

引文曲线的分析框架研究

——以诺贝尔奖得主的引文曲线为例^{*}

李 江 姜明利 李玥婷

摘 要 基于 341 位诺贝尔物理学、化学、生理学或医学、经济学奖获得者的引文曲线,借助曲线拟合方法构建引文曲线的分析框架,包括两种规则引文曲线——经典引文曲线、指数增长引文曲线,和三种不规则引文曲线——睡美人引文曲线、双峰引文曲线、波型引文曲线。以诺贝尔奖得主为例的实证分析表明:①引文曲线的分析框架是一种新的引文分析视角;②引文曲线的分析框架可用于将一组(有一定影响力的)论文或作者划分层次,也可用于分析不同学科论文或作者的引用差异。文章最后分析了该框架的适用性。图 4。表 1。参考文献 32。

关键词 引文曲线 诺贝尔奖 对数正态 睡美人 分析框架

分类号 G354.2

Citation Curves of Nobel Laureates

Li Jiang, Jiang Mingli & Li Yueting

ABSTRACT Applying curve fitting and sign test, this paper constructs a framework of citation curves based on citation curves of 341 Nobel laureates in the fields of physics, chemistry, physiology or medicine and economics. The framework involves lognormal and exponential curves as regular curves, and sleeping-beauty, double-peaks and wave-form as irregular ones. The empirical analysis shows that the framework, as a new perspective to citation analysis, can be used to group influential papers or authors, and to analyze the differences of citation behaviors in different disciplines. 4 figs. 1 tab. 32 refs.

KEY WORDS Citation curve. Nobel Prize. Lognormal function. Sleeping-beauty. Analysis framework.

1 介绍

1.1 引文曲线的类型

一般而言,一篇学术论文在发表之后的几年时间内会被其他论文引用,并逐渐达到被引次数高峰,之后被引次数慢慢减少,直到被人遗忘。被引次数反映在时间上的曲线被称为“引文曲线”(Citation curve)^[1],或者“引文模式”(citation pattern)^[2]、“引文历史”(citation history)^[3]、“引文

生命周期”(citation life cycle)^[4]。符合上述有生命周期特征的引文曲线被称为“经典引文曲线”^[5]。引文曲线属于文献老化的研究范畴。从引文的角度研究文献老化有两种方法:历时法(diachronous approach)与其时法(synchronous approach)^[6-7]。前者指定文献(集),统计分析其发表之后各年的被引数据;后者指定时间,统计分析该时间里发表的文献(集)引用过去历年文献的数据。二者之间的差异已被反复比较^[3,8-12]。作为历时法研究的产物,引文曲线从被引用的角度记录文献的老化轨

^{*} 本文系国家自然科学基金项目“跨学科知识扩散的规律研究(编号:71203193)”的研究成果之一。

通讯作者:李江, Email: Li-jiang@zju.edu.cn

迹,可以在一定程度上反映文献的学术影响力^[8]。

1979年,Avramescu在大量数据的基础上总结了五种引文曲线^[13]:①一发表就迅速被广泛认可的论文,引文曲线先单调递增,后单调递减,峰值较高;②受认可程度一般的论文,引文曲线先单调递增,后单调递减,峰值较前者低;③受认可程度较低的论文,引文曲线先单调递增,后单调递减,峰值较低;④天才型的论文,引文曲线一直单调递增;⑤刚开始被认可,但突然被否定并摒弃的论文,引文曲线中存在突变,被引次数突然降至0,该曲线极为罕见,至今尚未在其他文献中被提及。曲线①、②、③的特征与上文描述的“经典引文曲线”一致,具有明显的生命周期特征,而且是引文曲线中最常见的一种。对数正态函数(lognormal function)常用于从数学上描述“经典引文曲线”特征,表达式为(其中 y_0 是偏移量, x_c 指中心, w 指宽度, A 指波幅):

$$y = y_0 + \frac{A}{\sqrt{2\pi}wx} e^{-\frac{[\ln \frac{x}{x_c}]^2}{2w^2}}$$

除对数正态函数以外,“经典引文曲线”也有其他的数学表达式,如负二项式函数、韦布函数、反函数^[14]等,但Egghe和Rao认为,相比之下,对数正态函数的效果最优^[15]。曲线④显示,有些文献发表之后受关注程度逐年增长,引文曲线单调递增。Sangam提出用指数函数(exponential function)描述这种“指数增长引文曲线”^[16],其表达式为(其中 y_0 是偏移量, A 指初始值, R_0 指增长率):

