

数字图书馆资源聚合模式研究 ——基于社会网络分析的视角*

毕强¹, 王雨¹, 孙畅²

(1. 吉林大学管理学院, 长春 130022; 2. 沈阳市委党校图书馆, 沈阳 110036)

摘要: 面向数字图书馆资源深度开发与广度关联的应用需求, 从资源特征间关联和资源利用过程关联的视角出发, 对数字图书馆资源聚合的维度、路径、模式及其应用范围等问题进行系统研究, 综合运用社会网络分析、矩阵算法、网络图谱等工具、方法, 论证数字图书馆资源聚合的机理和途径, 以期为数字图书馆资源聚合的理论研究与实践提供借鉴。

关键词: 数字图书馆; 社会网络分析; 资源聚合

中图分类号: G250.7

DOI: 10.3772/j.issn.1673—2286.2014.06.001

1 研究背景和相关工作

自1994年以来, 有关数字图书馆的研究如雨后春笋般涌现, 研究主题包括数字图书馆发展^[1-3]、数字图书馆建设^[4-6]、数字图书馆管理^[7,8]等, 其成果为数字图书馆的建设提供了参考和借鉴。然而, 数字图书馆作为知识服务的载体, 对其研究单纯停留在宏观层面上已不能切实地满足信息环境下用户对信息资源的需求。针对不同的用户群体、用户需求, 如何处理海量的数字化信息资源, 将信息资源有效地聚合为用户所用, 从根本上深化知识服务, 成为亟待深入研究和全面解决的问题。

当前, 我国社会经济已从传统经济转变为知识经济。知识经济是信息产业、通信产业和文化产业的汇集物, 而数字图书馆作为知识的汇集地是知识经济的主要基础设施和物质形式^[9], 数字图书馆资源聚合在知识利用过程中发挥着重要的作用。本文以此为切入点, 基于社会网络分析的研究视角提出“数字图书馆资源聚合”这一研究主题。目前, 对该主题的研究仍处于探索

性阶段, 因此, 深入剖析数字图书馆资源聚合的机理, 形成资源聚合模式, 不仅是必要的, 而且具有重要的理论意义和现实意义。

2 数字资源聚合理论应用研究

2010年5月, 全国数字图书馆建设与服务联席会议发布了《数字图书馆资源建设指南》, 其中第二条提出: “数字图书馆资源是指图书馆以数字形式发布、存取和利用的信息资源的总称”。数字图书馆馆藏资源与传统图书馆馆藏资源不同, 它不仅包括不同的信息格式(如磁盘、光盘、磁带等), 还包括不同的信息类型(如书目信息、全文信息、图像、音频、视频等)。数字图书馆是指以数字形式贮存和处理信息的图书馆, 具体到数字图书馆资源是指经过收集、处理、加工的图像、文本、音频、视频、数据库等多媒体信息, 是数字化的信息资源。

聚合作为网络用语, 是指对互联网各种信息的集合^[10]。而资源聚合实质上是指聚集和融合多个资源特征的多

* 本研究得到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”(编号: 11&ZD152)和国家自然科学基金项目“语言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”(编号: 71273111)资助。

种不同类型的资源^[11]。资源聚合不是简单的信息堆积,而是对海量资源进行深度挖掘分析之后的分类和挑选。数字资源聚合的最终目的是针对图书馆用户需求为之提供切实的知识服务、更具针对性的信息,以提高用户的资源利用效率。

资源聚合作为一个新的概念,目前还没有明确统一的定义。有学者认为:信息资源聚合是指从不同的数据源汇集并分析相关信息,解决这些信息在语义方面的异构性,并提供基于数据源之间关系、业务过程的聚合等功能^[12]。也有学者认为:信息资源聚合简单地说,就是异构信息资源的语义集成和协同工作过程^[13]。此外,亦有研究表明,在大规模、动态的网络环境下,为了实现信息资源的深度共享和有效利用,需要提供一种信息聚合平台,它不仅能够实时地提供信息聚合的可视化视图,而且还能够根据信息的动态变化自动进行特征重组,调整聚合的路径以及模式。笔者认为,姜灵敏对信息资源聚合的概述较为全面、准确,他在《信息资源聚合与数据挖掘》一书中提到:信息资源聚合就是对互联网上的信息资源进行聚集、挖掘和融合的过程,目的是解决分布、异构信息资源的资源共享、语义融合和服务智能化等问题,最终根据用户需求提供相关联的、个性化的信息以及智能化的检索、查询服务^[14]。

