

图书馆服务场景下中文图书被引 特征的挖掘、分析与应用 ——以 G 类图书为例

彭贤哲 周海玲 石 进

(南京大学信息管理学院, 江苏 南京 210023)

摘 要: [目的/意义] 多角度分析图书被引行为, 可展现图书在学术应用场景的特点, 提高图书关联评价、采购入馆的准确性和科学性。[方法/过程] 本研究以南京大学图书馆 G 类中文图书为研究对象, 采集书目数据、点击量、借阅量及其在中国社会科学引文库中的引文数据, 从引用类型、引用规律、引用网络等方面, 展现了图书的自引、跨学科引用、被引半衰期、知识广度、知识深度等各项隐性指标的分布特征及相关性, 进一步剖析了图书采购、借阅行为的特点。[结果/结论] 结果表明, 当前读者借阅行为普遍缺乏方向感, 图书馆优先采购涉猎广泛的图书, 缺乏对图书知识深度及学界认可度的重视。为此, 有必要在图书推荐和采购过程中纳入引文评价、关联指标, 提升读者阅读体验, 优化图书馆馆藏质量。

关键词: 图书被引行为; 主题分布; 图书关联; 图书采购

DOI: 10.3969/j.issn.1008-0821.2023.10.010

[中图分类号] G250.7 [文献标识码] A [文章编号] 1008-0821 (2023) 10-0107-13

Mining, Analysis and Application of Citation Features from Chinese Books in Library Servicescape ——A Case Study of G-Class Books

Peng Xianzhe Zhou Hailing Shi Jin

(School of Information Management, Nanjing University, Nanjing 210023, China)

Abstract: [Purpose/Significance] The comprehensive analysis for the behavior of citing Chinese books, can show the characteristics of books in academic application scenarios, and improve the rationality of evaluating and purchasing books. [Method/Process] This study took G-class Chinese books in Nanjing University Library as the sample, then collected bibliographic data, clicks, borrowings and their citation data in the Chinese Science and Social Sciences Citation Library. From the aspects of citation type, citation law, citation network, etc., it showed the distribution characteristics and correlation of implicit indicators such as self-citation, interdisciplinary citation, cited half-life, knowledge breadth, and knowledge depth of books, and furtherly analyzed features of book purchasing and borrowing behavior. [Result/Conclusion] The results show that readers are usually aimless in browsing and borrowing books, and libraries give priority to purchasing a wide range of books, and lack attention to the depth of book knowledge and academic recognition. Therefore, it is necessary to include citation evaluation and related indicators in the process of book recommendation and procurement to improve readers' reading experience and optimize the quality of library collections.

Key words: the citing behavior of books; topic distribution; book association; book procurement

收稿日期: 2023-02-08

基金项目: 国家社会科学基金项目“面向国家安全的科技情报态势感知研究”(项目编号: 21BTQ012)。

作者简介: 彭贤哲(1995-), 男, 博士研究生, 研究方向: 数据挖掘与知识发现、信息组织与检索。周海玲(1987-), 女, 硕士研究生, 研究方向: 竞争情报, 科学计量。石进(1976-), 男, 教授, 博士, 博士生导师, 研究方向: 大数据分析与技术、情报学。

图书作为人类用以表达观点、保存信息、传播知识的载体,是学术研究和文化学习过程中不可或缺的资源,如何从众多良莠不齐的图书中快速识别并选取高价值的作品,是图书出版商、图书馆、学术界以及科研管理部门等共同面临的难题^[1]。当前的图书价值评价方法,主要从被引频次^[2]、馆藏数量^[3]、专家书评^[4]、引文内容^[5]、读者评论^[6]等方面展开。其中,学术文献对图书的引用行为,直观地展现了学术界对于图书的认识与流传程度,挖掘图书引证文献的隐藏特点,有助于了解图书的知识侧重方向、领域认可程度和学术应用价值。

本文拟通过由图书引文记录展现的引用类型、引用规律和引文网络,计量图书涵盖知识广度和深度、学界流传程度、应用时效性等,从该类视角下分析图书馆记录的借阅、浏览、采购行为的现状,并提出针对性的管理服务策略,优化服务模式,提供图书入藏参考建议。

1 相关研究

“选”“采”“用”是图书馆建设的3个闭环发展环节^[7],而“选”作为图书采购过程的依据,如何确定图书选择标准、检验入藏图书的优劣十分关键。图书质量评价为“选”的重要参考维度,其中脱胎于期刊评价的引证分析法,同样适用于学术图书评价。

通过引证分析法评价图书质量的实践工作,率先在外文图书中展开。2011年,Thomson Reuters^[8]正式发布类似SCI和SSCI的图书引文索引(Book Citation Index, BKCI),但收录择选图书仅2.5万种,数量相对较少。Gorraiz J等^[9]指出,BKCI尚不适用于文献计量分析和学术评价,齐东峰等^[10]不认可该观点,认为引文分析作为图书评价及参考工具具备可行性。刘晓娟等^[11]挑选了BKCI收录的出版于2005—2015年的学术著作,分析了图书出版量、被引频次的大小与分布以及学术著作的被引模式随出版年份和学科领域的变化情况。Zuccala A A等^[12]研究Scopus在2007—2011年收录的604种历史论文的参考文献中的图书,发现图书的被引量与Goodreads阅读评级为弱相关,支持引文分析可用于图书评价;为此,Zuccala A等^[4]将书评、

学术性引用指标作为学术图书的质量指标,采用机器学习的方法全面评估图书的影响力。Kousha K等^[13]比较了10个学科图书在Google Books、Google Scholar和Scopus3种数据来源的图书的引用数据差异,指出三者的引用频次高度相关。上海交通大学基于Scopus开发研制的《2019年外文学术图书引证报告》展示了近年来的高被引图书和高影响力出版社^[14]。张轶华等^[7]研究如何将高被引图书和高影响力出版社的引证分析与上海交通大学图书馆实际需求相结合,构建高品质外文学术图书保障体系。

对于中文图书引证行为的研究,起步较晚且相对较少。如苏新宁^[2]基于《中国社会科学引文索引》探讨人文社科各学科图书的被引情况。北京世纪读秀技术有限公司2017年发布的图书引用报告从高被引图书、高被引作者、出版社等方面分析了中文图书引用情况^[15]。肖宏等^[16]基于中国知网分析了哲学社科类图书的引证状况。叶继元^[17]阐述了我国自主研发《中文图书引文索引》的过程及意义。李明等^[18]探索性地分析了中文学术图书被引频次及其Altmetrics指标间的关系,为新型学术交流环境下中文学术图书影响力综合评价指标体系的合理构建提供了事实参考依据。章成志等^[1]从亚马逊中文网站上选取计算机、法律、医学、文学和体育5个学科领域的中文图书,从引文内容角度对图书被引行为进行分析,进一步提高了图书评价结果的准确性和科学性。为更好地利用图书引证文献的潜在信息,Zhou Q等^[19]收集了上述5个学科图书的引证文献记录,通过对引文文献细粒度分析,反映了被引图书的影响范围或主题,并将其与现有的图书影响力评估指标进行比较,印证了引文文献用于图书影响力评估的有效性。

总结已有研究,引文分析的研究视角从以往的引用频次分析,现已逐步扩展到对引文内容等深层次方面的研究;分析维度从之前专注于引证文献的数量分布、时间分布等直观指标,逐步转向深入挖掘图书被引行为的潜在特征,但深入挖掘引证文献的研究相对较少。外文图书引文数据的易获取性、详细性,使得早期研究的引文分析对象集中于外文

图书,对中文图书的类似探索于近些年方才展开。此外,已有研究多注重于如何完善引文分析方法深度、广度及准确度,少有如何将分析结果服务于现实场景的应用性探究,有待于将图书引文分析结果投入到中文图书管理服务应用场景。

本文以中文图书为研究对象,从引用类型、引用规律、引用网络等角度,全面揭示图书被引行为的隐性特点,如学界认可度、知识广度、知识深度、知识距离等,进而描述、评价、关联图书,以此角度分析当前浏览、借阅、采购过程中图书质量的分布状况,提出针对性建议,并将由引文分析构建的图书评价关联体系纳入图书“选”“采”“用”3个环节的参考依据,提升图书库建设质量。