$$y = y_0 + A e^{R_0 x}$$

在20世纪60年代,学者们注意到一种现象——有些论文发表初期很少被引用,但一段时间之后突然被大量引用,这些论文被认为是“不被认同的科学发现”(resisted discoveries)^[17]、“迟到的认同”(delayed recognition)^[18]、“过早的科学发现”(premature discoveries)^[19-20]。2004年, Van Raan将这种现象称为“睡美人(sleeping-beauty)”,并从数学上定义了沉睡深度(c_s)、沉睡长度(s)、唤醒强度(c_w)三个变量,睡美人数量的多少与三个变量之间的关系如下^[21]:

$$N = f\{s, c_s, c_w\} \sim s^{-2.7} \cdot c_s^{2.5} \cdot c_w^{-6.6}$$

之后,“睡美人”现象被广泛关注^[22-25]。符合

这种现象的论文的引文曲线被称为“睡美人引文曲线”。Li和Ye在获诺贝尔奖的论文的引文曲线中发现四个睡美人的特例,并在此基础上提出“全要素睡美人”(all-elements-sleeping-beauty)^[26]。从已经发现的睡美人案例来看,符合睡美人特征的论文的被引次数并不高,但是通常有重要发现。

1.2 引文曲线的影响因素

对于单篇论文而言,对引文曲线影响最大的因素是科学的发展与进步。一篇论文发表之后,逐渐被认可,不久之后因为新的理论与方法的出现而被遗忘,例如“经典引文曲线”;也可能随着科学的发展与进步,其中的理论与方法得到越来越广泛的认可与应用,例如“指数增长引文曲线”;也可能因为理论与方法超前而不被其他人认可,但一段时间之后,受到科学的发展与进步的推动而被广泛认可,例如“睡美人引文曲线”。此外,数据库收录范围的变化也会引文曲线产生影响。例如,SCI/SSCI/A&HCI数据库收录范围近年持续扩大,使得统计被引次数的基数增加,数据库中论文的被引次数上升的概率增大^[27]。对于科学家而言,除了上述影响因素之外,科学家学术成果的产量也会引文曲线产生影响。

获奖(如诺贝尔奖)是否对科学家的引文曲线产生显著影响仍有争议。Hugget选取获诺贝尔奖的科学家在获奖时间的前4年与后5年的引文曲线做比较,发现获奖后的被引次数明显增长,然后,他用这些引文曲线与对应领域其他未获诺贝尔奖的优秀科学家的引文曲线做比较,发现被引次数的增长并不能归因于获诺贝尔奖^[28]。但是,Mazlounian等人通过实证研究发现,诺贝尔奖得主有重大发现的标志性论文被引次数呈爆炸式增长,同时也会带动该科学家其他论文被引次数的增长^[29],这一结论与引用动机理论吻合^[30]。从作者的引用动机来看,引用获诺贝尔奖的科学家通常比引用其他科学家更有说服力,因此,获诺贝尔奖在一定程度上对科学家的被引次数产生正面影响,也因此对科学家的引文曲线产生影响。

除“经典引文曲线”、“指数增长引文曲线”和“睡美人引文曲线”之外,是否还存在其他类型的引文曲线?引文曲线中包含哪些信息?获奖对引文曲线是否产生影响?基于这些研究问题,本文拟

构建引文曲线的分析框架,然后,基于该分析框架研究引文曲线的特征、分布、学科差异、影响因素等,以此阐述该分析框架的理论与应用价值。

2 引文曲线的分析框架

2.1 数据与研究方法

笔者首先从诺贝尔奖官方网站(<http://www.nobelprize.org/>)上获取经济学、物理学、化学、生理学或医学的获奖名单,然后从汤森路透的 Web of Science 数据库中检索每位获奖者发表的期刊论文及其自 1900 年至 2011 年每年的被引次数。经济学奖自 1969 年开始颁发,为了便于比较,笔者只选取自 1969 年以来获得物理学、化学、生理学或医学奖的获得者,共计 341 位科学家。其中,经济学 71 位(共发表 4,592 篇论文)、物理学 94 位(共发表 10,127 篇论文)、化学 83 位(共发表 14,183 篇论文)、生理医学和药学 93 位(共发表 15,184 篇论文)。

