



中国农业大学图书馆  
“图书情报学”研究丛书

# 基于重叠结构的 知识演化研究

陈仕吉 著

 中国农业大学出版社  
CHINA AGRICULTURAL UNIVERSITY PRESS

中国农业大学图书馆“图书情报学”研究丛书

# 基于重叠结构的 知识演化研究

陈仕吉 著

中国农业大学出版社

· 北京 ·

## 内 容 简 介

本书提出一种基于重叠结构的知识演化分析方法,利用这种方法可以揭示科学研究中的重叠结构,有助于深入地分析知识演化过程中研究领域之间的交叉、渗透以及演化过程。本书首先综述了知识演化分析及基于复杂网络的重叠结构探测算法的研究进展,提出基于BGLL的重叠结构探测算法,并进一步对知识网络中涉及的社团重叠度、重叠结构动力学特性、重叠结构时间演变特征及基于重叠结构的跨学科领域监测展开较为深入的研究,通过重叠结构动力学特性和演变特征的研究,验证了基于重叠结构的知识演化分析方法用于分析学科或领域间交叉渗透的可行性。

### 图书在版编目(CIP)数据

基于重叠结构的知识演化研究/陈仕吉著.—北京:中国农业大学出版社,2018.12

(中国农业大学图书馆“图书情报学”研究丛书)

ISBN 978-7-5655-2135-5

I. ①基… II. ①陈… III. ①文献分析-知识管理-研究 IV. ①G252.8

中国版本图书馆 CIP 数据核字(2018)第 254392 号

书 名 基于重叠结构的知识演化研究

作 者 陈仕吉 著

策划编辑 潘晓丽

责任编辑 郑万萍

封面设计 郑 川

出版发行 中国农业大学出版社

社 址 北京市海淀区圆明园西路 2 号

邮政编码 100193

电 话 发行部 010-62818525,8625

读者服务部 010-62732336

编辑部 010-62732617,2618

出 版 部 010-62733440

网 址 <http://www.caupress.cn>

E-mail [cbsszs@cau.edu.cn](mailto:cbsszs@cau.edu.cn)

经 销 新华书店

刷 涿州市星河印刷有限公司

版 次 2018 年 12 月第 1 版 2018 年 12 月第 1 次印刷

规 格 787×1 092 16 开本 16 印张 210 千字

定 价 68.00 元

图书如有质量问题本社发行部负责调换

# 中国农业大学图书馆 “图书情报学”研究丛书编委会

总 编 何秀荣

副总编 潘 薇 李晨英

编 委 张 权 刘清水 韩明杰 王鲁燕  
左文革 王永进 黄兴建 陈仕吉  
孙德昊 师丽娟 赵 勇 徐 伟  
王宝济 袁永翠 张红伟 刘子瑜



## 总序

传统意义上的图书馆一直担负着文献收集、保存、传播的功能,图书馆也因此成为人们查找和阅读文献的场所。从最早的2500年前美索不达米亚的亚述巴尼拔图书馆,到20世纪80年代的图书馆,图书馆的主要变化表现为文献数量的不断增多、文献种类的不断丰富、馆舍空间的不断扩大、馆员队伍的不断壮大。图书介质虽然经历了泥版、羊皮版、竹简版到纸质版的革命,但自东汉蔡伦发明纸后的近2000年中基本上没有变化,即纸质图书一统天下。但随着人类普及使用计算机后,世界的方方面面都在以前所未有的速度发生着日新月异的变化,同样也打破了图书馆千古不变的状态。进入20世纪80年代后,源自摄影的微缩胶片成为保存和阅读文献的一种新手段,但很快又被电子数字化形式保存和阅读文献的手段所取代。数字化文献资源后来居上,日益挑战纸质文献,大有成为未来图书馆文献主要形式的趋势,世界上已经出现了多家完全是数字化文献资源的无书图书馆,例如美国德雷塞尔大学(Drexel University)的无书图书馆和得克萨斯州贝尔县(Bexar County)的无书公共图书馆。数字化文献资源打破了图书馆原有的文献积累模式,一家藏量丰富的图书馆以往需要数十年乃至数百上千年时间才能积累起丰富的文献,今天可能几天就能买到上百万种的电子文献,甚至包括纸质市场上已经绝卖的文献。20世纪90年代起,互联网和搜索引擎逐渐普及到了世界的各个角落,当数字化与互联网喜结良缘后,对于包括

