

运用重叠社群可视化软件 CFinder 分析学科交叉研究主题*

——以情报学和计算机科学为例

■ 李长玲 刘非凡 郭凤娇

【摘要】针对学科交叉领域知识重叠问题,以情报学和计算机科学为例,检索两学科主要交流期刊的互引论文,用 Bibexcel 对论文关键词进行处理,综合运用统计分析方法和重叠社群可视化软件工具 CFinder,分析两学科交叉关键词的词频分布和 k-clique 数量分布规律;对两学科中知识聚类 and 重叠社群网络进行可视化展示,并对两学科之交叉研究主题进行分析。

【关键词】重叠社群 CFinder 社会网络分析 学科交叉 研究热点

【分类号】G353

科学史表明,科学经历了综合、分化、再综合的过程,现代科学则既高度分化又高度综合。交叉科学集分化与综合于一体,实现了科学的整体化。学科交叉点往往是科学研究新的生长点,这里最有可能产生重大的科学突破,使科学发生革命性的变化^[1]。因此,在 multidisciplinary 融合的科学时代,学科交叉研究显得尤为重要,它不仅有利于挖掘新的科学生长点,而且还有助于进一步了解科学知识体系的变化和发展。近年来,国内外越来越多的学者关注学科交叉的研究。

A. L. Porter 于 1985 年提出了三个指标公式,用以分析学科的交叉程度^[2]。2009 年,他又通过文献计量指标和科学知识图谱对 1975 - 2005 年间 6 个研究领域的数据进行分析后发现:学科交叉程度随时间的推移越来越强^[3]。H. Small 通过对共引数据进行聚类,分析学科间的交叉性和相似性^[4]。张春美等从学科交叉研究出发,以诺贝尔奖自然科学奖获奖成果和获奖者为对象,运用统计分析方法,分析了诺贝尔奖中学科交叉的变化发展趋势^[5]。李春景等从学科交叉的构成要素出发,以学科为逻辑前提与理论始点,通过建立一种学科交叉模式的分析框架,试图揭示出学科交叉模式特征与规律^[6]。吴丹青等通过学科交叉的组织管理研究,初步总结出学科交叉的模式可分为交流型、方法型、项目型和平台型^[7]。朱蔚彤运用贝克尔 - 墨菲的

模型研究了影响学科交叉研究产出的因素,深入分析了影响学科交叉研究产出的协调成本问题,对学科交叉研究管理机制进行了实例研究^[8]。魏建香运用文档聚类技术手段,发现和展示学科之间的交叉知识^[9]。李长玲等以情报学等 5 个学科的期刊互引网络数据为样本,分析 5 个学科之间的交叉程度^[10]。张洪磊等利用社会网络分析方法和 Ucinet 工具,对情报学与计算机科学两个学科进行学科交叉实证研究^[11]。

可以看出,当前学者主要针对宏观或微观的学科交叉程度及变化趋势进行研究,鲜有文献进行学科交叉主题的分析。因此,笔者尝试应用复杂网络重叠社群发现软件工具 CFinder 挖掘情报学和计算机科学的交叉研究主题,试图更加准确、具体、清晰地呈现学科间交叉网络关系。这对于了解情报学与计算机科学的融合现状有重要意义,也为两学科的交叉主题研究提供了新的研究方法。

1 重叠社群与 CFinder 软件

1.1 重叠社群

社群是有共同特征的相关节点及连接关系形成的群落。真实网络中的每个节点往往同时属于多个具有不同特征的群落。另一方面,社群间的共同特征也导致社群重叠现象的产生。在网络结构中则表现为两个

* 本文系国家社会科学基金项目“基于社会网络分析的学科评价研究”(项目编号:11BTQ020)和山东理工大学人文社会科学基金发展基金项目“Web 信息挖掘与智能检索”(项目编号:2010GGTD05)研究成果之一。

【作者简介】李长玲,山东理工大学科技信息研究所教授,硕士生导师,E-mail:lichl@sdlu.edu.cn;刘非凡,山东理工大学科技信息研究所硕士研究生;郭凤娇,山东理工大学科技信息研究所硕士研究生。

