DOI:CNKI:11-2746/G2.20121210.1700.001

引文网络中文献深度聚合方法与实证研究
——以 WOS 数据库中 XML 研究论文为例*

邱均平　董　克

摘　要　科学文献之间通过引用关系构成了特定研究主题的知识网络,其单向无回路的特征揭示了学科主题的知识结构和发展过程。本文以 WOS 数据库中 XML 研究论文所构成的引文网络为例,利用引文关系权重与文献节点权重确定核心文献,并在此基础上从阈值和权值"高地"两个角度对核心文献进行聚合。研究发现:文献核心程度的确定过程充分考虑了不同引用实质上的重要程度区别,据此计算得到的引文和文献节点权重能够准确反映文献的质量;阈值聚合能够迅速发现整个学科发展过程中最核心的文献和引文;权值"高地"聚合分析结果更为多样,并能弥补阈值聚合在揭示次重要子结构方面的不足,发现整个知识体系发展过程中丰富的研究维度。图 6。表 4。参考文献 11。

关键词　引文网络　核心文献　深度聚合　权值"高地"

分类号　G354

Methods and Empirical Research on Deep Integration of Literature in Citation Network: Case Study on XML Research Literature from WOS

Qiu Junping & Dong Ke

ABSTRACT　Interconnected by citation, scientific literature construct knowledge network of researches on specific subjects. As an acyclic network, the knowledge network reveals the knowledge structure and development of research subjects. This paper uses the citation network formed by research literature on XML collected from WOS database in order to calculate the degree of core literature by weighing citation ties and literature vertices, and finally to present the cohesion of key literature according to the weights and highland. It is found that the core literature substantially reveals the importance for each citation, and the weights calculated from citation ties and literature vertices can accurately detect the quality of literature; cohesion based on the weights can help to fast discover the most core literature and citation during the developments of specific subjects; adequate results from analyzing weight highland in cohesion make up for drawback of analyzing weights in revealing substructure, showing profound research dimensions in the development of the whole knowledge system. 6 figs. 4 tabs. 11 refs.

KEY WORDS　Citation network.　Core literature.　Deep integration.　Weight highland.

1　引言

科学文献是科学知识的重要承载物,研究人员通过科学文献的发表,使自己的研究成果为人所知;对于整个科学的发展而言,每一篇科学文献的发表都具有相应的价值,科学文献作为知识节点,构成了特定研究主题的知识网络。

*　本文系国家社会科学基金重大项目"基于语义的馆藏资源深度聚合与可视化展示研究"(项目编号:1182D152)的研究成果之一。

通讯作者:董克,Email: dk8047@163.com

基于引用关系构成的科学文献网络是一种重要的知识网络,文献之间的引用网络是一个单向的无回路网络,体现了一个学科主题的知识发展和知识演化的过程。基于引文网络的整体结构,运用合适的方法确定文献的核心程度,能够更加客观地进行科学文献质量评价;通过揭示文献之间的关联,进一步将相互关联的文献进行深度聚合,通过可视化的方法将这些聚合结果展现出来,能够更为准确地揭示科学发展的结构和过程,其聚合结果也能提高用户信息获取的效率,更好地满足用户需求。

2 引文网络中的核心及其聚合

2.1 问题的由来

引文网络中文献的核心程度及其聚合研究主要出于以下两个方面的考虑:

(1)引文数据存在明显的集中和离散趋势,科学社会学家科尔兄弟利用引文数据,提出了著名的"奥尔特加"假说[1],并且以此证明了科学界存在的明显社会分层,即科学精英群体的存在及其在科学发展中的重要作用。Rousseau 等人也通过研究发现引文数据存在明显的集中趋势[2],其研究内容更多地集中于从整体的数学分布上进行揭示,但并未对具体是哪些文献属于知识体系核心进行进一步的研究。在确定科学文献事实重要程度的基础上采用相应的聚合方法,能够获取整个知识体系中的核心文献节点,实现这些核心文献的深度聚合,以利于揭示科学发展的结构,了解和分析科学发展过程。

