

基于引荐分析的国内引用认同领域分析*

周春雷, 李冰莹

(郑州大学信息管理学院, 郑州 450001)

摘要: 以国内引用认同研究者为研究对象, 借助CnkiRef、Gephi等工具, 从引荐分析视角分析了国内引用认同领域的研究景象, 揭示出领域经典文献、权威期刊、经典图书和研究主题分布等信息, 为相关学者了解该领域的研究现状提供参考, 也为丰富引文分析方法提供借鉴。

关键词: 文献计量; 引用认同; 引荐分析法; 领域分析

中图分类号: G350

DOI: 10.3772/j.issn.1673—2286.2015.01.005

自Garfield创立引文分析法以来, 被引问题就一直是图书情报领域关注的重点。2001年, White提出了“引用认同”(Citation Identity, CI)概念以描述施引者的引用特征, 并将是否进入重要学者的引用认同作为评估个体研究人员的指标^[1]。随后, 武夷山^[2]和屈宝强^[3]等介绍了引用认同概念, 但未引起国内学者的重视; 2009年, 武夷山在其科学网博客上再次推介这一概念, 开始引起国内科学计量界的关注; 马凤、武夷山^[4]以王崇德教授为例详细展示了运用引用认同思想进行引文分析的基本过程。由于多样化的引用动机使得引用行为具有很大的随意性, 所以偶然的引用行为不宜直接作为衡量作者学术影响力的标准, 因此, 周春雷于2009年^[5]在定义引荐的基础上提出了“引荐分析法”, 并进一步将其发展为“学术授信评价方法”^[6]。

简单来说, 引荐分析法是基于高影响力作者的有所限制的引用认同, 通过对多个作者及其施引对象集合的研究, 寻找领域同行、发现领域新秀, 最终勾勒出以重要研究者为节点的学科结构图。与常规引文分析技术相比, 引用认同分析更关注从施引者角度对被引者进行研究。通过对施引者和被引者的双重研究, 不仅能掌

握领域内的研究热点, 还能对施引者的引用风格及其与被引者之间的关系有所了解。本文从引荐视角挖掘国内引用认同研究者们的施引情况, 为读者深入了解引用认同领域提供帮助。

1 研究方法及数据

引荐分析法的前提假设是: 领域同行们对某一学者的频繁引用, 是将此作者视为领域专家的学术授信行为; 同一施引者对某人的多次引用反映了施引者对被引者领域同行身份的认可, 可视为对被引者学术成果的正面评价, 是真实推荐的表示。通过引用认同分析, 可以了解施引者与被引者在研究领域、学术思想上的联系, 从而发掘出领域内的同行学者, 而引荐分析法可以使我们借力学科专家们的学术洞察力分析出在领域内有影响的学者, 然后汇总专家们的引荐信息绘制较为全面的领域研究结构图。本研究以国内引用认同研究者的引用认同为研究对象, 将施引者对同一位候选人的多次引用仅视为一次引荐。本文在进行引荐分析时对施引者取第一作者, 对被引者取全部作者。

* 本研究得到国家自然科学基金资助项目“学术授信评价及其在学术博客评价中的应用”(编号: 11CTQ014)及“作者引用网络模式与功效研究”(编号: 12CTQ026)的资助。

引荐结果需要经过两次学术授信才能产生。第一次是提名者资格认定,以国内引用认同研究者的引用认同为依据,统计每位研究者引用的作者及次数,以确定有资格提供学术授信意见的作者名单。对于期刊论文作者,要求在本领域内至少被引用一次,其引荐信息才会被采信,而那些在领域内尚无影响力的作者即使多次引用某作者,其意见也不被采信。鉴于硕博论文的研究较为专业而深入,本文基于学术授信思想直接采信相关作者的引荐信息。第二次是有影响力作者引荐,这需要分析有一定影响力的作者们的引用认同,得到其频繁引用的作者名单。经试验,发现引用2次及以上可被视为一次有效的引荐。例如,武夷山被引用31次,获得的引荐数为8次,即有8名在本领域具有一定影响力的同行引用了他,这些人的学术地位为本领域硕博论文作者或至少在本领域内被引用1次。

本文利用CNKI、万方和CSSCI数据库,于2014年3月10日以“引用认同”为主题词进行检索,并对引文进行整理、合并。经人工规范,共获得40名作者在16种期刊上发表的31篇期刊论文和3篇学位论文,以及659条引文。其中,25篇在篇名中出现“引用认同”,31篇的关键词包含“引用认同”。

2 国内引用认同领域的引用认同分析

文献的年代分布有助于我们从时间上把握引用认同研究的发展状况和走势。表1介绍了2009年以来引用认同研究文献的年度分布。自武夷山于2002年向我国引进引用认同概念,我国学者在2009年才开始对引用认同研究进行关注,近年来才小规模兴起。

表1 文献的年度分布

年份	2009	2010	2011	2012	2013
数量/篇	1	8	10	11	4

2.1 高被引作者被引荐情况分析

统计高被引作者可以看出领域内学者研究成果的受关注程度。被国内引用认同研究者频繁引用的国内外学者可以认为是对其学术成果的肯定,可以被视为一种引荐。经统计,本领域40名作者共引用同行928人次,经汇总、去重后得到无重复作者451人,其中被引用10

