



外文期刊数据库学科相关评估指标的量化研究

——以同济大学理学部学科群为例

□马爽* 魏雯婕 李丽娟

摘要 高校“双一流”建设的核心任务是学科建设,图书馆的文献资源建设要与学科的发展需求紧密结合,围绕学科建设提供精准的文献资源和服务。文章基于 PageRank 算法,运用引文分析法和社会网络分析法,提出了三个与学科相关联的外文期刊数据库评估指标,即数据库的学科需求度、数据库的学科学术效益、数据库的学科匹配度,构建了数据库学科评估模型,并以同济大学理学部学科群为例对外文期刊数据库进行实证分析,构建了不同等级的理学部的文献资源集群。

关键词 数据库评估 引文分析 社会网络分析 学科资源 指标量化

分类号 G250

DOI 10.16603/j.issn1002-1027.2023.03.008

以学科建设为基础,是国家“双一流”建设的重要原则,新一轮的“双一流”建设意见对学科建设提出了更高的要求。高校图书馆的文献资源建设重点需要进一步紧密结合学科需求,学科文献资源建设的考核要更加注重使用效益^[1]。近年来,高校师生对数字资源的需求和关注度持续增加,外文期刊数据库作为学术研究的重要资源在高校图书馆数字资源建设中的比重不断上升。但由于外文期刊数据库价格逐年上涨,如何利用有限的经费保障学科外文期刊数据库的建设已成为众多高校图书馆急需解决的问题。建立科学的评估指标体系对学科资源的精确评估具有关键性作用,现有的评估指标体系虽然具有准确性、可获取性和可比性等优点,但存在与学科之间的关联度较弱的问题。本文在结合现有研究的基础上,提出了三个与学科相关的外文期刊数据库评估指标,即数据库的学科需求度、数据库的学科学术效益、数据库的学科匹配度,构建了数据库学科评估模型,并以同济大学理学部学科群为例对外文期刊数据库进行了实证分析,构建了不同等级的理学部的文献资源集群。

1 研究现状

基于数字资源评估标准、项目及实践经验,国内

外学者建立了包含资源内容、质量、成本和服务等多角度的数字资源评估指标体系^[2-6]。在学科数字资源评估方面,研究表明在期刊使用方面不同学科领域存在显著差异,应该根据学科领域来评估期刊资源^[7]。将引文分析法应用到数字资源的评估中,不但可以客观地反映资源被利用的情况,还可以有效地揭示师生对资源的实际需求和潜在需求^[8-9]。在学科数字资源评估中引文分析法备受青睐^[10-11]。莱特曼(Lightman H)等人利用引文分析法评估了美国西北大学经济学类馆藏期刊,结果显示被引期刊与教师研究兴趣相关联,对学术成果产生较大的影响^[12]。戴维斯(Davis P)通过分析美国康奈尔大学学者引用期刊的频次确定了核心馆藏,结果显示引用排名最高的期刊并非《期刊引用报告》(*Journal Citation Reports, JCR*)中排名最高的期刊,引用排名更加直观地反映了院系的研究重点^[13]。沃尔特斯(Walters W H)等人对期刊的保障情况、引用情况、来源等方面进行分析,确保了资源建设与师生需求的一致性,最大限度地提高了期刊订购的成本效益^[14]。国内学者张建勇等人建立了学科调节因子指标,其定义为“某学科购进期刊数量占用户引用期刊种数的比例”,该比例的高低反映了学科期刊满足度^[15]。花芳等人利用引文分析法构建了清华大学

* 通讯作者:马爽,邮箱:18101@tongji.edu.cn。



工程力学系的本地子集(学术群体频繁使用的期刊集合)和全球子集(学术群体所涉及的学科中具有影响力的期刊集合),表征了期刊资源的适用性和品质^[16]。张宇娥等人对电子科技大学生物与生物化学学科发文、引文、施引期刊和学科权威期刊合集进行分析,建立了期刊优先级模型,实现了单一学科期刊重要等级的排序^[17]。

上述研究多数是通过引用频次、被引期刊的数量或比重建立外文期刊的核心馆藏或等级排序。但在图书馆的实际资源采购工作中,大部分外文期刊资源是以数据库为单位进行购买的,同一种期刊可能被多个数据库收录,单一角度的期刊评估很难直接指导数据库的采购工作,需要针对数据库建立相关的学科评估指标。雅克索(Jacsó P)等人利用收录期刊数量和期刊包含的文章数量定义数据库覆盖范围的广度和深度,由此来反映数据库对特定主题或学科的覆盖情况^[18]。库格林(Coughlin D M)等人从内容、使用和成本等方面对数据库进行评估,在内容评估方面利用学科资源数量来评估数据库覆盖的研究领域^[19]。古森鲍尔(Gusenbauer M)指出数据库的主题覆盖范围可以反映数据库对特定主题或学科的覆盖程度,并通过关键词检索结果确定了数据库的主题覆盖率,实现了56个数据库学科覆盖范围的对比分析^[20]。国内学者张李义在数据库内容评估中引入了数据库学科特色指标,该指标通过数据库收录覆盖学校各级学科资源的比重来评估数据库是否与学校的学科发展相适应^[21]。肖珑等人在中国高等教育文献保障系统(China Academic Library & Information System, CALIS)数字资源评估指标体系的设计中建立了与学科结构相关的指标,如单个学科拥有资源的数量或比重,用来衡量数据库对学科专业设置以及读者专业知识结构的适用性^[22]。赵玮佳等人利用学科的期刊匹配率和核心期刊保障率指标分析了外文期刊数据库对复旦大学“双一流”学科建设的保障情况^[23]。

现有的数据库评估指标体系包含了多维度的评估指标^[24-26]。由以上研究可见,从学科角度开展的评估主要针对数据库包含的学科外文期刊数量、比重和质量,缺乏与学科资源使用数据之间的联系,因此不能完全反映数据库资源的学科价值和使用效率。为了解决这一问题,阳昕等人通过问卷调查法获取院系对数据库的使用行为,利用社会网络分析

