

doi:10.3772/j.issn.1000-0135.2015.001.001

# 超网络视域下的数字资源深度聚合研究<sup>1)</sup>

王传清<sup>1,2</sup> 毕强<sup>1</sup>

(1. 吉林大学管理学院, 长春 130022; 2. 中国科学院文献情报中心, 北京 100190)

**摘要** 基于关联关系维度的数字资源聚合是数字资源知识发现的重要基础和工具。超网络是由多个类型的同质和异质子网络组成的网络,通过多种关联维度聚合的数字资源即形成了拥有相同以及不同性质的结点和关系的数字资源超网络,这些不同性质的关联与链接是知识关联、挖掘、发现与创新的脉络线索。结合超网络理论,构建和描述数字资源超网络,并分析超网络中不同性质的关系类型,如引用关系、共现关系、耦合关系等,从关联维度探讨数字资源深度聚合的模式,进而分析利用数字资源超网络进行知识发现的具体应用方法,最后构建数字资源超网络聚合系统模型。

**关键词** 数字资源聚合 超网络 数字图书馆 关联维度 知识发现

## An Exploration of In-depth Aggregation of Digital Resources Based on Supernetworks

Wang Chuanqing<sup>1,2</sup> and Bi Qiang<sup>1</sup>

(1. School of Management, Jilin University, Changchun 130022;

2. National Science Library, Chinese Academy of Sciences, Beijing 100190)

**Abstract** The aggregation of digital resources from the relevant dimensions is the important foundation and tools for the knowledge discovery from digital resources. Supernetwork is composed of multiple types of network, the integration digital resources form a complicated digital resources supernetwork, which has different properties of nodes and relations. These different relations and links provide clues for knowledge link, mining, discovery and innovation. Combining with the supernetwork theory, the authors build and describe the digital resources supernetwork, analyze various types of relationship in the network, such as references, co-occurrence relation, coupling relations, and discuss the modes of in-depth integration of digital resources from the relevant dimension. The authors also analyze specific applications to knowledge discovery by using digital resources supernetwork. Finally, the authors propose the aggregation system model of digital resources supernetworks.

**Keywords** digital resources aggregation, supernetwork, digital library, relevant dimension, knowledge discovery

收稿日期:2014年8月28日

作者简介:王传清,男,1980年生,吉林大学管理学院博士研究生,中国科学院文献情报中心《图书情报工作》杂志社编辑,和要研究方向:信息资源管理,图情期刊编辑出版,E-mail:wangcq@mail.las.ac.cn。毕强,男,1954年生,吉林大学管理学院教授,博士生导师,主要研究方向:信息资源管理,E-mail:biqiang12345@163.com。

1) 本文系国家自然科学基金项目“语义网络环境下数字图书馆资源多维度聚合与可视化展示研究”(项目编号:71273111)和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研究”(项目编号:11&ZD152)的研究成果之一。

## 1 引言

随着信息技术和网络技术的不断发展,数字资源的种类和数量已大大超越传统的纸质资源,并且仍不断以几何级数的速度在增长,人们开始迈入大数据时代。Web 2.0 及相关应用的飞速发展更使得海量的数字资源分散在网络中的各个地方,给人们全面、快速、准确地获取和利用所需资源造成困难。同时,科学研究和交流环境的变化,也促使科研用户利用文献资源的方式发生改变,对数字资源组织和服务提出了更高的要求:不仅需要不同语言、不同载体、不同类型的数字资源之间建立起有机联系;还需要对这些数字资源及包含的丰富语义信息、实体等知识资源,实现从多种维度、多种途径、多种界面的揭示和链接,一站式满足用户语义化的知识发现和创新需求<sup>[1,2]</sup>。因此,近年来关于数字资源的深度揭示和聚合等研究,成为图书情报学者以及图书馆工作人员关注的热点领域。国家自然科学基金和国家社会科学基金也都设立了相关项目,重点支持这方面的研究<sup>[3,4]</sup>。

## 2 研究现状

### 2.1 数字资源聚合研究概述

数字资源聚合研究经历了从资源整合到资源聚合的发展过程<sup>[5]</sup>。早期的数字资源整合研究主要集中在联机公共访问目录(Online Public Access Catalogue, OPAC)<sup>[6-8]</sup>、跨库检索<sup>[9,10]</sup>、文献引用<sup>[11]</sup>以及元数据<sup>[12]</sup>等方面。随着文献计量、社会网络分析以及语义网等技术和方法的快速发展,基于语义的馆藏资源深度聚合研究逐渐兴起,并对数字资源聚合和知识发现相关研究产生了重要影响。

国外代表性的文献有:Brad等<sup>[13]</sup>将2D和3D信息化可视化技术引入数字图书馆资源建设,来支持多源和馆藏资源的展示与导航,用户可以方便地操作、检索和浏览这些资源。还有学者探讨利用GIS技术,将分布在各处的图书馆馆藏集中呈现和展示在三维空间中,为用户利用提供方便<sup>[14,15]</sup>,并进一步丰富GIS可视化系统中的非地理空间属性<sup>[16]</sup>。还有通过利用虚拟现实技术<sup>[17]</sup>、将可视化研究中的定性属性于数字图书馆资源的定量特征相结合<sup>[18]</sup>等技术和方法,来改造传统的数字资源目