(2)科学引文索引的发明人加菲尔德认为,"由于作者参考之前的研究成果材料,用以支持、列举或详细说明某一观点,引用行为反映了材料的重要性。"[3]利用被引情况确定文献的重要性,是科学研究质量评价的一个重要途径。原始引文数据的利用存在一些不可避免的问题,一般情况下计算被引次数采用实际累计被引次数的方法,但在实际研究中,被一篇普通文章和一篇重要文章引用的价值明显不同;对于整个知识体系而言,某些处于关键媒介位置的引用关系重要性要明显大于另外一些处于结构边缘位置的引用关系重要性,如何区分这些引用关系之间的重要程度,是目前没有解决的问题。引文网络是一个较为明显的多级层次结构,在确定特定文献的重要性时,应该从宏观的角度对引文网络及构成这些引文网络的文献进行深入的分析,这样才能使文献重要性评价更为接近事实。

2.2 核心文献聚合与核心引文聚合

核心文献,是指在科学发展的整个知识传承网络中具有重要意义的文献节点,具体而言,是指贯穿于知识传承的主要过程、对于整个知识体系发展而言具有核心的传递作用、作为知识传递主要媒介的一类文献[4]。核心引文的定义与之类似,是指在整个知识体系的发展过程中起到重要传递作用的引文关系。

在科学文献间存在诸多关联,如通过作者、引用、发表项等诸多其他因素之间产生的联系;其中,文献之间的引用关系最能体现知识的传承关系,科学文献对于发表较早文献的引用可以视作一种知识的继承。在引文网络中所构成的多级引用关系中,核心文献对已有的科学成果进行了较好的科学总结,并在此基础上对后续的研究提供了科学启发;具体到知识传承的实现,则是通过这些核心文献之间的引文链体现出来。

核心引文关系的确定方法比较单一。核心文献的确定方法有两种:一种是直接针对文献的核心程度确定核心文献,以图 1 为例,在 10 篇文献所构成的多级引文网络中,假设左面的 3 篇文献为早期的研究文献,右面的 4 篇文献为目前最新的文献,从结构中可以发现,文献节点 F 是这 10 篇文献中处于最重要媒介位置的文献节点,文献 F 就是一个核心文献,在一个大规模的引文网络中,存在大量与文献 F 类似处于关键位置的节点,这些文献的集合即构成核心文献聚合;另一种方法是在获得核心引文关系的基础上间接获取核心文献,以图 1 为例,其中引文 F→D,F→E 为这个引文网络中的核心引文,在此基础上,可以获得一个相应的核心文献聚合 D、E、F。间接获取核心文献及其聚合的方法具有一定的优势,在具体聚合过程中,间接获取

法的结果一般包括了直接获取法的结果,因此本研究中主要在核心引文的基础上用间接法获取核心文献及其聚合。

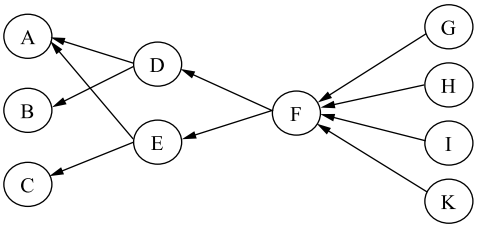

图1　多级引文网络中核心文献与核心引文

2.3　核心程度权值

定义一个引文网络 N = (A,C),其中 A(Articles)代表文献集合,C(Citations)代表引文的集合。引文 e = (v,u)表示文献 v ∈ A 出发至文献 u ∈ A,即文献 v 引用了文献 u。理论上,在这个引文网络中,文献按照其发表的时间分布,整个网络的引用关系不可逆。在这个网络 N 中,至少存在一个初始文献,其出度为 0(即没有引用文献集中的任何其他文献);同时也存在至少一个最新的文献,其入度为 0(即最新发表或发表后未被本主题内其他任何文献引用)。

定义 I 和 O 为起点和终点的集合,定义 W_e 为引用关系对 e 的权值,则有:

$$W_e = \frac{\Sigma e_{I\to O}}{\Sigma I\to O} \qquad \text{公式(1)}$$

如果从节点文献的视角出发,同样可以计算节点 v 的权值 W_v:

$$W_v = \frac{\Sigma v_{I\to O}}{\Sigma I\to O} \qquad \text{公式(2)}$$

引文和文献节点是共生关系,因此在获取引文聚合的同时,相应的文献节点聚合同时产生;通过一定方法获取文献节点聚合,引文也同时包含在其中。在公式(1)和公式(2)中对权值都进行了标准化处理,其得出的结果是实际通过的路径数量与整个网络中的路径比值,这样权值更能反映重要程度,节点 v 或者引用关系 e 的权值均处于 0 和 1 之间,其计算过程可以利用社会网络分析软件 Pajek 实现[5]。

