

基于文献计量学的 中国社会学学科半衰期分析

童国平

[摘要]半衰期是衡量文献老化的科学计量指标,通过对社会学学科半衰期的研究可以揭示本学科所处发展阶段和学科特性。本文按照中图分类号获取 CNKI 收录的社会学的相关文献,通过计算社会学各个研究领域的引用半衰期和被引半衰期,对比和分析了我国社会学下各研究领域的文献老化情况,以此揭示我国社会学研究的学科特性和文献老化的影响因素。

[关键词]学科半衰期; 引用半衰期; 被引半衰期; 文献老化; 社会学

中图分类号: G250.252

文献标识码: A

文章编号: 1004-3926(2014)11-0233-05

作者简介:童国平(1990-),男,浙江衢州人,南京大学信息管理学院,研究方向:数据挖掘和自然语言处理。江苏南京 210093

引言

半衰期是研究文献老化现象的科学计量指标,对文献的半衰期研究一直是科学计量界的重要研究内容。文献半衰期最早可以追溯到 1943 年戈斯内尔(C. F. Gosnell)的一篇博士论文,他提出用“文献老化(Obsolescence)”来表示文献资料的随着时间推移逐渐失去价值的过程^[1-2]。英国学者贝尔纳(J. D. Bernal)在 1958 年华盛顿召开的一次情报科学讨论会上借用物理学中的放射性物质的衰变现象,提出使用“半衰期”来衡量文献的老化程度^[3]。随后,一些学者开始对半衰期的定量计算进行了研究,最具代表性的有 1960 年巴尔顿(R. E. Burton)和凯普勒(R. W. Kebler)提出的文献老化方程和用其测定文献老化“半衰期”的方法^[4]; 1970 年布鲁克斯(B. C. Brookes)提出用一个负指数函数来作为文献老化方程,以此估算半衰期^[5]。

国内文献半衰期研究从 20 世纪 80 年代才开始逐渐兴起,王洵在 1980 年对半衰期的概念以及国外相关研究进行了综述^[6]。匡兴华提出了研究文献老化的三种方法:第一种是基于引文分析,第二种是基于情报中心流通的资料或图书馆内利用的资料,第三种是综合利用前两个方法^[7]。罗式胜对文献老化概念进行了进一步的探讨,并对文献使用价值和情报价值的关系进行了说明^[8]。金武刚等人通过分析我国社会科学论文的作者和引文,算出社会学引用半衰期为 5 年^[9]。白云对我国人文社会科学期刊(2000-2004)被引半衰期进

行了研究,其中社会学被引半衰期约为 3.3 年^[10]。之后,白云对社会学期刊的学术影响力进行了分析^[11]。近年来,半衰期的理论和方法被应用到网络信息资源上。马费成以社会书签网站为例对 Web2.0 信息半衰期的影响因素进行实证研究^[12]。高金菲使用半衰期的理论和计算方法对中文维基百科的演化进行分析^[13]。

国内对于社会学的半衰期研究比较少,尤其是很少有学者从文献老化和半衰期角度来研究社会学的整体情况。本文以我国社会学领域文献为对象,以学科分类和中图分类为划分依据,通过计算整个学科、各子学科、各研究领域的引用半衰期和被引半衰期,分析我国社会学领域的文献老化现象。

一、数据来源及指标说明

(一) 数据来源

为了研究社会学引用半衰期,本文从 CNKI 数据库利用中图分类号查询了社会学 2010-2012 年的所有文献,包括《中国学术期刊网络出版总库》、《中国博士学位论文全文数据库》、《中国优秀硕士学位论文全文数据库》、《国内会议》、《国际会议》、《报纸》等多个数据库,共计 57443 篇文献。在社会学被引半衰期的数据收集上,本文从 CNKI 的《中国引文数据库》查询了社会学的所有文献及其被引用情况。这里需要说明的是,《中国引文数据库》中的被引数据来自 CNKI 其它数据库的统计数据,也就是说由于 CNKI 没有包含书籍和外文文献,则《中国

引文数据库》缺少这部分的数据。

(二) 学科引用半衰期概念及计算

某学科某年的引用半衰期是指该学科在该年引用的所有参考文献中较新的一半是在最近多长时间发表的。其具体计算方法: 首先找到该学科该年的所有发表文献, 然后将其参考文献按出版年份从近到远排序, 其次计算各年引用百分比和引用累积百分比(以计算当年作为第0年), 最后根据计算公式(1)计算引用半衰期。

$$H_2 = A + \frac{(50\% - B)}{C} \quad (1)$$

表1 2012年社会学文献的所有参考文献的年代分布

年份	2012	2011	2010	2009	2008	2007	2006	2005	<2005
引用次数	1084	6730	9304	8559	8387	8238	7887	7159	54659
各年引用百分比	1.0%	6.0%	8.3%	7.6%	7.5%	7.4%	7.0%	6.4%	48.8%
引用累积百分比	1.0%	7.0%	15.3%	22.9%	30.4%	37.8%	44.8%	51.2%	100.0%