科学家的引文曲线是科学家每年所有论文累计被引用的次数形成的曲线。判断一条引文曲线属于哪种类型时,首先用到曲线拟合(curve fitting)的方法。曲线拟合是指选择适当的曲线类型来拟合观测数据,并用拟合的曲线方程分析两变量间的关系,常用软件有 Origin 8 等。曲线拟合的效果通常以 R^2 值给出,该值越趋近于 1,则拟合效果越好。本研究中要求曲线拟合结果中 $R^2 > 0.7$,若不满足该条件,则认为拟合效果较差或拟合失败。借助曲线拟合的方法,可用对数正态函数或指数函数判断科学家的引文曲线是否为“经典引文曲线”或“指数增长引文曲线”。如果一条呈对数正态式的单峰曲线,但仍处于下降阶段,即生命周期尚未结束,曲线拟合不成功,本研究中仍视其为经典引文曲线。其次,从不规则的引文曲线中筛选出睡美人引文曲线。参照 Van Raan 对睡美人现象的数学描述,笔者将“睡美人科学家”定义为:自发表第一篇论文之后,连续沉睡 6 年及以上时间,沉睡期间年均被引次数在 2 以内,唤醒期 4 年,年均被引次数在 20 以上的科学家。最后,从剩余不规则引文曲线中观察其特征,并从中发现新的引文曲线类型。

2.2 分析框架

在 341 位诺贝尔奖得主的引文曲线中,笔者发现“经典引文曲线”、“指数增长引文曲线”和“睡美人引文曲线”均存在,除此之外,还发现另外两种引文曲线——“双峰引文曲线”和“波型曲线”。“双峰引文曲线”存在明显的生命周期,处于衰退期的经典引文曲线在生命周期即将结束时,遇到环境变化而出现第二次生命周期。但是,双峰曲线中两峰的波幅不规则,曲线中存在突变,因此被视为不规则引文曲线。“波型曲线”一般呈上下振动,可能伴随上升或下降的趋势,无生命周期特征。基于此,笔者构建了图 1 所示的引文曲线的分析框架。图 1 中,曲线 a 对应于 Avramescu 的曲线①、②和③,为“经典引文曲线”;曲线 b 对应于 Avramescu 的曲线④,为“指数增长引文曲线”;曲线 c 为“双峰引文曲线”,即两个生命周期;曲线 d 为“波型引文曲线”,即无生命周期;曲线 e 为“睡美人引文曲线”。该引文曲线的分析框架中,曲线 1、2 为规则引文曲线,曲线 3、4、5 为不规则引文曲线。

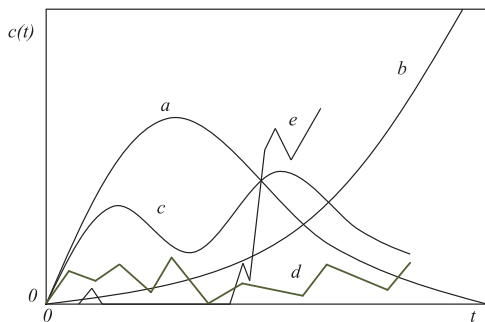


图 1 引文曲线的分析框架

为了验证图 1 中的分析框架不是孤例,笔者分析了 341 位诺贝尔奖得主的 44,086 篇论文的引文曲线,发现该分析框架同样适用,但引文曲线的特点与以科学家为对象的特点完全不同。以单篇论文为对象时,95%的论文年均被引次数小于 10,被引曲线呈不规则的波型或无法绘制引文曲线(被引次数为 0),这使得不规则引文曲线的数量远远超过规则引文曲线的数量。大量的波型引文曲线淹没了其他类型的引文曲线,也淹没了大量有价值的信息,使得统计分析的有效性大大降低。因此,笔

者认为,引文曲线的分析框架适用于有一定影响力的研究对象(单篇论文或单个作者)。获诺贝尔奖的科学家的学术影响力较大,所发表论文的被引次数高于学科平均水平^[31],这是本文选择该数据集的原因。