3　样本数据选择及其清洗

XML 的全称为 eXtensible Markup Language,中文为"可扩展的标记语言",XML 是由标准通用标记语言(SGML)发展而来,自 1996 年推出至今,在网络服务、电子商务以及语义网等技术和应用领域产生了巨大的影响。赵党志认为 XML 是一个最具发展前景的研究领域,他于 2003 年以 XML 研究领域的比较引文分析获得了佛罗里达州立大学的博士学位[6],此后的近 10 年时间中,XML 有了很大的发展,本文选择 WOS 数据库中的 XML 研究论文进行实证分析。

在 Web of Science 数据库中,以 XML 及其全称为主题进行检索①,文献类型主要选择 article、proceedings paper、review 三种,共检索到文献14,481篇,时间跨度为 1996—2012 年,检索时间为 2012 年 5 月 25 日。对数据进行清洗,合并相同文献后得到文献14,461篇,即整个网络中有节点14,461个;对文献自引等不规范引文进行清洗后共获得引文11,527对。

表1　子网规模分布

子网规模	1&2	3&4	5&6	7&8	9&11	12&16	5274
数量	8458	76	18	6	2	3	1

对网络进行成分分析发现,14,461 篇文献中最大的联通子网有 5274 篇文献,超过了全部文献的三分之一,无联系的孤立文献或者单一的引文数量为 8457,这两部分文献占据了文献总量的绝大多数,反映了较为明显的引文集中分布。大量的零星分布实质上同样表明,在文献群中,有大量的文献在发表后未被利用或在本研究主题文献集内部未被利用,这些文献的核心程度权值都很小,说明这

① 感谢 Henk F. Moed——2012 年 3 月在北京参加循证科学研讨会的茶歇期间,Moed 对于引文分析的一些内容,尤其是针对利用 Web of Science 进行引文数据收集的注意点给笔者提出了建议。

些文献对于整个知识体系发展而言的重要性很低，而构成大规模网络的文献则大多数处于知识体系发展的主干上。

4 基于阈值的聚合

4.1 阈值聚合方法

基于阈值的聚合方法是一种较为简单易行的聚合方法，通过公示(1)、(2)计算出引文和文献节点的权值后，只需要获得全部权值的分布情况，通过划定阈值可获得一定规模的文献聚合，其优点在于直接和易于根据需求调整，缺点在于获得的节点可能会是离散的，并且结果相对单一。

本文主要讨论从引文权值的角度探索文献节点的问题，定义引文网络 N = (A, C, w)，其中 w: C→R，在阈值 t 时获得的子网 N(t) = (A(C′), C′, w)，对于任意引用关系对 e，有：