法揭示数据库与学科的供需关系^[27]。但是该方法由于与研究对象的数量、覆盖范围以及问卷有效性等多方面因素相关,结论的代表性易受到影响。

综上所述,现有的研究从不同的角度对外文期刊数据库的学科需求和学科匹配度进行了分析,但并未明确定义相关的评估指标,且缺乏学术效益指标。本文在现有研究的基础上,运用PageRank算法、引文分析法和社会网络分析法,定义了三个与学科相关联的外文期刊数据库评估指标,即数据库的学科需求度、数据库的学科学术效益、数据库的学科匹配度,并构建了数据库学科评估模型。

2 数据库学科相关评估指标的理论研究

2.1 数据库学科相关评估指标建立的理论基础

社会网络是由多个节点和各点之间的连线组成的集合。社会网络分析(Social Network Analysis)是对社会网络中行为者之间的关系进行量化的研究,利用该方法可以构建学科之间以及学科和数据库之间的关系网络。PageRank算法由美国斯坦福大学的佩奇(Page L)和布林(Brin S)提出,通过网页之间链接的引用关系计算网页得分,从而判定网页的重要程度^[28-29]。PageRank算法在同质信息网络中取得了较好的结果,部分学者尝试将该算法应用到文献计量学中,在异质网络中成功实现了节点重要性的排序。赵(Zhao F)等人将PageRank算法应用到异质的引用网络中,通过在“作者—论文”两种不同类型的节点之间添加链接来处理非循环的网络问题,更好地考虑不同类型节点彼此间的影响,通过该方法实现了对作者重要性的排序^[30]。阳昕等人利用社会网络分析法构建了“院系—数据库”的使用网络,将PageRank算法应用到该异质网络中,得到每个节点在网络全局中的PageRank值,通过该方法实现了对数据库重要性的排序^[27]。

用户通常采用浏览、引用等方式使用外文期刊数据库的资源,由于用户隶属于不同学科,所以在使用过程中间接建立了学科和数据库之间的使用关系。利用数据可视化软件Gephi0.9.6获得学科与数据库之间的社会网络图,数据来源设置为学科,目标设置为数据库,方向从学科指向数据库。将PageRank算法应用到该异质有向网络中得到每个节点的重要程度,计算公式为:



$$PR(p_i) = (1-d) \frac{1}{N} + d \sum_{j=1}^k \frac{PR(p_j)}{L(p_j)} \quad (1)$$

其中, N 代表节点总数。 d 为阻尼系数, 表示用户访问一个页面后, 继续访问下一个页面的概率, 默认值为 0.85, 用户停止点击概率为 $1-d$ 。 $PR(p_i)$ 代表节点 p_i 的 PageRank 值, $PR(p_j)$ 代表节点 p_j 的 PageRank 值, $L(p_j)$ 代表节点 p_j 的出链数量, $PR(p_j)/L(p_j)$ 代表节点 p_i 的链入节点 p_j 给予 p_i 的 PageRank 值, 求和项代表节点 p_i 从全部链接到它的节点获取的 PageRank 值之和。

公式(1)适用于无权网络, 在计算中默认节点之间链接的重要性是相同的, 即节点之间边的权重相同。但是在学科和数据库构成的异质有向网络中, 不同的学科对数据库的使用力度存在差异, 学科节点与数据库节点之间链接的权重不同。节点获得的链接权重越大, 节点的重要性就越高, 在计算中增加权重有利于区分不同节点的层次结构。因此, 将边的权重加入到公式(1)中, 获得有向加权网络节点重要度的计算公式^[31]:

$$PR_w(p_i) = (1-d) \frac{1}{N} + d \sum_{j=1}^k PR_w(p_j) * \frac{w(p_j, p_i)}{\sum_m w(p_j, p_m)} \quad (2)$$

其中, $w(p_j, p_i)$ 表示节点 p_j 对节点 p_i 的链接权重(即边的权重), $\sum_m w(p_j, p_m)$ 表示节点 p_j 所有出链(共有 m 个出链)的权重之和。将公式(2)应用到学科与数据库之间构建的异质有向加权网络以及学科与学科之间构建的同质有向加权网络中, 计算得到各节点的重要度。

2.2 数据库学科相关评估指标的定义

数据库常用的统计指标为数据库的使用次数、引用频次, 以及学科发文数量等, 目前数据库商提供的数据库使用次数大部分是针对整个机构的, 无法按学科区分。引用频次和发文数量可以对应到所属学科的期刊, 只需要建立学科、期刊与数据库之间的对应关系, 即可实现学科和数据库间权重的设定。利用引用频次和发文数量可建立两个与学科相关的数据库评估指标。

2.2.1 数据库的学科需求度指标

学术论文的产生需要调研大量文献资源, 不同学科的学术论文所引用的期刊可以客观地揭示出该学科对文献资源的直接需求, 通过引文分析建立学科与数据库之间的权重关系可以反映出不同学科对

数据库的需求度。数据库的学科需求度揭示了学科对数据库需求的迫切程度, 通过数据库节点在学科与数据库之间构建的引文网络中的加权 PageRank 值来表示, 基于公式(2)得到数据库的学科需求度指标公式为:

$$PR_{w,c}(Database_i) = (1-d) \frac{1}{N} + d \sum_{j=1}^k PR_{w,c}(p_j) * \frac{w_c(p_j, Database_i)}{\sum_m w_c(p_j, p_m)} \quad (3)$$

其中, 下角标 C 为引用(Citation)的缩写, 代表节点之间边的权重值为引用频次。 $PR_{w,c}(Database_i)$ 代表数据库节点 $Database_i$ 的 PageRank 值, $PR_{w,c}(p_j)$ 代表与数据库节点相链接的节点 p_j 的 PageRank 值, $w_c(p_j, database_i)$ 表示节点 p_j 对数据库节点 $Database_i$ 的链接权重, $\sum_m w_c(p_j, p_m)$ 表示节点 p_j 所有出链(共有 m 个出链)的权重之和。