2 数据来源与研究方法

2.1 数据来源

2.1.1 南京大学图书馆书目资源

图书馆数字化与信息化进程中产生的读者活动记录、书目记录、馆藏记录等服务记录^[20],为研究图书引用行为提供了充分的数据支撑。南京大学图书馆馆藏丰富,经长期积累,形成了比较系统、完整的综合性的藏书体系,截至2021年12月,该馆馆藏中文图书444.64万册。其中,G类(文化、科学、教育、体育)图书由于文理知识兼具、学科交叉性强、种类齐全、借阅频繁、馆藏丰富,产生了大量的图书服务采购记录,为研究图书管理服务现状提供了详实的数据参考。

本文采集了南京大学图书馆纸质馆藏资源中的G类图书总计46 976本,包括书目数据、全国馆藏数据、读者借阅数据、读者点击数据。其中,书目数据包括书名、主题词、图书摘要、作者、中图分类号、图书馆馆藏数量、出版年份、版次信息等;读者借阅数据主要为该类图书近10年被借阅的次数分布,直接反映了图书不同时间段在读者群中的热度变化;全国馆藏数据指该类图书入藏各大图书馆的分布情况,侧面表现了图书的传播辐射范围。

2.1.2 中文学术图书引文索引库

图书作为知识传播最重要的媒介之一,完整的引文分析体系有必要纳入图书引文的被引分析内容,图书引文索引数据库的建立是对已有引文数据库平

台数据的一种补充,对于完善已有引文分析体系具有重要意义^[21]。南京大学中国社会科学研究评价中心2015年发布了“中文学术图书引文索引”(Chinese Book Citation Index, CBKCI)入选图书数据,涉及11个学科,近600家出版社出版的图书,该数据库覆盖人文、社会科学的全部学科,可用于检索中文社会科学领域的图书收录和图书被引用情况^[22]。

本文根据已采集的南京大学图书馆G类纸质图书书目数据,依据书名、作者、出版年份,构建了图书在中文学术图书引文索引库检索平台的检索式,共获取5 200本图书的引文数据,总计32 689条引文记录。

2.2 研究方法

2.2.1 图书半衰期计算

某学科图书在 T 年的引用半衰期,是指该学科图书在 T 年(1年时间内)所引用的全部参考文献中较新一半是在最近(以 $T-1$ 年为最近第一年)多长时间内出版或发表的^[23]。学科图书引用半衰期反映了学科图书利用文献的新颖度,体现了图书近来被参考利用的频率,可用于测定学科图书的老化速度。根据其定义得出对应的计算公式如式(1)所示:

$$HCited = Y + \frac{50\% - C}{D - C} \quad (1)$$

其中, C 为累计被引频次最接近并小于50%的那年的累计被引百分比, D 为累计被引频次最接近并大于50%的那年的累计被引百分, Y 为累计被引百分比为 C 的那年到最近被引那年的间隔年数。

2.2.2 图书主题分布衡量

1) LDA 主题聚类

LDA模型指使用贝叶斯估计的统计学方法,将文本集中每篇文本的主题按照概率分布的形式表示的方法。作为一种无监督学习算法,LDA聚类模型的优势在于无需标注训练集,仅需指定最优主题聚类数量。在本文中,主要利用了文献的标题和关键词,通过限定主题聚类个数的变化范围,构建困惑度与主题数的折线图,在此基础上,利用“折肘法”确定主题数 N ,通过Gensim的LDA模型对文献进行主题表示,进而获得每篇文献隶属于各个

主题的概率, 构建文献—主题概率矩阵, 得到每本被引图书或每篇引证论文的主题概率表示。

2) 文献主题分布情况衡量

一篇完整的文献一般涵盖多个主题, 尤其是知识面较为广泛的图书载体, 如何衡量文献的知识面广度, 一定程度上决定了图书的应用面广度。文献的主题概率分布曲线直观体现了文献的主题侧重点, 单峰主题分布曲线表示内容点单一, 而多峰曲线则展现了文献内容的多方侧重特点。因此, 主题分布曲线的差异分布特点, 可侧面表现文献的知识广度^[24]。本文纳入主题概率分布方差衡量曲线的差异分布, 进而评估图书或论文的知识广度, 具体如下:

$$R^2 = \frac{1}{M} \times \sum_{i=1}^M \left(p_i - \frac{1}{2} \times (p_{\max} + p_{\min}) \right)^2 \quad (2)$$

式(2)中, M 为 LDA 聚类主题总数, p_i 为图书属于第 i 个主题的概率, p_{\max} 、 p_{\min} 分别为图书主题概率分布中的最大值和最小值, 该值越小, 图书涵盖知识越广。

此外, 图书应用面的广度亦可采用类似的计量方法, 由图书引证文献的主题概率分布状态所表示, 公式如下:

$$R^2 = \frac{1}{NM} \times \sum_{i=1}^M \left(p_i - \frac{1}{2} \times (p_{\max} + p_{\min}) \right)^2 \quad (3)$$

$$p_i = \frac{1}{N} \times \sum_{j=1}^N p_{ij} \quad p_{\max}, p_{\min} = \max_{1 \leq i \leq N} p_i, \min_{1 \leq i \leq N} p_i$$

式(3)中, M 为 LDA 聚类主题总数, N 为图书的引证文献篇数, p_{ij} 为图书的第 j 个引证文献属于第 i 个主题的概率, p_{\max} 、 p_{\min} 分别为图书主题概率分布中的最大值和最小值, 该值越小, 图书应用面越广。

单一主题的图书与多主题的图书相比, 在获得相同引用数基础上, 谈及的内容通常更为深入^[24]。为此, 图书的主题概率分布状况通过结合图书引文数, 可体现图书谈及知识的深度, 公式如下:

$$Depth = \frac{N}{M} \quad (4)$$

式(4)中, M 为 LDA 聚类主题总数, N 为图书的引证文献篇数, 该值越大, 图书谈及知识越

深入。

3) 文献主题相似度计算

在获取完图书的主题概率分布表达式之后, 可引入余弦相似度评估两本图书的内容相似度, 计算式为:

$$d_1 = \frac{\sum_{i=1}^M p_{1i} \times p_{2i}}{\sqrt{\sum_{i=1}^M p_{1i}^2} \times \sqrt{\sum_{i=1}^M p_{2i}^2}} \quad (5)$$

M 为 LDA 聚类主题总数, p_{1i} 、 p_{2i} 分别为第 1 本和第 2 本图书属于第 i 个主题的概率, 该值越高, 两本图书越相似。

此外, KL/JS 散度作为常用的衡量数据概率分布的数值指标^[25], 可用于衡量两本图书概率分布特征的相似度, 其具体计算公式如下:

$$D_{KL}(P \parallel Q) = \sum_{1 \leq x \leq M} P(x) \log \frac{P(x)}{Q(x)} \quad (6)$$

$$D_{JS}(P \parallel Q) = 1 - \frac{1}{2} D_{KL} \left(P \parallel \frac{P+Q}{2} \right) - \frac{1}{2} D_{KL} \left(Q \parallel \frac{P+Q}{2} \right) \quad (7)$$

M 为 LDA 聚类主题总数, P 、 Q 为两本图书的主题概率列表, $P(x)$ 、 $Q(x)$ 为两本图书属于第 x 个主题的概率, 该值越大, 两本图书内容越相似。

2.2.3 图书关联程度评估

将采集的图书书目结构化数据导入 Neo4j 图数据库之后, 图书与图书之间并非孤立单一的, 均有相应的关联路径, 图书产生关联的原因包括但不限于谈及相同内容、来自相同作者、涉及知识具备承接性或邻近性等。为此, 可引入图书关联路径相关指标, 如路径条数、路径距离、平均路径距离、关联度, 进而评估图书之间的关联程度, 具体含义如表 1 所示。

3 图书被引行为特征分析

引用行为是指在各种因素影响下, 学术文献作者在科研写作时参考或者借鉴其他学术文献并加以标注说明的学术行为^[26]。分析图书被引行为特征, 可展现学科领域对图书的利用、反馈、传播情况, 揭示图书之间的关联程度, 为图书的评价推荐提供参考依据。

表 1 图书之间的路径距离和主题相似度衡量指标

类 型	指 标	指标含义
图书关联指标	路径条数	两本图书相连路径的条数，条数越多，关联性越强
	路径距离	两本图书相连路径中途经图书节点数量，数量越小，关联性越强
	平均路径距离	两本图书所有相连路径距离的平均值，距离越短，关联性越强
	关联度	两本图书所有相连路径距离平方的倒数和，该值越大，关联性越强