收稿日期:2012-11-26 修回日期:2013-03-01 本文起止页码:75-80 本文责任编辑:刘远颖

或者多个社群共同拥有一个或者多个节点,具有这种特征的社群称之为重叠社群。重叠节点与各个社群内部的关系都相对紧密,可以很直观地观察到,而重叠社群之间因为重叠节点发生的联系,却不易被察觉^[12]。所以,重叠节点在网络中往往起黏合或者过渡的作用。

1.2 CFinder 软件及其算法

CFinder 是 G. Palla 等基于 CPM 算法开发的一个自由软件^[13-14],它不仅能非常有效地定位和可视化处理大规模稀疏网络社群,而且还可用于定量描述社会网络的演变。

发现重叠社群的派系过滤 CPM 算法 (clique percolation method),可以快速高效处理并呈现复杂网络数据。派系是任意两点都相连的顶点的集合,即完全子图。在社群内部节点之间连接密切,边密度高,容易形成派系 (clique)。因此,社群内部的边有较大可能形成大的完全子图,而社群之间的边却几乎不可能形成较大的完全子图,从而可以通过找出网络中的派系来发现社群。同时,一个节点所在的派系可能属于不同的社群,因此派系过滤方法可以发现重叠社群。CPM 算法的主要思想是先找到事先指定规模的完全子图,然后通过扩展或相邻派系进行合并,得到重叠的社群结构。

一方面,派系稀疏的网络,派系过滤难以给出有效的社群发现结果。另一方面,如果网络中的派系非常密集,该类算法将把整个网络视为一个社群。对于稀疏网络与密集网络,CFinder 软件通过选取合适的查找社群的派系规模参数 K ,使社群网络达到较好的可视化效果。

关键词是表征论文研究主题的词汇,能够简单地反映论文核心研究内容。学科关键词的集合涵括某一学科领域主要研究内容,因此,两个或多个学科的交叉关键词能够反映学科交叉的研究内容。笔者在情报学与计算机科学的主要交流期刊中,检索两学科的交叉文献,对交叉文献的关键词网络,用重叠社群分析软件 CFinder 进行可视化,寻找情报学与计算机科学的主要交叉研究领域与潜在研究主题,为情报学的深入发展与计算机科学的广泛应用提供借鉴。

2 运用 CFinder 分析学科交叉研究主题

2.1 数据采集和预处理

笔者在文献[15]中,通过社会网络中心性分析方法,研究发现情报学与计算机科学交流中起桥梁作用的6种主要期刊——《情报学报》、《现代图书情报技术》、《情报杂志》、《计算机工程与应用》、《计算机工

程》和《计算机应用研究》。在 CNKI 山东理工大学镜像版数据库中,时间跨度选择 2006-2010 年,分别在“刊名”与“参考文献”检索项输入两学科的不同刊物名称,逻辑运算用“并且”,采用“精确”检索,检索时间为 2012 年 3 月 2 日,检索情报学与计算机学科主要期刊之间的互引文献,共 889 篇,以此为研究对象,选用复杂网络重叠社群可视化分析软件 CFinder,对检索结果进行处理、分析与可视化,更形象、直观地研究情报学与计算机科学的交叉研究主题与潜在研究方向。

对检索到的文献的处理过程如下:①由于同一篇文章会被相同或不同期刊的不同文献所引用,因此,需要将检索到的文献进行去重处理。对检索到的 889 篇学科交叉文献(其中情报学引用计算机科学文献 703 篇,计算机科学引用情报学文献 186 篇)进行去重处理,最后得到情报学文献 687 篇、计算机科学文献 179 篇。②对去重后的文献集合,分别提取情报学与计算机科学交叉文献的关键词,用 Bibexcel 进行关键词处理与统计分析,共得到情报学研究文献关键词 1 491 个,计算机文献关键词 366 个。③分别对两学科关键词及共现关键词进行分组和编号。通过自编程序找出同时出现在计算机科学和情报学两学科的关键词共 106 个,相关数据见表 1 第 1 栏;将剩余情报学学科关键词 1 385 个(1 491-106)分为 B 组,并按词频降序排列并编号,见表 1 第 2 栏;将计算机科学剩余学科关键词 260 个分为 C 组并编号,见表 1 第 3 栏。④由于样本数据来源的原因,B 组与 C 组关键词数量差别较大,如果采用绝对数进行分析,可能会出现偏差。为了更好地呈现两学科的交叉情况,使网络可视化效果较好地反映整体情况,对 A 组内的两学科交叉关键词进行加权处理:选取情报学关键词数量 $n_b = 1\,491$ 和计算机科学关键词数量 $n_c = 336$ 占总关键词数量的比重作为权值,即:

$$\alpha = \frac{n_c}{n_b + n_c} \quad \beta = \frac{n_b}{n_b + n_c} \quad (1)$$

再将权重分别乘以各自绝对词频数量 $FinB$ 和 $FinC$ 得到相对词频 $\alpha FinB$ 和 $\beta FinC$,再求和得到每个关键词的相对词频数值:

$$FsumA2 = \alpha * FinB + \beta * FinC \quad (2)$$

根据 $FsumA2$ 降序依次给 A 组关键词排序编号为 $A1, A2, \dots, A106$,见表 1 第 1 列。

最后,由于 CFinder 软件不能处理中文数据,因此将去重后所有文献中的关键词用表 1 中对应的编号替换。运用 Bibexcel 软件统计两学科文献关键词共现词

表1 情报学和计算机科学文献关键词分组编号(部分)

学科共现关键词 (A 组)								情报学 (B 组)				计算机科学 (C 组)			
编号	关键词	FinB	FinC	FsumA1	α FinB	β FinC	FsumA2	编号	关键词	词频	编号	关键词	词频		
A1	本体	47	10	57	9.4	8	17.4	B1	领域本体	14	C1	供应商选择	3		
A2	信息检索	23	4	27	4.6	3.2	7.8	B2	图书馆	8	C2	组合预测	3		
A3	搜索引擎	14	6	20	2.8	4.8	7.6	B3	语义相似度	7	C3	多属性群决策	3		
A4	数据挖掘	23	3	26	4.6	2.4	7	B4	知识组织	6	C4	检索	2		
A5	文本分类	13	5	18	2.6	4	6.6	B5	个性化服务	6	C5	图书采购	2		
A6	数字图书馆	29	1	30	5.8	0.8	6.6	B6	语义	5	C6	推理机	2		
A7	文本聚类	8	5	13	1.6	4	5.6	B7	网络	5	C7	TFIDF	2		
A8	向量空间模型	14	3	17	2.8	2.4	5.2	B8	服务质量	5	C8	分词算法	2		
A9	知识管理	18	2	20	3.6	1.6	5.2	B9	WEB挖掘	5	C9	软件项目	2		
A10	XML	18	2	20	3.6	1.6	5.2	B10	知识服务	5	C10	协作过滤	2		
A11	信息抽取	15	2	17	3	1.6	4.6	B11	知识网络	5	C11	面向服务的体系结构	2		
A12	相似度	2	5	7	0.4	4	4.4	B12	图像检索	5	C12	ELUCENE	2		
A13	中文分词	5	4	9	1	3.2	4.2	B13	RSS	5	C13	PAGERANK算法	2		
A14	特征选择	5	4	9	1	3.2	4.2	B14	数据清洗	4	C14	描文本	2		
A15	AGENT	5	4	9	1	3.2	4.2	B15	知识本体	4	C15	网络演化模型	1		
A16	聚类	9	3	12	1.8	2.4	4.2	B16	信息组织	4	C16	网页分块	1		
A17	电子商务	13	2	15	2.6	1.6	4.2	B17	系统设计	4	C17	网页排名	1		
A18	支持向量机	4	4	8	0.8	3.2	4	B18	知识抽取	4	C18	推荐算法	1		
A19	WEB服务	7	3	10	1.4	2.4	3.8	B19	P2P	4	C19	稀疏	1		
A20	粗糙集	4	3	7	0.8	2.4	3.2	B20	条件随机场	4	C20	网络服务	1		
A21	LUCENE	4	3	7	0.8	2.4	3.2	B21	个性化信息服务	4	C21	切分路径	1		
A22	知识管理系统	3	3	6	0.6	2.4	3	B22	查询扩展	4	C22	网页排序	1		
...		
A106	语义元数据	1	1	2	0.2	0.8	1	B1385	灰色系统理论	1	C260	基于高层语义的视频检索	1		

