

国内数字图书馆领域作者知识扩散研究

吴新宇

上海大学文化遗产与信息管理学院 上海 201900

摘要: 以我国数字图书馆领域高被引文献作者及引证作者为研究对象, 结合复杂社会网络软件Gephi, 对该领域引证文献的作者、机构、学科分布进行分析, 构建作者知识扩散网络并对其进行测度和可视化, 分析作者知识扩散网络情况。

关键词: 数字图书馆; 知识扩散; GEPHI; 社会网络分析

A Study on the Knowledge Diffusion of Authors in the Field of Digital Libraries in China

Xinyu Wu

School of Cultural Heritage and Information Management, Shanghai University, Shanghai, 201900

Abstract: This paper focuses on highly cited authors and citing authors in the field of digital libraries in China. Using the complex social network software Gephi, the study analyzes the distribution of authors, institutions, and disciplines in the domain of cited literature. It constructs a knowledge diffusion network of authors, measures it, and visualizes it to analyze the dissemination of knowledge among authors in this field.

Keywords: Digital Library; Knowledge Diffusion; GEPHI; Social Network Analysis

引言:

科学文献并不是孤立的, 文献之间存在各种各样的联系, 这些联系体现在科学文献间的相互引证, 从而形成了科学知识之间的交流^[1], 文献的引证与被引证常用于描述和测度学科领域知识扩散的模式或规律^[2]。本研究基于作者间的引证与被引证关系, 结合社会网络分析软件构建我国数字图书馆领域高被引和高引证作者的引文网络, 研究作者间的知识扩散规律。

一、知识扩散测度及作者知识扩散关系构建

1. 数据来源

本文所用数据来自中国引文数据库以“数字图书馆”或“智慧图书馆”为题名进行检索, 被引文献类型为“核心期刊”和“CSSCI收录刊”, 出版时间设置为2002年至2022年, 总共检索出含有引证文献的文献共有4960篇, 共被引次数为46427, 篇均被引数为9.36, 其中最高的含有370篇引证文献, 最低的含有1篇引证文献。

2. 作者知识扩散关系构建

国内外主要从引文与合作这两个角度出发, 利用网络分析方法以及知识扩散特征测度指标对知识扩散进行

表征测度研究^[3]。基于作者间的引用关系来构建知识扩散关系网络。作者间知识扩散关系由图1所示, 被引证和引证文献作者间存在关联, 每位作者可能引证不同数目的相关文献, 基于此搭建相关领域的作者集合。

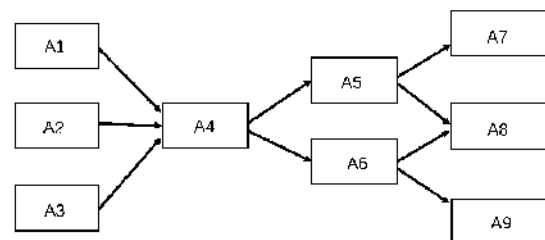


图1 作者知识扩散关系

知识扩散过程基于上述搭建的相关领域作者集合而展开, 同领域的不同作者通过相互之间的学术引证和学术观点交流, 构建起该学科的理论体系^[4]。本文在文献引证关系和相关领域作者集合的基础上, 利用Python工具构建作者间的知识扩散矩阵, 从而形成知识扩散网络。

3. 作者知识扩散测度指标

作者知识扩散测度指标分为广度和深度两种。作者知识扩散广度 (Author Diffusion Breadth, ADB) 指标的

测度对象为文献作者的引证关系，具体来说，即一篇文献作者对应的被引数量，如某篇文献被5名同领域作者分别引用1次，则该篇文献作者的知识扩散广度为5；作者知识扩散深度（Author Diffusion Intensity, ADI）则涉及每篇被引文献数量的平均值，涉及的指标包括该作者的文献数量和引证作者的数量和引证次数。

二、作者知识扩散网络构建与分析

1. 作者知识扩散网络可视化

将构建的被引频次排名前30位的作者及其引证作者间知识扩散矩阵导入Gephi软件并将其可视化，从中心性分析、网络模块化分析、网络直径和平均路径长度分析、平均聚类分析和作者知识扩散类型分析等方面描绘数字图书馆领域作者知识扩散网络。本研究社会网络分析为了保留有价值的作者节点，在绘制可视化图谱过程中将边的权重值设定为6.982，从而更直观地呈现出数字图书馆领域作者知识扩散情况（图2）。