图书1—图书4 路径参数
路径条数: 3
路径列表: 【3,4,4】
平均路径距离: 3.67
关联度: 0.23

Neo4j 数据库存储的书目数据

3.1 被引图书概况分析

从 G 类图书被引数据的学科分布特点来看，不同学科对图书的认可比例差异较大，且并不受该学科图书总数的影响。G2 类(信息与知识传播)馆藏图书数量最多，其被引用图书数量也相应最多，图书被引比例仅约 12%；而 G1 类(世界各国文化与文化事业)、G6 类(各级教育)图书馆藏数量较多，但被引比例较低；G0 类(文化理论)图书虽然图书馆藏数量较少，但被引比例最高。图书引用数量符合“二八定律”，南大图书馆馆藏 G 类图书总体引用比约为 11.1%，细究其下辖的二级学科分布情况，引用比在 6%~17% 范围内。由于出版图书在内容上存在一定比例的同质化，冗余的存在使得在学术研究过程中图书的借鉴参考比例相对较低，一般仅选择 20% 以内的出版图书作为该领域代表性的参考。

3.2 被引行为类型分布

图书引用行为分为自引和他引两类，其中自引根据其动机可分为合理自引和不当自引两类^[27]，二者并非泾渭分明，一般难以识别。自引的影响力相比他引是有限的，为此，科学评价是否应纳入自引行为，其权重该如何设定，均有待估量^[28]。研究自引行为在引文系统中的比例，衡量自引行为的普遍性特点，有助于展现图书自引现象在图书引文评

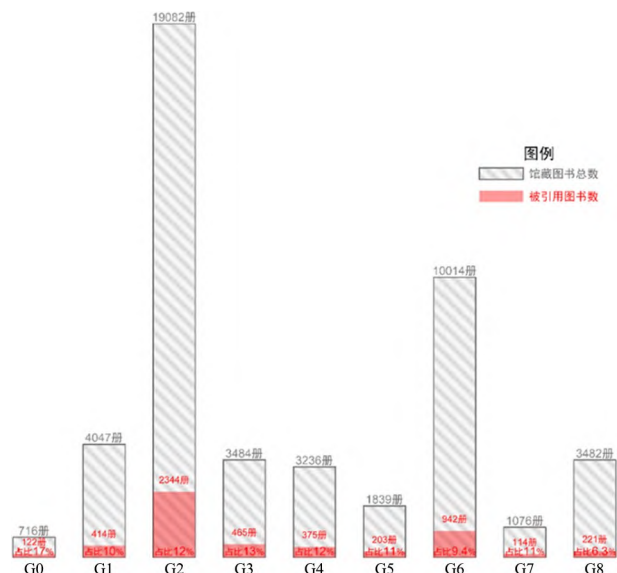


图 1 不同学科被引用图书数量占比

价中的地位。

他引行为和自引行为的比例均符合“二八定律”，应将自引行为纳入图书引文评价。分析从中文学术图书引文索引库引文记录发现，自引发生比例约为 13.2%，从不同二级学科自引图书分布比例来看，自引图书占比基本保持在 10%~14% 以内，差异较小。他引行为依旧是图书引文评价的主体依据，由此构成的图书评价体系仍具有一定代表性和客观性；自引行为在图书引用现象中占有一定比例，应将其作为图书引文评价体系的一环。

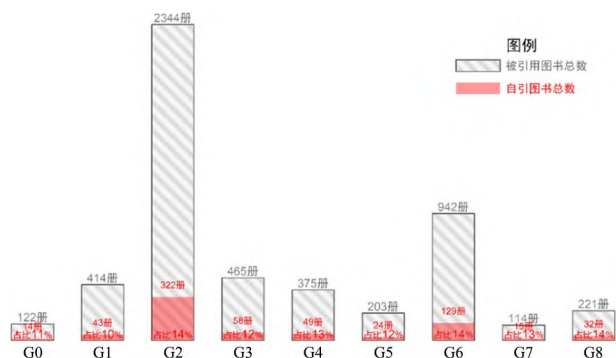


图2 不同学科图书自引率

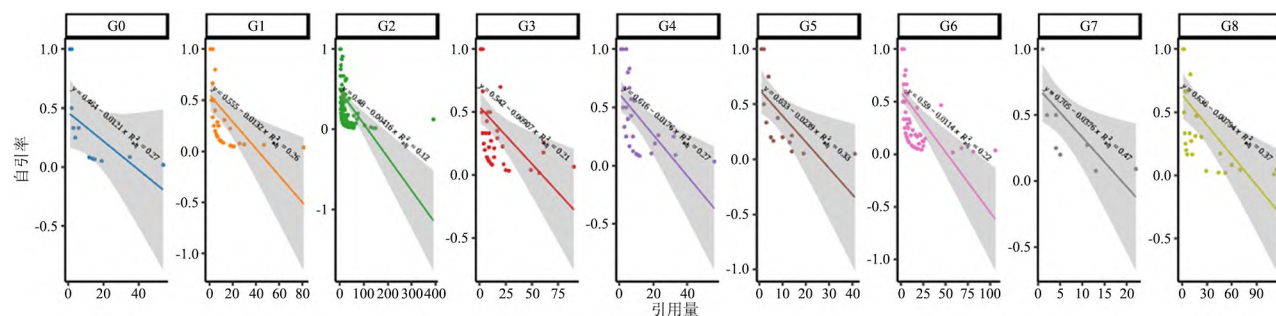


图3 不同二级学科图书自引率—引用量分布

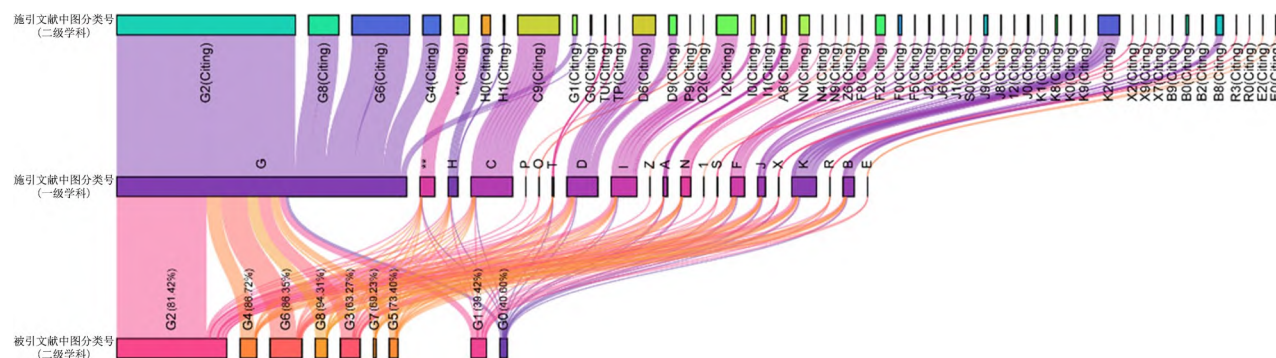


图4 被引图书—施引文献的学科对应关系

跨学科引用侧面反映 G 类学科与其他学科的扩散距离，分析引文记录发现，不同二级学科图书同学科引用现象占比变化范围为 39%~96%，G8 类(体育)学科 94.31% 的引证文献属于 G 类，发散性最低；G0 类(文化理论)学科 40.6% 的引证文献属于 G 类，发散性最高。整体分析 G 类图书引证文献的跨学科分布特征，C 类(社会科学总论)、D 类(政治、法律)、I 类(文学)和 K 类(历史、地理)文献对其的引用现象最为显著。

每本图书跨学科引用率定义为图书跨学科引证期刊论文数量与图书引用量的比值。结果表明，在不同二级学科所辖图书中，跨学科引用率较高的图书的引用量一般较低，以 G2、G4、G8 类图书为代

图书自引率为图书自引期刊论文数量与图书引用量的比值，计算筛选存在自引现象的 686 本图书的自引率后发现，所有二级学科图书的自引率与引用量存在一定的负相关关系，高自引率图书的引用量一般较少，说明用引文分析的方法评价低引用量的图书难免会受到自引现象的干扰，而高引用量的图书自引率较低，由此展开的引文分析虽然涵盖了少量自引现象，但仍是较为客观公正的。

表，当然个别二级学科图书中不乏同时具备较高的引用量和跨学科引用率的现象，如 G3、G7 类图书。由此说明，兼具高度学科发散性和高认可度的图书仅是个例，发生跨学科引用行为的图书的引用量较低，并未受到学界广泛的认可。

3.3 被引行为特征挖掘

图书被引行为特点，反映出图书在科学交流、知识继承与学科发展过程中的普遍规律，也可作为科学评价、科技管理和科技政策制定等行为提供科学依据。为此，本文从图书的引用量、引用半衰期、主题数、图书主题概率分布方差、引证文献主题概率分布方差、深度值 6 个角度，展现了图书被引行为的特点。