3 以诺贝尔奖得主为例的实证分析

本节以诺贝尔奖得主为例,通过实证分析阐释上文所述引文曲线分析框架的理论与应用价值。实证分析的数据来源于上文中提到的 341 位诺贝尔奖得主的引文曲线。首先,通过引文曲线的数量分布、被引次数、年龄等指标揭示科学家学术生命

周期的特征;其次,比较引文曲线的学科差异;最后,分析获诺贝尔奖是否对引文曲线产生显著影响。

3.1 引文曲线的分布特征

基于上文中引文曲线的分析框架,341 位获诺贝尔奖的科学家的引文曲线类型统计如表 1 所示。表中也显示了各类引文曲线的平均被引次数与平均年龄,以便揭示引文曲线的特征。平均年均被引次数是一组科学家的年均被引次数的均值。一条引文曲线的年龄是一位科学家自发表第一篇论文起截止到 2011 年的年龄。

表 1 获诺贝尔奖科学家的引文曲线分布

学科	引文曲线	数量	平均年均被引次数	平均年龄
化学	对数正态	44	343.3	24.6
	指数增长	18	486.5	30.8
	睡美人	7	187.6	27.7
	双峰	7	168.1	22.9
	波型	7	185.2	28.0
物理学	对数正态	51	502.5	26.8
	指数增长	14	590.9	24.4
	睡美人	4	487.7	30.8
	双峰	15	377.3	24.9
	波型	9	171.1	19.8
生理学或医学	对数正态	39	270.9	18.3
	指数增长	13	483.8	23.0
	睡美人	7	194.0	33.6
	双峰	17	223.6	20.7
	波型	18	68.1	19.3
经济学	对数正态	4	54.2	38.5
	指数增长	44	122.8	35.8
	睡美人	1	168.4	36.0
	双峰	6	29.1	34.3
	波型	16	11.7	31.7

(1) 规则引文曲线的数量更多

每个学科诺贝尔奖得主的引文曲线中,规则引文曲线(经典引文曲线与指数增长引文曲线)的数量都远远超过不规则引文曲线(睡美人、双峰、波型)的数量,如表1所示。化学、物理学、生理学或医学、经济学中规则引文曲线分别是不规则引文曲线的3.0倍、1.2倍、2.3倍与2.1倍。

(2) 规则引文曲线的被引次数更高

化学、物理学、生理学或医学科学家中,规则引文曲线的被引次数均高于不规则引文曲线的被引次数;经济学家中,只有一条睡美人曲线,无法计算其被引次数均值,但规则引文曲线的平均被引次数为117.1,不规则引文曲线的平均被引次数仅为23.0。

3.2 引文曲线的学科差异

在分析引文曲线的学科差异时,物理学、化学、生理学或医学科学家合并代表获诺贝尔奖的自然科学家,经济学家代表获诺贝尔奖的社会科学家。

(1) 自然科学家学术影响力的生命周期更短

在规则的引文曲线中,经典引文曲线显示出生命周期,而指数增长引文曲线可认为是经典引文曲线中处于增长阶段的部分。在学术年龄相近的前提下,指数增长引文曲线明显比经典引文曲线的生命周期更长。表1数据显示,在自然科学家的引文曲线中,经典引文曲线的数量均最多,且远远超过其他曲线的数量;而在社会科学家的引文曲线中,指数增长引文曲线的数量远远高于其他曲线的数量。因此,可认为自然科学家学术影响力的生命周期比社会科学家更短,社会科学家的研究成果的影响更长远。

(2) 自然科学家中睡美人更多

自然科学家中睡美人的比例分别为8.0%、6.7%、4.0%,而社会科学家中睡美人的比例仅为1.4%。睡美人曲线中公主有一段沉睡期,沉睡期内,极少被引用,这被认为是“过早的科学发现”,未能得到认可,一旦被唤醒之后,状态突变,被引次数短时间内激增,这种现象在自然科学家中较常见,Campanario的研究证实了这一观点^[32]。社会科学家的引文曲线相对平滑,睡美人现象较罕见。