2.2.2 数据库的学科学术效益指标

科研产出是评估数字资源使用效益的重要指标, 也是数字资源学术效益评估的主要依据^[32]。通过分析学科在数据库收录期刊上的发文数量, 可建立学科与数据库之间的权重关系, 以此来评估数据库的学术效益。数据库的学科学术效益通过数据库节点在学科与数据库之间构建的发文网络中的加权 PageRank 值来表示, 基于公式(2)得到数据库的学科学术效益指标公式为:

$$PR_{w,p}(Database_i) = (1-d) \frac{1}{N} + d \sum_{j=1}^k PR_{w,p}(p_j) * \frac{w_p(p_j, Database_i)}{\sum_m w_p(p_j, p_m)} \quad (4)$$

其中, 下角标 p 为出版物(publication)的缩写, 代表节点之间边的权重值为发文数量。其余参数解释参照公式(3)。

2.2.3 数据库的学科匹配度指标

现有对数据库的学科匹配程度的研究往往仅针对所研究的学科(以下简称本学科)进行期刊数量的分析。但在某一学科发文引用的期刊中, 除了本学科领域的期刊外, 通常还包含了其他学科领域的期刊。发文学科引用其他学科的期刊数量越大, 代表对被引学科的关注度越高。因此在数据库学科匹配度的定义中, 不但要考虑本学科领域资源, 还要对该学科领域高度关注的学科进行研究。为了获得本学科领域对其他学科领域的关注情况, 可将学科之间的相互引用关系替换网络之间的相互引用关系, 利用 Gephi0.9.6 获得学科之间的网络关系图, 通过公式

(2)得到学科的 PageRank 值,其中数据来源节点和目标节点均为学科,学科之间的边的权重为学科之间引用频次之和,由此可以获得与本学科领域关系最紧密的其他学科。以本学科及其关系最紧密学科资源为基础,建立第三个与学科相关的数据库评估指标,具体如下:

数据库的学科匹配度定义为本学科以及其关注学科的期刊数量占比与引用频次占比的乘积。期刊数量占比反映了数据库的学科数量,引用频次占比反映了数据库的学科关注度。数据库学科匹配度公式如下:

$$\delta_y = \sum \left(\frac{J_x}{J_y} * \frac{t_x}{T} \right) \quad (5)$$

数据库 y 的期刊总量为 J_y ,其中学科 x 的期刊数量 J_x 。 T 为本学科及其关注学科总引用频次, t_x 为学科 x 的引用频次。

数据库的学科学术效益指标与发文数量相关,学科需求度指标和学科匹配度指标中均利用了引用频次,但两者之间的侧重点有所不同。学科需求度指标是学科对数据库的单向引用,学科需求度指标精确地定位到本学科的直接需求,反映出所建设资源是否符合学科当下需求。学科匹配度指标是学科之间的双向引用且指标覆盖的范围更加宽泛,由本学科领域扩展到其关注的学科领域,由直接需求扩展到潜在需求,反映了数据库与学科发展的契合度和相关度。三个指标较全面地反映出关于数据库学科相关评估的主要信息,具有一定的完备性和较好的独立性,可以将三个指标添加到已有外文期刊数据库综合评估模型中作为学科相关评估的补充。

以上评估指标同样适用于包含多学科的学院与数据库关系的评估,以《研究生教育学科专业目录(2022年)》中的一级学科类别为例,每一种期刊对应至少一个一级学科,一级学科与学院相对应,通过学院对期刊的引用可以引申出不同学院之间的引证关系。

2.3 学科与数据库之间关系的建立

学科与数据库之间社会网络关系的建立是以期刊为媒介实现的。由于同一期刊可能来源于多个数据库,需要对来源数据库进行权重处理。基于外文电子期刊数据库的 COUNTER 标准报告,可以获取一种期刊在不同来源数据库的使用次数占比,无标准报告的数据采取等权重分配。以图 1 给出的关系图

为例,学科 1 引用了期刊 1 和期刊 2,引用频次分别为 TC_1 和 TC_2 ,期刊 1 来源于数据库 1 和数据库 2,期刊 1 在两个数据库中的使用次数分别为 U_{11} 和 U_{12} ,期刊 2 来源于数据库 2 和数据库 3,使用次数分别为 U_{22} 和 U_{23} 。每个数据库获得的引用频次通过使用次数的占比来分配,由此建立了学科 1 与三个数据库之间的权重关系,如学科 1 与数据库 2 之间的权重值包含两个部分,分别为期刊 1 和期刊 2 分配给数据库 2 的引用频次,其结果为 $TC_1 \left(\frac{U_{12}}{U_{11} + U_{12}} \right) + TC_2 \left(\frac{U_{22}}{U_{22} + U_{23}} \right)$ 。在计算数据库的学科学术效益指标时,将学科与期刊之间的引用频次替换成发文数量即可,该方法同样适用于包含多学科的学院与数据库关系的建立。

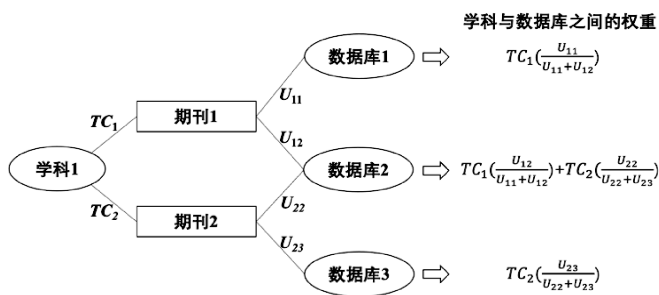


图 1 学科与数据库之间的关系

3 数据库学科相关评估指标体系的实证分析

为了验证上述指标,本文选择同济大学理学部学科群及相关数据库进行实证研究和量化分析,以期构建适合高校学科建设与学科发展的文献资源保障体系。同济大学 2022 年发布的《同济大学新一轮“双一流”建设方案》中明确提出了学校学科建设总体目标,按照“优势工科引领带动、厚重理科融合推动、特色医科协同驱动、精品文科共享联动、前沿交叉创新互动”的思路,集中力量进行一流学科和一流学科交叉领域的建设,全面提升学科的整体水平和国际竞争力。理学部作为学校“厚重理学融合推动”发展的重要支撑学部,包含了物理、化学、数学等基础学科,对该学部文献资源的精准建设有助于提升基础学科的竞争力、推动原创性引领性科技攻关。本研究以同济大学理学部的五个学院为研究对象开展外文期刊数据库的学科相关评估指标体系的实证研究。