$$C' = \{e \in C : w(e) \geq t\} \qquad \text{公式(3)}$$

A(C′)为构成子网中所有引用关系对的节点，即在阈值 t 水平上的核心文献聚合。

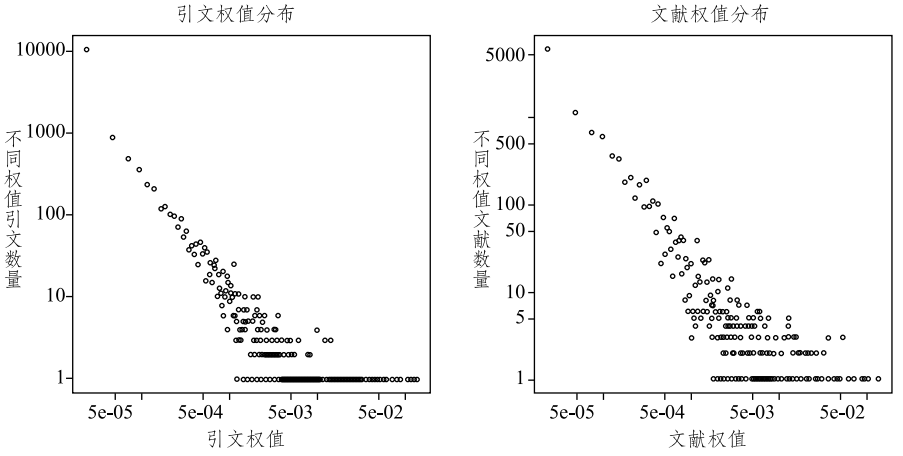

图 2 文献权值与引文权值分布

4.2 阈值聚合结果

利用社会网络分析软件 Pajek 中的 SPC 程序[7]获得的整个引文网络中文献权值分布与引文的分布情况(见图 2)。X 轴表示文献和引文的权值大小，Y 轴则表示具有不同权值的文献或引文频数。从分析结果来看，随着权值的增加，文献权值与引文权值的频数表现出了幂律分布的特征。进一步对引文权值进行分析发现，引文权值中的最低值为 0.000023，最大值为 0.09046，将其分成十段后得到 11 组聚合数量分布结果(见表 2)。

从引文权值的阈值聚合分布来看，选取不同的阈值获得的文献聚合大小差别很大。尤其是最后一个部分，引文权值[0.000023, 0.0082)之间集中了大量的文献。选取引文权值的最低值 0.000023 为阈值时获得了 6275 个文献节点，这是由于在 WOS 中以 XML 为主题的 14,461 篇文献中其他的

表 2 基于引文权值的阈值聚合分布

阈值	引文	文献量	百分比(%)
0.0822	1	2	0.01
0.0739	1	2	0.01
0.0657	2	4	0.03
0.0575	4	7	0.05
0.0493	4	7	0.05
0.0411	6	10	0.07
0.0328	11	15	0.10
0.0246	18	22	0.15
0.0164	35	38	0.26
0.0082	90	76	0.53
0.000023	11527	6275	100.00

8186 篇文献均未与其他文献发生引用关系。8186 篇文献中某些文献的绝对被引值很高，在利用 XML 作为主题在 WOS 中获取时虽然被收录进来，但与研究主题 XML 之间的实际关联很小，通过阈值聚合，这些文献都将被排除在分析结果外，因此，阈值聚合的结果无论选择多少都将大大提高最终获取文献的相关性程度。

选取阈值 t 为 0.0246 获得的文献聚合结构图如图 3 所示，其具体文献如表 3 所示，共包含 4 个

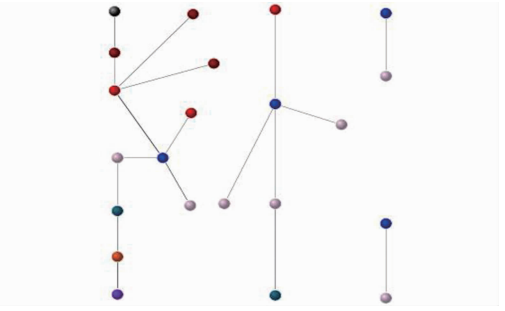

图 3　引文权值 t = 0.0246 水平上的文献聚合结构图

表 3　引文权值 t = 0.0246 水平上的文献聚合

聚合	文献序号	文献节点
1	6	Link S, 2012, J COMPUT SYST SCI, V78, P1026
	76	Sengupta A, 2012, DATA KNOWL ENG, V72, P219
	237	Hartmann S, 2011, COMPUT J, V54, P1166
	520	Link S, 2011, FUND INFORM, V106, P233
	776	Koehler H, 2010, INFORM PROCESS LETT, V110, P717
	840	Langeveldt WD, 2010, INFORM SYST, V35, P352
	1598	Hartmann S, 2009, DATA KNOWL ENG, V68, P1128
	2134	Link S, 2009, LECT NOTES ARTIF INT, V5514, P256
	2364	Hartmann S, 2009, FUND INFORM, V92, P83
	3086	Link S, 2008, ACTA INFORM, V45, P565
	3222	Link S, 2008, INT J FOUND COMPUT S, V19, P691
	6102	Hartmann S, 2006, THEOR COMPUT SCI, V364, P212
2	4695	Hartmann S, 2007, ANN MATH ARTIF INTEL, V50, P195
	6370	Hartmann S, 2006, ANN MATH ARTIF INTEL, V46, P114
3	9354	Vincent MW, 2004, ACM T DATABASE SYST, V29, P445
	10797	Buneman P, 2003, INFORM SYST, V28, P1037
	12191	Buneman P, 2002, COMPUT NETW, V39, P473
	12235	Fan WF, 2002, J ACM, V49, P368
	13261	Buneman P, 2001, SIGMOD RECORD, V30, P47
	14288	Calvanese D, 1999, J LOGIC COMPUT, V9, P295
4	9484	Arenas M, 2004, ACM T DATABASE SYST, V29, P195
	14334	Shanmugasundaram J, 1999, PROCEEDINGS OF THETWENTY-FIF, P302