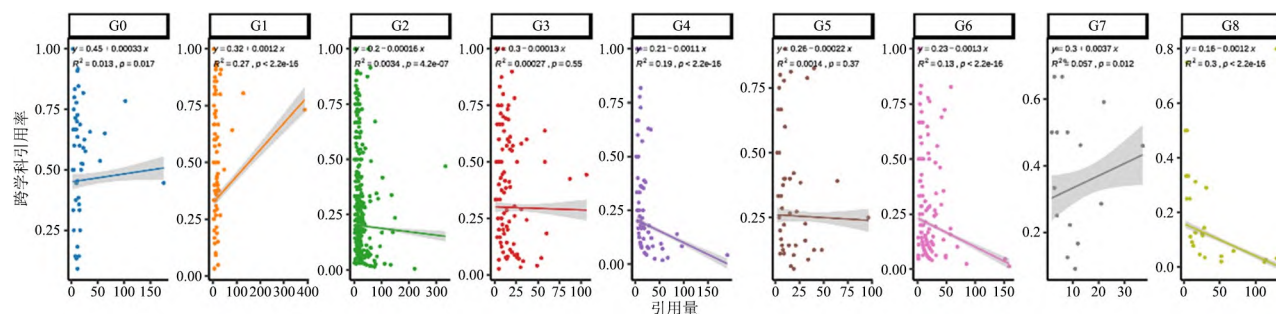


图 5 不同二级学科图书跨学科引用率—引用量分布

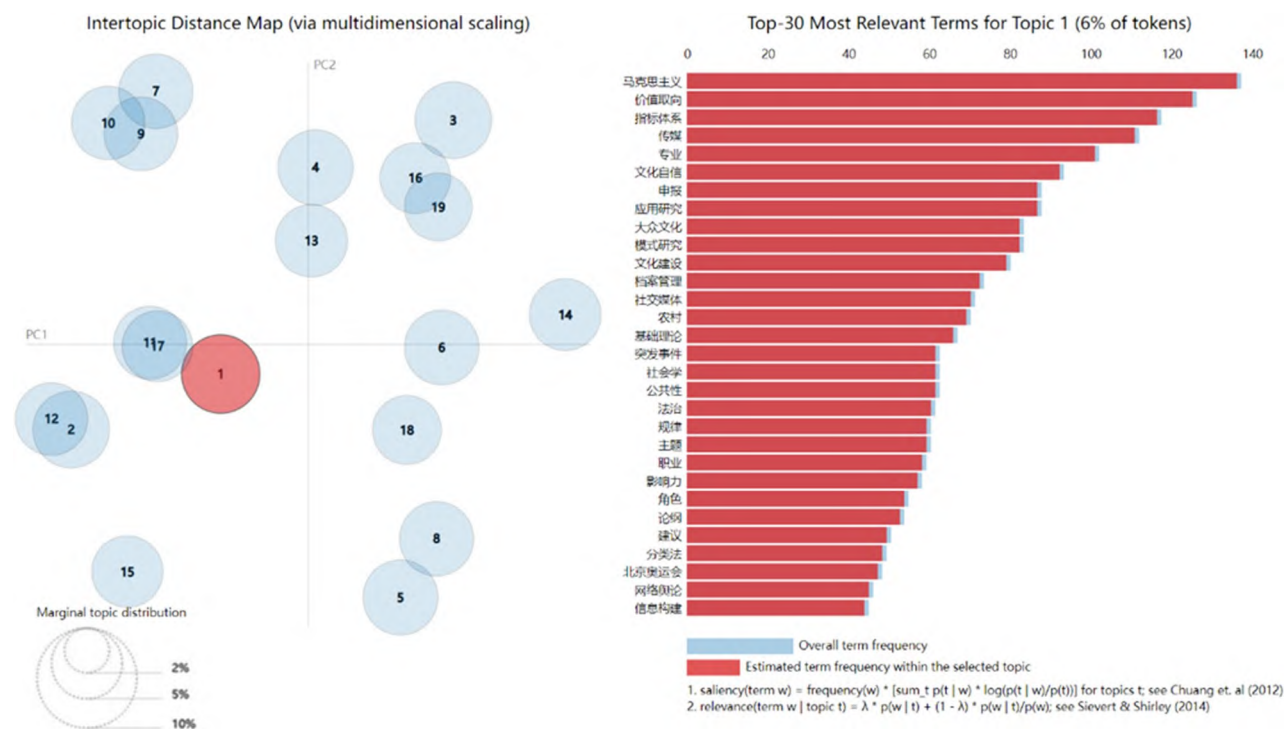


图 6 基于图书和引证文献的标题和关键词构建的 LDA 模型聚类结果

图书的被引特征,可从数量、时间、知识广度、知识深度 4 个层面加以限定。其中引用量即为其数量特征,引用半衰期为引用数在时间上的分布特征,图书主题概率分布方差代表图书内容涉及主题的广度,图书引证文献涉及主题数、主题概率分布方差指示图书应用方向的广度,图书深度借助引用量反映了图书在特定主题的深度。

利用 LDA 主题聚类分析 5 200 本图书和 27 834 篇引证论文的标题和关键词,获得共计 19 个主题,如图 6 所示。此外,由于图书半衰期的计算方法,无法衡量施引文献集中在同一年(短暂热度)和最近施引文献占比超过 50%(未表现衰退趋势)的图书,故对该两类图书做了初步的统计,如图 7 所示,未表现被引衰退趋势的图书很少,短暂热度图

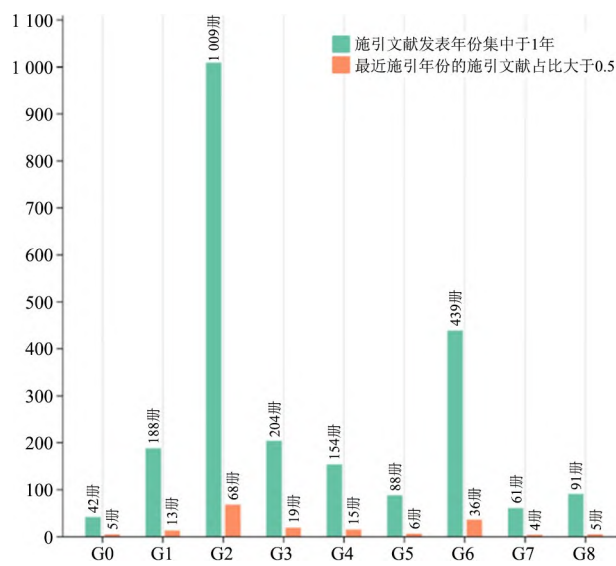


图 7 不同学科无法计算半衰期的图书数量

书数量不菲,说明图书“昙花一现”的现象较为

普遍。

从 G 类图书被引行为 6 类特征分布特点来看,如图 8 所示,图书的被引量主要分布于 50 篇以内。不同学科图书的半衰期分布特点较为一致,均保持在 2~6 年的范围内,最长约在 10 年左右。图书涉及主题的范围较广,单一主题图书较少,侧面印证

了图书这种信息载体涵盖知识的系统性和全面性。图书引证文献主题概率相比图书主题概率分布的方差增大明显,说明图书的引证文献较多地关注图书的某一方面主题,使得主题概率分布出现峰值突出的现象,增大了方差值。图书每个主题的综合认可度并不高,单个主题的平均被引量一般小于 5 篇。