3.1 数据收集

利用 Web of Science 数据库,检索同济大学理



学部近十年发表的 SCI 论文,理学部包含海洋与地球科学学院(以下简称海洋)、航空航天与力学学院(以下简称航力)、数学科学学院(以下简称数学)、物理科学与工程学院(以下简称物理)、化学科学与工程学院(以下简称化学)。选择 Web of Science Core Collections 数据库的 SCI-EXPANDED,发表时间为 2012—2021 年,检索地址为五个学院和下设机构的地址,数据下载方式为“全纪录并且包含所引用的参考文献”,检索时间为 2022 年 6 月 13 日。对检索结果进行人工清洗,共获得 8637 篇文章。各学院发文和引用来源情况详见表 1。利用 VOSviewer 软件获取 8637 篇文章参考文献的引用来源,去重后共计 24180 个引用来源。

引用频次 TC 阈值的选择依据:(1)论文引用频次的分布是不对称的,符合帕累托法则规律(即“二八现象”),即 80% 的引用来源于 20% 的论文,根据此法则,保障各学院所选择引用来源的引用频次总量占比超过其全部引用来源的 80%,确保分析结果中包含了绝大部分信息。(2)为了对比各学院的情况,各学院 TC 阈值的取值相同。根据以上原则,当引用频次 ≥ 7 时,各学院的引用来源均超过其全部引用来源的 80%,理学部整体的引用来源占全部引用来源的 91%。去重后获得 2641 种期刊,标记为引用来源期刊集合。8637 篇文章涉及到的发文期刊共计 1849 个,去重后为 1423 种,标记为发文来源期刊集合。

表 1 同济大学理学部各学院发文和引用来源情况

	海洋 学院	航力 学院	数学 学院	物理 学院	化学 学院
发文数量(篇)	1521	1231	1273	2517	2095
全部引用来源(个)	8718	5518	5488	6002	4813
引用频次 ≥ 7 时引用 来源期刊(种)	1261	1405	1127	1488	1428
发文来源期刊(种)	280	398	378	404	387

3.2 期刊保障和质量分析

将同济大学图书馆馆藏期刊清单及 JCR 核心期刊清单分别与引用来源期刊集合和发文来源期刊集合进行匹配,结果见表 2。理学部引用了 2641 种期刊,其中馆藏已保障期刊数量 2356 种(馆藏保障率为 89.2%),2317 种期刊被 JCR 收录(JCR 匹配率为 87.7%)。理学部在 1423 种期刊上有发文,其中馆藏保障期刊 1325 种(馆藏保障率为 93.1%),1378

种期刊被 JCR 收录(JCR 匹配率为 96.8%)。说明同济大学图书馆对理学部所需资源保障情况良好,且保障期刊质量较高,资源建设较为完备。

表 2 同济大学理学部发文和引用期刊的保障及质量情况

数据集	期刊数量 (种)	馆藏保障率 (%)	JCR 匹配率 (%)
引用来源期刊集合	2641	89.2	87.7
发文来源期刊集合	1423	93.1	96.8

3.3 数据库的学科相关评估结果

3.3.1 数据库的学科需求度

基于引文分析建立学院与数据库之间的权重关系,利用 Gephi0.9.6 绘制学院与数据库的社会网络关系图(见图 2),布局采用双圆布局法(Dual Circle Layout)。双圆布局法是将节点按照一定的参数要求均匀地分布在两个同心圆上,根据不同的设置在圆上按顺时针或逆时针方向绘制节点,该布局呈现效果直观,可以清晰展示社会网络关系图中节点之间的等级关系。为了更加直观地反映学院与数据库之间的关系,将代表学院的五个节点放置在图形的中央,由节点围成的两个环形代表学院使用的数据库。

(1)网络密度分析。该网络图包含节点数量 60 个,其中学院节点数 5 个,数据库节点数 55 个,边数量 252 条。平均度 4.2,平均加权度 5345.325,网络密度 0.071,为稀疏性网络(网络密度反映了网络中各节点之间的邻接程度,当网络密度 ≤ 1 时,对应的网络为稀疏性网络)。

(2)数据库的学科需求度。基于公式(3)计算该社会网络图中各节点的 $PR_{w,c}(Database_i)$,节点的大小与该值成正比。内部环形代表 PageRank 排名前 20 的数据库,其详情见表 3,以 ScienceDirect(以下简称 SD)数据库为起点逆时针方向从大到小分布,外部环形以 DOAJ 为起点。节点内侧数字代表节点连入度,未标注数字的节点连入度为 5。由排名前 10 的数据库可知,同济大学理学部对知名出版社的综合类(SD、Wiley、Springer)、自然科学类(Nature)以及专业类学/协会数据库需求度较高,其中专业类学/协会数据库占比 60%,分别偏重化学学科(美国化学学会的 ACS 和英国皇家化学学会的 RSC 数据库)、物理学科(美国物理学会的 APS、美国物理联合会的 AIP 和英国物理学会的 IOP 数据库)以及海洋和物理学科(美国地球物理学会的

AGU 数据库)。

表 3 PageRank 排名前 20 的数据库(引文分析)

排名	数据库	连入度	PageRank	排名	数据库	连入度	PageRank
1	SD	5	0.036016	11	IEEE	5	0.016939
2	ACS	5	0.021068	12	Science	5	0.016736
3	Wiley	5	0.020652	13	OSA	5	0.016616
4	Springer	5	0.020046	14	JSTOR	5	0.016552
5	APS	5	0.019707	15	T&F	5	0.016549
6	Nature	5	0.018803	16	SIAM	5	0.016383
7	AIP	5	0.018003	17	OJC 牛津	5	0.016361
8	RSC	5	0.017980	18	SEG	4	0.016257
9	AGU	5	0.017176	19	EBSCO (ASC+BSC)	5	0.016160
10	IOP	5	0.017045	20	NSTL 现刊	5	0.016143