录,对数字资源等级结构进行可视化<sup>[19]</sup>,为用户提供了全新的检索和浏览数字资源方式。英国联合信息系统委员会(JISC)更是发布了《数字资源的聚合和维护》报告,分别研究了机构的技术和战略以及数字内容的聚合和加强两个方面<sup>[20]</sup>。

国内代表性的有:贺德方和曾建勋<sup>[1]</sup>从基于概念及概念之间关系、基于引证关系、基于科研本体等角度探讨了基于语义的馆藏资源聚合方式,并阐述了基于语义的馆藏资源聚合服务的内容,包括馆藏资源语义描述框架、对聚合结果进行可视化呈现以及开展聚合服务等;邱均平和王菲菲<sup>[21]</sup>探讨了文献计量学中的共现与耦合理论与方法在馆藏资源集合中的应用;李劲和程秀峰等<sup>[22]</sup>对基于本体集成的馆藏资源聚合方法进行了探讨;陈晓美和毕强等<sup>[23]</sup>从关联关系维度、传递扩散维度、组织结构维度构建了基于多维度聚合的数字图书馆资源知识发现理论框架;邱均平<sup>[24]</sup>以专题的形式,引入分面组配思想和信息计量学理论与方法,研究了馆藏资源语义化与多种方法的融合;毕强<sup>[25]</sup>以“数字资源聚合”专题论文探讨了大数据环境下数字资源整合到聚合的组织模式变革,分析资源整合与资源聚合的异同点,并从社会网络分析的视角对数字图书馆资源聚合模式开展了研究,还进行了相关的实证分析。

概括起来,目前的研究主要探讨两个方面的内容,一是数字资源的聚集,即将各种类型、不同性质的数字资源集聚在一起;二是数字资源的内容的融合,即打破数字资源存储载体限制,利用本体等,基于语义关联发现和计量分析等将数字资源聚合在一起。这些研究成果为语义网环境下数字资源的聚合与知识发现提供了很好的理论指导和研究思路。同时也可以发现,虽然目前对于数字资源整合的理论和实践已经较为丰富,但对于数字资源聚合的研究主要还是理论和模型探索阶段,且有待于从单一维度,如共现与耦合维度、文献引用维度等,向多维度聚合发展和演进,并实现同质和异质数字资源的聚合。

### 2.2 超网络研究概述

最早引入超网络概念并加以应用的是Sheffi<sup>[26]</sup>。之后,Nagurney<sup>[27]</sup>在处理复杂交织网络时把超网络定义为“above and beyond existing networks”,使超网络的概念更为明晰。此后,基于超网络理论的知识网络研究逐渐引起学者们的关注。Nagurney和Dong<sup>[28]</sup>基于知识密集型系统及其组织基础管理理

论建立了知识的超网络模型;席运江和党延忠<sup>[29,30]</sup>基于组织知识系统中知识、人和存储载体三类要素及其复杂关系进行详细分析基础上构建了知识超网络模型,并就有关知识网络鲁棒性进行分析讨论。王众托等<sup>[31-33]</sup>认为网络社会中的知识网络无处不在,并通过研究认为知识网络是由存在一种特定关系的人际关系网络、知识资源网络和技术网络三个层次构成,是一个复杂的超网络,即网络的网络,并对这样的复杂知识超网络模型及其一些特性进行了研究。李晟<sup>[34]</sup>以软件行业为研究实例,从知识超网络视角研究知识协同的方法和技术,构建了SNKC知识协同系统;徐升华等<sup>[35]</sup>基于社会与知识网络复合而成超网络模型,针对网络环境下知识转移行为展开剖析。另外,还有许多研究者利用超网络相关理论就知识传播、知识流动和知识转移相关方面做了分析研究<sup>[36,37]</sup>。

数字资源的聚合实际上是利用多种方法对数字资源以及所包括的知识进行整合和融合,其最终目的是为了向用户提供知识服务。不同载体、不同类型的数字资源之间的多维关联和聚合,形成了一个又一个性质各异、关系不同而又存在联系的多个网络,这些网络共同构成了一个庞大的数字资源超网络。综合分析已有研究文献发现,鲜见从超网络视角来研究数字资源聚合与知识发现的成果。因此,笔者尝试引入超网络的理论与方法,从数字资源关联维度来总结和探讨数字资源聚合广度整合和深度聚合的主要模式,构建数字资源超网络聚合系统模型,并分析基于数字资源超网络的知识发现应用,为语义网环境下数字图书馆从提供文献信息服务向提供知识服务转变提供参考。

### 3 数字资源的广度关联

#### 3.1 超网络的基本理论和特征

在研究超大规模的网络系统时,会出现信息网络、物流网络、资金网络等相交织的问题,或网络中网络的问题。如果仅用单用简单网络或异质网络的方法来分别处理各个网络的问题,就很难理清这些网络之间的联系。因此出现了“超网络”的概念,超网络即“高于而又超于现存网络的网络”,或者说“超网络”由是多个网络组成的网络<sup>[19]</sup>。超网络中的节点表示给定集合的网络,边表示在给定集合中的结合移动和结合偏好,超网络惟一地表示了由规