(二) 学科被引半衰期的概念及计算

某学科某年的被引半衰期是指该学科在该年被引用的所有文献中较新的一半是在最近多长时间发表的。其具体计算方法和引用半衰期很像, 首先找到该学科在该年所有被引用的文献, 然后将这些文献按出版年份从近到远排序, 接着计算各年被引百分比和被引累积百分比(以计算当年作为第0年), 最后根据公式(2)计算被引半衰期。

$$H_1 = A + \frac{(50\% - B)}{C} \quad (2)$$

表2 2012年的社会学文献的被引用情况

年份	2012	2011	2010	2009	2008	2007	2006	2005	<2005
被引次数	894	6736	9530	9175	8015	9416	7437	6824	23769
各年被引百分比	1.1%	8.2%	11.7%	11.2%	9.8%	11.5%	9.1%	8.3%	29.1%
被引累积百分比	1.1%	9.3%	21.0%	32.2%	42.0%	53.5%	62.6%	70.9%	100.0%

二、社会学引用半衰期分析

社会学(一级学科, 学科代码0303)有四个二级学科: 社会学(学科代码030301)、人口学(学科代码030302)、人类学(学科代码030303)、民俗学(学科代码030304)。其中社会学和人口学占主要部分。统计社会学及其子学科2010-2012年的引用半衰期, 发现社会学及其子学科的引用半衰期都在缓慢增加(见表3)。社会学和人口学的引用半衰期基本相等, 约为6.5。

为了探讨社会学各个研究领域的引用半衰期

其中, A是指不大于50%的最大引用累计百分比所在年份距离统计年份的年数, B是指不大于50%的最大引用累计百分比, C是指引用累计百分比第一次大于50%的那年的当年引用百分比。

例如2012年社会学文献的所有参考文献的年代分布如表1所示。则不大于50%的最大引用累计百分比所在年份是2006年, 即A=6, B=44.8%, C=6.4%。那么社会学学科的2012年的引用半衰期为: $6 + (50\% - 44.8\%) / 6.4\% = 6.81$ (年)。

其中, A是指不大于50%的最大被引累计百分比所在年份距离统计年份的年数, B是指不大于50%的最大被引累计百分比, C是指被引累计百分比第一次大于50%的那年的当年被引百分比。

例如2012年的社会学文献的被引用情况如表2所示。则不大于50%的最大被引累计百分比所在年份是2008年, 即A=4, B=42.0%, C=11.5%。那么社会学学科的2012年的被引半衰期为: $4 + (50\% - 42.0\%) / 11.5\% = 4.70$ (年)。

情况, 文章以中图分类为依据, 对社会学各个研究领域的引用半衰期进行分析。表3详细列举了社会学各个研究领域从2010-2012年的引用半衰期情况。一些研究领域由于发文及引文较少, 因而引用半衰期的波动较大。如“国际社会生活与社会问题”领域的发文量大, 但引文很少(近三年发文6222篇, 引文379篇); “社会发展和变迁”、“社会调查和社会分析”领域的发文量小(近三年发文量分别为55, 226篇)。除此以外, 其余社会学各研究领域引用半衰期基本呈现上升趋势。

表3 2010-2012年的社会学各个研究领域及其子学科的引用半衰期和发文量

中图分类号	研究领域/学科名称	2010		2011		2012		三年发文总量
		半衰期	发文量	半衰期	发文量	半衰期	发文量	
C91-0	社会学理论与方法论	8.25	258	9.46	287	10.25	302	847
C911	社会发展和变迁	4.58	11	8.50	9	9.43	35	55

C912	社会结构和社会关系	6.29	3850	6.81	3654	7.36	2989	10493
C913	社会生活与社会问题	6.21	9554	6.72	8148	7.23	8421	26123
C914	社会利益	5.65	110	6.24	104	6.55	87	301
C915	社会调查和社会分析	4.94	67	7.11	60	5.55	99	226
C916	社会工作、社会管理、社会规划	4.64	1104	4.84	1248	5.06	1443	3795
C92	人口学	5.91	2708	6.17	2930	6.61	3388	9026
D57	国际社会保障与社会福利	8.63	75	7.77	58	7.59	79	212
D58	国际社会生活与社会问题	6.37	1881	6.85	1923	0.91	2418	6222
/	其它	9.04	37	8.75	18	10.60	88	143
/	社会学	6.06	16947	6.58	15509	6.86	15961	48417
/	人口学	5.91	2708	6.17	2930	6.61	3388	9026
/	社会学(所有领域)	6.03	19655	6.49	18439	6.81	19349	57443