频,再将数据转换成 CFinder 软件可识别的纯文本文档格式,为下一步进行的可视化分析做好数据准备。关键词共现词频统计部分数据如表2所示:

表2 关键词共现词频统计(部分)

关键词1	关键词2	词频	编号1	编号2	词频
关联规则	数据挖掘	8	A24	A4	8
数字图书馆	个性化服务	6	A6	B5	6
本体	信息检索	6	A1	A2	6
数据挖掘	聚类	5	A4	A16	5
XML	半结构化	5	A10	B69	5
本体	知识库	5	A1	A41	5
信息体系结构	技术体系结构	5	B154	B101	5
本体论	信息检索	5	A43	A2	5
信息网络	研究进展	4	B151	B76	4
知识管理	知识本体	4	A9	B15	4
数字图书馆	信息抽取	4	A6	A11	4
GATE	信息抽取	4	B168	A11	4
本体	语义相似度	4	A1	B3	4
信息抽取	本体	4	A11	A1	4
粗糙集	属性约简	4	A20	A48	4
本体	信息检索系统	4	A1	B54	4
知识抽取	自然语言处理	4	B18	A34	4
文本表示	向量空间模型	4	A42	A8	4
语义 WEB	本体	4	B36	A1	4
电子商务	数据挖掘	4	A17	A4	4
本体	知识组织	4	A1	B4	4
本体	领域本体	4	A1	B1	4
向量空间模型	信息检索	4	A8	A2	4

2.2 数据结果分析

本文从情报学与计算机科学交叉关键词词频分析、k-clique 数量分析和网络可视化分析三个方面探索发现学科交叉研究内容。

2.2.1 两学科交叉关键词词频分析 对情报学和计算机科学文献中共同出现的 A 组 106 个关键词按标准化后的数据(表1中第1栏的6-8列)进行词频分析,并比较 相关结果见图1-图3。

情报学和计算机科学交叉关键词词频分布规律:在被检两学科文献中,两交叉学科共同出现关键词

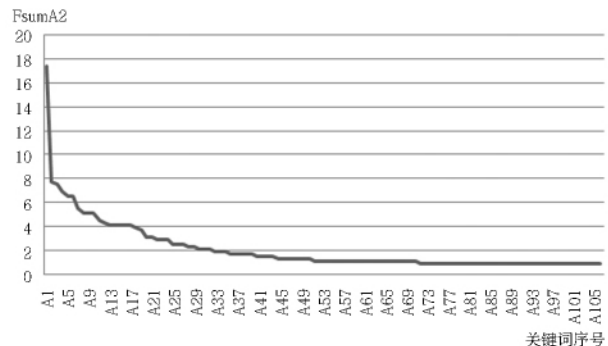


图1 词频分布

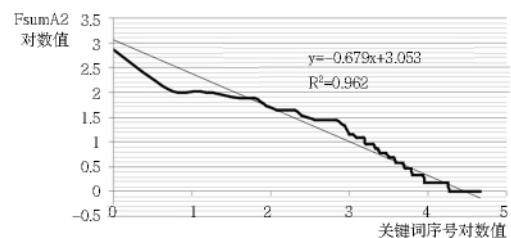


图2 词频-序号对数分布

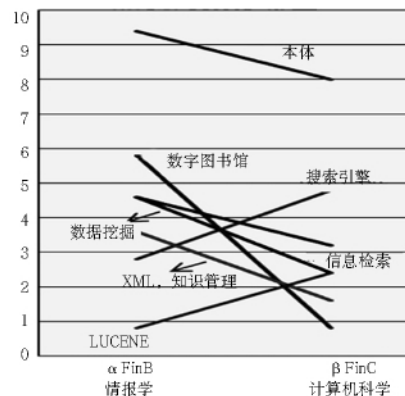


图3 两学科共现关键词加权词频分布

FsumA2 词频分布图见图1,词频-序号对数分布见图

2. 从图2中可以看出,词频-序号对数分布曲线近似一条直线,回归系数为0.962,接近于1,说明学科交叉关键词的词频分布基本符合齐普夫定律,与学科内部关键词的分布一样存在词频分布的集中与离散规律。