则支配的所有结合移动和结合偏好,即边之间的加、减或替换。

超网络的特征主要包括:①多网络性,超网络中包含了多个网络,网络中嵌套着网络;②多层性,如超网络中包含用户层、数据层、物理层等,层的内部和层与层之间都存在连接;③多级性,如超网络中涉及个体、部门、全机构等级别,同一级别的内部和级与级之间也同样存在连接;④多维性,超网络中节点之间的边是多维度的,节点之间的边可能相同,也可能不同;⑤多准则性,超网络中可以通过多种途径连接到某个节点,需要综合考虑连接成本,如时间、费用、安全等;⑥协调性,对超网络进行优化,有时个体与整体网络之间会产生冲突,需要进行统筹协调。超网络的这些特征和架构为描述和表示网络的结构,研究网络之间的相互作用和影响提供了方法和工具,可以利用数学方法、可视化工具等对网络中的流量、关联、聚合情况进行定量分析和计算。

数字资源聚合过程中涉及三类重要的实体:文献作者(如个人或机构)、文献内容所包含的知识以及文献的物质载体(如期刊或专著)等。作为一个有机整体,数字资源体系中存在相互关联和相互链接的人、文献知识和物质载体这三种不同类型的知识资源<sup>[20,23]</sup>。对数字资源进行有效聚合并实现知识发现,需要对这三种类型的知识资源进行系统分析和研究,需要对数字资源体系进行建模和描述。人、文献知识和载体要素之间具有复杂的关联关系,三个具有不同性质结点的复杂网络(即人际网络、文献知识网络和载体网络)相互关联,构成的具有多层次、多级、多维、多属性、多规则的更为复杂的复杂网络。超网络最重要的特点就是同时可以描述同质事物以及异质事物间的联系,构建同质网络之间、不同质网络之间以及同质与不同质网络间的关联关系。因此,可以借助超网络的理论和模型来建构和描述数字资源体系,将数字资源广度关联在一起。

#### 3.2 子网络构建与描述

数字资源体系中存在三种类型的知识子网络构建和描述如图1所示:

(1)文献知识与文献知识的网络。即文献及其包含知识之间的网络,可以简称为 $K-K$ 网络,其节点为数字资源中所包含的不同知识点,其模型为:

$$G_k = (K, E_{K-K})$$

其中, $K = \{k_1, k_2, \dots, k_n\}$ 为知识点的集合; $E_{K-K} = \{(k_i, k_j) | k_i, k_j \in K\}$ 为边的集合,边 $(k_i, k_j)$ 表示

知识点  $k_i$  与  $k_j$  之间存在知识联系,如文献知识点共现等。

(2)人与人的网络。即著作权人之间的网络,可简称为  $P-P$  网络,其节点为数字资源体系中的著作权人,可以是个人,也可以是机构,边表示著作权人之间的知识联系。该网络模型可表示为:

$$G_p = (P, E_{p-p})$$

其中,  $P = \{p_1, p_2, \dots, p_b\}$  为著作权人的集合;  $E_{p-p} = \{(p_i, p_j) \mid p_i, p_j \in P\}$  为边的集合,边  $(p_i, p_j)$  表示著作权人  $p_i$  与  $p_j$  之间存在知识联系,如著作权人合作研究,发表合著等。

(3)文献知识载体之间的网络。即存储文献的物质载体之间形成的网络,可以简称为  $M-M$  网络,其节点为期刊、图书等各种物质载体,边表示他们之间的关系,如期刊的分类关系等。

该网络可表示为:

$$G_m = (M, E_{m-m})$$

其中,  $M = \{m_1, m_2, \dots, m_c\}$  为各类文献知识存储媒介载体的集合,  $E_{m-m} = \{(m_i, m_j) \mid m_i, m_j \in M\}$  为边的集合。边  $(m_i, m_j)$  表示  $m_i$  与  $m_j$  之间存在知识联系,如  $m_i$  与  $m_j$  存储的文献知识有相同点或存储的文献有相同的分类关系等。

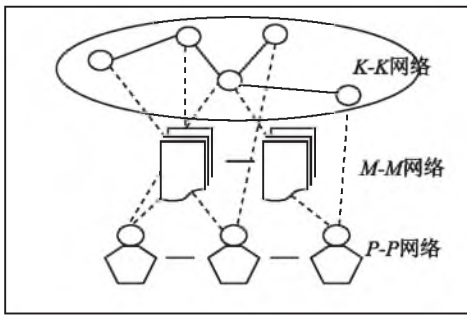


图1 数字资源体系中3种子网类型及关联示意

### 3.3 子网络间映射关系的构建与描述

数字资源体系中人、文献知识及其存储载体之间存在着多种映射关联关系,归纳起来主要有以下几种:

(1)子网  $G_p$  与  $G_k$  间的映射关系。人与文献知识之间的映射关系集合用  $E_{p-k}$

进行描述,表示存在两种关联:①著作权人提出或拥有哪些文献知识;②文献知识是由哪些著作权人提出或拥有的。定义表达式为:

$$E_{p-k} = P(k_j) + K(p_i)$$

其中,  $P(k_j)$  表示提出或拥有某一文献知识的著作

权人集合,

$$P(k_j) = \{p_i \mid p_i \in P, \alpha(k_j, p_i) = Q, i = 1, \dots, n\}$$

其中,  $k_j$  表示文献知识集合中某一知识元,  $p_1, \dots, p_n$  表示著作权人,  $Q$  取值如下:

$$Q = \{1, k_j, p_i\}$$

其中,  $K(p_i)$  表示著作权人提出或拥有文献知识的集合,

$$K(p_i) = \{k_j \mid k_j \in K, \alpha(p_i, k_j) = Q, j = 1, \dots, n\}$$

其中,  $p_i$  表示著作权人集合中某一著作权人,  $k_1, \dots, k_n$  表示知识元,  $Q$  取值如下:

$$Q = \{1, p_i, k_j\}$$

(2)子网  $G_p$  与  $G_m$  间的映射关系。著作权人与物质载体之间的映射关系集合用  $E_{p-m}$

进行描述,也表示两种关联:著作权人产生或被利用了哪些物质载体;物质载体被哪些著作权人产生或被利用。定义表达式为:

$$E_{p-m} = M(p_i) + P(m_l)$$

$$= \{m_l \mid m_l \in M, \beta(p_i, m_l) = Q, l = 1, \dots, n\} + \{p_i \mid p_i \in P, \beta(m_l, p_i) = Q, j = 1, \dots, n\}$$

其中,  $p_i$  表示著作权人集合中某一著作权人,  $m_l$  表示知识载体集合中某一载体,  $Q$  的取值与上文类似,取 1 表示  $p_i$  与  $m_l$  存在链接关系,取 0 表示不存在链接关系。

(3)子网  $G_k$  与  $G_m$  间的映射关系。文献知识与物质载体之间的映射关系集合用  $E_{k-m}$

进行描述,表示存在两种关联:文献知识存储在哪些物质载体中;物质载体中存储了哪些文献知识。定义表达式为:

$$E_{k-m} = M(k_j) + K(m_l)$$

$$= \{m_l \mid m_l \in M, \gamma(k_i, m_l) = Q, l = 1, \dots, n\} + \{k_j \mid k_j \in K, \gamma(m_l, k_i) = Q, j = 1, \dots, n\}$$

其中,  $m_l$  表示知识载体集合中某一载体,  $k_j$  表示文献知识集合中某一知识元,  $Q$  的取值与上文类似,取 1 表示  $k_i$  与  $m_l$  存在链接关系,取 0 表示不存在链接关系。

### 3.4 数字资源的广度关联

根据上述映射关系,可以在子网络  $G_k$ 、 $G_p$ 、 $G_m$  的节点间建立映射关系,当不同类型的节点之间存在链接关系,就在这三个子网络之间增加一条边,即对于  $G_k$ 、 $G_p$ 、 $G_m$  子网中每一个节点分别在另外子网中进行一对多映射,利用关联描述公式  $E_{p-k}$ 、 $E_{p-m}$  和  $E_{m-k}$ ,获取显性对应的著作权人集合  $P$ 、文献知识集合  $K$  或物质载体集合  $M$ ,最终建立子网络之间所有

的关联关系,从而将原先孤立的这三种不同类型的子网络聚合在一起,形成一个包含三种类型节点的超网络,即为数字资源超网络。

该过程可表示为:

令  $p_i \in P$ ,  $k_j \in K$ ,  $m_l \in M$  分别表示  $P-P$  网络、 $K-K$  网络和  $M-M$  网络中任一节点,布尔变量  $\alpha(p_i, k_j)$ 、 $\beta(p_i, m_l)$ 、 $\gamma(m_l, k_j)$  分别表示不同类型节点之间是否相关,当取值为 1 时表示相关,为 0 时表示不相关,则数字资源超网络  $KSN$  有:

$$KSN = f(G_k, G_p, G_m) = G_k + G_p + G_m + E_{p-k} + E_{p-m} + E_{m-k} = (K, P, M, E_{K-K}, E_{P-P}, E_{M-M}, E_{p-k}, E_{p-m}, E_{m-k})$$

其中,  $E_{p-k}$ 、 $E_{p-m}$ 、 $E_{m-k}$  的计算方法如 3.3 小节所示。

从以上  $KSN$  模型的构建和描述可以看出,数字资源超网络集聚了人、文献知识、载体这常用的 3 种类型的文献知识子网络,包括 3 种类型节点和 6 种关联关系形成的节点与节点之间的连线,比较丰富

地揭示和展现了数字资源知识体系的复杂构成和结构形态,实现了数字资源的广度整合。

#### 4 数字资源超网络的深度聚合模式

数字资源超网络中存在多种关联关系,例如,文献之间的引用关系,体现了文献之间知识的流动和继承;文献作者之间的共现关系,体现了同一或不同科研机构学者之间的合作研究或知识交流等。数字资源超网络可以从不同角度、不同模式实现数字资源的多维度深度聚合,可以充分呈现与揭示不同知识维度之间的相互关联,从而进一步实现不同模式和维度的知识发现。笔者以最常见的期刊论文为例,从引用、共现和耦合等关联关系分析数字资源超网络中的各种聚合模式<sup>[12]</sup>,探讨不同模式下的文献分析方法和知识发现的具体应用,总结如表 1 所示:

表 1 数字资源超网络聚合模式与知识发现应用

子网类型	聚合模式	关联关系	分析方法	知识发现具体应用
$P-P$ 子网	基于学者关联的聚合	文献作者—文献作者共现	作者合作分析 中心性分析 凝聚子群分析	发现作者知识交流情况,揭示科研领域合作网络
		文献作者—引文作者共现	作者直引—互引分析 中心性分析 凝聚子群分析	发现作者知识交流与知识传承情况,揭示学科进展
		引文作者—引文作者共现	作者共被引分析 凝聚子群分析	发现作者学术贡献与影响力情况揭示学术共同体
	基于机构关联的聚合	文献机构—文献机构共现	机构合作分析 中心性分析 凝聚子群分析	发现机构交流情况
$K-K$ 子网	基于主题关联的聚合	关键词—关键词共现	共词分析 凝聚子群分析	发现领域研究主题、研究热点与趋势
	基于文献关联的聚合	引文—引文共现	文献共被引分析 中心性分析 凝聚子群分析	揭示学科研究主题结构
	基于文献—主题的聚合	关键词—引文共现	引文关键词耦合分析	发现文献研究主题,识别核心文献
	基于知识关联的聚合	知识单元—知识单元共现	内容分析 知识单元共现分析	分析学科知识的构成,进行知识挖掘与知识发现

续表

子网类型	聚合模式	关联关系	分析方法	知识发现具体应用
$M-M$ 子网	基于期刊关联的聚合	文献期刊—被引期刊共现	期刊直引—互引分析 中心性分析 凝聚子群分析	寻找核心期刊,进行期刊评价
		被引期刊—被引期刊共现	期刊共被引分析 凝聚子群分析	揭示期刊主题相关性及其对学科发展的影响
	基于用户需求关联的聚合	资源标签—资源标签共现	标签共现分析 中心性分析 凝聚子群分析	发现用户需求,引导资源推荐
$P-K$ 子网	基于学者—主题关联的聚合	文献作者—关键词共现	作者关键词耦合分析 2-模网 2-模量表因子分析 核心-边缘分析	发现作者知识扩散与分布情况,揭示潜在合作群体与作者研究集中领域
	基于学者—文献关联的聚合	文献作者—引文共现	作者文献耦合分析 2-模量表因子分析	发现研究活跃群体以及当前研究主题结构
	基于机构—主题关联的聚合	文献机构—关键词共现	机构关键词耦合分析 2-模量表因子分析 核心-边缘分析	揭示机构研究主题相关性,寻找潜在合作伙伴
$P-M$ 子网	基于学者—期刊关联的聚合	文献作者—文献期刊共现	作者期刊耦合分析 2-模量表因子分析	发现期刊核心作者,揭示期刊对学者的吸引力
	基于机构—期刊关联的聚合	文献机构—文献期刊共现	机构期刊耦合分析 2-模量表因子分析	发现相似研究水平的科研机构
	基于学者兴趣关联的聚合	文献作者—资源标签共现	作者标签耦合分析 2-模网 2-模量表因子分析	进行相关学者聚类,发现研究兴趣相似的学者
$M-K$ 子网	基于期刊—主题关联的聚合	文献期刊—关键词共现	期刊关键词耦合分析 2-模网 2-模量表因子分析 核心-边缘分析	揭示报道主题相似的期刊,寻找某主题研究的核心期刊
	基于期刊—引用关联的聚合	文献期刊—引文共现	期刊引用耦合分析 2-模网 2-模量表因子分析	揭示主题相似研究的期刊分布和知识传承网络

实际应用中,假如用户利用学者的姓名开始检索,则可以选择基于学者关联的聚合模式,利用合作作者分析、作者关键词耦合分析、作者共被引分析等方法,从引用、共现和耦合等关联维度展现数字资源聚合结果,即呈现相互关联的 $P-P$ 子网络、 $P-K$ 子网络和 $P-M$ 子网络。还可以进一步设定和调整设定的聚合参数,选择不同的聚合层次和聚合深度的阈值,如选定单层共现,共现强度的阈值为6,则

在聚合层中将利用作者合作分析或作者共被引分析方法,将共现频次大于或等于6的共现网络呈现出来,方便用户对相关作者数字资源的筛选和利用。很多情况下,用户在检索时采用了多条件搜索,输入多个检索项或是包含布尔逻辑的组合项,这时就需要采用交叉关联聚合模式进行数字资源的聚合和分析。

## 5 数字资源超网络聚合系统模型

根据以上对于数字资源超网络聚合模式和应用的分析,笔者进一步构建数字资源超网络聚合系统模型如图2所示。



图2 数字资源超网络聚合系统模型

从图2中可以看出,数字资源超网络聚合系统模型从下至上分为基础层、关联层、聚合层和应用层4个层次,从最原始的数字文献资源到互相关联的数字资源网络,再到经过多维聚合模式形成的数字资源超网络,这4个层次相互串联,贯穿于整个过程之中,且处于内容和体系不断更新变化的状态。

其中,基础层主要功能是负责采集文献数据、用户数据、文献知识数据等,存储数字资源中的人、知识、载体等知识要素信息,形成多种类型的资源数据库;关联层通过对各种类型数据进行处理,提取不同类型知识间的多维关联关系,生成P-P、K-K、M-M、P-K、K-M、M-P等数字资源子网络,实现数字资源的广度关联;聚合层在完成数字资源子网络关联链接的基础上,从主题关联、机构关联、学者-

主题关联、用户需求关联等维度对数字资源进行多维深度聚合,利用语义内容分析、引用分析、共现和耦合分析、2-模量表因子分析、核心-边缘分析等方法,实现数字资源超网络的知识聚合与发现;应用层是用户与数字资源聚合超网络系统的交互层,提供友好的使用界面,用户输入检索需求,系统自动分析检索项,进行超网络计算,将相关联的计算结果以可视化的方式呈现给用户,并提供易操作的参数调节选项,供用户对输出的结果进行调整,此外,还通过相关数字资源推荐提供个性化服务和知识服务,帮助用户预测未来的发展趋势。