由此可见,资源聚合的研究主要包括两个方面:一是信息的聚集,即如何收集和挖掘互联网上的信息资源;二是信息的融合,即如何解决异构资源互操作和自主协同问题。如何使信息资源能够互相理解,并根据用户的需求有效、动态、智能地聚集融合各种资源,是信息聚合研究的核心问题。只有信息资源的有序化聚合才能深度共享,才便于可控可管,才利于动态、跨领域的网络信息资源集成,构造个性化的智能应用服务,从而辅助实现知识创新、协同工作、问题解决和决策支持。

在资源聚合方面,图书情报领域的学者也有不同程度的探索,如针对数量巨大的异构馆藏资源进行聚合与重组、建立基于语义的馆藏资源聚合模型^[15],利用关联数据进行文献信息聚合^[16],基于OPAC、跨库检索、引证关系和元数据的聚合^[17],数字信息资源整合技术的研究^[18],将文献计量分析方法应用到馆藏资源语义化^[19]等等,而从社会网络分析的角度去审视,通过构建资源间的社会网络如何聚合资源还需要进一步研究,具有很大的挖掘和研究空间。

3 基于社会网络分析的数字图书馆资源聚合的机理及模式

3.1 馆藏资源聚合的机理

对数字图书馆服务而言,已由传统的服务转变为现代的信息服务,兼及知识服务,但知识服务只是一种认识和组织服务的观念。信息服务的深化,其有效途径是馆藏资源的聚合。目前,馆藏资源的聚合所要解决的主要问题有:资源重构、协调规范化和自组织、语义互联和智能聚集融合^[20]。社会网络分析为数字图书馆资源聚合提供了新的契机,特别是借助SNA在可视化方面的优势把数字图书馆资源聚合和可视化推向了一个新的高度。社会网络分析应用于数字图书馆资源聚合,即构建一个网络系统能够识别的结构化的社会网络,当用户检索时系统能够根据检索条件提供更多相关联的资源,实现更加智能化的应用。基于社会网络分析的数字图书馆资源聚合主要体现在以馆藏数字资源的内、外部特征(如关键词、作者、主题、出版机构)为基础,通过作者互引关系分析、作者合作关系分析、作者-关键词分析等关系分析,实现相关领域资源在某种方面的聚合;其次体现在根据资源内、外部特征聚合呈现出的效用特征,以揭示研究主题、学术影响力、核心作者群等。

3.2 馆藏资源聚合的模式

通过对资源的内、外部特征分析,结合资源特征间的社会网络关系以及资源利用过程中的社会网络关系,将数字图书馆资源聚合模式凝练为6种:基于作者互引关系的资源聚合、基于作者合作关系的资源聚合、基于作者-关键词关系的资源聚合、基于多作者-关键词关系的资源聚合、基于多关键词-作者关系的资源聚合,以及资源特征间交叉关联的聚合。结合资源聚合的模式,笔者将对各聚合模式下的应用范围作进一步延伸,如表1所示。其中资源特征间的社会网络关系只是根据现有的研究得出,并未涵盖所有的聚合模式,如基于关键词共现关系的资源聚合、基于机构间合作关系的资源聚合、基于机构-作者间的合作关系的资源聚合等等。

表1 基于社会网络分析的数字图书馆资源聚合模式^[21]

