄亚楼224次,相较其作品数量而言,被引次数较多。

群体E中,9位作者的作品多属同一时期,集中在该学科萌芽阶段,侧重论证相关算法的形成过程。胡学钢、刘宗田关注关联规则,王秀峰和刘明吉研究决策树,焦李成、商琳以及林士敏探索数据挖掘,石纯一和史开泉则注重粗糙集。从该群体的同被引频次看,刘明吉和王秀峰最多(47次),其次是石纯一和刘宗田(33次)、石纯一和胡学钢(21次)、石纯一和王秀峰(14次),剩余任意两位作者共同被引次数较少(均在10次以下),表明其研究成果关联度不大。从同被引频次总和看,石纯一最多(561次),其次是刘宗田、胡学钢、王秀峰、焦李成、刘明吉、林士敏、商琳、史开泉。

群体F有18位作者,在整个网络中占比最大。该群体研究范围广泛,在数据库、算法、应用、关联规则、数据挖掘等方面均有所涉足,研究方向有所交叉,又有所侧重。如唐常杰、程继华、郑诚、刘君强、李增智、欧阳为民、王清毅、陈恩红均涉及关联规则,但唐常杰围绕基因表达式展开,欧阳为民以数据库为主,陈恩红则借助贝叶斯方法进行研究。从同被引频次来看,次数较少,甚至多数人之间都没有共被引关系;从同被引频次总和来看,欧阳为民最高,其次是高文、程继华、吉根林、郑诚、陈恩红、王清毅、潘云鹤、唐常杰、吕安民、朱绍文、刘君强、周傲英、卢正鼎、谢榕、张天庆、殷国富、李增智。

群体G中,4位作者的共同研究方向是图书馆的数字资源整合、图书馆的变革方向,以及与数字化结合的图书馆如何快速实现知识发现。温有奎和毕强从语义检索方向研究检索方法的改进,张晓林和朱东华从数据处理方面进行研究。从该群体的同被引频次来看,张晓林与毕强和温有奎次数较多,有16次和13次,剩余彼此之间次数很少。从被引频次总和而言,张晓林最多,其次是毕强、温有奎。

群体H的3位作者主要研究知识发现在医药领域的

应用。同被引频次显示,只有蒋永光与吴朝晖之间有共被引关系(8次),其余任意两位作者之间没有关联,只是共同将知识发现作为工具,在熟悉领域加以应用。3人的总被引频次偏少,表明3人的研究方向与群体其他人员之间的一致程度较低。

47位作者同被引网络的可视化成果分析,揭示了我国知识发现领域学术群体结构分布、成员地位,明确主要学者之间的关联程度。

2.2 网络整体属性分析

利用Ucinet软件,对我国知识发现研究群体同被引网络进行分析,计算出该网络的可达矩阵密度是0.744,即该网络关联度为0.744。由该数据可以看出我国知识发现研究群体同被引网络的关联度较高,学者之间的联系较为紧密。且该网络中节点之间的平均距离为1.256,表明在该网络体系中,每两位作者之间建立相应的学术关联只需通过1.256个人就能实现。

2.3 网络中心性分析

2.3.1 程度中心性

我国知识发现研究群体同被引网络程度中心性分析结果显示(见表4),张文修在该团体中程度中心性最高,达0.978 3,节点关系数是45,表明其中有45位学者与他存在同被引关系。同样可知,节点数超过40且程度中心性超过0.850 0有17位,说明该部分作者在该网络中处在核心地位,是我国知识发现学术研究群体的核心成员。

表4 部分学者的程度中心性

作 者	节点 关系数	程度 中心性	作 者	节点 关系数	程度 中心性
张文修	45	0.978 3	潘云鹤	41	0.891 3
陆玉昌	44	0.956 5	吉根林	40	0.869 6
蔡庆生	43	0.956 5	高文	40	0.869 6
李德毅	43	0.934 8	邸凯昌	40	0.869 6
石纯一	42	0.934 8	焦李成	40	0.869 6
胡可云	41	0.913 0	施鹏飞	40	0.869 6
李德仁	41	0.891 3	刘宗田	40	0.869 6
欧阳为民	41	0.891 3	胡学钢	39	0.847 8
唐常杰	41	0.891 3	王树良	39	0.847 8
郑诚	41	0.891 3	刘君强	39	0.847 8