图书馆在内的各行各业都意味着一场革命,无论是主动参与还是被迫卷入,都将经历这场科技革命带来的洗礼。图书馆面临着转型发展,不能进入转型发展的图书馆必将逐步沦落;而实现转型发展的图书馆才能把丰富、高效、温馨的优质服务提供给用户。

目前高校图书馆面临着多方面的发展挑战,首先是以 IT 技术和信息通信技术为标志的第五次科技革命带来了生活和工作的可分散性和多样化,这些变化特征要求图书馆服务突破时间和空间限制,方兴未艾的数字化和网络化技术为此提供了可行性,图书馆开始向数字图书馆和移动图书馆发展,使得图书馆突破了时空限制,时时刻刻在用户身边。这方面的变化发展主要还是技术手段层面的,即由传统技术手段的图书馆向现代化电子技术手段的图书馆发展,由此当然也引发了适应新技术手段的管理模式变化,如操作流程简便化、主动推送服务、借还书提醒服务、手机端文献查询等。我国大多数高校图书馆已经处于这一技术进步及其管理优化的过程中。

其次是以网络搜索引擎为代表的社会公开信息网站对图书馆的挑战,越来越多的读者从谷歌、百度、新浪、雅虎等网站上搜寻和阅读文献,以前人们认为这些网络公开资源主要是非学术性的,学术性的资源还是需要依靠高校图书馆,但随着网络搜索技术的发展和全球信息资源的爆炸式累积,网络搜索引擎日益显示出巨大的优势,原有的认识和局限性正在逐渐被破除。绝大多数高校图书馆资金的有限性和资源种类的狭窄性导致的文献局限问题使得读者越来越多地转向社会网络资源网站,因为高校图书馆永远无法提供广泛多样的信息,而互联网却是一个开放的、包罗万象的无限信息空间。教育部 2012 年高校图书馆经费统计显示,532 家提供统计数据的高校图书馆中,有 207 家的年经费在 200 万元以下,扣除行政管理和硬件维修等非文献资源开支后,又有多少经费可以用于文献资源购置? 面对谷歌这样

的巨鳄,高校图书馆应当做出怎样的适应性变化来完成自己的使命?资源丰富、实力强大、效率快捷的网络搜索引擎网站对高校图书馆作用的挑战意味着:高校图书馆再也不能局限于简单的文献收藏、保存和传播功能作用,甚至不能局限于文献资源中心这一传统优势的功能作用,必须拓展功能,图书馆转型发展的过程实际上也是一个拓展功能作用的过程。高校图书馆不仅应当依然具有文献资源中心和学习中心的功能,而且还应当成为高校的学术会议中心,成为文化展览中心,成为师生创意信息交流中心;不仅是一个文化活动中心,而且还应当是一个休闲和人际交往的活动中心,很多高校图书馆附设的咖啡屋已经成为师生最爱去的校园场所之一。这也就是说,传统图书馆一直以安静为基本特征,但现代图书馆不仅是一个仍然能找到安静的个人学习场所,同时还是一个能找到互动活力的学习场所和人际交流场所。近几年我国很多新建的高校图书馆在空间布局上都考虑了这些新的功能需求,尤其是高校图书馆中面对面的温馨人际氛围是社会网站不具备的,这也是图书馆受人喜欢的重要原因之一。高校图书馆应当利用自己的优势拓展教育、咨询、科研等功能,很多高校图书馆提供的科技查新、信息素质教育、学习共享、学科咨询等服务正是这些方面功能拓展的具体体现。有条件的高校图书馆应当进一步发挥自己所具有的图情研究优势,进行学科发展状况分析和科研动向研究,向校院领导和广大师生提供深层次的学科发展和科研动向分析结果,作为他们进行学校管理决策和学术决策的有效参考信息。