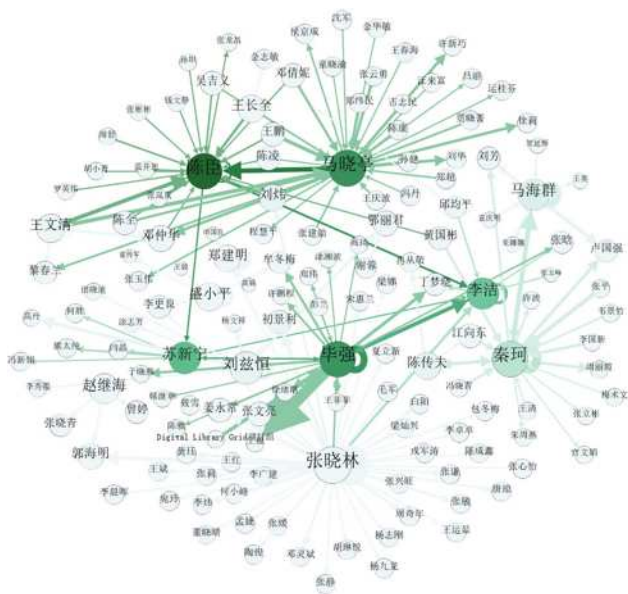


图2 作者知识扩散可视化图谱

2. 作者知识扩散网络分析

(1) 中心性分析

使用Gephi度统计功能计算出该网络节点平均度为1.243、平均加权度为15.474。度和加权度最大的节点对应作者是陈臣，节点度为310，加权度为436。同时分析图3可知，陈臣、毕强等作者的颜色较深、节点较大，表明诸位作者的被引或引用文献频次较高，同时也包括自引的情况。张晓林、毕强等作者节点的连线箭头大多是由其他作者指向诸位作者，即上述作者的被引的次数较多，具有较高的点度中心性。作者王福、李洁较多地引用作者毕强的文献，作者马海群和秦珂之间存在多次

互相引用关系。作者陈臣较多地引证作者马晓亭和王文清的文献，作者马晓亭较多的引用作者王文清的文献，作者邓仲华同时引用了作者马晓亭和王文清的文献，而作者邓仲华的文献又被作者陈臣所引用，作者吴吉义、王长全、邓倩妮同时引用了作者马晓亭和陈臣的文献，由此可见以上几位作者形成了小范围的知识扩散网络。

(2) 网络模块化分析

使用Gephi的社群发现算法对作者知识扩散网络进行社群发现，可以找到5个社群，说明在数字图书馆领域被引频次排名前30的作者网络中有5个合作团体，用5中不同的颜色来表示，如图3所示。同时得到模块化度量值为0.675，说明这些合作团体之间的连接较多，但作者群体内部连接并不是非常紧密，形成的知识扩散关系不是非常稳固。

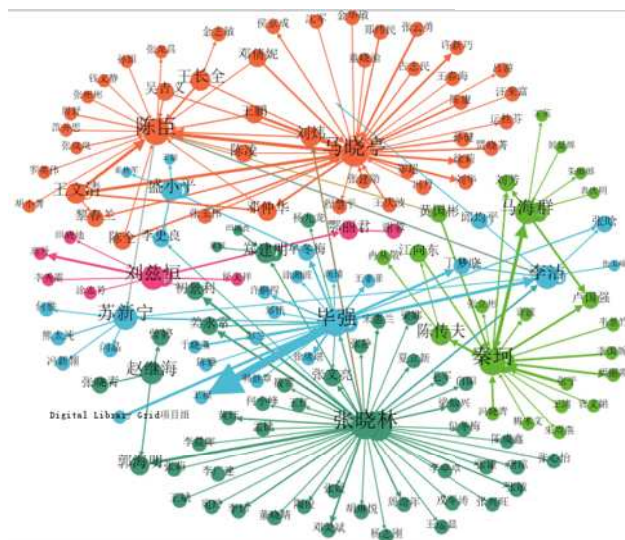


图3 作者知识扩散网络模块化分组图

(3) 网络直径和平均路径长度分析

作者知识扩散社会网络分析中两个作者节点间的最大距离即为网络直径，使用Gephi计算模块功能统计出数字图书馆领域高被引作者网络的直径为5。平均路径长度则代表两个作者节点间的路径平均值，该值大小与节点间关联性呈现负相关，经Gephi计算模块功能统计为2.932，表明其具有小世界性的特征。