	关系路径	社会网络分析方法	聚合模式	应用范围
资源特征关联	作者-作者互引	中心性分析 距离和密度分析 凝聚子群分析	基于作者互引关系的聚合	发现作者间引用情况, 识别核心资源
	作者-作者合作	中心性分析 凝聚子群分析	基于作者合作关系的聚合	挖掘作者间的相同的研究主题, 揭示外在的合作群体
资源利用过程关联	作者-关键词交叉关联	2-模网 2-模量表因子分析 核心-边缘分析 分派分析	作者-关键词交叉关联	以关键词为纽带, 揭示潜在合作群体
	多作者-关键词交叉关联	2-模量表因子分析 核心-边缘分析 分派分析	多作者-关键词交叉关联	寻找关键词共同体, 识别相同研究主题
	多关键词-作者交叉关联	2-模网 2-模量表因子分析 核心-边缘分析 分派分析	多关键词-作者交叉关联	识别同一研究领域的核心作者
用户需求关联	检索内容-检索条件匹配	调用资源 聚合结果	检索内容-检索条件匹配	发现用户需求, 提高用户检索效率及资源的利用率, 满足个性化服务需求

4 社会网络分析在数字图书馆资源聚合中的应用

4.1 数字图书馆资源社会网络关系的创建

馆藏资源社会网络关系的创建是实现数字图书馆资源聚合的重要一环, 创建的过程应满足如下要求^[22]: 根据资源的内外部特征, 将资源分类, 具备相似特征的资源可根据资源的中心性、整体网的凝聚性、子网络的小团体性等建立社会网络关系。关系的紧密程度取决于资源本身的特征及资源间的强弱关系。对多元化的馆藏资源进行结构属性分析和关联关系分析, 以此来建立资源群落、聚合相关资源。关系创建阶段关键在于对各类馆藏资源的分布状况的拓扑结构进行充分的研究。只有了解馆藏资源的结构及内、外在特征关联程度, 所创建的才是资源的社会网络关系。以互联网平台存储所创建的馆藏资源关联信息, 建立资源群落、聚合相关资源, 进而识别用户的知识诉求和发现知识热点。

4.2 数字图书馆资源社会网络关系的形成

资源之间的关联构建主要包括三个层面: 资源数据提取、社会网络关系的形成、聚合模式。其中, 网络关系的形成是最关键的环节。社会网络最突出的优势是结构关系清晰、明朗, 支持资源间特征化的关联, 并通过“节点”和“有向性的线段”连接整个网络, 其中, 线段的粗细、节点的大小显现出的便是资源间关系的紧密程度及资源所处的地位和影响力。通过社会网络关系中心性、凝聚子群、派系等建立不同馆藏资源间的关联, 以此提高检索结果的价值性和关联性, 优化查询结果和规范信息资源管理。

4.3 馆藏资源社会网络可视化

从宏观的层面来看, 社会网络关系的形成、聚合模式的构建的最大好处是可以提供多个聚合应用的数据关联访问, 将不同资源的关联特征或不同资源间建立

的社会网络进行整合、提取, 遵照用户的检索条件和检索内容, 提供给用户关于该匹配条件的相关资源的可视化视图。从某种程度上来说, 这就避免了检索结果的单一化、无序化, 为用户的检索提供了更智能化的导航, 这无疑是信息检索和知识发现的有力途径。但不容忽视的问题是, 馆藏资源的不断增加导致聚合模式不是固定不变的, 新资源的收录, 资源之间的社会网络关系会发生变化, 检索结果也会随之改变。因此, 数字图书馆系统应定时对资源间的社会网络关系进行检查, 当有关系发生变化时, 检索系统自动跟踪, 根据检索结果调整可视化内容。

5 基于社会网络分析的数字图书馆资源聚合模型构建

从上述数字图书馆资源聚合的社会网络关系的揭示可以看出, 针对数字图书馆资源聚合下的社会网络分析, 通过深度揭示资源之间的社会网络关系, 将能更好地挖掘、建立、利用资源之间的联系, 从而提升检索质量, 优化检索结果, 将繁杂、无序的检索内容通过资源的聚合, 实现检索结果的关联性, 在此基础上, 根据社会网络分析的结果, 开发新的检索方法, 增加检索的数字图书馆资源的数量^[23]。