数据库为起点,外部环形以 SAGE 现刊为起点。排名前 3 的综合类数据库对理学部的学术效益贡献最大,排名前 10 的数据库中包含 5 个专业类学/协会数据库:RSC、ACS、IOP、AIP 和 OSA。

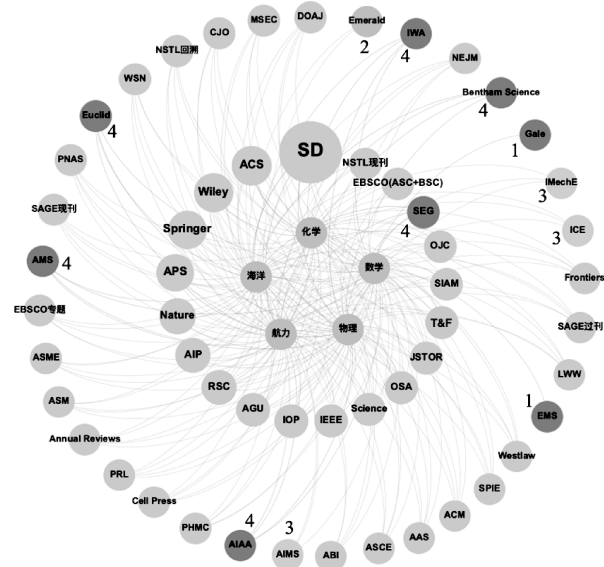


图 2 同济大学理学部各学院与数据库之间的社会网络关系图(引文分析)

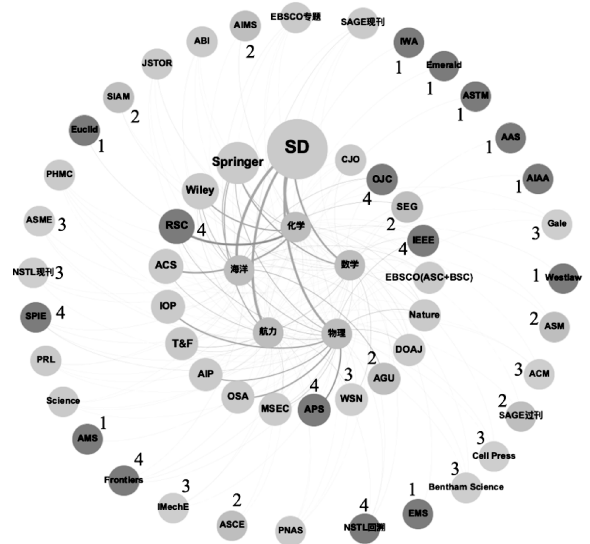


图 3 理学部各学院与数据库之间的社会网络关系图(发文分析)

表 4 PageRank 排名前 20 的数据库(发文分析)

排名	数据库	连入度	PageRank	排名	数据库	连入度	PageRank
1	SD	5	0.038575	11	APS	4	0.018145
2	Springer	5	0.025312	12	WSN	3	0.017834
3	Wiley	5	0.020866	13	AGU	2	0.017825
4	RSC	4	0.019955	14	DOAJ	5	0.017688
5	ACS	5	0.019482	15	Nature	5	0.017604
6	IOP	5	0.018740	16	EBSCO (ASC+BSC)	5	0.017432
7	T&F	5	0.018325	17	IEEE	4	0.017222
8	AIP	5	0.018260	18	SEG	2	0.017129
9	OSA	5	0.018247	19	OJC 牛津	4	0.017012
10	MSEC	5	0.018182	20	CJO 剑桥	5	0.016893

3.3.3 数据库的学科匹配度

通过 Gephi0.9.6 软件获得了同济大学理学部五个学院引用其他学院学科期刊的关系图(见图 4),基于公式(2)获得学院的 $PR_w(p_i)$ 值。以物理学院为例,物理学院周围较粗的半圆环代表物理学院发文引用本学院学科期刊的情况,物理学院引用最多的为化学学院学科的期刊,由指向化学的粗箭头表示。选取权重比例大于 1% 以及各学院权重排名前

3.3.2 数据库的学科学术效益

与引文分析的方法相同,通过分析发文数据获取各学院对应的数据库发文数量权重值,学院与数据库之间的社会网络关系见图 3。

(1)网络密度分析。该网络图包含节点数量 57 个,其中学院节点数 5 个,数据库节点数 52 个,边数量 180 条。平均度 3.158,平均加权度 143.182,网络密度 0.056,为稀疏性网络。

(2)数据库的学科学术效益。基于公式(4)计算网络图中各节点的 $PR_{w,p}(Database_i)$,PageRank 排名前 20 的数据库如表 4 所示。内部环形以 SD 数



5 的学院作为理学部重点关注的学院,并整理该部分学院覆盖的一级学科,共筛选出 11 个学院涉及到 23 个一级学科,除理学部的学院外,其他六个学院分别为:生命科学、医学、材料、电信、机械和软件等学院,通过引用频次获得各一级学科的使用频次占比。