聚合,其中两个聚合文献量较大,另外两个仅包含了2篇文献。这4个聚合中的22篇文献由整个网络中核心程度权值最高的18个引文联系起来,是14,461篇文献中核心度最高的4个聚合,反映了在整个XML发展过程中影响最为重要的主题内容。

聚合1中的11篇文献主题集中于两组关键词,第一组关键词以知识获取为核心,包括反馈关系、函数相关性和数据库依赖与多值依赖;第二组关键词以命题逻辑为中心,包括了数据的关系模型、数据库分解、数据库约束和公理化的内容。两部分的内容通过知识获取和命题逻辑两个主题词的共现联系起来。

聚合2中的2篇文献都是由Sven Hartmann及其合作者完成的,其研究主题集中在数学与人工智能方面,主要研究了关系数据库嵌套中的类关系依赖,深入挖掘其中的层次依赖,其中涉及了面向对象的数据模型包括XML支持的表构造程序。聚合3中有5篇文献,Buneman P是其中的主要作者,该聚合的主题相对集中,主要研究XML文档的类型定义,进一步讨论了利用XML进行推理时的约束问题。聚合4包括了2篇文献,其主题集中于利用传统的关系数据库模型处理XML文档的规范化方法,并研究了相关的扩展模型,使用户能够有效地挖掘利用XML文档存储的数据。

从时间发展上来看,聚合3和聚合4的文献发表较早,其次是聚合2,聚合1中的文献绝大多数是近5年的研究成果。如果说早期的研究内容更多的是围绕对于XML文档本身的结构、特征及应用环境等内容的话,聚合1的研究表现出了明显的深化,这种深化充分体现在聚合1中文献的研究主题上。

5 基于权值"高地"的聚合

5.1 权值"高地"聚合方法

在一个连通的引文网络中,其平面的X、Y坐标分布结构由整个网络的结构决定,可以通过一定的算法(如弹簧算法)使网络整体的二维布局更为合理,在此基础上以各文献节点或引文关系的核心程度权值作为Z坐标,利用R软件绘制引文网络的三维立体图形[8](见图4)。所谓权值"高地",是指在这样的三维结构中一些特殊的聚合,其中文献节点或引文关系核心程度权值要大于其周边文献节点或引文关系核心程度权值,从而产生一个相对"高地",即这些聚合产生的原因是其中的文献节点或引文核心权值相对较高。

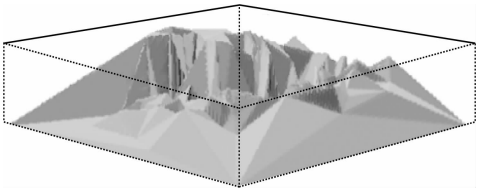

图4 整体文献的等高线三维分布

本文主要通过获取引文"高地"聚合间接获取文献节点"高地"聚合。令网络N=(A,C,l),网络中的引用关系权值l:C→R,有非空文献子集P⊆A,如果存在最大生成树T,其中引用关系权值的最小值大于或等于该聚合中文献P到邻近其他文献集合之间引用关系的权值,则称该最大生成树T为引文权值"高地"聚合,即

$$\min_{(u,v)\in C(T)} w(u,v) \geqslant \max_{(e,v)\in L,e\in P,v\in P} w(e,v)$$

公式(4)

构成最大生成树T的文献子集P生成了一个文献权值"高地"聚合[9]。

5.2 聚合结果

对于权值"高地"聚合而言,聚合中文献数量上限的设定十分重要,如果规模过大则聚合结果反映主题结构的功能将被大大削弱,通过不断测试后,选取40作为文献数量的上限。通过计算,共获得引文相对"高地"聚合735个,其中大多数聚合文献数量都很小,聚合中文献数量为2的聚合有531个,包含文献数量最多的聚合中有34篇文献。权值"高地"聚合规模的分布图与结构图分别如图5、图6所示。