发文量前三的研究领域分别是“社会生活与社会问题”(三年发文 26123 篇)、“社会结构和社会关系”(三年发文 10493 篇)、“人口学”(三年发文 9026 篇),占本学科发文的 79%,它们对社会学引用半衰期的贡献最大。而且这三个领域的引用半衰期都在增加(增幅分别约为 8%、9%、6%),故而社会学引用半衰期是增长的。中国社会学自 19 世纪末 20 世纪初从西方传入中国后至今已发展了百余年,学科的核心理论知识发展相对比较成熟,经典文献众多,文献增长缓慢(最近几年基本没有增长),故而引用半衰期变长。

对比引用半衰期较长的和较短的领域,发现引用半衰期较长的社会学研究领域多偏向理论研究,往往需要借鉴成熟的理论和研究成果,其引文很多都是很老的经典文献。而引用半衰期较短的社会学研究领域多偏向实证(或调查)研究,这类文献经常关注当前社会热点问题,引用的文献普遍较新。引用半衰期较长的社会学研究领域有“社会学理论与方法论”(三年平均 9.32 年)、“国际社会保障与社会福利”(三年平均 8.00 年);引用半衰期较短的社会学研究领域有“国际社会生活与社会问题”(三年平均 4.71 年)、“社会调查和社会分析”(三年平均 5.87 年)。两者相差 3~4 年。除此之外,“国际社会生活与社会问题”、“社会调查

和社会分析”的引用半衰期较短是因为它们的学位论文比重很小,分别只占本领域发文的 0% 和 6%,明显低于平均的 9.9%(2012 年的数据),而学位论文的引用半衰期会比期刊论文的半衰期长。

社会学 2012 年引用的文献中有 56% 的中文期刊论文、31% 的外文文献、7% 的中文图书、4% 的中文学位论文。统计引文最多的 1000 篇文献(占总文献的 5%),总引文数达到 52653 篇(占有引文的 46%)。统计引文少于 10 篇的文献(有 16069 篇,占总文献的 83%),其引文只有 21443 篇(占有引文的 19%)

三、社会学被引半衰期分析

社会学各个研究领域及其子学科 2010 - 2012 年的被引半衰期和被引情况见表 4。由此可见,社会学被引半衰期在逐年增加,增幅达到 7.6%,说明社会学学科的影响力正在不断加强。社会学两个子学科的被引半衰期也在缓慢增加。社会学各个研究领域除了“国际社会生活与社会问题”外,其余领域的被引半衰期也在增加。“国际社会生活与社会问题”的被引半衰期的波动较大是因为其发文量大,但被引次数很少(近三年发文 6222 篇,被引 954 次)。

表 4 2010 - 2012 年的社会学各个研究领域及其子学科的被引半衰期和被引次数

中图分类号	研究领域/学科名称	2010		2011		2012		三年被引总量
		半衰期	被引量	半衰期	被引量	半衰期	被引量	
C91-0	社会学理论与方法论	3.92	1432	4.05	1716	4.25	2048	5196
C911	社会发展和变迁	4.01	191	4.88	197	5.29	204	592
C912	社会结构和社会关系	4.33	19501	4.66	23963	5.09	26284	69748
C913	社会生活与社会问题	3.71	17648	3.87	22279	4.34	25601	65528
C914	社会利益	3.12	789	3.47	761	4.12	780	2330
C915	社会调查和社会分析	4.83	457	4.78	544	5.04	577	1578
C916	社会工作、社会管理、社会规划	3.97	5018	4.27	6080	4.01	8912	20010
C92	人口学	4.26	11390	4.5	12968	4.77	14545	38903
D57	国际社会保障与社会福利	3.43	549	4.27	648	5.08	720	1917

D58	国际社会生活与社会问题	5.03	308	6.06	306	4.94	340	954
/	其它	6.98	1582	8.02	1668	9.12	1785	5035
/	社会学	4.05	47475	4.36	58162	4.68	67251	172888
/	人口学	4.26	11390	4.5	12968	4.5	14545	38903
/	社会学(所有领域)	4.08	58865	4.39	71130	4.70	81796	211791