在图3中,每一根线段代表一个关键词。线段的端点分别代表情报学与计算机科学的该关键词标准化后的加权词频 αFinB 和 βFinC ,线段的陡峭程度和加权词频数值能够在一定程度上反映这一概念在两学科中的研究程度。例如,“本体”线段远高于其他线段,表明本体在两学科当中的交叉研究程度高于其他关键词。“数字图书馆”线段的斜率在所有线段斜率中最高,说明这一概念在情报学学科中的研究热度远远高于计算机科学的研究热度,而搜索引擎、LUCENE 在计算机科学领域的研究比情报学领域的研究更有优势。

2.2.2 两学科交叉关键词的 k-clique 数量统计分析

k-clique 派系表示网络中含有 k 个节点的完全子图,如果一个 k 派系通过若干个相邻的 k 派系到达另一个 k 派系,则称这两个 k 派系是连通的。由所有彼此连通的 k 派系构成的集合就是一个 k 派系社群。学科交叉关键词 k-clique 的数量表示该关键词连接派系的数量,是学科交叉程度的重要标志,某关键词的 k-clique 数量越多,表明其在学科交叉中的作用越大。

将情报学与计算机科学两学科文献关键词标准化后的关系数据矩阵(表2数据)导入 CFinder 软件,对 A 组学科交叉节点关键词所形成的 k-clique 派系数量进行统计分析,结果如表3、图4-图5所示:

表3 学科共现关键词 k-clique 的数量

关键词	k-clique 数量	关键词	k-clique 数量
本体	69	聚类	22
数字图书馆	40	文本聚类	19
数据挖掘	36	关联规则	19
XML	34	WEB 服务	15
信息检索	31	信息系统	15
知识管理	28	相似度	14
电子商务	27	中文分词	14
信息抽取	25	特征选择	14
搜索引擎	24	支持向量机	14
文本分类	24	神经网络	13
向量空间模型	22	元数据	13

情报学和计算机科学共现关键词 k-clique 数量分布规律:在被检索到的两学科间起桥梁作用的文献中,两交叉学科共现关键词 k-clique 的数量按递减的顺序排列,见表3。从表中可以看出本体、数字图书馆、数据挖掘、XML、信息检索等关键词形成的 k-clique 的数量相对较多,说明它们在学科与学科交叉中的作用较大。但结合图3可以看出:情报学对“数字图书馆”的研究更为深入。图4为关键词 k-clique 的数量定律,图

5为序号与 k-clique 的数量的对数分布,从分布结果可以看出:关键词的 k-clique 的数量分布也近似服从齐普夫分布,并存在集中与离散的规律。

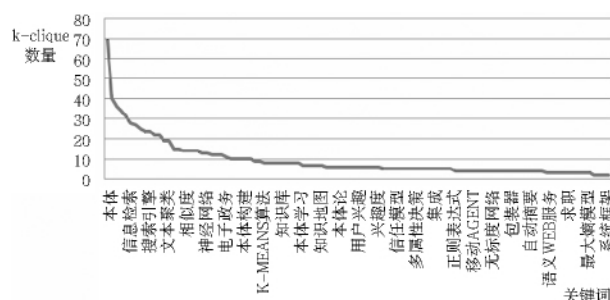


图4 学科共现关键词 k-clique 数量分布

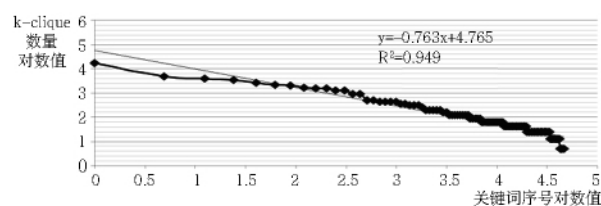


图5 学科共现关键词 k-clique 数量对数分布

2.2.3 两学科交叉关键词重叠社群网络可视化 根据 CFinder 软件算法原理,重叠的社群的形成实质上是将有重叠节点的社群进行扩展与合并。对于共词网络来说,由于社群之间节点关键词共现,社群与社群形成重叠关系,最终形成一个关系网络图。