通过对聚合结果的进一步分析发现,以聚合1和聚合113为代表的一类聚合特点突出,与其他的

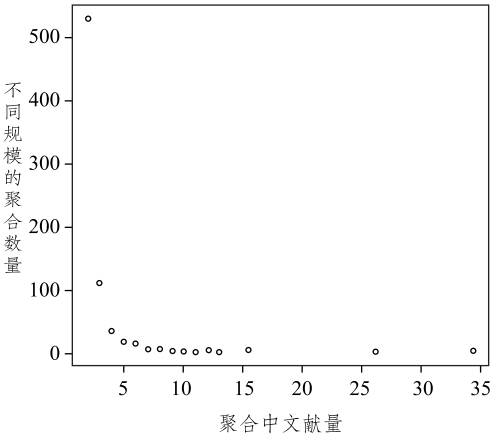

图 5　权值“高地”聚合的规模分布

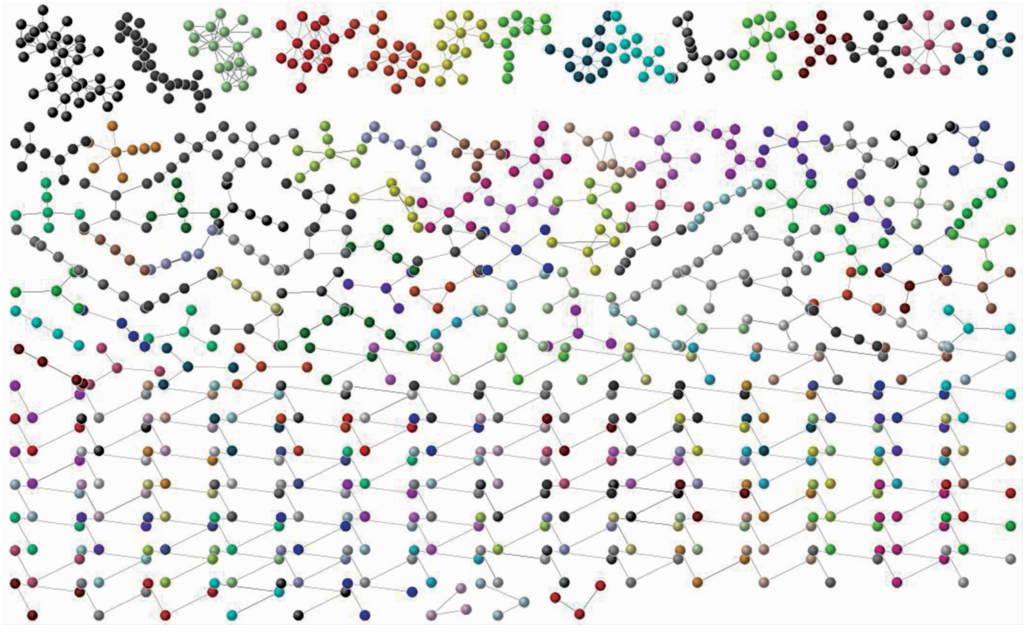

图 6　权值“高地”聚合结构图(局部)

文献聚合相比,这两个聚合中文献数量多,且文献权值极高,整体的重要性很高(见表 4)。如果从整体的三维效果上来看,这类聚合的可视化效果是“山脉式”的权值“高地”聚合。

从研究主题上来看,聚合 1 中的文献囊括了阈值聚合 1 中的全部文献,数量进一步增加到 33 篇,由于更多文献的加入,该聚合中研究主题表现得更为集中,主要是关系数据库环境下的 XML 研究,出现频率较高的关键词有功能依赖与多值依赖(共 35 次)、约束(共 21 次)、关系数据库和关系数据集(共 18 次)、公理化(共 19 次)。聚合 2 中 26 篇文献出现最多的关键词有查询(共 27 次),包括查询语言、查询优化、查询算法;半结构化(11 次),包括半结构化数据实例、半结构化数据模型和半结构化

数据查询;数据(16 次),包括数据提取、数据合并、数据共享和数据挖掘。语义(10 次),包括语义功能、语义模式匹配、关系语义;Web(8 次),Tamino(7 次)。聚合 2 中的文献关于 XML 应用的相关研究较多。