第三种发展挑战来自师生对图书馆服务的需求变化。过去高校师生对传统图书馆的要求基本上是可以借到所需要的图书文献就行了,并且高校图书馆几乎是师生基本的文献信息来源。但今天学术信息来源已经大大拓展,甚至很多文献信息首先是通过网络学术搜索或从开放资源网站获取;过去人们称图书馆是知识的宝库,今天更多的知识却是来自互联网。过去高

校图书馆是学生除教室以外的主要学习场所,但随着居住条件的改善以及住所的分散化趋向,咖啡馆等阅读环境舒适的社会场所的普及,图书馆作为学习场所的作用明显降低。尽管人们依然普遍把图书馆视作学术宝藏的知识殿堂,但进入图书馆的读者却在不断减少,因为新一代师生的信息利用方式和生活方式已经有别于传统方式。在这些变化中,师生对图书馆简单服务的需求已经大大降低,而对图书馆高层次服务的需求却不断增强,特别是需要图书馆提供经过二次甚至多次加工后的信息和文献服务、提供满足读者个性化需求的信息和文献服务,这就要求图书馆能相应提升服务层次和水平。学术研究型图书馆过去只是少数实力雄厚的高校图书馆的选择,但今天将成为大多数高校图书馆的必然选择,因为简单的书刊文献服务已经到处可获。应对这种需求挑战已经大大超越了物质层面的技术和手段,而是要求图书馆馆员具有更高的综合素质和业务能力,这意味着高校图书馆馆员要从以往简单型管理服务的馆员转变为研究型学科服务的馆员,知识创造也成了时代赋予图书馆的新功能。

高校图书馆顺应这场革命性的转型发展需要很多的条件,比如观念、制度、资金、技术,等等,但其中图书馆馆员素质是一个极其关键的因素,没有一流的馆员,就难以提供一流的服务。如何提高现有馆员的综合素质和业务能力是很多图书馆在改善服务质量、拓宽服务范围、提升服务水平过程中所面临的瓶颈之一。中国农业大学图书馆一直把培养和提高馆员综合素质和业务能力作为重点工作之一,采取了多种方式方法来提高馆员的素质和能力,以“请进来”的方式让专家们向馆员展示学术高台或研究前沿;以“走出去”的方式拓宽馆员的业务视野,带着自己的问题向兄弟院校图书馆学习业务长处和先进经验;以设立研究课题的方式鼓励馆员针对自身岗位工作中的问题展开调查研究,研以致用来帮助自己提高服务质量和帮助图书馆领导班子提高决策水平,同时增强自己的研究能力和提高研究水平。

任何一位伟人都是从无知婴童起步成长的,任何一名科学家都是从莘莘学子开始成长的,从以往简单型管理服务的馆员转变到研究型学科服务的馆员,同样需要一个循序渐进、逐步提高的过程。本丛书中的作者全是中国农业大学图书馆馆员,本丛书中的研究成果都是他们立足自身岗位工作的思考结晶,他们都在各自的基础上迈出了成长的步伐,他们的研究成果不仅对于中国农业大学图书馆提高服务质量和科学管理水平是重要的,而且对于很多遇到同类问题并在思考的图书馆馆员也是有参考价值的。我相信,在学习进取型的状态中,他们的后续研究会提供更好的研究成果,他们的综合素质和服务质量也会在潜移默化中得到提升。

何秀荣

中国农业大学图书馆馆长

二〇一七年春节于绿苑

# 前言

《国家中长期科学和技术发展规划纲要》明确指出“全面推进中国特色国家创新体系建设”，科技创新已经成为国家科技发展的基本战略。科技创新需要了解 and 把握现有领域的现状、趋势、结构、变化以及潜在的问题等，而这些任务正是科技情报工作的职责所在，从而也凸显了科技情报工作的必要性和重要性。

由于现代信息技术的飞速发展，科技情报工作面临机遇与挑战。一方面，信息技术的发展、信息环境的变化为情报工作提供了许多工具与资源，如数据挖掘与可视化技术为科学计量学的发展提高了数据处理效率和质量；另一方面，面对信息爆炸和情报不足并存的现象，如何有效利用海量信息，是迫切需要解决的问题。另外，科学知识体系逐步庞大、细化，各类学科之间相互交叉渗透、汇聚融合的频率增加，联系不断增强，这些都使得科技情报研究的对象和需求日益复杂，因此，科学情报工作必须重视学科或领域之间的知识演化关系的分析揭示。