综合考虑 CPM 算法的性能、CFinder 在不同阈值下所形成的重叠社群数量、重叠社群网络规模等因素,本文选择关键词共现网络阈值为3、派系规模 $k=5$ 进行研究。将两学科共词矩阵(数据见表2)输入 CFinder 软件,设定阈值3、 $k=5$,形成的最大社群网络如图6所示:

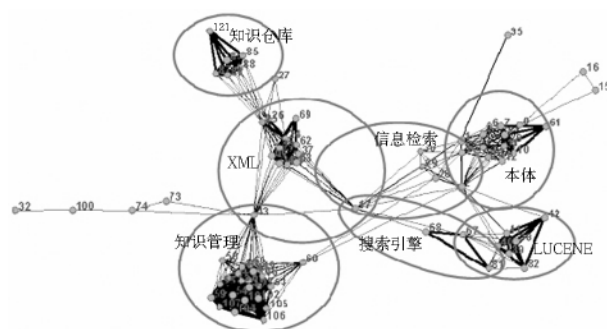


图6 阈值为3的5-clique最大重叠社群网络

图6中,一个节点代表一个5核社群。较多重叠节点(关键词)所形成的社群,表明其研究内容更具有相关性,在图6中表现为高度或相对集中在一起,形成一个重点研究主题,即知识管理、信息检索、本体、XML、知识仓库、LUCENE、搜索引擎等是情报学与计算机科学的主要交叉研究内容。这些研究主题之间通过

桥梁社群相连接,例如社群节点17、43、2等,这些主要5核桥梁社群的关键词节点构成如图7所示:



图7 桥梁社群的关键词节点组成

由图7可知,A1(本体)、A2(信息检索)、A3(搜索引擎)、A6(数字图书馆)、A8(向量空间模型)、A9(知识管理)、A10(XML)等是桥梁社群的主要构成元素,也是情报学与计算机科学交叉研究的主要内容。

对图6的社群网络图进行关键词节点展示,如图8所示:

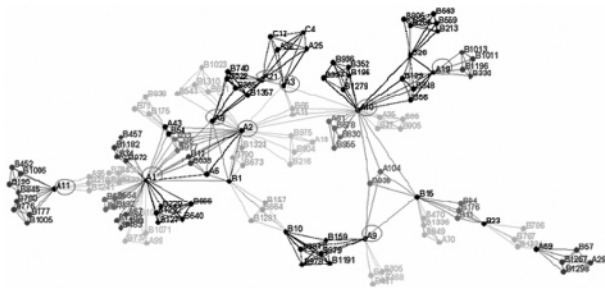


图8 阈值为3时5-clique最大重叠社群
关键词节点网络

图8为情报学与计算机科学交叉共现关键词节点社群网络,图中圈注了起到桥梁作用的学科关键词节点,例如A1、A2、A10、B1、B10等,由于A组关键词是情报学和计算机科学共同出现的关键词,同时,本文分析的重点是情报学与计算机科学交叉研究内容,因此以下的分析也以A组关键词为主。图中A组核心关键词共9个,分别为:A1(本体)、A2(信息检索)、A3(搜索引擎)、A8(向量空间模型)、A9(知识管理)、A10(XML)、A11(信息抽取)、A19(WEB服务)、A21(LUCENE)。一方面,从网络上能够明显地看出,这些关键词的周围往往连接着多个相关社群,并且在这些社群中处于中心位置,体现出这些研究主题内在的相关性。另一方面,核心关键词之间也存在连接关系,比如,A1(本体)与A2(信息检索)、A8(向量空间模型)、A11(信息抽取)之间有直接连接关系,说明本体与这些研究主题之间内容关联更紧密。