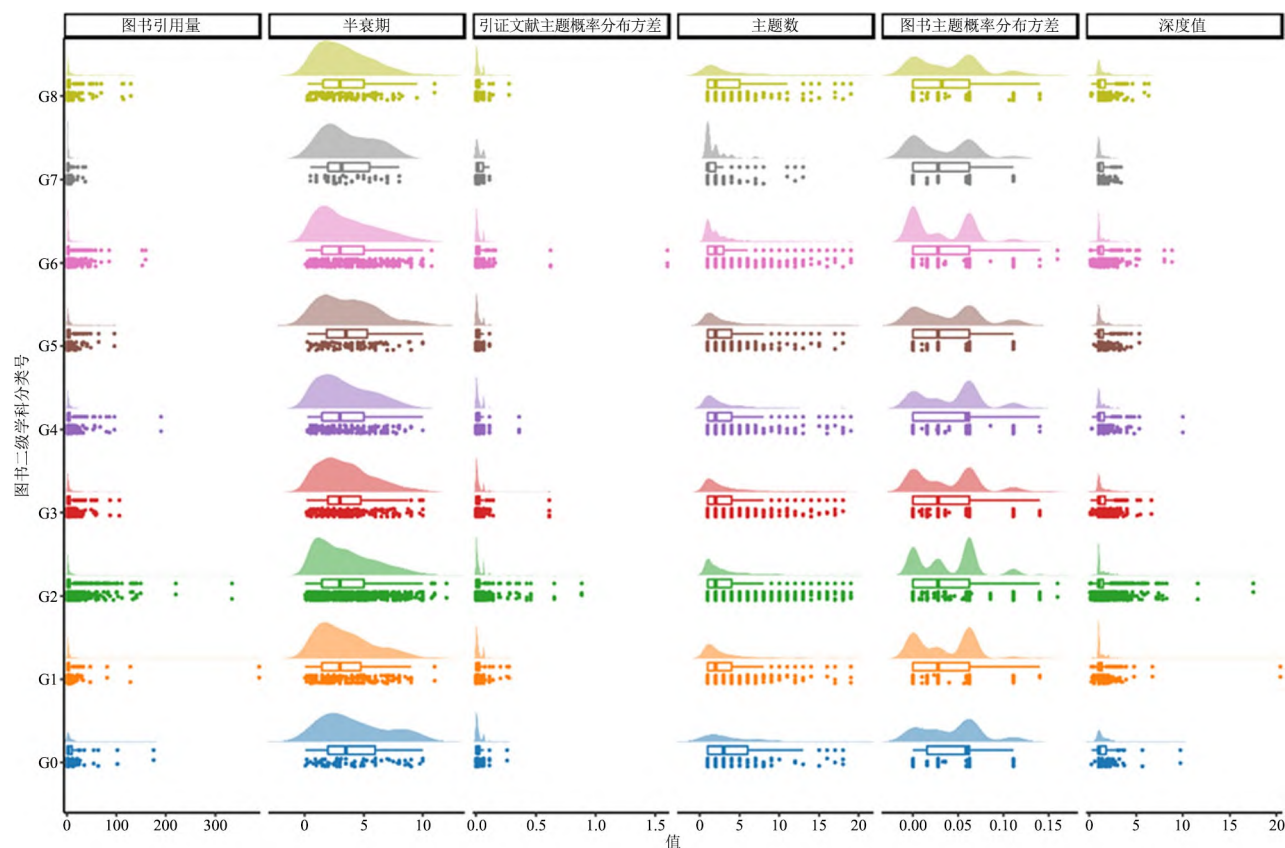


图 8 不同二级学科图书被引行为特征分布

进一步分析引用行为特征的关联性,可全面了解图书被引行为的“全貌”。结果表明,图书被引半衰期与引用量关系较弱,图书引用量的提高同时印证了图书应用主题的高度纵深;而图书主题概率分布方差作为图书固有的知识属性,与引用量关系并不相关;随着引用量的提升,该书的主题发散程度轻微加强,致使引证文献主题概率分布方差发生轻微程度地降低,涵盖主题数相应增多。

3.4 引文耦合网络解析

图书之间的耦合关系,在一定程度上反映了两本图书的相似性或者关联度大小^[29],引入图书路径距离和图书主题相似度等指标,可展现耦合次数对图书相似性或关联性的反映状况。

结果如图 10 所示,两本图书的耦合次数一般

限定在 10 次以内,二者之间的主题相似度和距离关联度并不确定。耦合次数在 10 次以上的两本图书之间,主题余弦相似度和 JS 相似度为 1,二者反映的主题内容高度匹配,二者相连路径条数不定;平均路径距离约为 3,一般仅间隔 1 个图书节点,关联度变化范围为 0.5~1.5,在图数据库中表现为图书之间至少存在两条距离小于 2 的路径。这个结果表明,存在耦合关系的两本图书之间关联较为紧密,有助于实现图书的精准关联推荐,但对于发散性的图书扩展推荐服务略显不足。

图书引文耦合网络的特征表明,图书之间内容越相似,期刊论文在引用多本图书时,存在综合引用相似内容图书的习惯。总结而论,图书耦合关系展现的是二者高度的主题相似度、关联度,主题分

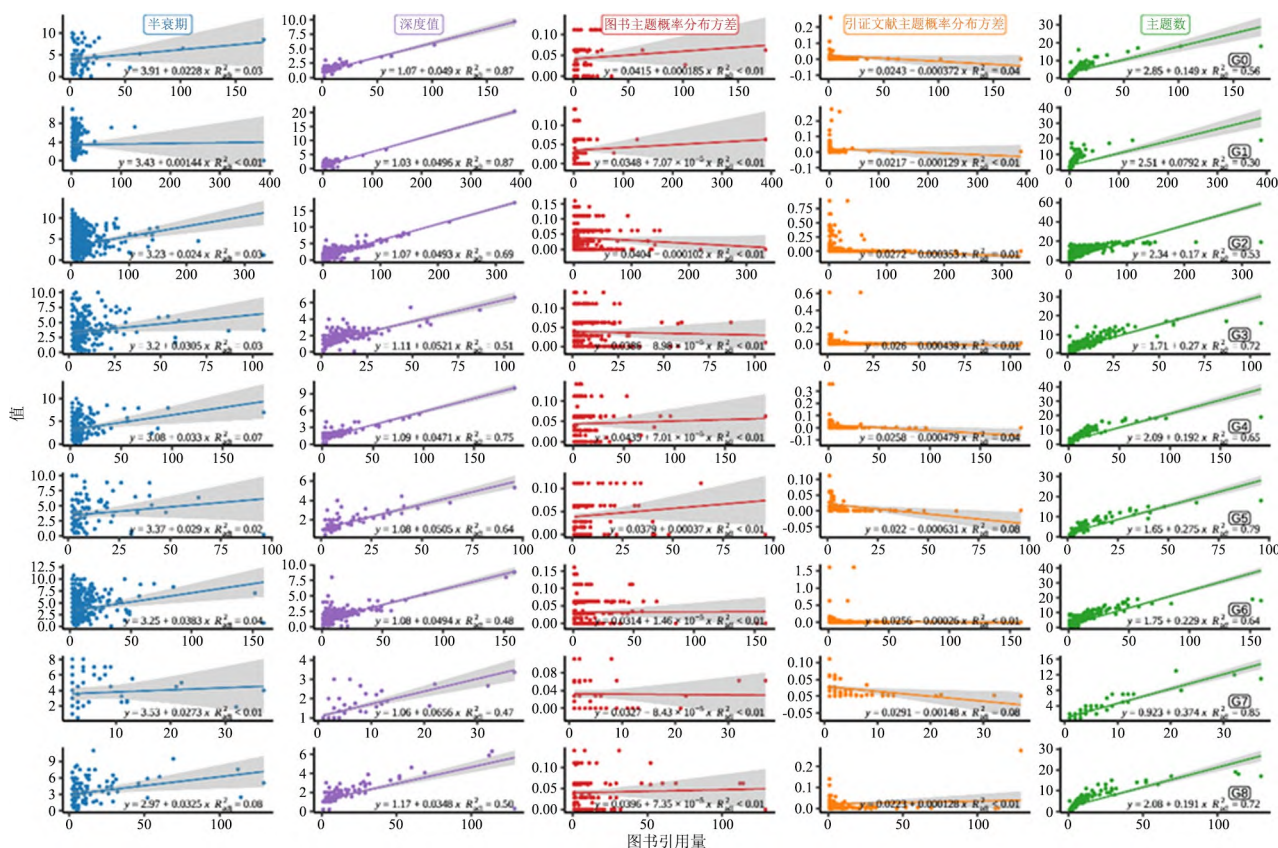


图9 图书被引行为特征之间的关系

布存在重合、在知识库中距离相近是两本图书存在耦合的必要条件。这说明核心文献对多本图书的引用行为注重相同内容的对比参照，而非相关知识的整合补全，更多强调挖掘图书知识深度，可利用图书的耦合关系为读者提供深入式阅读的推荐服务。

4 图书被引行为应用探讨

科学文献之间通过参考与引证行为相互联系，其中，参考现象展现了一篇科学文献研究工作的背景和依据，而引证现象则反映了一篇科学文献研究工作的继续、应用、发展或用于评价该篇科学文献^[30]。期刊论文对图书的参考行为与引证效应，从一定程度上直接表现了图书的学术质量、应用价值、知识深度、应用范围等，因而可为读者阅览选择图书、管理者采购引进图书、系统关联推荐图书提供丰富的参考信息。