3.3 引文曲线与获诺贝尔奖

笔者认同Hugget的观点,引文曲线中的单调递增阶段不能归因于获奖。根据常理,引文曲线中的单调递减阶段不能归因于获奖。因此,如果能证明引文曲线中的突变(单调性发生变化或被引次数大幅度跳跃)不能归因于获奖,那么,就可以认为获奖没有对引文曲线产生显著影响。本文所构建的引文曲线的分析框架中,指数增长引文曲线是平滑的单调递增曲线,波型引文曲线是无规则的振动,只有经典引文曲线和双峰引文曲线的单调性发生了变化,只有睡美人引文曲线中被引次数出现大幅度跳跃。

基于上述分析,笔者采用符号检验的方法分析引文曲线中突变的时间与获诺贝尔奖时间的关系,进而判断获奖是否对引文曲线产生显著影响。如果引文曲线中的突变发生在获诺贝尔奖之后,那么可以认为突变与获奖有关联,否则不能将突变归因于获奖。符号检验(sign test)是最为常用的非参数假设检验方法,通过两个相关样本的每对数据之差的符号进行检验,以判断两个样本之间的差异是否显著。用“+”、“-”、“0”记录每对数据中前者与后者之差的符号, $N = n_+ + n_-$, $r = \min(n_+, n_-)$,查符号检验表中N值对应的表值,如果 $n_+ >$ 表值并且 $n_- >$ 表值,则两个样本差异不显著;在单边检验中,如果 $n_+ \leq$ 表值,则样本中前者显著高于后者;如果 $n_- \leq$ 表值,则样本中前者显著低于后者。

图2中,左侧76个散点表明76位科学家获诺贝尔奖之后引文曲线才出现峰值,右侧54个散点表明54位科学家在引文曲线出现峰值之后才获奖。符号检验结果:两者无显著差异。也就是说,经典引文曲线中单调性的变化不能归因于获诺贝尔奖。图3中,左侧23个散点表明23位科学家获奖之后才出现第二次生命周期,右侧20个散点表明20位科学家出现第二次生命周期之后才获奖。符号检验结果:两者无显著差异。也就是说,双峰引文曲线中单调性的变化不能归因于获诺贝尔奖。图4中,左侧1个散点表明1位科学家获奖之后才被王子唤醒,右侧17个散点表明17位科学家被王子唤醒之后才获奖。符号检验结果:两者有显著差异。进而用单边符号检验发现,睡美人引文曲线中被引次数的跳跃发生在获诺贝尔奖之前。因此,睡美人引文曲线中的跳跃不能归因于获诺贝尔奖。