在跨学科研究中，学科交叉点往往就是科学的生长点或新的科学前沿，这里最有可能产生重大的科学突破。因此，学科或领域间的交叉渗透可能会对知识演化过程和演化关系造成重要的影响，揭示这种演化关系有可能对研究交叉学科的形成和演变机制以及探索研究前沿的产生有一定的启示作用，同时对科技创新也有重要的意义。目前的知识演化分析方法（如同被

引分析方法、共词分析方法等)大多是将研究领域聚类划分,而现实的科学网络或知识网络中,科学家、主题词汇或科技文献都可能同时属于多个学科或研究领域(即本书所提的重叠结构)。因此,目前知识演化分析方法并不能真实地反映科学网络或知识网络的特性。虽然跨学科研究(或交叉学科)间接地体现重叠结构现象,但跨学科研究(或交叉学科)都是基于现有学科分类进行的,而现有学科分类本身并不能实时地反映科学发展的状态。知识结构中的重叠结构本身就是学科交叉的体现,探索从重叠结构的角度研究学科或领域之间的交叉渗透,结合知识演化和跨学科研究优势,是一个值得解决的问题。因此,揭示学科或领域之间的知识演化关系,发现可能的交叉重叠及其影响因素和演化规律等,对于挖掘科研动态、发现潜在变化以及规划科技创新具有重要的意义。

重叠现象并不是一种新现象,在模糊聚类分析和复杂网络中已有研究,并且跨学科研究中的知识转移也间接地体现知识演化中的重叠现象。但对于知识结构中重叠结构的形成过程及影响因素,重叠结构的时间演变特征及重叠结构与学科交叉领域的关系并未有深入的研究。

本书的研究目标在于提出一种基于重叠结构的知识演化分析方法,利用这种方法可以揭示科学研究中的重叠结构,并进一步深入地分析知识演化过程中研究领域之间的交叉、渗透以及演化过程。研究内容主要涉及重叠结构探测、重叠度计算、重叠结构动力学特性、重叠结构时间演变特征及基于重叠结构的跨学科研究,其中重叠结构动力学、重叠结构时间演变特征和基于重叠结构的跨学科研究是本书的研究重点,重叠结构探测和重叠度计算是研究基础。

本书内容共分8章,主要内容如下:

第1章主要综述了知识演化研究现状。在知识演化分析中,多数研究通常只关注知识演化分析的部分演化特征,特别是知识的新增、转移和扩散在知识演化的研究中相对较多。我们通常所说的研究前沿的探测与分析,实际上包含知识的新增和融合的演化特征,知识转移和知识传播则包含

着知识的转移、扩散及融合的相关特征。因此,本章主要从研究前沿探测、知识转移和知识传播 3 个方面分析知识演化研究相关进展。

第 2 章主要介绍复杂网络中重叠结构探测算法的相关进展和基于 BGLL 重叠探测算法的实现。重叠结构已成为社团结构探测研究中一个新兴的研究方向,其探测算法主要分为两类:以非重叠社团结构探测算法为基础的重叠社团结构探测算法和原生重叠社团结构探测算法。由于前一种方式可以充分利用现阶段最优秀的、最新的社团结构探测算法,因此本章对优秀的社团结构探测算法 BGLL 算法进行改进,实现了重叠结构探测。

第 3 章主要内容是社团重叠度计算。社团重叠度是重叠结构中的基础指标,其可以反映社团间的相似程度和关系远近,直接影响知识结构的表现。目前的重叠社团结构研究对重叠度计算仅考虑了节点数,不能反映出重叠节点与关联社团的链接强度及链接强度差异对重叠度的影响。本章提出了一种综合的、可以反映重叠节点数、重叠节点对关联社团贡献程度和重叠节点对关联社团贡献程度差异的重叠度计算方法。

第 4 章主要探索重叠结构动力学特性。重叠动力学指重叠结构的形成过程及影响因素。重叠结构的形成可能是由于一个社团借用(或引入)另一个社团的理论或方法,也有可能是综合多个社团的理论和方法构建一个新框架而形成的。当然,在重叠结构的形成过程中,可能涉及时间、主题、机构、地区等多种因素,本书从重叠方向、重叠速度和知识流动 3 个方面考虑。