数字资源超网络系统中的关联关系采用RDF(Resource Description Framework)数据来进行描述、存储和发布,包括节点、关联属性和属性值组成的三元组( $RDF = \langle \text{subject}, \text{predicate}, \text{object} \rangle$ ),由于RDF使用的是结构化的XML数据,为超网络中的节点及它们之间的相互关系定义了一个简单数据模型和语法,节点之间关联关系信息可以轻易地在使用不同类型的操作系统和应用语言的计算机之间进行互操作;搜索引擎也可以理解元数据的精确含义,使得搜索变得更为智能和准确,便于分散数据的机器自动交换和使用。

系统模型另外还包括政策与法规体系、数字资源保障体系、安全保障体系和分工与协作体系这4个系统运行保障和支撑体系。政策与法规体系对数字资源超网络发展要达到的目标、遵循的原则等进行明确的规定,为数字资源超网络聚合发展提供了指导,明确了方向;数字资源保障体系为数字资源超网络提供源源不断的知识原料,使得整个资源网络体系处于动态变化过程中,不断更新;安全保障体系提供用户权限控制、资源异地镜像备份、容灾系统、系统危机管理等功能,为数字资源超网络系统的稳定运行保驾护航;分工与协作体系通过云计算、云服务指明分布式资源系统建设的方案和标准,避免资源的重复建设和浪费,消除信息孤岛,充分发挥各协作机构的资源优势和特长,提升数字资源聚合的效率和超网络系统的投入产出比率,促进数字资源超网络聚合系统的生态化。这四大体系形成一个有机整体,从政策、资源、安全、实施等方面维持系统的运行和不断优化,保障数字资源超网络聚合系统的良性循环发展。

另外,在具体实施数字资源超网络构建时,建议由一个专门的机构,如CALIS、NSTL等,实行统一管理的模式,采用统一的数据库格式和标准,根据各单

位资源特点和自身优势,联合行动,避免不同数据库之间的后期转换合并以及重复建设,降低成本,提高效率。

## 6 结 论

本文在对数字资源聚合涉及的要素进行分析的基础上,通过构建数字资源  $P-P$  子网络、 $K-K$  子网络、 $M-M$  子网络,并建立网络中人、文献知识、物质载体等 3 种类型节点的复杂关联关系,最终将多个类型的子网络链接起来构建了数字资源超网络。与以往单一类型节点及关系组成的文献知识网络,如文献被引网络、关键词共现网络等相比,数字资源超网络更丰富地揭示和呈现了文献知识体系的复杂结构,对于更全面、深入地开展数字资源知识系统研究提供了路径。具体表现在:

### 6.1 全面深入揭示文献知识体系结构

因为数字资源超网络中集合了人、文献知识和存储载体,不仅可以多维度表现文献知识结构,还可以反映其存储情况,较全面描述了文献知识体系。例如,用户不仅可以在  $P-P$  子网络中采用基于学者关联的聚合模式,利用作者被引分析得到某研究领域中的研究学者聚类,发现核心学者,还可以进一步选择聚合方式和网络细节,得到文献主题的研究学者群体以及他们之间的知识流动情况,使得用户可以清晰了解该文献主题及细分领域的研究学者和知识结构,迅速掌握该主题研究现状和空白点,为用户开展科研活动提供强力支撑。

### 6.2 更好的利用相关数字资源

数字资源超网络不仅反映了文献知识结构,还揭示了著作权人、文献知识以及存储载体之间的多维度关联,这些关联和链接可以以关系数据库的形式存储,当某一节点发生变化,只要对该节点本身进行修改或维护即可,不影响到超网络中的其他节点。另外,这些关联和链接通过可视化的形式呈现给用户,直观形象,易于理解。用户在检索文献知识时,选择任一途径,如著作权人等,即可检索到研究者、文献知识以及存储知识的资料,如拥有某主题文献知识的著作权人可以表示为:

$$P(k_j) = f(p_i, k_j) = \{p_i \mid p_i \in P, \alpha(p_i, k_i) = 1\}$$

拥有某种知识点的文档资料可表示为:

$$M(k_j) = g(m_i, k_i) = \{m_i \mid m_i \in M, \gamma(m_i, k_i) = 1\}$$

用户还可以通过设置不同阈值或参数对检索结果进行筛选,极大地方便了用户有效获取和利用相关文献知识,降低使用成本,节约科研经费。

### 6.3 便于文献知识的创新和发现

利用超网络聚合不同语言、不同载体、不同类型的数字资源,为各类科研和知识管理活动的开展提供了方便,有利于文献知识的创新和发现,包括知识学习、知识传播、知识创新等。

(1)文献知识学习。个人或机构学习文献知识反映在数字资源超网络中,即  $K-K$  子网络中文献知识节点增加或者是原文献知识点的权重增加,同时,在  $P-K$  网络中节点之间的连线增加或是关联关系增强,从而明确反映出用户开展知识学习的成果。

(2)文献知识传播。个人或机构在接受知识后对外传播,体现在数字资源超网络中,不仅是  $K-K$  子网络中知识点权重的变化,还反映在  $P-P$  子网络中节点间的连线增多和关联关系增强,以及  $K-M$  网络中边的增加。从多个层次和纬度呈现出知识传播的效果。