基于上述从引文曲线视角的分析可以看出,引文曲线中的各种变化均不能归因于获诺贝尔奖,因此,

可以认为科学家的引文曲线并没有因为获诺贝尔奖而受到显著影响。

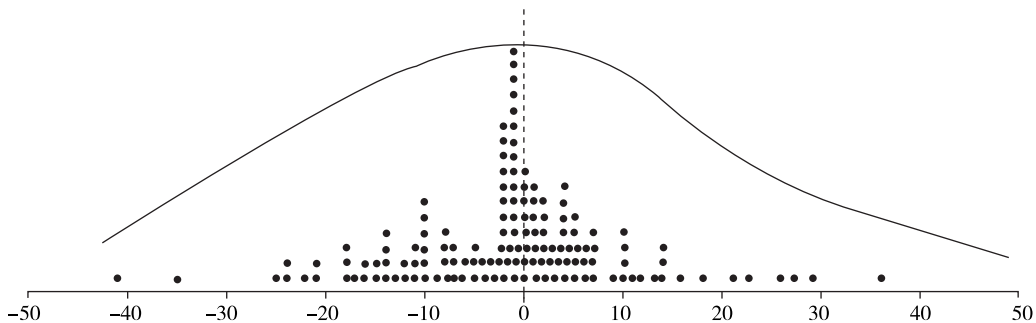


图2 获奖时间在经典引文曲线中峰值时间前后的分布散点图

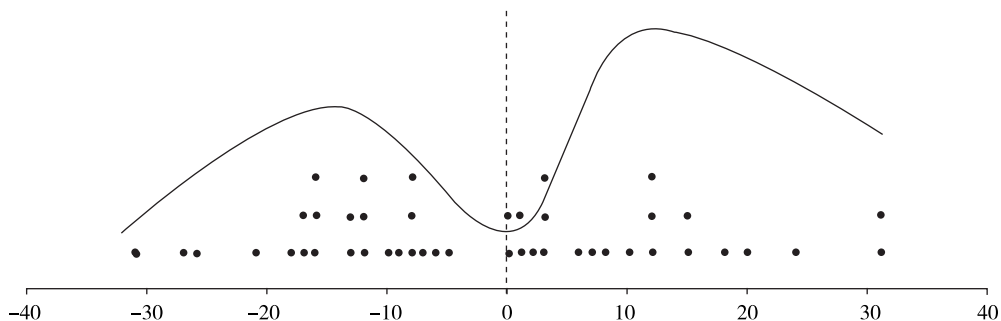


图3 获奖时间在双峰曲线中第二次生命周期起始时间前后的分布散点图

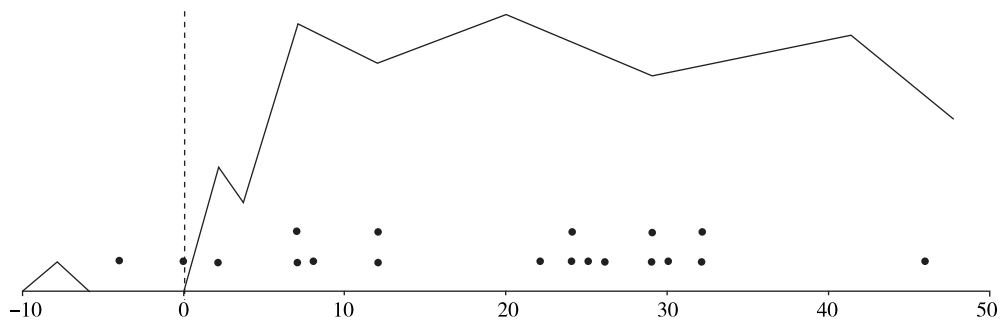


图4 获奖时间在睡美人曲线中王子出现时间前后的分布散点图

4 结论

本文在引文曲线相关研究的基础上,构建了

引文曲线的分析框架,包括两种规则引文曲线——经典引文曲线、指数增长引文曲线,和三种不规则引文曲线——睡美人引文曲线、双峰引文曲线、波型引文曲线。以诺贝尔奖得主为例的实证分析表

明:①引文曲线的分析框架是一种新的引文分析视角。传统引文分析以被引次数为核心,基于被引次数设计测度指标,例如,期刊影响因子、即年指标、 h 指数等,而不考虑引用的时间分布。引文曲线则从引用的时间分布入手,分析引用的历史过程,以此反映论文的影响,例如,睡美人引文曲线通常暗含重要发现,但不能被同时代的学者理解。②引文曲线的分析框架可用于将一组(有一定影响力的)论文或作者划分层次,例如,诺贝尔奖得主中规则引文曲线的数量多,被引次数高,而不规则引文曲线的数量少,被引次数低;也可用于分析不同学科论文(或作者)的引用差异,例如,诺贝尔奖得主中,自然科学家比社会科学家学术影响力的生命周期更短、睡美人现象更多。需要注意的是,引文曲线的类型与被引用对象的质量没有直接关联,所以,通过引文曲线的类型推断被引用对象的质量缺乏科学依据。

本文所构建的引文曲线分析框架具有一定的开放性。分析框架中的5种引文曲线来源于341

位诺贝尔奖得主,5种引文曲线涵盖了各种见诸文献的引文现象,其合理性在341位科学家的44,086篇论文中也得到验证。随着引文曲线研究的深入,新发现的引文现象可以补充到该分析框架中。

本文所构建的引文曲线分析框架适用于从事科研有一定年限、有一定影响的学者。对于从业时间太短的学者而言,例如入行两年的研究人员,其被引次数无法在时间轴上绘制成引文曲线。对于被引次数较低(例如年均被引次数低于10)的学者而言,可以预见其被引曲线呈不规则的波型,不具备研究意义,如果被引次数低至0,则不存在被引曲线。同样,该分析框架也适用于发表有一定年限、有一定影响的论文。该分析框架是否适用于期刊、机构、国家/地区等研究对象尚未得到论证,但不可从论文、作者推而广之。因为论文与作者的生命周期特征明显,而期刊、机构、国家/地区的新知识源源不断,没有明显的生命周期。