表 4 两种类型的"高地"聚合结果

聚合序号	聚合规模	文 献	文献权重最大值	文献权重最小值
1	34	6、76、151、237、520、776、840、1087、1598、2134、2364、3086、3222、4677、4695、6102、6294、6370、6993、6994、8105、8773、9354、10174、10797、11008、11323、11713、11923、12191、12235、13261、13748、14288	0.1359	0.0108
113	26	3107、4708、7905、8494、9484、9614、10313、10554、11009、11679、11710、12137、12271、13182、13784、13785、13787、13788、13791、13797、14178、14217、14327、14333、14334、14345	0.106	0.0081
6	12	45、171、194、687、3627、4701、7486、7665、8384、10891、13279、13281	0.0003	0.00004
7	16	49、2048、2768、3809、5255、6326、6869、7513、7555、7735、8060、8488、8860、9844、10076、11528	0.0002	0.00002

以聚合 6 与聚合 7 为代表的一类聚合同样具有鲜明的特点,从绝对数量上来说,这一类聚合中包含了一定数量的文献节点,然而其绝对高度相对较低,如果从整体的三维效果上来看,这类聚合的可视化效果是"丘陵式"的权值"高地"聚合。

聚合 6 中包含了 12 篇文献,出现最多的关键词有注释(15 次),包括语言注释、注释规范、注释框架和注释工具;语料库(12 次);语言(8 次);韵律学(8 次);NITE XML Toolkit(7 次)。聚合 6 中文章研究的主题主要围绕语音语料库的构建展开,因此其关键词体现了极大的关联,其中文献权值最大的两篇文献主要分析两个基于 XML 开发的功能强大的语料库工具包[10-11],其他的文章或利用这些工具或在此基础上进行进一步开发。

聚合 7 中有 16 篇文献,研究内容相对集中于基于 XML 的服务和管理研究。出现最多的关键词是服务(39 次),包括 Web 服务;管理(25 次),包括管理构建、报表管理、管理系统;测试用例(14 次),包括策略依赖的测试用例、测试用例生成方法;QoS 网络服务质量(13 次),包括 QoS 特性、QoS 支持、网格 QoS 管理框架、网格 QoS 描述语言等。

与聚合 1、聚合 113 比较,聚合 6 和聚合 7 中的文献绝对权值和事实上的被引都明显较低,从整个 XML 发展来看,聚合 6 和聚合 7 中的研究主题虽然产生的影响不是特别巨大,但是充分体现了 XML 发展过程中学者研究的不同视角,大量类似存在的权值"高地"聚合能够充分反映 XML 研究的整体状况和多元化的主题结构。

6 讨论

在计算核心程度权值时本文给出了两个计算公式,事实上,无论是在阈值聚合还是在权值"高地"聚合的过程中,都可以直接从文献节点的核心程度权值出发进行聚合,但本文主要是从引文核心程度权值的角度出发,间接获取了核心文献聚合,一个重要的原因是,本文的分析过程主要是一种"网络观"视角的分析,通过网络关系入手再延伸到网络中的节点上。在网络中,作为其中衔接存在的"关系"所处的位置十分重要,本文的研究将其作为一个分析的基本单元,是为了贯彻"网络观"的分析视角。

阈值聚合与权值“高地”聚合的基础都是基于引文网络中文献与引文的核心程度权值，其计算过程与结果充分考虑了不同引用实质上的重要程度区别，所得到的引文和文献节点的权重能够更加准确地反映文献的质量，是对质量文献评价的深度探索，最终的聚合结果也是一种在目前水平上的深度聚合。

两种不同类型的聚合方法及其结果之间既有联系又有区别。基于阈值的聚合实质上是针对绝对核心程度的一种聚合方法，其结果主要揭示的是文献整体中的最核心文献和引文，4.2 小节中获取的四组核心文献及引文聚合，其研究主题明显是整个 XML 发展中的主流。但阈值聚合的缺点也显而易见，由于整个网络中通过运算得到的文献及引文核心程度集中趋势相当明显，因此其结果一般较为单一；阈值聚合的另一个缺点是，忽视了在整个知识体系发展中具有明显特点的一些群体，这些群体的绝对重要程度可能不高，但是在整个知识体系中明显占有一席之地。基于阈值的聚合往往不能获取这些潜在发展能力较强的子结构。