对此, 本文提出基于社会网络分析的数字图书馆资源聚合模型。该模型从微观到宏观的层面上共设数据层、分析层、聚合层、应用层四个层次, 从原始的文献数据甄选到选取社会网络指标分析资源关系网络, 再到资源内容的聚合, 最终实现用户对资源的挖掘与定位, 有效获取聚合的资源, 提高资源的利用率。

(1) 数据层: 首先依据文献的内外部特征从馆藏资源数据库中选取文献资源数据。其次选取某一学科领域为研究样本, 根据所要研究的对象确定作者群、关键词、机构、文献来源等要素, 对于数据采取严格筛选、剔除重复的处理方式, 进行数据甄别。随后根据社会网络分析各指标的特性, 完成数据库各要素内部以及不同要素之间关系的建立。关系的建立是后续进行社会网络关系分析的关键。

(2) 分析层: 依据选取的研究样本, 分析各要素之间的相关性, 构建关系矩阵, 结合社会网络分析中的中心性、凝聚子群、引文结构洞以及2-模网络指标, 对作者潜在合作关系、外在合作关系、引文关系、机构合作关系、关键词聚类、共词关系等进行深层次分析, 从而

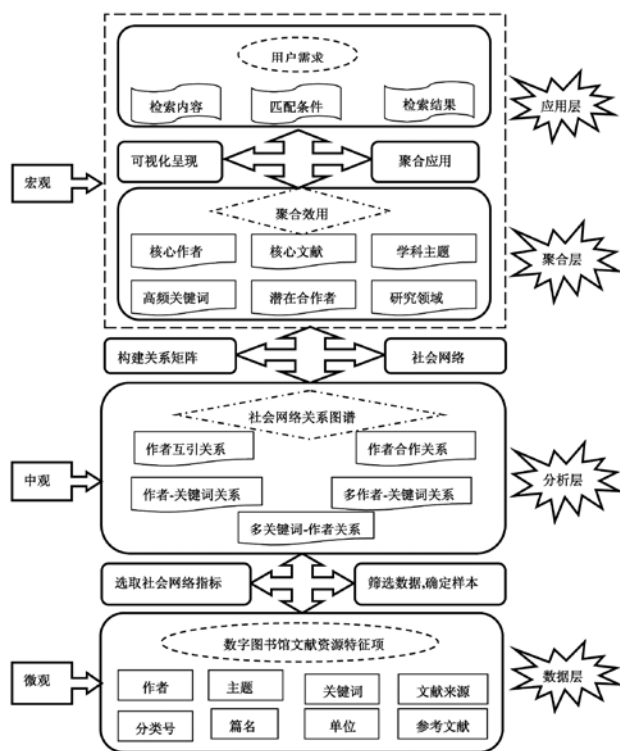


图1 基于社会网络分析的数字图书馆资源聚合模型

对聚合模式提供一定参考。

(3) 聚合层: 该层是对分析结果的分组聚合, 也是实现用户对馆藏数据应用的过渡层, 起到承上启下的作用。它既要根据用户的需求提供完善的检索结果, 还要在社会网络分析的基础上, 利用节点、线条将各要素相关联, 节点的大小、线条的粗细表明了要素之间关系的强弱以及所处的地位, 最终呈现可视化的视觉符号, 准确定位核心作者、核心机构、学科前沿、热点领域等, 为用户提供有价值的检索内容, 这对用户了解某一学科领域的研究现状以及发展动态大有裨益。聚合模式主要是根据用户的检索内容捕捉数据库中的聚合资源, 例如用户的检索字段是关键词, 则相对应的便是聚合模式中的关键词的聚合, 即会调用分析层中的关键词聚类分析、共词分析、关键词中心性分析与关键词相关的社会网络分析结果; 深层次的便是根据用户特定的检索条件, 调用聚合阈值与之相匹配的资源, 提供核心资源, 进而实现用户的检索需求。此外, 当用户的检索条件由多个检索字段组成, 则需交叉资源聚合模式, 例如: 机构-关键词、作者-关键词、作者-机构等。