2.3.2 中介中心性

Ucinet软件分析结果显示(见表5),知识发现领域研究群体同被引矩阵网络的平均中介性是5.894,平均标准化中介性为0.569;张文修的中介中心性最高,达23.349,标准化中介性为2.256;其中有6位作者的标准化中介中心性大于1,说明在该领域这6位作者领导性程度较高,掌握资源能力较强,能够作为成员间联系的主要纽带,推动该领域的交流进程,加速创新发展。

表5 部分学者的中介中心性

作者	中介性	标准化的中介性	作者	中介性	标准化的中介性
张文修	23.349	2.256	胡可云	9.095	0.879
唐常杰	16.230	1.568	焦李成	8.805	0.851
欧阳为民	16.098	1.555	吉根林	8.453	0.817
陆玉昌	13.990	1.352	胡学钢	6.939	0.670
蔡庆生	12.829	1.240	陈恩红	6.815	0.658
石纯一	11.204	1.082	周傲英	6.576	0.635
刘明吉	10.157	0.981	李德仁	6.266	0.605
李德毅	9.863	0.953	刘宗田	6.054	0.585
郑诚	9.529	0.921	施鹏飞	6.004	0.580
潘云鹤	9.251	0.894	王树良	5.825	0.563

2.4 基于引证关系的作者与主题多重共现网络构建

选定某一学科的科技文献集合作为样本,用 $A=\{a_1, a_2, a_3, \dots, a_m\}$ 表示作者群体集合, $K=\{k_1, k_2, k_3, \dots, k_n\}$ 表示主题集合, $C=\{c_1, c_2, c_3, \dots, c_p\}$ 表示引文集合,则基于作者群体、主题和引文之间的关联关系可做出如下定义: $R=\{(a_i, k_t) | 1 \leq i \leq m, 1 \leq t \leq n\} \cup \{(a_i, c_v) | 1 \leq i \leq m, 1 \leq v \leq p\} \cup \{(k_t, c_v) | 1 \leq t \leq n, 1 \leq v \leq p\} \cup \{(a_i, k_t, c_v) | 1 \leq i \leq m, 1 \leq t \leq n, 1 \leq v \leq p\}$ 。该定义描述4种共现情况:作者群体 a_i 与主题 k_t 的共现;作者群体 a_i 与引文 c_v 的共现;主题 k_t 与引文 c_v 的共现;作者群体 a_i 与主题 k_t 、引文 c_v 的共现。

2.4.1 基于引证关系的作者与主题多重共现网络构建步骤

第一,选定某一领域,按照作者群体、主题和引文之间彼此的对应关系,识别每篇文献对应作者所属

子群得到每篇文献对应的作者群体、主题和引文关系表;第二,通过Bib Excel整理得到文献标号,分别与作者群体、主题和引文之间的对应关系,以及作者群体共现矩阵、主题共现矩阵和引文共现矩阵;第三,导入SQL Server,生成作者群体表、主题表、引文表,并借助SQL Server的查询功能和Excel的统计功能,对作者群体-主题-引文之间的关联关系进整理记录,得到作者群体-主题-引文共现频次表;第四,与主题-引文共现频次表、作者群体-主题共现频次表、作者群体-引文共现频次表逐一对应进行转换,生成作者群体-主题-引文共现矩阵;第五,共现矩阵导入Ucinet软件,对作者群体-主题-引文矩阵进行可视化操作,生成多重共现超网络。

2.4.2 实例应用

选取群体C陆玉昌、张文修、胡可云3位作者以及其在知识发现领域相关的主题词和引文进行实例展示。将陆玉昌、张文修和胡可云作为单个个体进行研究,形成的作者-主题-引文的多重共现超网络模型的可视化图谱(见图1),在异构层中,通过不同基础特征项之间的联系,构成局部网络,以揭示异质节点之间的关联。