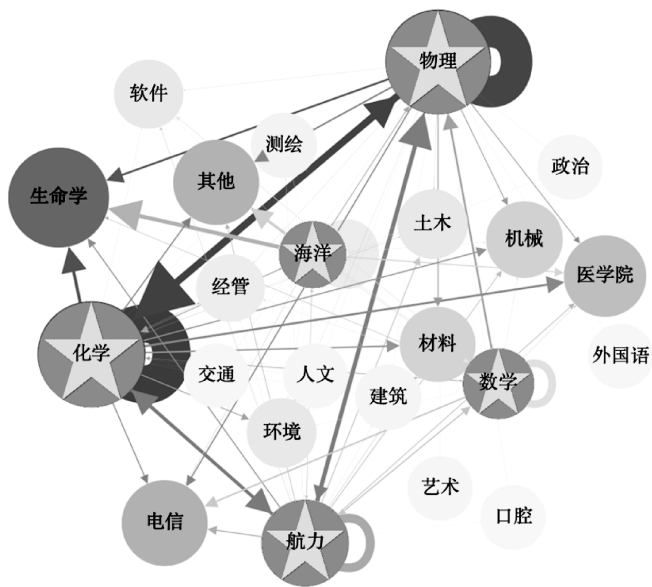


图 4 同济大学理学部引用其他学院学科期刊的关系图

在以上分析中,涉及到的外文期刊数据库共计 56 个,根据 2021 年同济大学馆藏期刊的统计结果,得到 56 个数据库分别映射到 23 个学科的期刊数量,由此得到不同学科在不同数据库的期刊数量占比。根据公式(5)获得数据库的学科匹配度,匹配度排名前 20 的数据库见表 5。排名前 5 的数据库均偏重数学学科,分别为:美国数学学会的 AMS、欧洲数学学会的 EMS 和欧几里德项目 Euclid、美国数学科学研究所的 AIMS、工业和应用数学学会的 SIAM。5 个数据库对理学部的学科匹配度均超过 50%,其他学科则低于 20%,说明 5 个数据库与理学部的学科高度相关,从另一个角度也说明数学类的 5 个数据库专业性较强,学科覆盖面较窄。

3.4 数据库学科评估模型的构建与应用

数据库学科评估模型的构建过程包括:(1)评估指标的确定。如上文所述确定了三个与学科相关联的外文期刊数据库评估指标,分别为:数据库的学科需求度、数据库的学科学术效益、数据库的学科匹配度。通过相关公式计算得到三个指标的数值;(2)指标的百分比排名计算。百分比排名可以横向衡量数

表 5 学科匹配度排名前 20 的数据库

排名	数据库	学科匹配度 (%)	排名	数据库	学科匹配度 (%)
1	AMS	67.50	11	SEG	13.56
2	EMS	60.08	12	SPIE	12.51
3	Euclid	54.64	13	AIP	12.20
4	AIMS	53.21	14	AGU	10.55
5	SIAM	52.58	15	Springer	9.62
6	RSC	19.39	16	APS	9.50
7	ACS	17.57	17	IMechE	8.20
8	OSA	14.65	18	MSEC	8.06
9	WSN	14.48	19	NSTL 回溯	7.97
10	IOP	14.03	20	AIAA	7.18

据库在当前所有数据库中的位置,利用该算法将数值结果转换为百分制得到三个指标下数据库的优劣排序;(3)指标权重的设置。本文将三个指标重要程度设置为相同,平均分配三个指标的权重,利用等权重计算各个数据库的总分。

基于学科评估模型和三个指标的数值结果,对 56 个外文期刊数据库进行综合评分及等级划分,设置 20 个分值为一个档位,大于 80 分的数据库级别定义为 5 级,排名前 10 的数据库见表 6。理学部外文期刊数据库的 5 个集群结果见图 5。

(1)5 级分类包含数据库 8 个,这些数据库在学科需求度、学科学术效益和学科匹配度方面均表现较好。学/协会专业类数据库 6 个,占比 75%,其中化学相关数据库 2 个(ACS、RSC),物理相关数据库 4 个(IOP、AIP、OSA、APS);综合类数据库 2 个(Springer、SD)。

(2)4 级分类包含数据库 11 个,其中专业类数据库 6 个,占比 54.5%,海洋与物理相关数据库 2 个(AGU、SEG),数学相关的 4 个数据库(SIAM、AIMS、AMS、Euclid)虽然在学科匹配度方面得分较高,但是其学科需求度和学科学术效益指标得分相对较低;自然科学类数据库 3 个,即 WSN 世界科技期刊、ProQuest Materials Science & Engineering Collection(MSEC)数据库和 Nature,其中 Nature 数据库在学科匹配度方面得分较低,因该数据库资源偏重生物和医学,与理学部的学科匹配度较低;综合类数据库 2 个,Taylor & Francis(简称 T&F)和 Wiley。

(3)3 级分类包含数据库 14 个,含学/协会专业

类数据库 4 个,即国际光学工程学会的 SPIE、欧洲数学学会的 EMS、电气电子工程师学会的 IEEE 和美国机械工程师学会的 ASME,其中物理和数学相关的数据库 SPIE 和 EMS 在学科匹配度方面得分较高,但是其学科需求度和学科学术效益指标得分相对较低;Science Online 数据库在该分类中得分最低,该数据库的需求度较高,但是理学部在该数据库的发文数量较低,且其学科偏重于工科;该分类中还包含了国家科技图书文献中心的外文回溯数据库(NSTL 回溯)和外文现刊数据库(NSTL 现刊)。

(4)2 级分类包含数据库 14 个,这些数据库在学科需求度、学科学术效益和学科匹配度方面均表现欠佳。其中学/协会数据库 5 个,即英国机械工程师学会的 IMechE、美国航空航天学会的 AIAA、美国土木工程师学会的 ASCE、美国微生物学会的 ASM、美国计算机协会的 ACM;集成商数据库有 3 个,分别为 ABI、ProQuest Research Library (PRL) 和 ProQuest Health & Medical Collection (PHMC),均为 ProQuest 公司的数据库。

(5)1 级分类包含数据库 9 个,其中学/协会数据库 5 个,含美国试验与材料协会的 ASTM、国际水协会的 IWA、英国土木工程师协会的 ICE、美国天文学会的 AAS、马萨诸塞州医学学会的新英格兰医学 (NEJM)。在该分类中,理学部在 ICE、Lippincott Williams & Wilkins 出版公司出版的医学全文期刊数据库 (LWW) 和 NEJM 数据库上均未产生学术效益。

表 6 学科评估模型得分排名前 10 的数据库

排名	数据库	学科需求度 (%)	学科学术效益 (%)	学科匹配度 (%)	总值
1	ACS	96.4	90.5	87.7	91.5
2	RSC	85.7	92.4	89.4	89.2
3	Springer	92.8	96.2	73.6	87.5
4	SD	98.2	98.1	63.1	86.5
5	IOP	82.1	88.6	82.4	84.4
6	AIP	87.5	84.9	77.1	83.2
7	OSA	76.7	83.0	85.9	81.9
8	APS	91.0	79.2	71.9	80.7
9	Wiley	94.6	94.3	49.1	79.3
10	AGU	83.9	75.4	75.4	78.2