(4) 应用层: 这一层主要发挥着需求检索资源的作用。其中, 资源的利用率以及检索效果取决于聚合层的资源聚合程度。通过资源聚合可视化的导航平台以及个性化检索服务, 让用户对所需资源尽览无余, 并能从聚合的资源中预测到未来的研究趋势。

聚合呈现的是根据相关联要素的特征组建的关系, 从某种程度上说, 聚合的关系正是要素之间相似性的体现, 聚合的程度可以通过社会网络关系的强弱来反映。上述分析将聚合的过程细化到数据层、分析层、聚合层、应用层, 但从整体上来看, 社会网络分析在资源聚合中的应用主要体现在微观、中观、宏观的层面上, 各个层面揭示了不同的聚合结果, 见表2。

表2 社会网络关系在各个层次聚合的应用

微观层面	揭示资源间的特征关系, 主要是对数字图书馆文献资源数据的提取, 从海量的文献中提取资源的特征, 按照知识单元的特点分类, 根据社会网络的指标匹配检索项。
中观层面	揭示资源实体间的社会网络关系, 提供个性化的信息推荐服务。基于社会网络分析原理, 匹配资源间的关系, 从用户的角度出发, 尽可能地减少在传统文本匹配检索上面的繁琐。利用学者的信息需求偏好, 来分析资源主体间的关联, 从而能够挖掘出具有相似特征关联的资源。
宏观层面	揭示学科领域知识网络结构和学科前沿。社会网络关系在客观上将表面毫无关系的数字图书馆资源实体聚合为一个知识网络体系。具有社会网络关系的知识节点之间存在某种关联性, 这种基于社会网络关系的检索方式把众多知识节点聚合成一种有序结构 ^[24] , 通过分析这些结构能够揭示学科领域的知识结构与发展规律, 为科研人员的学术工作提供全面的检索方式。

6 结论与展望

本文将社会网络分析方法应用到数字图书馆资源聚合中, 阐述了数字图书馆资源聚合方式以及基于社会网络分析的数字图书馆资源聚合的机理, 提出了基于社会网络分析的数字图书馆资源聚合的模式, 构建了数字图书馆资源社会网络关系的聚合模型, 旨在实现建立在资源特征关联和资源利用过程关联基础上的馆藏资源的社会网络分析, 最终实现满足用户的个性化需求, 提高用户检索效率及准确率。本文得到的主要结论如下:

(1) 基于社会网络分析的数字图书馆资源聚合机理。社会网络分析作为数字图书馆资源聚合的技术方法, 通过网络把看似无序的、无关联的资源搭建起来, 从根本上说, 即构建一个网络系统能够识别的结构化的社会网络。对此, 基于社会网络分析中资源内、外部特征间的关联以及资源利用过程的关联, 结合社会网络分析的中心性、距离、密度、凝聚子群等指标测度, 构建作者互引关系、作者合作关系、作者-关键词关系, 以揭示潜在和外在合作者、核心作者群、研究主题、学科领域等, 实现相关领域资源的聚合, 为用户检索提供更多相关联的资源, 实现更加智能化的应用。

(2) 基于社会网络分析的数字图书馆资源聚合模型。综合馆藏资源聚合机理以及聚合模式, 融合社会网络分析方法, 构建了基于社会网络分析的数字图书馆资源聚合模型。该模型由微观、中观、宏观三大层面组成, 聚合的过程细化到数据层、分析层、聚合层、应用层四个层次。四个层次串联了从数字图书馆馆藏资源数据收集到社会网络关系图谱构建再到聚合效用的呈现直到用户根据需求获取聚合资源的整个过程。

基于社会网络分析的数字图书馆资源聚合属于一个多角度、创新性、探索性的研究主题, 社会网络分析的所有指标测度是否适合应用于资源特征间关系的构建还需要进一步实践。此外, 由于时间和个人能力有限, 对资源特征关联和资源利用过程关联的资源聚合模式的研究还未穷尽, 文中有些内容还有待推敲和完善。目前, 本研究还处于理论阶段, 对于具体应用于数字图书馆资源聚合平台的构建还有待于进一步尝试和实践。