(4) 平均聚类系数分析

平均聚类系数是社会网络分析中节点间关联强度的重要指标，本研究基于数字图书馆领域作者引证与被引证关系的测度探究知识扩散情况^[5]。使用Gephi计算出数字图书馆领域高被引作者网络的平均聚类系数为0.079，其范围更接近于1，表明其具有小世界性的特征。

(5) 作者知识扩散类型分析

数字图书馆领域作者知识扩散规律符合当前主流研究结论,即可将领域内作者知识扩散类型分为以下三种类型^[6]:一是知识输出型作者,作为知识扩散源节点,随着数字图书馆研究的深入,将会继续作为知识扩散源节点出现在网络中,或是引证其他作者的研究成果,传播学术新思想,从而在源节点的角色之外加入了中间节点的角色;二是知识输入型作者,作者知识扩散终节点,广泛吸收和引证其他作者的学术观点,随着数字图书馆研究的深入,发表前沿学术观点,由此作为知识扩散中间节点出现;三是知识综合型作者,作为前两种作者类型的综合体,能够同时作为知识扩散的源节点、中间节点和终节点出现,在奠基数字图书馆研究领域发展的同时,广泛吸收和接纳其他作者的研究成果,又创新性地通过知识传播和交流集大成式地推动数字图书馆相关领域的研究。因此,在知识扩散过程中,综合型作者可以看作知识扩散的桥梁。如图2中的毕强、马晓亭、秦珂、张晓林、陈臣、刘兹恒、苏新宁、马海群等作者。

3. 作者知识扩散水平测度

(1) 作者知识扩散广度

利用Gephi软件统计功能板块,计算出数字图书馆领域的6743位作者的引证文献作者数量,即ADB(表1)

表1 作者知识扩散广度统计表

作者	ADB	作者	ADB	作者	ADB
赵继海	411	王文清	189	邱均平	96
毕强	335	卢共平	176	艾雯	78
陈臣	310	马文峰	153	王长全	74
张晓林	299	郑建明	132	盛小平	66
陈凌	236	张晓林	130	陈传夫	66
马晓亭	212	莫少强	117	朱庆华	63
黄晓斌	207	苏新宁	102

赵继海、毕强、陈臣、张晓林、陈凌等作者ADB值较大,根据作者知识扩散广度的定义,这几位作者文献研究成果被引证数量大,表明他们在数字图书馆研究领域具有较强的代表性和权威性。本文对数字图书馆研究领域ADB分布量进行统计分析,根据测算出的幂函数拟合值 $R^2=0.8977$,认为其符合幂函数指数特征,具体函数图像如图4所示,可见ADB值小于10的作者数量呈现广泛分布的特征。

根据幂函数图像的分布特点可知,作者知识扩散广度较低的作者数量居多,比例达到整体的86.06%,具体作者数量为5083人。从学术研究的角度来说,文献发表的质量高低与研究成果的发布时间先后决定了作者知识

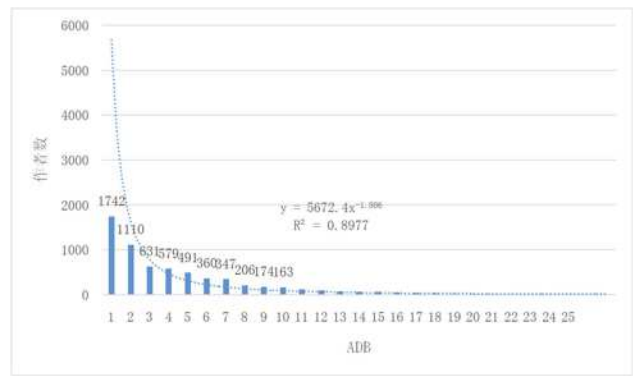


图4 作者知识扩散广度分布图

扩散广度,高质量论文和新发布的研究成果能够在知识扩散中占据先机,被同领域的学者引证学习和赢得他们的青睐。在上世纪60年代就有文献指出,随着对某一研究领域的不断深入,后续研究往往将会引证出版时间较早和高质量的文献,这与本文对数字图书馆领域作者知识扩散广度的测度结果一致^[19]。

(2) 作者知识扩散深度

借助中心性分析——点度中性分析结果,来确定网络节点的入度(In Degree)和出度(Out Degree)。对某一网络中某一作者节点来说:该作者的入度表示其在科研过程中引证其他作者的次数,出度表示该作者被引证的次数。所获得的作者点度中心性测度值见表2。

表2 知识扩散网络作者点度中心性测度值
(前10位作者)