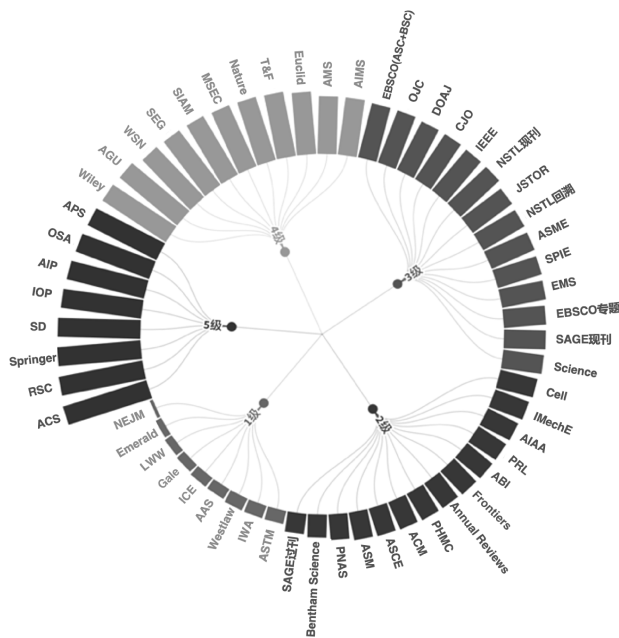


图 5 学科评估模型对数据库的等级划分

4 结论

本研究基于 PageRank 算法和现有研究成果提出了三个与学科相关的外文期刊数据库评估指标,即数据库的学科需求度、数据库的学科学术效益、数据库的学科匹配度,搭建了数据库学科评估模型,并通过实证分析建立了不同等级的理学部的文献资源集群。在数据库的学科匹配度指标设定中,利用社会网络分析获得了理学部引用其他学科资源的情况,拓展了理学部所关注资源的学科范围,通过该指标可以进一步建立院系或学科之间的学科群,有利于院系资源的共建共享和交叉学科资源的建设与服务。三个指标适用于不同的资源服务对象,如学科、院系和不同学科群,数据来源具有客观性、可获取性和适用性,可以作为外文期刊数据库综合评估中关于学科相关评估的补充,为学科资源建设和精准化的资源推广提供了有力依据。关于该项研究,有几点建议可供参考。

(1)数据准确性的验证

学院教师发文作为本研究的基础数据,其准确性直接影响该项研究的可行性和可靠性。为了保证数据的准确性,对发文数据进行了如下处理:①与院系、学科馆员合作确定同济大学理学院教师名单,避免了由于院系官网人员信息更新滞后带来的误



差;②基于WOS数据库获取的教师发文信息结果,通过人工逐条验证是否为所属学院教师发文,对有异议部分联系对应学院进行确认,其次利用同济大学机构知识库中教师成果进行了双重验证。

(2)引用来源词表的建立和不同类型引用来源的分析

利用VOSviewer软件获取发文参考文献的引用来源结果为引用来源信息的缩写,大部分的期刊缩写可以通过引用来源词表进行匹配,少部分期刊和非期刊的引用来源结果需要逐条查询验证,在对多个学部或全校进行研究时,建议逐步收集和完善的引用来源缩写和全称的引用来源词表,以提高研究的准确性和提升研究效率。

在本文的实证分析中虽然选取的引用来源中包含了超过80%的引用信息,但是未选取的引用来源中也可能包含有价值的信息,如高价值的图书资源可能受到引用规范的影响未产生有效的引用,针对该部分引用来源需要进行深入的研究。

(3)一级学科分类划分和匹配依据

各学院一级学科的划分主要依据《同济大学授予博士硕士学位和培养研究生学科专业目录》(2022年编制),并结合院系专家意见对未列入的学科进行补充,保证了学科划分的准确性和完整性。期刊的一级学科匹配数据来源于电子资源绩效分析平台,目前该平台实现了9万多种外文期刊的学科映射,包含了Web of Science Core Collections数据库的SCI-EXPANDED的全部内容。

受篇幅所限,本论文仅对学科相关数据库评估指标进行了理论研究和实证分析,尚有相关问题需后续深入探究:①通过问卷调查、用户访谈、专家座谈会等方式进一步获取以下两方面的信息,一是同济大学理学部的教师对外文期刊数据库的使用信息,通过教师的选择和排序验证三个指标的正确性和适用性;二是专家对三个指标的权重打分,对现有的权重设定进行修正;②逐步收集和完善的引用来源词表;③对不同类型的引用来源进行分类分析;④不同学科的科研论文产出和引文特征不同,如人文学科引用文献来源丰富,外文期刊数据库在人文学科引用文献的比重会降低,本文提出的三个指标的评估方法和理念具有普适性,但是针对不同的学科还需要结合学科发文及引文特点对评估算法进行修正和补充;⑤评估模型有效性的验证,一方面通过问卷

调查的方式判断不同等级的学科资源群是否与教师实际需求相吻合,另一方面通过对使用数据的动态跟踪,判断随着学科发展和需求的变化,评估结果是否能够反映出学科对文献资源需求的变化。