2.2.4 情报学和计算机科学交叉学科研究主题分析

纵观以上分析,不论是表1、图1-图5所示的情报学与计算机科学交叉研究关键词的词频与社群数量分布,还是图6-图8所示的社群网络可视化结果,关于

两学科的交叉研究内容都可以得出如下结论:①本体与信息检索是两学科交叉研究的主要内容;②知识管理与XML虽然两学科都有研究,但就本文的样本数据来看,情报学研究更广泛;③计算机科学研究人员在搜索引擎与LUCENE方面的研究比情报学研究学者更加深入。通过对上文结论与数据进行分析得出各研究主题在不同领域的研究侧重点,具体内容如表4所示:

表4 情报学和计算机科学交叉学科研究主题分析

主题	含义	情报学研究领域	计算机科学研究领域
本体	本体是对一个领域里的概念及其关系的清晰描述,是人与人之间、人与计算机之间对领域知识达成共同理解的桥梁 ^[16] 。	知识组织、领域本体、数字图书馆个性化等。	知识共享、异构数据源语义互操作、供应链系统知识集成、信息检索和本体构建等。
信息检索	信息检索是从大量被存贮的信息中加工、检索出需要的信息以及向计算机用户提供一整套信息的工作。	语义网检索,信息检索的可视化、本科信息检索课程设置等。	WEB信息检索排序函数及标引词加权技术、信息检索模型构建、信息检索算法的查找速度和查准率等。
XML	可扩展标记语言(Extensible Markup Language,XML),用于标记电子文件使其具有结构性的标记语言,可以用来标记数据、定义数据类型,是一种允许用户对自己的标记语言进行定义的源语言 ^[17] 。	XML数据的转换与查询;XML语言在WEB数据挖掘中的应用。	在LDAP目录服务上实现对XML数据的XPath查询,提高XML数据加密的效率和通用性。
知识管理	我国学者邱均平、段宇锋认为,“知识管理”的定义可以从狭义和广义角度理解。狭义的知识管理主要是针对知识本身的管理;广义的知识管理不仅包括知识本身,还包括对与知识有关的各种资源和无形资产的管理,涉及知识组织、知识设施、知识活动、知识人员等的全方位、全过程的管理 ^[18] 。	图书馆的知识服务模式、知识管理与知识创新、知识管理与竞争情报研究等。	动态知识管理模型和系统、知识集成技术和知识管理平台构建等。
搜索引擎	搜索引擎是计算机研究人员开发出来的对互联网上的海量资源进行快速而准确的检索的工具。	个性化信息服务、网络信息资源组织、学术搜索引擎和链接分析等。	搜索引擎的原理探讨、搜索引擎设计和技术实现、排序算法和搜索策略研究。
LUCENE	LUCENE是一套由Apache软件基金会支持和提供,用于全文检索和搜索的开源程式库 ^[19] 。	搜索引擎设计和效率测试,处理中文信息的中文分词模块等。	全文搜索引擎的研究和开发、信息检索、语义检索。

3 结 语

本文在前期研究成果的基础上,以情报学与计算机科学为例,在两学科主要的交流期刊中检索起桥梁作用的文献,通过对检索到的文献中情报学和计算机科学共现关键词进行词频统计分析,运用CFinder软件进行k-clique数量分析和社群网络可视化,发现本体、信息检索、搜索引擎、XML、LUCENE、知识管理等是两

学科共同的研究主题,但两学科的研究内容各有侧重,这与文献[11]的研究结论类似。因此,用重叠社群分析软件 CFinder 分析学科交叉研究主题应该是可行、有效的。

从国内关于学科交叉方面的研究来看,鲜有社群可视化软件 CFinder 的应用,本文对该软件在这方面研究的尝试尚属首次。因为样本文献的来源仅限于两学科交流的主要期刊,不能涵盖学科交叉的全部文献并且阈值的选取造成部分数据值的损失,软件自身处理数据所需的硬件环境等因素,可能会使分析结论有一定的片面性。但从另一个角度来说,由于文献具有高度针对性,可能不仅不会影响总体评价结果,还会使学科交叉分析结果更加明了。