4.1 图书阅览

图书借阅是图书馆最传统、最基础的服务^[31]，在该过程中会产生点击量、借阅量等相关数据，展现了广大读者的阅读倾向和兴趣所在。为此，本文

探析了南京大学图书馆服务系统记录的点击、借阅数据，通过将其与图书被引特征关联分析，探析读者在选择图书时是否了解该图书的应用广度、知识深度、学术认可度。

从图书点击、借阅数据与被引特征的关联性分析来看，读者对图书的点击量、借阅量与图书被引半衰期、深度值、引用量、图书主题概率分布方差、引证文献概率分布方差、主题数相关性差。这说明大部分读者在服务端口略览、借阅图书资源时，具备很强的随意性，缺乏对该方向图书分布的认识，在择书时缺乏参考依据，一定程度上提高了读者找寻目标图书付出的各项成本，易出现“找错书、读错书”的问题。

鉴于目前读者择书的困境，有必要采集读者个性化数据提供精准服务，依据读者的学术背景信息、过往阅览历史等，结合图书应用主题的方向，为初学者提供主题分布广、知识深度浅、体系相对成熟的图书，为科研人员推荐主题分布集中、前沿性强的图书。衡量图书被引特征的这些指标，通过描述图书蕴含知识的广度、深度、认可度，为图书的推

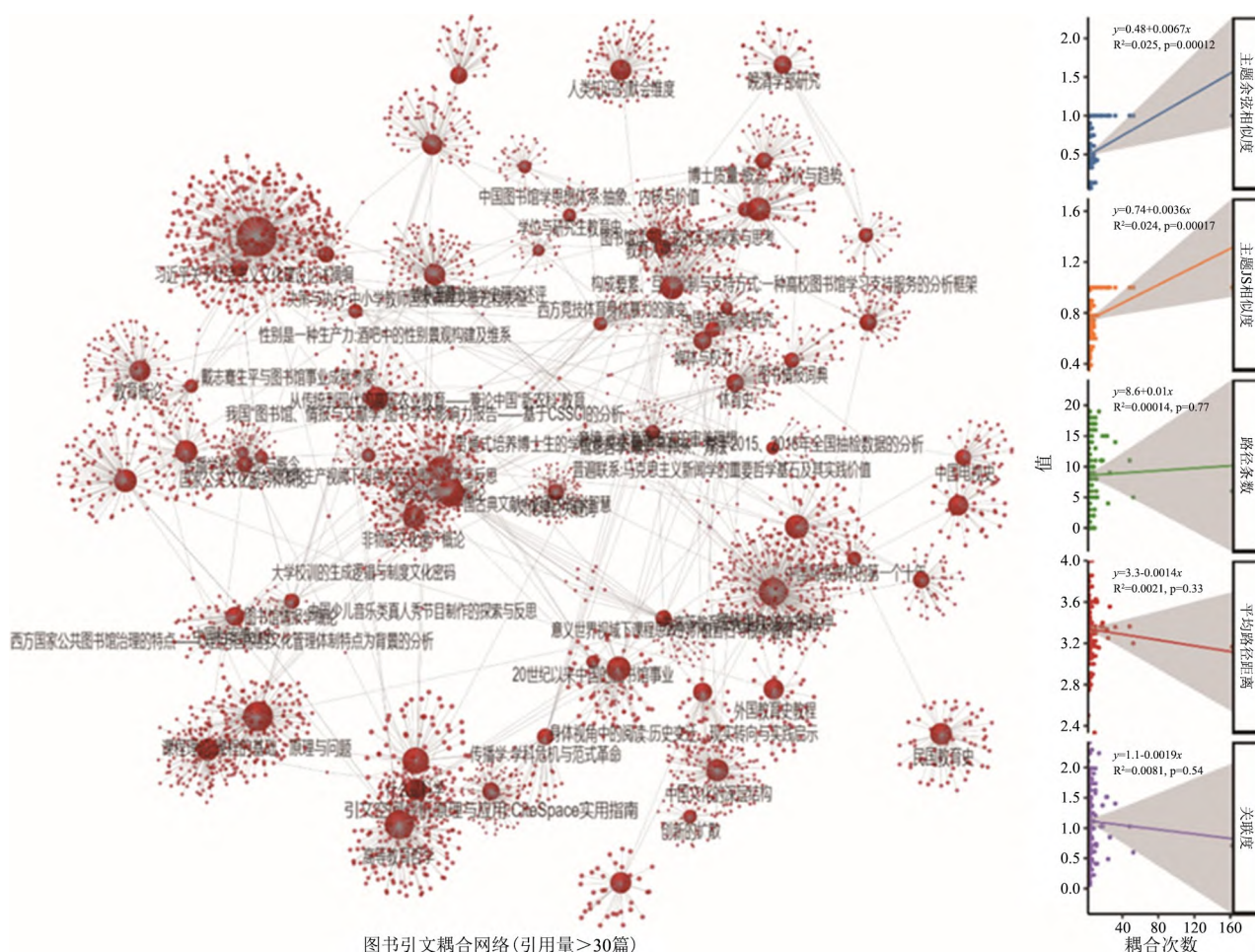


图 10 图书引文耦合网络及参数特征

荐过程提供了一系列参考依据，有助于提高图书馆系统服务的精准性。

4.2 图书采购

图书在全国各大图书馆的馆藏分布，即现有多少家图书馆入藏该本图书，一定程度上说明了图书被收藏入馆的意愿，代表了图书馆管理人员采购图书的倾向。了解图书馆倾向采购的图书被引特征，可展现目前广泛入藏图书的整体质量，有助于后续进一步采购图书入馆、优化馆藏资源，提高图书馆服务质量。

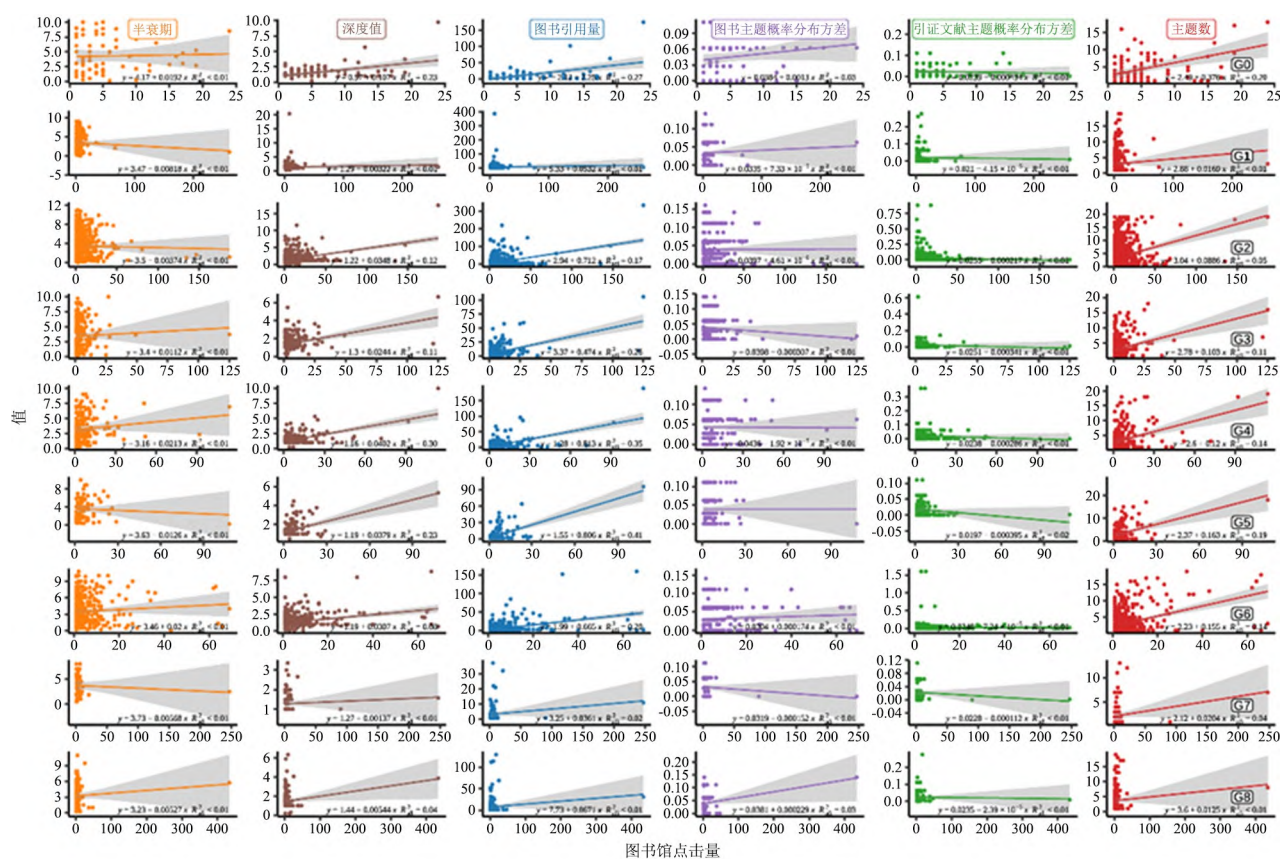
分析图书全国馆藏数与被引特征的关联性，被全国图书馆广泛入藏(≥ 500 家)的图书被引半衰期、深度值、引用量、主题数分布未呈现规律性，但图书主题概率分布方差、引证文献主题概率分布方差普遍较低。这个结果表明，各大图书馆在采购图书入馆时，优先考虑图书知识广度和应用范围，未过多权衡图书的知识深度和在科研领域的认可度，便