社会学和人口学的被引半衰期比较接近,只相差0.21年。社会学各研究领域,被引半衰期较长的是“国际社会生活与社会问题”(三年平均5.34年)、“社会调查和社会分析”(三年平均4.88年)。被引半衰期较短的是“社会利益”(三年平均3.57年)、“社会生活和社会问题”(三年平均3.97年)、“社会工作、社会管理、社会规划”(三年平均4.08年)。两者被引半衰期相差不大,在1~2年。“国际社会生活与社会问题”的被引半衰期最长,是因为它的文献大都是评论性的短文,其被引次数很少,而那些存在被引次数的文献几乎都是优秀的文献,被引时间较长,故其被引半衰期最长。“社会调查和社会分析”中很多文献都是实证文献,文中涉及了很多模型和方法,而这些模型和方法经常被其它学科借鉴,由于学科之间的知识交流速度比本学科间知识交流的慢,故其被引半衰期较长。“社会利益”、“社会生活和社会问题”、“社会工作、社会管理、社会规划”这些领域的文献都是研究具体的社会问题,它们不仅仅需要理论上的研究,更多需要的是依据现实情况来研究特定的主题。由于现实问题变化很快,即研究

主题转变很快,导致文献老化速度很快,被引半衰期较短。

被引次数前三的研究领域是“社会结构和社会关系”(三年被引69748次)、“社会生活与社会问题”(三年被引65528次)、“人口学”(三年被引38903次),占所有领域的82%。它们对社会学被引半衰期的贡献最大。而且这三个研究领域的被引半衰期都在增加(增幅分别约为9%、8%、6%),故而社会学被引半衰期是增长的。

四、学科引用半衰期与被引半衰期的对比

学科引用半衰期反映的是该学科的文献对其相关领域的不同年龄的文献的感兴趣程度,即它衡量的是该学科相关领域的文献老化速度;而学科被引半衰期反映的是该学科的不同年龄的文献吸引其相关领域文献的能力,即它衡量的是该学科自身的文献老化速度。为了更好地对比社会学各个研究领域引用半衰期和被引半衰期,表5列举了2010-2012年社会学各个研究领域及其子学科的引用半衰期和被引半衰期。

表5 2010-2012年社会学各个研究领域及其子学科的引用半衰期和被引半衰期

中图分类号	研究领域	2010		2011		2012	
		引用半衰期	被引半衰期	引用半衰期	被引半衰期	引用半衰期	被引半衰期
C91-0	社会学理论与方法论	8.25	3.92	9.46	4.05	10.25	4.25
C911	社会发展和变迁	4.58	4.01	8.50	4.88	9.43	5.29
C912	社会结构和社会关系	6.29	4.33	6.81	4.66	7.36	5.09
C913	社会生活与社会问题	6.21	3.71	6.72	3.87	7.23	4.34
C914	社会利益	5.65	3.12	6.24	3.47	6.55	4.12
C915	社会调查和社会分析	4.94	4.83	7.11	4.78	5.55	5.04
C916	社会工作、社会管理、社会规划	4.64	3.97	4.84	4.27	5.06	4.01
C92	人口学	5.91	5.03	6.17	6.06	6.61	4.94
D57	国际社会保障与社会福利	8.63	3.43	7.77	4.27	7.59	5.08
D58	国际社会生活与社会问题	6.37	5.03	6.85	6.06	0.91	4.94
/	其它	9.04	6.98	8.75	8.02	10.60	9.12
/	社会学	6.06	4.05	6.58	4.36	6.86	4.68
/	人口学	5.91	4.26	6.17	4.5	6.61	4.5
/	社会学(所有领域)	6.03	4.08	6.49	4.39	6.81	4.70

由表5可知,社会学及其子学科、社会学各研究领域的引用半衰期均要大于被引半衰期。产生这个现象的一个重要原因是CNKI数据库的局限

性,CNKI的被引库中没有书籍和外文文献,而在引用文献中(即参考文献里)却有很多书籍和外文引文。书籍和外文文献的老化速度比一般中文期

刊论文的要慢,即参考文献中的书籍和外文引文无疑会增大引用半衰期。在剔除这些书籍和外文引文之后,引用半衰期和引半衰期就比较接近了,见图1。但是引用半衰期和引半衰期并不相等,因为除了最常见的也是最多的学科内部引用之外还存在学科之间的引用(包括其它学科引用本学科和本学科引用其它学科),故而引用半衰期和引半衰期会比较接近但不相等。