参考文献

- 1 阎亚矢.“双一流”建设背景下高校学科文献资源建设与绩效考核[J].大学图书馆学报,2020,38(4):36-39,49.
- 2 Kyriallidou M, Giersch S. Qualitative analysis of association of research libraries' e-metrics participant feedback about the evolution of measures for networked electronic resources[J]. Library Quarterly, 2004, 74(4): 423-440.
- 3 California Digital Library. Calculating scholarly journal value through objective metrics[EB/OL]. [2022-10-23]. <https://cdlib.org/cdlinfo/2012/02/13/calculating-scholarly-journal-value-through-objective-metrics/>.
- 4 肖琼,张宇红.电子资源评价指标体系的建立初探[J].大学图书馆学报,2002,20(3):35-42,91.
- 5 唐琼.采用因子分析法构建数字资源选择标准体系[J].大学图书馆学报,2011,29(2):56-63.
- 6 熊霞,高凡,李睦,等.高校图书馆电子资源统计指标体系设计与构建[J].大学图书馆学报,2021,39(3):59-66.
- 7 Coughlin D M, Campbell M C, Jansen B J. A web analytics approach for appraising electronic resources in academic libraries[J]. Journal of the Association for Information Science and Technology, 2016, 67(3): 518-534.
- 8 Coughlin D M, Jansen B J. Modeling journal bibliometrics to predict downloads and inform purchase decisions at university research libraries[J]. Journal of the Association for Information Science and Technology, 2016, 67(9): 2263-2273.
- 9 Becker L T, Gould E M, Mezick J A. Assessing the value of library resources for renewal negotiations[J]. Qualitative and Quantitative Methods in Libraries, 2021, 10(3): 389-405.
- 10 王丽珍,张幸芝,雷润玲.基于SCI的图书馆外文期刊资源建设可视化分析——以西安交通大学图书馆为例[J].农业图书馆情报学刊,2015,27(12):58-61.
- 11 刘凤仪,叶继元.我国区域高校图书馆西文期刊保障体系实证研究——以华东五校图书馆情报学为例[J].大学图书馆学报,2021,39(2):28-34.
- 12 Lightman H, Manilov S. A simple method for evaluating a journal collection: a case study of northwestern university library's economics collection[J]. Journal of Academic Librarianship, 2000, 26(3): 183-190.
- 13 Davis P. Where to spend our e-journal money? defining a university library's core collection through citation analysis[J]. Portal Libraries and the Academy, 2002, 2(1): 155-66.
- 14 Walters W H, Markgren S. Zero-based serials review: an objective, comprehensive method of selecting full-text journal resources in response to local needs[J]. Journal of Academic Librarianship, 2000, 46(5): 102189.
- 15 张建勇,刘筱敏.基于引文统计分析的西文期刊馆藏发展策略



- [J].图书情报工作,2007,354(5):106-109.
- 16 花芳,赵军平,管翠中,等.目标馆藏及其在馆藏期刊评价中的应用[J].图书情报工作,2014,58(15):94-98.
- 17 张宇娥,柯佳秀,杜丽.高校学科期刊资源保障优先级方法与实证分析[J].图书情报工作,2020,64(8):71-77.
- 18 Jacsó P. Database source coverage: hypes, vital signs and reality checks[J]. *Online Information Review*,2009,33(5):997-1007.
- 19 Coughlin D M, Campbell M C, Jansen B J. A web analytics approach for appraising electronic resources in academic libraries [J]. *Journal of the Association for Information Science and Technology*,2016,67(3):518-534.
- 20 Gusenbauer M. Search where you will find most: comparing the disciplinary coverage of 56 bibliographic databases[J]. *Scientometrics*,2022,127(5):2683-2745.
- 21 张李义.高校图书馆商业数据库评价指标体系研究[J].*中国图书馆学报*,2004,30(1):68-71.
- 22 肖珑,李浩凌,徐成.CALIS数字资源评估指标体系及其应用指南[J].*大学图书馆学报*,2008,161(3):2-8,17.
- 23 赵玮佳,郝群,张立彬.高校图书馆外文数据库资源建设的使用统计和分析思考——以复旦大学4个数据库的实践为例[J].*图书情报工作*,2022,66(4):73-86.
- 24 张轶华,李芳,宋海艳,等.基于实证比较的外文期刊数据库评价模型构建与应用[J].*图书馆杂志*,2019,38(8):65-73,16.
- 25 钱婧,王猛,陈雅.国外一流高校图书馆馆藏资源评估分析及启示[J].*国家图书馆学报*,2022,31(5):95-103.
- 26 宋海艳,张轶华,黄毓,等.高校图书馆文献信息资源绩效管理体系统构建研究[J].*大学图书馆学报*,2021,39(4):22-27.
- 27 阳昕,张敏,廖剑岚,等.社会网络视角下的高校图书馆电子资源利用研究——以复旦大学图书馆为例[J].*图书情报工作*,2021,65(15):91-99.
- 28 Brin S, Page L. The anatomy of a large-scale hypertextual web search engine [J]. *Computer Networks and ISDN Systems*, 1998,30(1-7):107-117.
- 29 Ding Y, Yan E J, Frazho A, et al. PageRank for ranking authors in co-citation networks[J]. *Journal of the American Society for Information Science and Technology*, 2009, 60 (11): 2229-2243.
- 30 Zhao F, Zhang Y, Lu J, et al. Measuring academic influence using heterogeneous author-citation networks[J]. *Scientometrics*, 2019,118(3):1119-1140.
- 31 Bollen J, Rodriguez M A, Van D. Journal status[J]. *Scientometrics*,2006,69(3):669-687.
- 32 任传荣,李娟娟.高校图书馆电子资源学术效益评估指标体系构建[J].*图书馆论坛*,2013,33(1):160-163,76.

作者单位:同济大学图书馆,上海,200092

收稿日期:2022年11月8日

修回日期:2023年3月2日

(责任编辑:关志英)

Quantitative Research on Subject-Related Evaluation Indexes of the Foreign Journal Database

—Taking the Science Division of Tongji University as an Example

Ma Shuang Wei Wenjie Li Lijuan

Abstract: The core mission of Double First-Class construction is discipline construction. The construction of library resources should closely match the development needs of disciplines, provide accurate literature resources and services for disciplines, and make contributions to the discipline's construction and development in colleges and universities. Based on the PageRank algorithm, using citation analysis and social network analysis, this paper puts forward three subject-related evaluation indexes of foreign periodicals database, namely, the subject demand degree of the database, the academic benefits of the database, and the subject matching degree of the database, and constructs the subject evaluation model of the database. Taking the Science Division of Tongji University as an example, this paper makes an empirical analysis of the database of foreign periodicals and establishes the literature resource clusters of different levels of the Science Division.

Keywords: Database Evaluation; Citation Analysis; Social Network Analysis; Discipline Resources; Indicator Quantification