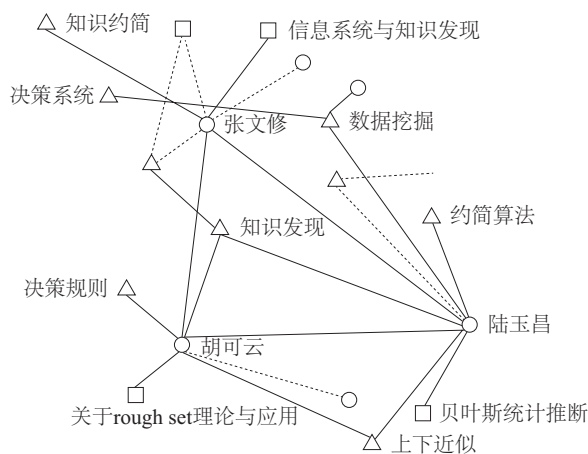


图1 作者-主题-引文网络

从图1可直接发现作者、主题与引文之间的关联关系,能够表示彼此之间在研究方向上的关联程度。借助该图谱进行检索,当出现网络中的某一节点时,需通过关联关系,层层筛选得到全部文献。可见,利用多重共现超网络的层次性结构,在检索过程中能够实现检索信息的快速聚集。当输入“知识发现”“胡可云”,检索结果见表6所示。

表6 基于作者-主题-引文网络检索结果

文献名称	作者
《粗糙集理论及其应用进展》	胡可云、陆玉昌、石纯一
《基于粗糙集理论的知识发现综述》	王志海、胡可云、胡学钢、徐本柱、刘宗田

以上两篇文献为初次检索所得结果,利用图1可关联得到陆玉昌的《贝叶斯统计推断》、张文修的《信息系统与知识发现》,但需要层层发现。

若将陆玉昌、张文修和胡可云可作为一个研究群体,将其与主题和引文关联进行可视化操作,得到基于作者群体-主题-引文的多重共现超网络,如图2所示。作者群体替代作者作为构建网络的基础单元,与主题、引文进行关联,直接展示整个群体的研究方向。

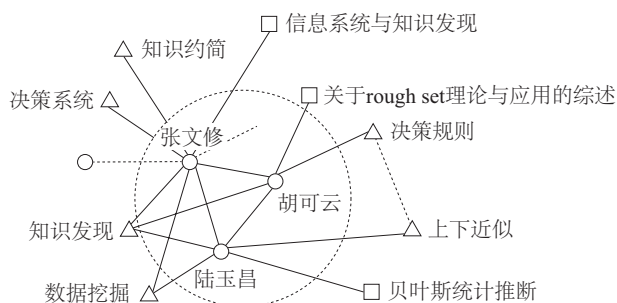


图2 作者群体-主题-引文网络

利用图2检索时,检索类目“作者”以群体作为检索单元,结果以“群体模式”呈现,省去层层筛选的时间,能够对检索过程起到优化作用。当输入“知识发现”“胡可云”,检索结果见表7所示。

表7 基于作者群体-主题-引文网络检索结果

文献名称	作者
《粗糙集理论及其应用进展》	胡可云、陆玉昌、石纯一
《基于粗糙集理论的知识发现综述》	王志海、胡可云、胡学钢、徐本柱、刘宗田
《基于支持向量机的决策系统知识发现》	魏玲、祁建军、张文修
《知识系统与知识发现》	张文修、徐萍、米据生、吴伟志
《贝叶斯网络的建造及其在数据挖掘中的应用》	林士敏、田凤占、陆玉昌
《Bayesian方法的计算学习机制和问题求解》	林士敏、王双成、陆玉昌
《数据挖掘与知识发现》	陆玉昌

以上检索结果均为初次所得,由于3位作者是一个整体,在同一领域检索时,将按照其关联关系主动显示全部结果,避免盲目查找,一次实现该领域相关作者以及其著作的集合。

3 总结

基于引证关系,本文对知识发现领域研究群体的可视化知识图谱进行分析,直观呈现该群体成员的研究方向分布、学术地位等级及信息关联等情况。随后,在此基础上基于引证关系构建作者群体与主题的多重共现超网络模型,验证将二次凝练的因子作为基础特征项进行应用的效果,实现检索信息的快速聚集。