从分析结果来看，权值“高地”聚合恰好弥补了阈值聚合存在的这两个问题，核心程度权值“高地”在结果上能够涵盖阈值聚合的结果，且对于这个网络而言，核心程度权值“高地”聚合结果更为多样，较好地揭示了整个知识体系发展过程中具有极为活跃发展潜力或在某个阶段的研究中居于重要位置的子结构；对整个引文网络获得的 735 个权值“高地”聚合进行更为深入的主题分析可以更好地把握整个 XML 研究领域的内容和学科布局，体现了整个知识体系发展过程中丰富的研究维度。

在讨论不同的权值“高地”聚合类型时，文中主要列出了两种类型，还可能存在其他类型，如由极少数核心程度权值极高的文献和引文所构成的“尖峰式”权值“高地”聚合，由大量核心程度权值很低的文献和引文所构成的“沙滩式”权值“高地”聚合等，不同类型的聚合在整个知识体系结构中的作用不尽相同，限于篇幅本文未作过多的讨论；此外，引文网络的一个重要特征是其在时间上的结构，如果能够将时间维度更为深入地应用到整个聚合结果的分析中，可以在学科知识结构的基础上进行学科演化结构的分析，在以后的研究中我们将做作进一步的探索。

参考文献：

[1] Cole S, Cole J R. The ortega hypothesis[J]. Science, 1972, 178(4059): 368 - 375.

[2] Rousseau R. Concentration and diversity of availability anduse in information systems: A positive reinforcement model [J]. Journal of the American Society for Information Science, 1992, 43(5): 391 - 395.

[3] Garfield E. Citation indexing: Its theory and application in science, technology and humanities[M]. New York: Wiley, 1979: 23 - 24.

[4] De Nooy W, Mrvar A, Batagelj V. Exploratory social network analysis with Pajek[M]. New York: Cambridge University Press, 2005: 244.

[5] Batagelj V, Mrvar A. Pajek——Program for large network analysis[J]. Connections, 1998, 21(2): 47 - 57.

[6] Zhao Dangzhi. A comparative citation analysis study of Web-based and print journal-based scholarly communication in the XML research field[EB/OL]. [2012 - 05 - 20]. http://diginole.lib.fsu.edu/etd/530.

[7] Batagelj V. Efficient algorithms for citation network analysis[EB/OL]. [2012 - 05 - 28]. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.121.3008&rep=rep1&type=pdf.

[8] R Development Core Team. R: A language and environment for statistical computing[CP/OL]. [2012 - 06 - 01]. http://www.R-project.org/.

本刊征订启事

《中国图书馆学报》是首批国家社会科学基金资助的图书情报学领域专业期刊,由文化部主管、中国图书馆学会和中国国家图书馆共同主办。每年6期,逢单月15日出版。面向国内外公开发行,全国邮局均可订阅,中国国际贸易图书总公司负责国外发行。国内代号2-408,国外代号BM184。

每册单价26元,全年156元。欢迎广大读者订阅。

通过本刊编辑部订阅者,全年整订可享受8.5折优惠(含邮寄费),单册零购可享受免费邮寄服务。

邮购汇款地址:北京市西城区文津街7号,中国图书馆学报编辑部收。

邮编:100034。

请订阅者在汇款单上注明订阅者姓名、地址、邮编、所购期次、份数。

电话:010-88545234

邮箱:tsgxb@nlc.gov.cn

《中国图书馆学报》编辑部

[9] Zaverˇsnik M, Batagelj V. Islands[EB/OL]. [2012-06-10]. www.pajek.imfm.si/lib/exe/fetch.php?media=dl:wsxxixa.pdf.

[10] Barras C, Geoffrois E, Wu Z B, et al. Transcriber: Development and use of a tool for assisting speech corpora production[J]. Speech Communication, 2001, 33(1-2): 5-22.

[11] Carletta J, Evert S, Heid U, et al. The NITE XML Toolkit: Flexible annotation for multimodal language data[J]. Behavior Research Methods, Instruments, and Computers, 2003, 35(3): 353-363.

邱均平 武汉大学信息管理学院教授、博士生导师。

通讯地址:湖北省武汉市珞珈山武汉大学信息管理学院。邮编:430072。

董 克 武汉大学信息管理学院博士研究生。通讯地址同上。

(收稿日期:2012-06-25)