参考文献

- [1] 杨沛超, 魏来论中国数字图书馆发展策略[J]. 情报资料工作, 2001(3): 34-37.
- [2] 刘新周. 数字图书馆发展现状及趋势分析[J]. 图书馆界, 2004(3): 24-27.
- [3] 夏立新, 刘启强. 数字图书馆馆藏发展的中外比较[J]. 图书与情报, 2007(1): 14-18.
- [4] 黄晓斌. 论数字图书馆建设的几个管理问题[J]. 图书馆学研究, 2008(3): 18-26.
- [5] 孙青, 耿筹. 数字图书馆建设及其知识产权[J]. 情报科学, 2007(7): 987-991.
- [6] 孙正东. 我国数字图书馆建设现状分析与思考[J]. 情报资料工作, 2002(2): 42-45.
- [7] 宋海艳, 郑建明, 陆宝益. 数字图书馆管理模式探析[J]. 情报理论与实践, 2007(3): 331-335.

- [8] 徐建华,付娇.近五年来我国图书馆管理研究现状分析[J].中国图书馆学报,2001(6):67-70.
- [9] 孙立红.数字图书馆资源组织[J].现代图书情报技术,2002(1):3-5.
- [10] 百度百科.聚合[EB/OL]. [2013-08-25]. http://baike.baidu.com/link?url=KadW55XOmoUrwH4pFS295jLRLpXLdWS_sdPDxyWP_2h5lt2kqIUCBNUtgcdiBPG.
- [11] 孟宇龙.基于本体的多远异构安全数据聚合[D].哈尔滨工程大学,2010.
- [12] ETZIONI O. The World Wide Web: Quagmire or gold mine [J]. Communications of the ACM, 1996(11): 65-68.
- [13] 恽爽,韩立新,董浚,等.KDW综述:基于Web的数据挖掘[J].计算机工程,2003(1):284-286.
- [14] 姜灵敏.信息资源聚合与数据挖掘[M].广东:华南理工大学出版社,2007:1-184.
- [15] 李劲,程秀峰,宋红文.基于语义的馆藏资源深度聚合模型探析[J].湖北民族学院学报(自然科学版),2013(6):212-215.
- [16] 丁楠,潘有能.基于关联数据的图书馆信息聚合研究[J].图书与情报,2011(6):50-53.
- [17] 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.
- [18] 蒲筱哥.图书馆数字信息资源整合技术研究进展[J].现代情报,2008(7):31-35.
- [19] 邱均平,余凡.基于计量分析的馆藏资源语义化理论[J].中国图书馆学报,2012(7):71-78.
- [20] 信息资源聚合与数据挖掘[EB/OL]. [2013-10-25]. http://blog.sina.com.cn/s/blog_4c683f990100bp7x.html.
- [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(3):25-33.
- [22] 王涛.基于关联数据的馆藏信息资源聚合研究[J].图书馆学刊,2012(8):44-46.
- [23] 伍革新,程秀峰.面向语义检索的数字图书馆资源聚合模式分析[J].图书馆建设,2013(10):22-26.
- [24] 杜晖.基于耦合关系的学术信息资源深度聚合研究[D].武汉大学,2013.

作者简介

毕强,男,1954年生,吉林大学管理学院教授。研究方向:知识管理。E-mail: biqiang12345@163.com。

Digital Library Resources Aggregation Research Based on Social Network Analysis

BI Qiang¹, WANG Yu¹, SUN Chang²

(1. School of Management, Jilin University, Changchun 130022, China; 2. Library, Shenyang Municipal Party School, Shenyang 110036, China)

Abstract: To meet the needs of the development of digital library resources and the associated application requirements in both width and depth, from the resource characteristics of correlation between the perspective of the process and resource utilization associated dimension, for digital library resources aggregation of path, pattern and its application scope and other issues were studied deeply and systematically. Social network analysis, matrix algorithm, the network mapping tools, such as method, aggregation mechanism and the way of demonstration of digital library resources were used comprehensively, so as to provide reference for theory research and practice of digital library resource aggregation.

Keywords: Digital library; SNA; Resources aggregation

(收稿日期: 2014-04-30)