致谢:感谢叶鹰教授、蒋弘研究馆员为论文修改提出的宝贵建议。

参考文献

- [1] Garfield E. More delayed recognition. Part 1.Examples from the genetics of color blindness, the entropy of short-term memory, phosphoinositides, and polymer Rheology[J]. Current Contents, 1989, 38: 3-8.
- [2] Aversa E S. Citation patterns of highly cited papers and their relationship to literature aging: A study of the working literature[J].Scientometrics, 1985,7(3-6):383-389.
- [3] Redner S. Citation statistics for more than a century of physical review[J]. Physics Today, 2004, 58(6): 49-54.
- [4] Cano V, Lind N C. Citation life cycles of ten citation classics[J]. Scientometrics, 1991,22(2): 297-312.
- [5] Cunningham S J. An empirical investigation of the obsolescence rate for information systems literature[J/OL]. Library and Information Science Research.1995[2012-07-20]. <http://library.fgcu.edu/iclc/lisrissu.htm>.
- [6] 邱均平. 信息计量学[M]. 武汉:武汉大学出版社,2007:76.(Qiu J P. Informetrics[M]. Wuhan: Wuhan University Press, 2007: 76.)
- [7] Nakamoto H. Synchronous and dyachronous citation distributions[G] // Egghe L, Rousseau R. Informetrics 87/88. Amsterdam: Elsevier Science Publishers,1988: 157-163.
- [8] Line M B, Sandison A. "Obsolescence" and changes in the use of literature with time[J]. Journal of Documentation, 1974, 30(3): 283-350.
- [9] Griffith B C, Servi P N, Anker A, et al. The aging of scientific literature: A citation analysis[J]. Journal of Documentation, 1979, 35(3): 179-196.

- [10] Stinson E R, Lancaster F W. Synchronous versus diachronous methods in the measurement of obsolescence by citation studies[J]. *Journal of Information Science*, 1987, 13(2): 65-74.
- [11] Gupta U. Obsolescence of physics literature; Exponential decrease of the density of citations to physical review articles with age[J]. *Journal of the American Society for Information Science*, 1990, 41(4): 282-287.
- [12] Egghe L. On the influence of growth on obsolescence[J]. *Scientometrics*, 1993, 27(2): 195-214.
- [13] Avramescu A. Actuality and obsolescence of scientific literature[J]. *Journal of the American Society for Information Science*, 1979, 30(5): 296-303.
- [14] Pollmann T. Forgetting and the ageing of scientific publications[J]. *Scientometrics*, 2000, 47(1): 43-54.
- [15] Egghe L, Rao I K R. Citation age data and the obsolescence function: Fits and explanations[J]. *Information and Processing Management*, 1992, 28(2): 201-217.
- [16] Sangam S L. Obsolescence of literature in the field of psychology[J]. *Scientometrics*, 1999, 44(1): 33-46.
- [17] Barher B. Resistance by scientists to scientific discovery[J]. *Science*, 1961, 134: 596-602.
- [18] Cole S. Professional standing and the reception of scientific discoveries[J]. *American Journal of Sociology*, 1970, 76: 286-306.
- [19] Stent G S. Prematurity and uniqueness in scientific discovery[J]. *Scientific American*, 1972, 227(6): 84-93.
- [20] Wyatt H V. Knowledge and prematurity-journey from transformation to DNA[J]. *Perspectives in Biology and Medicine*, 1975, 18(2): 149-156.
- [21] Van Raan A F J. Sleeping beauties in science[J]. *Scientometrics*, 2004, 59: 467-472.
- [22] Glanzel W, Garfield E. The myth of delayed recognition[J]. *Scientist*, 2004, 18(11): 8-9.
- [23] Burrell Q L. Are "Sleeping Beauties" to be expected[J]. *Scientometrics*, 2005, 65(3): 381-389.
- [24] Braun T, Glanzel W, Schubert A. On sleeping beauties, princes and other tales of citation distributions[J]. *Research Evaluation*, 2010, 19(3): 195-202.
- [25] Egghe L, Guns R, Rousseau R. Thoughts on uncitedness: Nobel laureates and fields medalists as case studies[J]. *Journal of the American Society for Information Science and Technology*, 2011, 62(8): 1637-1644.
- [26] Li J, Ye F Y. The phenomenon of all-elements-sleeping-beauties in scientific literature[J]. *Scientometrics*, 2012, 92(3): 795-799.
- [27] Li J, Willett P. Bibliometric analysis of Chinese research on cyclisation, Maldi-Tof and antibiotics[J]. *Journal of Chemical Information and Modeling*, 2010, 50(1): 22-29.
- [28] Hugget S. Does a Nobel Prize lead to more citations[J/OL]. *Research trends*, 2010 (20) [2013-06-18]. <http://www.researchtrends.com/issue20-november-2010/does-a-nobel-prize-lead-to-more-citations>.
- [29] Mazloumian A, Eom Y-H, Helbing D, et al. How citation boosts promote scientific paradigm shifts and Nobel Prizes[J/OL]. *PLoS ONE*, 2011, 6(5) [2013-06-18]. <http://www.plosone.org/article/info%3Adoi%2F10.1371%2Fjournal.pone.0018975>.
- [30] Willett P. Readers' perceptions of authors' citation behavior[J]. *Journal of Documentation*, 2013, 69(1): 145-156.
- [31] 叶鹰. 高品质论文被引数据及其对学术评价的启示[J]. *中国图书馆学报*, 2010, 36(1): 101-104. (Ye Ying. An outline of academic assessment with the citation data of high-quality papers[J]. *Journal of Library Science in China*, 2010, 36(1): 101-104.)