次以上的有12人,被引用5次以上的有33人。表2列出了部分累计被引次数在5次以上的作者名单,同时给出了对应的被引荐次数。

表2 引用认同领域学者被引用、引荐(≥ 2)次数表

序号	作者	被引用次数	被有影响力作者引荐数	被全部作者引荐数
1	White HD	58	11	12
2	武夷山	31	7	8
3	Nebelong BE	30	7	8
4	邱均平	23	5	5
5	马凤	26	4	5
6	Cronin B	20	4	5
7	Bar-Ilan J	11	3	3
8	Garfield E	10	3	3
9	郑德俊	16	2	2
10	屈宝强	15	2	2
11	周春雷	9	2	2
12	Katherine WM	6	2	2
13	Rousseau R	6	2	2
14	李晓辉	6	2	2
15	苏新宁	6	2	2

根据前文约定,提名者(施引者)资格为本领域硕博论文作者或在本领域内被引1次及以上,表2中有被引荐记录的学者被这些有资格提名的作者至少引用了2次。由表2可以看出,国内外知名学者的被引用次数和被引荐次数都很高,如WhiteHD、武夷山、NebelongBE、邱均平等,他们被视为引用认同领域的重要学者。引荐分析法是基于引用认同的评估方法,能够减弱自引以及来自特定施引人的干扰,从而更加合理地反映研究者的学术影响力。

通过对引用数据的分析,不仅能发掘国内外领域专家,还可以看出研究者在引用过程中体现出的如下特点:(1)从引用频次来看,总体符合集中-离散分布规律。大量作者只被引用一次,只有少数作者被多次引用,而被同一施引者多次引用的作者更是稀少。(2)自引是引用认同的重要组成部分,自引的频次往往比引用他人的次数要高出很多。引荐分析法将作者对自己的学术授信投票处理为1次引荐,这就大大削弱了自引的影响。

(3)作者在引用过程中会受到其所处社会网络的潜在影响,通常会多次引用熟人。例如,武夷山和马凤是师生合作关系,邱均平与周春雷亦如此。(4)国内学者的引用认同主要为国内学者,对国外学者的关注仅限于少

数知名学者。这也反映出引用认同分析与引荐分析在揭示信息方面的差异。

2.3 作者关系分析

引荐分析不仅能从被引者角度发现业界主流学者，还可以从施引者与被引者之间的关系，以及施引者之间的互引关系挖掘引用认同研究者的引用关系网络。2010年，苏芳荔、孙建军^[7]认为，引用认同从严格意义上来说只包括作者自己真正引用过的作者，多人合著文章的参考文献并不一定被每位合著者所认同。笔者认同这一看法，故本文只分析第一作者间的引用关系。



图1 作者引用关系可视化

利用Gephi软件可得到作者引用关系图，见图1。图中，作者节点标签的大小是其被引次数多少的体现，即被同行引荐的次数越多作者标签越大。节点标签较大的White HD、Nebelong BE、Cronin B、武夷山、邱均平、马凤、屈宝强等是引用认同研究领域的核心作者，是得到国内研究者们以引荐方式认可的领域内学者。图中的核心作者可以将引用关系自然分割为以他们为中心的引荐作者群。同时，图中连线是作者引用关系的体现，连线的粗细是引用频次的体现，即连线越粗表示对该作者的引用越频繁。从图中我们还能看出，向国内学术界引荐国外学者的主要研究者有武夷山、马凤、屈宝强、邱均平、苏芳荔、宿瑞芳等，这有助于找出领域内对外交流的核心学者。

3 关键词分析

3.1 被引文献高频关键词统计

关键词集合可以揭示出研究领域的总体内容特征，并反映出它们之间的相互联系。笔者对国内引用认同研究文献所引关键词的分析，是基于引荐分析视角对研究者关注主题的分析，能够更加客观真实地描绘出引用认同研究的发展脉络与方向。通过对659条引文（除去72本专著和25条网页信息）补充关键词，并对关键词进行规范化处理，共得到2369个关键词，表3是被引文献中出现频次大于20的关键词汇总。为避免中英文关键词重复统计，英文关键词已经预先翻译为中文。对于未明确标记关键词的外文文献，笔者在Web of Science中检索引用该论文的文献，在统计出高频关键词的基础上得出该文献的最终关键词。

表3 被引文献高频关键词汇总（前20）

关键词	频次	关键词	频次
引文分析	101	Ortega假说	32
引用认同	81	开发研究	30
自引	57	科学计量学	30
模型	49	实证分析	27
文献计量学	49	文献计量	27
情报科学	40	施引者	27
期刊	37	科学	26
参考文献	34	信息科学	25
动机	34	社会学	22
分类	34	出版物	21

被引文献中出现频次较高的关键词为核心关键词，指一定时期内研究人员关注较多且较能体现学科特征的关键词。表3列出了词频较高的Top 20关键词。引文分析、引用认同、自引、模型、文献计量学和情报科学出现频次较高，是与引用认同研究关系密切的词汇；期刊、期刊评价、期刊引用认同的高频出现率也体现出相关学者对期刊引用认同研究的关注；此外，研究者对于引用认同的研究多采用实证分析法，对作者引用动机的分析较为关注；同时，在研究引用认同时还会涉及其他一些问题，如对作者社会网络关系的分析、知识地图的构建等。

4.2 作者引用关键词分析

作者对关键词的引用是对作者关注主题的集中反