①重叠方向:重叠方向指重叠区域由一个社团转移至另一个社团或多个社团主题流向重叠区域所表现出来的方向特性。本书提出同被引时间差来探测重叠方向,即根据重叠区域与不同关联社团发生同被引关系的时间分布来判断重叠方向。

②重叠速度:重叠结构的形成有一定的过程,不同重叠结构的形成可能有不同的能量。这种能量体现重叠结构形成速度的快慢,即重叠速度。本书定义重叠速度指标来表示重叠结构形成的快慢。

③重叠结构中的知识流动:重叠结构中的知识流动指重叠结构中各关联社团之间的知识融汇和交互过程。重叠主题有可能是关联社团中的某一个社团贡献

的,也有可能是多个社团共同贡献的。知识流动可以揭示出重叠结构是由哪些主题构成的,这些主题是从一个社团流向另一个社团还是多个社团流向重叠区域。本书定义了重叠主题关联差异度来判断重叠结构中的知识流动。

第5章主要探讨重叠结构时间演变特征。重叠结构的时间演化特征主要揭示不同时间窗中重叠结构的变化情况。重叠结构的时间演变特征可能有:①新重叠结构的出现:指重叠结构在前一个时间窗中找不到匹配的重叠结构。这种变化可能意味着一个新的交叉领域的产生和形成。②显著变化的重叠结构:指重叠度发生显著变化的重叠结构,这种变化可能意味着一个交叉领域的发展趋势。③重叠结构消失:指前一时间窗中的重叠结构在下一时间窗找不到匹配的重叠结构,重叠结构可能融入下一时间窗中的某个或某些社团,或者重叠结构独立发展成为一个社团。这种变化可能意味着一个交叉领域的成熟。由于重叠结构中重叠节点的变化与重叠结构的变化有密切关系,本书尝试通过重叠结构中重叠节点的演变来探测重叠结构的演变特征。

第6章主要分析重叠结构与跨学科研究的关系。知识结构中的重叠结构本身就是学科交叉的体现,因此,从重叠结构的角度研究学科或领域之间的交叉渗透具有天然的优势。本章通过构建 ESI(Essential Science Indicators)高被引论文的同被引网络,利用改进 BGLL 算法实现了重叠结构探测,并结合 Guimera 和 Amaral 定义的 Hub 节点指标,进一步分析了具有跨学科属性的重叠关系链接。研究表明,重叠结构可以直观地反映学科交叉的主题,并可以用于分析和探测潜在的跨学科领域。

第7章主要利用重叠结构分析农业学科群主题结构构建及领域交叉特性。本章选择 ESI 学科分类中与农业紧密相关的农业科学、动物与植物科学和环境与生态科学 3 个学科的 ESI 高被引论文构建农业学科群的引文网络,利用改进的 BGLL 算法获取农业学科群的主题结构。同时,根据第6章提出的学科交叉领域识别方法,在农业学科群主题结构的基础上,筛选重要

的跨学科重叠关系链接,并进一步比较分析跨学科重叠关系链接的形成及合理性,为发现新的交叉领域提供了可能。最后,本研究进一步比较分析农业学科群主题结构与中国农业大学学科布局,揭示中国农业大学在农业学科群中的优势领域和未涉及的重要领域。

第8章对本书的研究进行总结和展望。概括性地介绍取得的研究成果、创新之处、研究存在的问题、不足及后续工作,最后就基于重叠结构的知识演化分析方法在应用中可能存在的问题进行讨论。

本书由国家社会科学基金项目“跨学科研究影响力的计量分析研究”(批准号:15BTQ070)、教育部人文社科青年项目“基于重叠结构的学科交叉演化研究”(批准号:13YJC870001)、大北农青年学者研究计划“农业学科群中交叉领域探测及分析研究”(批准号:3801-2413001)资助。