于不同知识背景的读者均可在图书馆获取到与其相匹配的图书，快速拓展知识视野。

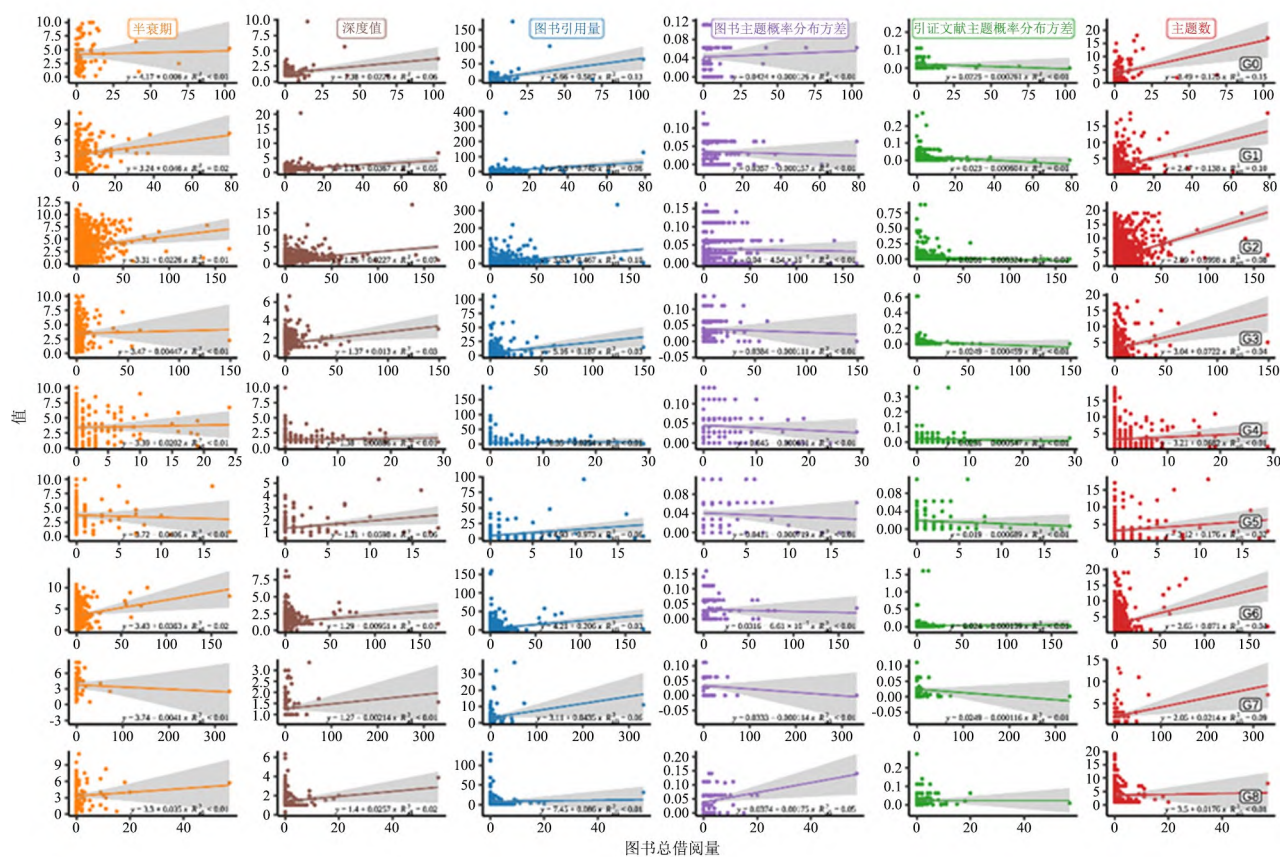
图书馆采购知识覆盖面广泛的图书，虽能促进阅读推广的实施，但对于具备科研应用场景的图书馆，如高校图书馆、科研机构图书馆等，在采购时仍需入藏一定比例深入前沿的图书。因此，不同性质图书馆由于侧重点不同，在采购图书时可动态性考量图书被引特征，保证图书馆馆藏图书的分布特点适于其应用场景。

4.3 图书关联

图书的关联在知识的延展性上可分为横向关联和纵向关联，前者注重从图书谈及内容的邻近知识予以关联，帮助读者扩展知识视野，后者强调不同深度但关联相似知识的图书，辅助读者深入掌握知识。依据引证文献构建的图书耦合网络，结合图书的主题相似度参数、距离参数、深度值、广度值，可实现图书在横向和纵向上的关联。横向关联注重