参考文献

- [1] 庞弘桑,方曙. 基于多重共现的可视化分析工具设计及其知识发现方法研究[J]. 图书情报知识, 2012(2): 100-107.
- [2] 曹树金,吴育冰,韦景竹,等. 知识图谱研究的脉络、流派与趋势——基于SSCI与CSSCI期刊论文的计量与可视化[J]. 中国图书馆学报, 2015, 41(5): 16-34.
- [3] BORNER K, CHEN C, BOYACK K W. Visualizing knowledge domains[J]. Annual Review of Information Science and Technology, 2003, 37(1): 179-255.
- [4] O'DONNELL A M, DANSEREAU D F, HALL R H. Knowledge maps as scaffolds for cognitive processing[J]. Educational Psychology Review, 2002, 14(1): 71-86.
- [5] VAN ECK N J, WALTMAN L. Software survey: VOS viewer, a computer program for bibliometric mapping[J]. Scientometrics, 2010, 84(2): 523-538.
- [6] BOYACK K W, KLAIVANS R. Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately?[J]. Journal of the American Society for Information Science and Technology, 2010, 61(12): 2389-2404.
- [7] PORTER A L, RAFOLS I. Is science becoming more interdisciplinary? Measuring and mapping six research fields overtime[J]. Scientometrics, 2009, 81(3): 719-745.
- [8] 陈悦,刘则渊. 悄然兴起的科学知识图谱[J]. 科学学研究, 2005, 23(2): 149-154.
- [9] 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.

- [10] 陈悦, 刘则渊, 陈劲, 等. 科学知识图谱的发展历程 [J]. 科学学研究, 2008, 26 (3): 449-460.
- [11] 秦长江, 侯汉清. 知识图谱——信息管理与知识管理的新领域 [J]. 大学图书馆学报, 2009, 27 (1): 30-37.
- [12] 赵蓉英, 王菊. 图书馆学知识图谱分析 [J]. 中国图书馆学报, 2011, 37 (2): 40-50.
- [13] 庞弘燊, 方曙, 杨志刚, 等. 研究领域的主题发展趋势分析方法研究——基于多重共现的视角 [J]. 情报理论与实践, 2012, 35 (8): 44-47, 73.
- [14] 郭秋萍, 梁梦丽, 刘秀丽, 等. 基于作者-关键词-引文多重共现的超网络知识关联研究 [J]. 情报理论与实践, 2016, 39 (7): 20-26.
- [15] 郭红梅, 张智雄. 基于多重文本术语关系叠加识别文本核心主题的有效性探索 [J]. 情报学报, 2017, 36 (11): 1157-1164.
- [16] LEE K, JUNG H, SONG M. Subject-method topic network analysis in communication studies [J]. Scientometrics, 2016, 109 (3): 1761-1787.
- [17] LEYDESDORF L. What Can Heterogeneity Add to the Scientometric Map? Steps towards algorithmic historiography [EB/OL]. [2019-09-01]. <https://arxiv.org/abs/1002.0532>.
- [18] 张静. 引文、引文分析与学术论文评价 [J]. 社会科学管理与评论, 2008 (1): 33-38.

作者简介

刘爱琴, 女, 1974年生, 博士, 副教授, 硕士生导师, 研究方向: 信息技术与信息服务, E-mail: km_aql@sina.com。
吴瑞瑞, 女, 1996年生, 硕士研究生, 研究方向: 信息技术与信息服务。

Construction of Multiple Co-occurrence Network of Author and Subject Based on Citation Relationship

LIU AiQin WU RuiRui
(School of Economics and Management, Shanxi University, Taiyuan 030006, China)

Abstract: Using visualization tools to describe knowledge resources and their carriers, deeply mine knowledge content and its structural relationship, and represent the association between knowledge are important ways to effectively improve the efficiency and accuracy of information retrieval and realize the innovation of database service mode. Based on the data of CNKI Chinese Citation Database and the citation relationship, this paper constructs the knowledge map and cluster analysis of the research groups in the field of knowledge discovery. On the one hand, it shows the relationship between groups and the overall knowledge structure, and excavates the hypernetwork model among authors, keywords and citations in the field of knowledge discovery. It proves the effect of using the secondary condensation factor as the basic feature item, improves the recall and precision of information retrieval, and makes up for the shortcomings of information retrieval.

Keyword: Knowledge Map; Knowledge Association; Multiple Co-occurrence; Citation Relationship

(收稿日期: 2019-09-05)