最后特别指出,本书的内容有部分来源于我的博士学位论文,在此对给予我悉心指导的博士生导师张晓林教授致以最诚挚的谢意!同时,我还要感谢给予我指导和帮助的老师 and 同学们,感谢冷伏海教授、金碧辉教授、张智雄教授、李广建教授、孟连生教授等各位老师提供的建议,感谢中国科学院科技战略咨询研究院的王小梅研究员,中国科学院文献情报中心韩涛副研究员,是他们指引我进入情报研究这个原本陌生的学科领域,为我提供了良好的工作环境和丰富的科研实践机会;感谢中国农业大学图书馆馆长何秀荣教授、副馆长左文革教授以及信息服务部全体同事的支持和帮助;感谢我的妻子何欣和女儿陈若溪(Dorothy Chen),是她们给了我一个图书馆馆员从事情报学研究的支柱和动力。

陈仕吉

2018年1月于北京海淀



# 目录

第 1 章 知识演化研究现状 .....	1
1.1 研究前沿探测研究 .....	2
1.1.1 基于引文的研究前沿探测研究 .....	3
1.1.2 基于主题词汇的研究前沿探测研究 .....	12
1.1.3 ETD 与前沿探测 .....	15
1.2 学科间知识转移与知识演化研究 .....	19
1.2.1 基于引文的知识转移研究 .....	20
1.2.2 基于科学家迁移或合作的知识转移研究 .....	22
1.2.3 跨学科知识交流的“知识贸易” .....	26
1.3 基于网络模型的知识传播研究 .....	28
第 2 章 重叠结构理论及探测 .....	32
2.1 重叠结构相关概念 .....	32
2.1.1 社团结构 .....	32
2.1.2 重叠结构 .....	33
2.1.3 模块化 $Q$ 函数 .....	34
2.2 重叠结构探测研究 .....	36
2.2.1 以非重叠探测算法为基础的重叠结构算法 .....	37

2.2.2	原生重叠社团结构探测算法研究	41
2.2.3	基于重叠结构的社团演化研究	43
2.3	基于 BGLL 的重叠社团结构探测算法	45
2.3.1	BGLL 算法原理	46
2.3.2	基于 BGLL 算法的重叠结构探测	48
2.3.3	实验与分析	50
<b>第 3 章</b>	<b>社团重叠度计算</b>	<b>52</b>
3.1	模糊聚类中的重叠度	52
3.2	社团重叠度	56
3.2.1	重叠度计算方法	57
3.2.2	实验与分析	59
3.2.3	评测	62
3.3	小结	64
<b>第 4 章</b>	<b>重叠结构动力学研究</b>	<b>65</b>
4.1	重叠方向	66
4.1.1	同被引时间差	66
4.1.2	实验与分析	68
4.2	重叠速度	74
4.2.1	重叠速度的概念	74
4.2.2	实验与分析	75
4.3	重叠结构中知识流动	77
4.3.1	重叠主题识别	78
4.3.2	重叠主题的来源社团判断	78
4.3.3	实验与分析	81
4.4	重叠结构动力学的综合分析 & 评测	83
4.4.1	Ov#25 重叠结构动力学分析	83

4.4.2	Ov#17 重叠结构动力学分析	89
<b>第5章</b>	<b>重叠结构时间演变特征</b>	<b>96</b>
5.1	重叠结构演变	96
5.1.1	新重叠结构出现	97
5.1.2	重叠结构消失	98
5.1.3	显著变化的重叠结构	98
5.2	时间窗及重叠节点演变特征	99
5.2.1	时间窗	99
5.2.2	重叠节点的演变特征	100
5.3	重叠结构消失	102
5.3.1	重叠结构消失计算指标	102
5.3.2	实验结果	105
5.3.3	典型社团分析	108
5.3.4	评测	116
5.4	重叠结构突现	116
5.4.1	重叠结构突现计算指标	116
5.4.2	实验结果	118
5.4.3	典型重叠结构分析	120
5.4.4	评测	133
5.5	小结	137
<b>第6章</b>	<b>重叠结构与跨学科研究</b>	<b>138</b>
6.1	科学计量学中的跨学科研究	138
6.2	数据与研究方法	140
6.2.1	研究数据	140
6.2.2	重叠结构探测及可视化	141
6.2.3	跨学科链接识别	142