观察表5,发现引用半衰期波动较大的是“社会发展和变迁”、“国际社会生活与社会问题”。前一个因为文献量太少所以波动较大;后一个因为很多都是评论性短文,引文很少,所以引用半衰期波动也很大。相比之下,社会学各个研究领域的引半衰期波动就小很多。对比社会学各个研究领域的引用半衰期和引半衰期大小,发现“社会学理论与方法论”的引用半衰期很长,但引半衰期较短。产生这个现象是因为其引文中有19%的中文书籍和46%的外文文献引文(2012年的引文情况),这些引文都比较老,故其引用半衰期很长。而在CNKI引半库中没有书籍和外文文献数据,所以其引半衰期较短。

对比每个研究领域的引用半衰期和引半衰期,发现引用半衰期并不和引半衰期成正相关。即引用半衰期越长,引半衰期并不一定越长。如前面提到的“社会学理论与方法论”,其引用半衰期很长,但引半衰期较短。再如“国际社会生活与社会问题”的引用半衰期很短,但引半衰期较长。之所以会出现这种研究领域的引用半衰期和引半衰期的不一致性现象,是因为该领域与其它相关领域之间存在大量的引用情况,而且再加上CNKI的引半库缺失了部分数据的原因。

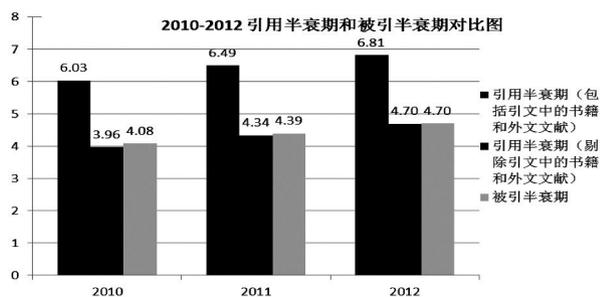


图1 引用半衰期和引半衰期对比图

结语

本文按照中图分类号从CNKI数据库中搜索

出社会学各个研究领域的文献及其引文和引证文献。通过计算得到社会学引用半衰期约为6.5年,社会学引半衰期约为4.5年。社会学各研究领域的引用半衰期均大于引半衰期。对比最近几年的半衰期,发现社会学及其子学科、各研究领域的引用半衰期和引半衰期都在缓慢增加。比较长半衰期领域和短半衰期领域,发现偏向理论的研究领域的引用半衰期和引半衰期都较长,这些领域的知识更新比较慢,文献引用的都是成熟的理论和方法,相应的知识理论也需要经时间检验才能被其它文献引用。偏向实证的或研究社会热点的研究领域的引用半衰期和引半衰期都较短。这类文献更新速度快,知识、主题变化频繁。此外,学科间的引用也会减缓文献老化的速度。

参考文献:

- [1] C. F. Gosnell. The Rate of Obsolescence in College Library Book Collections as Determined by an Analysis of Three Select Lists of Books for College Libraries [D]. New York University, 1943.
- [2] 俞培果. 文献老化现象研究的历史评说 [J]. 中国图书馆学报, 1997(4).
- [3] 郭红梅, 邵艳娥, 何钦成. 文献半衰期研究现状 [J]. 医学信息学杂志, 2011(1).
- [4] R. E. Burton, R. W. Kebler. The Half-life of some Scientific and Technical Literatures [J]. American Documentation, 1960, 11(1): 18-22.
- [5] B. C. Brookes. The Growth, Utility and Obsolescence of Scientific Periodical Literature [J]. Journal of Documentation, 1970, 26(4): 283-294.
- [6] 王洵. 科技文献的“半衰期” [J]. 情报科学, 1980(4).
- [7] 匡兴华. 科学文献老化的研究 [J]. 情报学刊, 1983(4).
- [8] 罗式胜. 关于文献老化概念的讨论 [J]. 情报科学, 1991(6).
- [9] 金武刚, 詹华清. 中国社会科学论文的作者和引文分析 [J]. 社会科学管理与评论, 2001(3).
- [10] 白云. 中国人文社会科学期刊引半衰期分析研究 [J]. 云南师范大学学报(哲学社会科学版), 2006(4).
- [11] 白云. 社会学学术期刊学术影响力分析——基于CSSCI (2004-2006) [J]. 学海, 2008(4).
- [12] 马费成, 高静. Web2.0 信息半衰期影响因素实证研究——以社会书签网站为例 [J]. 情报理论与实践, 2010(11).
- [13] 高金菲, 梁晗晗. 基于半衰期的维基百科演化分析 [J]. 图书情报与工作, 2011(5).

收稿日期 2014-07-17 责任编辑 吴俊