(3)文献知识创新。用户开展知识创新活动时,会产生新的知识,如提出观点或发表文献。这些新的文献知识可能是由已有文献知识相互关联和作用产生,形成新的文献知识子网络,也可能与已有知识没有关联。反映在构建的数字资源超网络中:① $K-K$ 子网络会出现一些新的节点和节点之间的连线,这些节点和连线本身可能形成一个更小的  $K-K$ 子网络。另外,这些节点可能与原先节点存在关联链接,也可能没有,形成孤立节点。② $P-K$ 子网络也可能出现新的节点和连线,表明著作权人知识的增加。③ $K-M$ 子网络也会有新的节点和边出现,说明该文献知识以某种形式的载体被存储。此外,还会促使  $P-P$ 子网络、 $M-M$ 子网络、 $P-M$ 子网络也出现新的关联链接关系。如此,文献知识创新的成果的多种要素、多种关联都一一反映在数字资源超网络的多个层面上。

另外,数字资源超网络的构建还有利于对网络中的文献知识进行编码,如形成树形结构等,方便利用计算机存储、读取和检索。还可以结合社会网络分析等方法,探讨多种因素影响下文献知识的演化路径,以及数字资源超网络或其子网络的结构稳定性等。

总之,数字资源超网络的构建能够弥补领域本



体中层级结构知识关系的不足,同时也将数字图书馆资源聚合与知识发现扩展到一个更广阔的复杂系统空间。由于数字资源超网络是一个非常复杂的网络体系,本文研究仅从理论上进行了探索,对于实践中数字资源超网络的具体构建尚待进一步实现和完善。另外,本文探讨的构建偏向于文献知识,下一步将针对知识单元,特别是经语义标注的知识单元的超网络聚合进行研究。

### 参 考 文 献

- [1] 贺德方,曾建勋. 基于语义的馆藏资源深度聚合研究[J]. 中国图书馆学报,2012(7):79-87.
- [2] 贺德方. 基于语义的数字图书馆服务创新[J]. 情报学报,2012(10):1012-1017.
- [3] 2011年度国家社科基金重大项目开题报告之六:基于语义的馆藏资源深度聚合与可视化展示研究[EB/OL]. [2014-07-16]. <http://www.npopss-cn.gov.cn/GB/219506/219507/16887825.html>.
- [4] 科学基金共享服务网[EB/OL]. [2014-07-16]. <http://npd.nsf.gov.cn/>.
- [5] 张云中. 从整合到聚合:国内数字资源再组织模式的变革[J]. 数字图书馆论坛,2014(6):16-20.
- [6] 范亚芳,边佳平. 对高校图书馆虚拟馆藏资源整合的分析研究[J]. 图书情报工作,2002(9):60-63.
- [7] 章成志,苏新宁. 信息资源整合的建模与实现方法研究[J]. 现代图书情报技术,2005(10):60-63.
- [8] 庞跃霞,曹丽娟,丁申桃. 高校图书馆馆藏目录整合方法探讨[J]. 图书馆杂志,2006(4):40-42.
- [9] Chang Chun, Lu Wenlin. Agricultural cross languages information retrieval schema based on muti-thesaurus mapping[C]//Computer and Computing Technologies in Agriculture II, Volume 1. New York: Springer, 2009: 357-364.
- [10] NSTL. 跨库检索[DB/OL]. [2014-07-16]. <http://cds.nstl.gov.cn/>.
- [11] ISI web of knowledge[EB/OL]. [2014-07-16]. <http://wokinfo.com/>.
- [12] Collections in the age of e-research; realizing potential through curation and aggregation[EB/OL]. [2014-07-10]. <http://www.clir.org/dlf/forums/fall2010/22PalmerDLF.pdf>.
- [13] Brad E. Practical applications of 2D and 3D information visualization for information organizations[J]. Library Technology Reports, 2005, 41(1):55-64.
- [14] Buckland M, Chen A, Gey F C, et al. Geographic search: Catalogs, gazetteers, and aps[J]. College & Research Libraries, 2007, 68(5):376-387.
- [15] Coyle A. Interior library GIS[J]. Library Hi Tech, 2011, 29(3):529-549.
- [16] Buchel O, Sedig K. Extending map-based visualizations to support visual tasks: The role of ontological properties[J]. Knowledge Organization, 2011, 38(3):204-229.
- [17] Julien C A, Guastavino C, Bouthillier F. Capitalizing on information organization and information visualization for a new-generation catalogue[J]. Library Trends, 2012, 61(1):149-161.
- [18] Osinska V. Fractal analysis of knowledge organization in digital library[J]. [EB/OL]. [2014-09-30]. <https://repozytorium.umk.pl/handle/item/467>.
- [19] Fagan J C. Usability testing of a large, multidisciplinary library database: Basic search and visual search[J]. Information Technology and Libraries, 2013, 25(3):140-150.
- [20] Clustering And Sustaining Digital Resources[EB/OL]. [2014-09-30]. <http://www.jisc.ac.uk/media/documents/publications/general/2011/JISCeContentClusteringAndSustainingDigitalResources.pdf>.
- [21] 邱均平,王菲菲. 基于共现与耦合的馆藏文献资源深度聚合研究探析[J]. 中国图书馆学报,2013(5):25-33.
- [22] 李劲,程秀峰,宋红文. 基于语义的馆藏资源深度聚合方法研究[J]. 情报科学,2013(11):100-103.
- [23] 陈晓美,毕强,滕丹青,等. 语义网环境下数字图书馆知识发现的维度框架研究[J]. 情报学报,2014(2):148-157.
- [24] 邱均平. 馆藏资源语义化与多种方法的融合研究[J]. 图书情报工作,2014,58(7):5-24.
- [25] 毕强. 数字资源:从整合到聚合的转变[J]. 数字图书馆论坛,2014(6):1.
- [26] Sheffi Y. Urban transportation networks: equilibrium analysis with mathematical programming methods[M]. NJ: Printice-Hall, 1995.
- [27] Nagurney A, Dong J. Supernetworks: decision-making for the information age[M]. Cheltenham: Edward Publishing, 2002.
- [28] Nagurney A, Dong J. Management of knowledge intensive systems as super-networks: modeling, analysis, computations, and application[J]. Mathematical and Computer Modeling, 2005, (42):397-419.
- [29] Xi Yunjiang, Dang Yanzhong. Method to analyze robustness of knowledge network based on weighted supernetwork model and its application[J]. Systems Engineering Theory & Practice, 2007, 27(4):134-140.
- [30] 席运江,党延忠,廖开际. 组织知识系统的知识超网络模型及应用[J]. 管理科学学报,2009(3):12-21.