因此,扩大样本范围,考虑时间因素,探索学科交叉关键词社群网络的延伸和演化规律,是本文后续研究的主要内容。

参考文献:

- [1] 路甬祥. 学科交叉与交叉科学的意义[J]. 中国科学院院刊, 2005, 20(1): 63-65.
- [2] Porter A L, Chubin D E. An indicator of cross-disciplinary research [J]. *Scientometrics*, 1985, 8(3-4): 161-176.
- [3] Porter A L, Rafols I. Is science becoming more interdisciplinary? Measuring and mapping six research fields over time [J]. *Scientometrics*, 2009, 81(3): 719-745.
- [4] Small H. Maps of science as interdisciplinary discourse: Co-citation contexts and the role of analogy [J]. *Scientometrics*, 2010, 83(3): 835-849.
- [5] 张春美, 郝凤霞, 闫宏秀. 学科交叉研究的神韵——百年诺贝尔自然科学奖探析[J]. *科学技术与辩证法*, 2001, 18(6): 63-67.
- [6] 李春景, 刘仲林. 现代科学发展学科交叉模式探析——一种学科交叉模式的分析框架[J]. *科学学研究*, 2004, 22(3): 244-248.
- [7] 吴丹青, 张菊, 赵杭丽, 等. 学科交叉模式及发展条件[J]. *科研管理*, 2005, 26(5): 159-162.
- [8] 朱蔚彤. 学科交叉研究的协调成本分析[J]. *中国软科学*, 2008(8): 152-158.
- [9] 魏建香. 学科交叉知识发现及其可视化研究[D]. 南京: 南京大学, 2010.
- [10] 李长玲, 纪雪梅, 支岭. 基于 E-I 指数的学科交叉程度分析——以情报学等 5 个学科为例[J]. *图书情报工作*, 2011, 55(16): 33-36.
- [11] 张洪磊, 魏建香, 杜振东, 等. 基于社会复杂网络的学科交叉研究[J]. *情报杂志*, 2011, 30(10): 29-33, 58.
- [12] 王熙. 复杂网络中的层次重叠社区发现及可视化[D]. 北京: 北京交通大学, 2010.
- [13] Palla G, Derényi I, Farkas I. Uncovering the overlapping community structure of complex networks in nature and society [J]. *Nature*, 2005, 435(7043): 814-818.
- [14] Adamcsek B, Palla G, Farkas I J. CFinder: Locating cliques and overlapping modules in biological networks [J]. *Bioinformatics*, 2006, 22(8): 1021-1023.
- [15] 李长玲, 支岭, 纪雪梅. 基于中心性分析的学科期刊地位评价——以情报学等 3 学科为例[J]. *情报理论与实践*, 2012, 35(6): 49-53.
- [16] 谷琦. 网络信息资源组织管理与利用[M]. 北京: 科学出版社, 2008.
- [17] XML[OL]. [2012-03-22]. <http://baike.baidu.com/view/63.htm>.
- [18] 邱均平, 段宇锋. 论知识管理与竞争情报[J]. *图书情报工作*, 2000, 44(4): 11-14.
- [19] Lucene[OL]. [2012-03-22]. <http://baike.baidu.com/view/371811.htm>.

Analysis on Interdisciplinary Research Topics with Cinder of Overlapping Communities Visualization Software ——Taking the Information Science and Computer Science for Example

Li Changling Liu Feifan Guo Fengjiao

Science and Technology Information Research Institute of Shandong University of Technology, Zibo 255049

[Abstract] According to the problem of knowledge overlapping in the interdisciplinary field, this paper takes the information science and computer science for example, retrieves the cited papers of their main exchanging journals and deals with the keywords by Bibexcel. By the methods of statistical analysis and CFinder of overlapping community visualization software, it analyzes the keywords frequency distribution of two interdisciplinary and the k-clique distribution law, and visualizes the knowledge clustering and overlapping social networks. Finally, this paper analyzes their interdisciplinary research topics.

[Keywords] overlapping community CFinder SNA interdisciplinary research topic