作者	出度 (Out Degree)	入度 (Indegree)	度 (Degree)
赵继海	548	13	561
毕强	480	404	884
陈臣	463	456	919
张晓林	379	66	445
陈凌	366	0	372
马晓亭	322	372	694
黄晓斌	308	123	431
王文清	288	0	296
卢共平	285	19	304
马文峰	278	43	321

从表2中的出度和入度值可知作者类型。由于获取的数字图书馆领域的4960条文献至少有1篇引证文献,由数据统计可知6743位作者的出度值均大于一,其中知识输出型作者有陈凌、王文清等2113位作者,知识综合型作者有赵继海、毕强、陈臣、张晓林等4630位作者。出度值虽然能在总体上反映出作者在知识扩散网络中的知识贡献程度,但是无法避免作者文献量对引证作者数

的影响。因此，笔者依据作者知识扩散深度指标的测度，得出表3数据。

表3 作者知识扩散深度测度表

作者	ADI	作者	ADI	作者	ADI
王文清	288.00	苏新宁	126.50	艾雯	80.66
谢蓉	175.00	陈凌	122.00	王长全	80.66
刘焕成	149.00	史国祥	111.00	王世伟	74.50
沙勇忠	149.00	杨明芳	99.00	赵继海	68.50
张晓林	130.00	郭海明	95.00	陈传夫	63.16
代钰珠	129.00	谭祥金	94.00	卢共平	57.00
钱鸥	129.00	杨沛超	86.00

由表3可知，样本中作者知识扩散深度最高的作者是王文清，其测度值高达288，其次是谢蓉、刘焕成和沙勇忠等。同时结合表1可以看出，作者王文清、艾雯、王长全、苏新宁、陈凌、赵继海、张晓林、卢共平等的知识扩散广度和深度值均较高，可见诸位作者的被引证作者数和单篇文献引证作者数都居于前列，即在知识扩散网络中有较为重要的地位，为了更为直观地展现作者知识扩散深度的人员分布，笔者对其进行了区间统计（图5）。

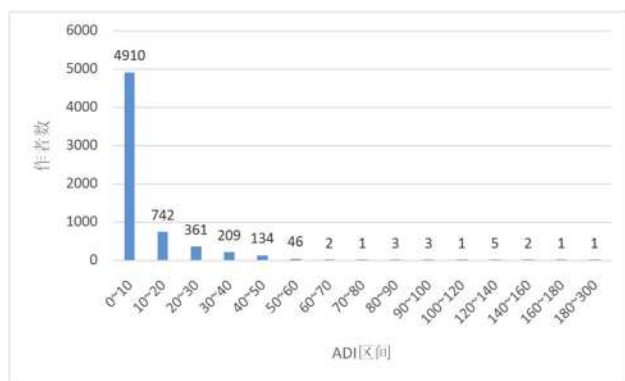


图5 作者知识扩散深度区间分布图

在数字图书馆研究领域作者知识扩散深度分布图像符合幂函数特征，该研究领域72.82%的作者知识扩散深度小于10，表明数字图书馆领域作者知识扩散深度较低，权威作者和有影响力的研究成果较少。相较于ADB，ADI的测度结果除研究成果质量、出版时间外，其引证文献的合作者数量和引证次数同样为重要参数，也就是说，同一作者多次引证某一文献或出现自引情况时，该作者文献被引数量增加，ADI值也会随之上升。

三、结语

本文以我国数字图书馆领域4960篇被引文献为研究对象，从作者互引的角度，将高被引作者之间的互引网络作为研究对象，结合复杂社会网络软件Gephi，对数字图书馆领域高被引作者及引证作者间的知识流动情况进行分析，以期促进该研究领域学术交流。

参考文献：

[1] 庞景安.科学计量研究方法论[M].第二版.北京：科学技术文献出版社，2002：213.

[2] 赵蓉英，魏绪秋.引证视角下的作者知识扩散规律探析[J].情报理论与实践，2016，39（08）：12-17.

[3] 李力，刘德洪，董克.我国竞争情报领域知识扩散结构研究[J].情报科学，2015，33（12）：13-19.

[4] PRICE D J. NETWORKS OF SCIENTIFIC PAPERS. [J].Science(New York, N.Y.),1965,149(3683).

[5] 王龙.合作网络模型结构研究与应用[D].济南：山东师范大学，2015：19-21.

[6] 章忠志，荣莉莉，周涛.一类无标度合作网络的演化模型[J].系统工程理论与实践，2005（11）：57-62.