- [31] 王众托. 无处不在的网络社会中的知识网络[J]. 信息系统学报, 2007, 1(1): 1-7.
- [32] 王众托, 王志平. 超网络初探[J]. 管理学报, 2008(1): 1-8.
- [33] 王志平, 王众托. 超网络理论及应用[M]. 北京: 科学出版社, 2008.
- [34] 李晟. 基于知识超网络视角的知识协同方法和技术的研究——以软件行业为研究实例[D]. 杭州: 浙江理工大学, 2011.
- [35] 徐升华, 邹红. 基于超网络模型的知识转移动力分析[J]. 情报杂志, 2011(7): 94-98.
- [36] 沈秋英, 王文平. 基于社会网络与知识传播网络互动的集群超网络模型[J]. 东南大学学报(自然科学学报), 2009(2): 413-418.
- [37] 孙薇, 马钦海, 于洋. 基于知识超网络的科技创新团队的组建方法[J]. 科学学与科学技术管理, 2013, 34(8): 166-171.

(责任编辑 魏瑞斌)

doi:10.3772/j.issn.1000-0135.2015.001.002

## 基于复杂网络演化模型的三元闭环合著网络研究<sup>1)</sup>

张洋 麦江萌

(中山大学资讯管理学院, 广州 510006)

**摘要** 本文通过构建和分析国外图书情报领域的三元闭环合著网络,改进已有的复杂网络演化模型,提出严格三元闭环概念和扩张系数用以研究影响合著网络扩张速度的关键因素。研究表明,学者合著集群的扩张速度,与他跟集群成员合著的均匀程度成正相关关系。扩张系数是对目前合著网络测度指标的一个有效补充。

**关键词** 合著网络 三元闭环 复杂网络 学术引荐人 信息计量学

### A Research of Triadic Closure in Coauthorship Network Based on Evolution Model of Complex Network

Zhang Yang and Mai Jiangmeng

(School of Information Management, Sun Yat-Sen University, Guangzhou 510006)

**Abstract** The concepts of Strictly Triadic Closure and Expanding Coefficient were proposed in the research of Coauthorship network expansion, based on the analysis of current Library & information science Coauthorship network and an optimized evolution model of Complex Network. It is proven that the expansion of a scholar's coauthorship clique is positively correlated with the homogeneous degree of his/her coauthoring with other scholars. Expanding Coefficient is considered as a useful supplementary index to the evaluation of Coauthorship network.

**Keywords** coauthorship network, triadic closure, complex network, research facilitator, informetrics

在信息计量和科学评价领域内,合著和合著网络一直是众多学者关注的重点。通过引入社会网络的分析方法,合著和合著网络研究在近年来取得了显著进展,各种关于合著网络的理论研究和实证研究不断出现,推动合著研究的深入发展。然而,基于社会网络分析方法的合著网络研究大多以静态网络为对象,学者与学者间的合著关系客观存在,分析结果则侧重于描述某特定时间点上学者在整个网络中所处的位置。对于合著网络的动态描述、学者在合著网络中扮演角色的发展演化等问题,在目前的合

著网络研究中较少有成果报道。三元闭环合著现象和学术引荐人,是国外学界近年对合著网络研究的新发现,是一种在合著关系下的隐性现象的描述。本研究尝试将动态的复杂网络演化模型用于三元闭环合著现象的研究中,探讨影响合著网络发展速度的因素及其度量指标。

## 1 文献回顾

1999年,Barabasi和Albert提出了“BA网络模

收稿日期:2014年9月18日

作者简介:张洋,男,1975年生,博士,中山大学资讯管理学院教授,主要研究方向:科学计量学、网络信息计量学、科学评价,E-mail: zhyang2@mail.sysu.edu.cn。麦江萌,男,1990年生,中山大学资讯管理学院硕士研究生,主要研究方向:科学计量学。

1) 本文为国家社会科学基金项目“新型网络环境下学术期刊影响力的计量分析与评价研究”(批准号:14BTQ067)和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”(批准号:11&ZD152)的研究成果之一。