(a) 图书点击量



(b) 图书借阅量

图 11 图书点击、借阅数据与被引特征的关联性

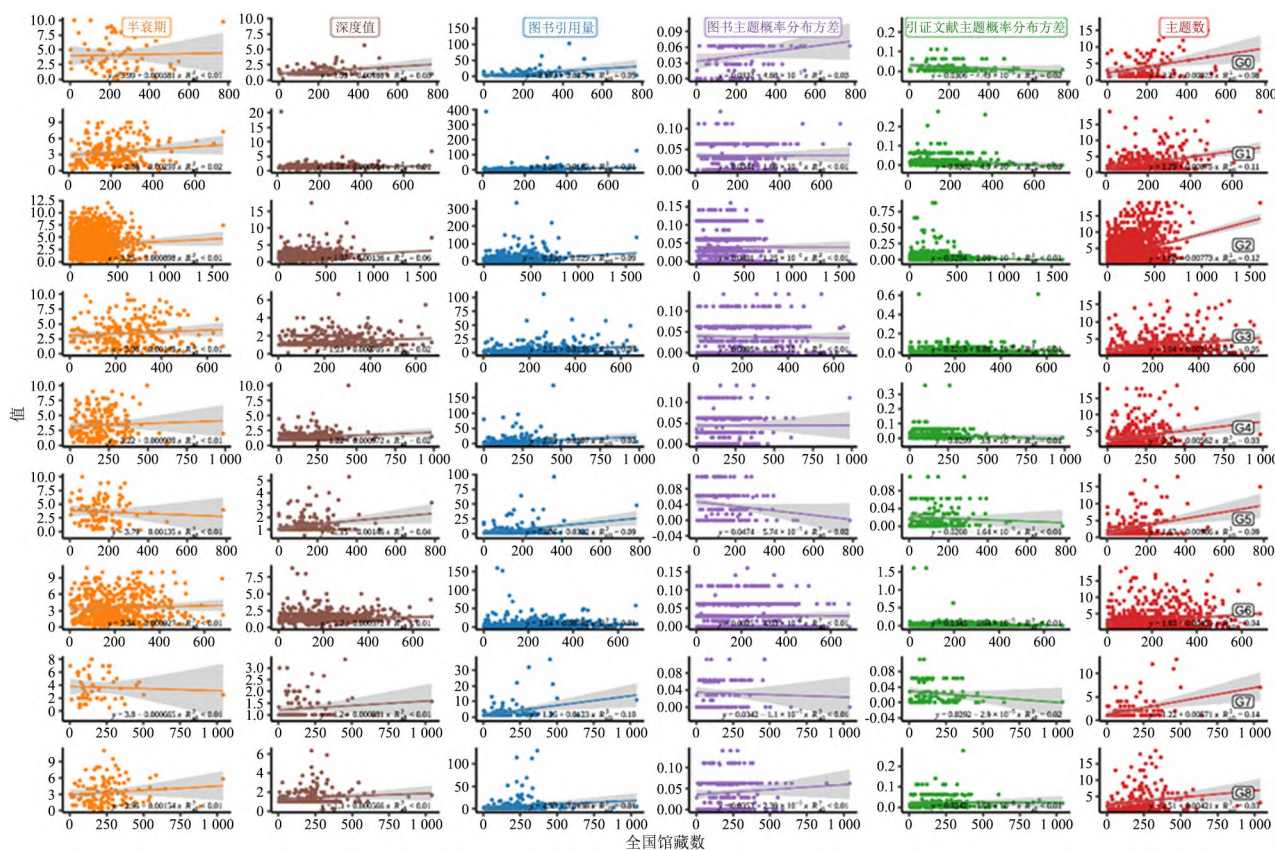


图 12 图书全国馆藏数与被引特征的关联性

放宽图书之间的距离参数、主题相似度，优选主题泛化图书，而纵向关联则需限定图书之间的紧密程度、主题相似度，过滤出专业化图书。

此外，图书内容在横向和纵向上的关联，亦可映射在图书应用方向上，根据图书引证文献主题数、主题概率分布参数，可展现图书应用主题分布特征。据此，在横向上推荐与此相关的多主题图书，保证知识广度，在纵向上关联核心主题的高引用量图书，确保知识深度。基于图书被引特征参数，可为图书多维度的关联提供数据支撑，增强关联路径的发散性和纵深性，构建一个系统、全面、周密的图书关联网络。

5 结 语

不同于以往专注图书的引文评价研究，本文的落脚点更加关注引文评价结果的应用探究，将引文对图书内容、应用、价值的呈现，反映到图书管理服务过程。为此，本文首先综合了前人在数量和内容上的引文分析方法，利用图书丰富的引证文献，从引文挖掘视角分析了当下图书服务管理模式的特点。在图书的引文数量分布、时间分布、主题分布、

关联程度等方面，通过揭示 G 类图书的被引行为特征，展现了该类图书的内容广度、知识深度、应用范围和关联性等隐藏信息。

之后，以此视角切入当下图书浏览点击、借阅采购等具体应用场景，发现图书资源的借阅浏览行为具备很强的随意性，读者对该领域图书内容的广度和深度分布认识不足，图书服务模式应依据读者学术背景提高精准个性化指导；此外，管理员倾向于采购入藏知识面广泛的图书，在未来应根据图书馆性质适当引入深入前沿的图书，并结合图书的主题分布重合度和知识深度，提供个性化的推荐服务。因此，从引文挖掘视角看待当下图书服务管理模式，有利于为图书馆精准推荐、采购入藏图书提供学术依据，降低读者择书成本，优化馆藏资源质量。

然而，本文仅以南京大学图书馆馆藏资源中的 G 类图书为研究样本，针对不同地区、不同类型图书馆、不同学科图书引证文献的特征以及应用价值，仍有待于后续进一步扩展研究范围予以验证探讨，斟酌将图书引证数据全面纳入图书评价关联体系的可行性。

参 考 文 献

- [1] 章成志, 李卓, 赵梦圆, 等. 基于引文内容的中文图书被引行为研究 [J]. 中国图书馆学报, 2019, 45 (3): 96-109.
- [2] 苏新宁. 我国人文社会科学图书被引概况分析——基于 CSSCI 数据库 [J]. 东岳论丛, 2009, 30 (7): 5-13.
- [3] White H D, Boell S K, Yu H, et al. Libcitations: A Measure for Comparative Assessment of Book Publications in the Humanities and Social Sciences [J]. Journal of the American Society for Information Science and Technology, 2009, 60 (6): 1083-1096.
- [4] Zuccala A, Someren M, Bellen M. A Machine-learning Approach to Coding Book Reviews as Quality Indicators: Toward a Theory of Megacitation [J]. Journal of the Association for Information Science and Technology, 2014, 65 (11): 2248-2260.
- [5] 胡志刚. 全文引文分析方法与应用 [D]. 大连: 大连理工大学, 2014.
- [6] 任静. 基于网络数据的学术图书评价研究 [D]. 南京: 南京大学, 2013.
- [7] 张轶华, 黄楠, 郭晶, 等. 基于引证分析的高校图书馆外文图书建设实证研究——以上海交通大学图书馆为例 [J]. 图书馆学研究, 2021, (21): 32-42, 17.
- [8] Adams J, Testa J. Thomson Reuters Book Citation Index [C] // The 13th Conference of the International Society for Scientometrics and Informetrics, 2011.
- [9] Gorraiz J, Purnell P J, Glänzel W. Opportunities for and Limitations of the Book Citation Index [J]. Journal of the American Society for Information Science and Technology, 2013, 64 (7): 1388-1398.
- [10] 齐东峰, 陈文珏. 图书引文索引(BKCI)——新的图书评价及参考工具 [J]. 图书馆杂志, 2013, 32 (4): 45-48.
- [11] 刘晓娟, 马梁. 基于BKCI的学术著作引文分布研究 [J]. 图书情报工作, 2017, 61 (24): 105-113.
- [12] Zuccala A A, Verleysen F T, Cornacchia R, et al. Altmetrics for the Humanities: Comparing Goodreads Reader Ratings with Citations to History Books [J]. Aslib Journal of Information Management, 2015, 67 (3): 320-336.
- [13] Kousha K, Thelwall M, Rezaie S. Assessing the Citation Impact of Books: The Role of Google Books, Google Scholar, and Scopus [J]. Journal of the American Society for Information Science and Technology, 2011, 62 (11): 2147-2164.
- [14] 上海交通大学出版传媒研究院. 2019年外文学学术图书引证报告 [M]. 2020.
- [15] 北京世纪读秀技术有限公司. 图书被引用情况报告 [R]. 北京: 北京世纪读秀技术有限公司, 2017.
- [16] 肖宏, 伍军红, 郭文涛, 等. 我国哲学社会科学领域图书文献引证大数据研究 [J]. 中国社会科学评价, 2018, (3): 88-115, 128.
- [17] 叶继元. 《中文图书引文索引·人文社会科学》示范数据库研制过程、意义及其启示 [J]. 大学图书馆学报, 2013, 31 (1): 48-53.
- [18] 李明, 李江, 陈铭, 等. 中文学术图书引文量与 Altmetrics 指标探索性分析及其启示 [J]. 情报学报, 2019, 38 (6): 557-567.
- [19] Zhou Q, Zhang C. Evaluating Wider Impacts of Books Via Fine-grained Mining on Citation Literatures [J]. Scientometrics, 2020, 125 (3): 1923-1948.
- [20] 程静, 于亚秀, 汪志莉. 高校图书馆数据可视化表达研究——以华东师范大学图书馆为例 [J]. 图书馆学刊, 2022, 44 (6): 47-52.
- [21] 孙奇, 张晓梅. 国内外图书引文索引库发展探析 [J]. 图书馆杂志, 2017, 36 (6): 84-87, 38.
- [22] 任全娥. 用于出版学术影响力分析的三类文献计量指标探讨 [J]. 大学图书馆学报, 2016, 34 (5): 110-119.
- [23] 宋雯雯, 刘晓娟. 基于BKCI的图书半衰期分析——以图书情报学学科为例 [J]. 图书情报工作, 2016, 60 (12): 124-129.
- [24] 周清清. 基于多源异构数据的图书影响力评价研究 [D]. 南京: 南京理工大学, 2018.
- [25] Zareie A, Sheikahmadi A, Jalili M. Identification of Influential Users in Social Networks Based on Users' Interest [J]. Information Sciences, 2019, 493: 217-231.
- [26] 肖香龙. 基于最省力法则的引用行为研究 [D]. 武汉: 武汉大学, 2018.
- [27] Hyland K, Jiang F K. Changing Patterns of Self-citation: Cumulative Inquiry or Self-promotion? [J]. Text & Talk, 2018, 38 (3): 365-387.
- [28] Pichappan P. A Dual Refinement of Journal Self-citation Measures [J]. Scientometrics, 1995, 33 (1): 13-21.
- [29] 李成赞, 黎建辉, 王学志, 等. 基于引文网络社区发现的数据推荐研究 [J]. 情报学报, 2021, 40 (8): 879-886.
- [30] 魏绪秋, 李长玲, 郭凤娇, 等. 基于引证数据的单篇论文学术生命力研究 [J]. 情报杂志, 2020, 39 (1): 154-161.
- [31] 陈雯. 我国高校图书馆社会服务调查研究——基于南京地区25所高校图书馆社会服务调查 [J]. 图书情报工作, 2017, 61 (4): 76-82.

(责任编辑: 陈 媛)