小词条

“睡美人”释义

科学计量学中的“睡美人”是指发表后历经多年低被引而后转为高被引的那些学术论文。“睡美人”术语起源于格林童话中的睡美人故事:美丽的公主受女巫之咒接触纺锤而陷入昏睡,一百年后有位英俊的王子找到了沉睡的公主,用真爱之吻唤醒了睡美人。2004年,荷兰莱顿大学的定量科学家 van Raan 借用睡美人童话把发表后很少引用但若干年后被引却迅速上升的特殊引文现象命名为“睡美人”(sleeping beauty)^[1],明确某些发表后很少被引用(“沉睡”)的论文(“公主”)一直要等到被某篇论文(“王子”)引用(“亲吻”)后才会被大量引用(“苏醒”),并提出了定量标准——“沉睡”期年均被引 ≤ 2 ,“苏醒”后四年内总被引 >20 。

虽然“睡美人”现象早在20世纪70年代就作为“延迟认可”(delayed recognition)现象被认识和研究, van Raan 也不是第一位研究者,但他形象而有趣的命名为科学计量学研究注入了趣味和活力,激发了中外学者进行相关研究^[2-3];Braun、Glänzel 和 Schubert 把“王子”设定为“公主”沉睡后第一次引用、被引次数相对较高、与“公主”共同被引达到一定次数的论文,依此标准,他们发现有些“公主”先后被多位“王子”亲吻,也发现存在一位“王子”同时亲吻多位“公主”的现象^[4];李江和叶鹰合作发现高品质论文中存在“公主”、“纺锤”、“王子”同时出现的“全要素睡美人”(all-elements-sleeping-beauty)^[5]。这些研究均揭示出“睡美人”论文往往是具有原创性发现当时却未被认识、而沉寂若干年后终被学界肯定的重要文献。

科学“睡美人”现象与睡美人童话的类比是科学与文学成功交融的一个范例。

资料来源

- [1] Van Raan A F J. Sleeping beauties in science[J]. Scientometrics, 2004, 59: 467-472.
- [2] 梁立明,林晓锦,钟镇,薛晓舟. 迟滞承认:科学中的睡美人现象——以一篇被迟滞承认的超弦理论论文为例[J]. 自然辩证法通讯,2009,31(1):39-45.
- [3] Burrell Q L. Are “Sleeping Beauties” to be expected [J]. Scientometrics, 2005, 65(3): 381-389.
- [4] Braun T, Glänzel W, Schubert A. On sleeping beauties, princes and other tales of citation distributions [J]. Research Evaluation, 2010, 19(3): 195-202.
- [5] Li J, Ye F Y. The phenomenon of all-elements-sleeping-beauties in scientific literature[J]. Scientometrics, 2012, 92(3): 795-799.

(叶鹰整理)

-
- [32] Campanario J M. Rejecting and resisting Nobel class discoveries: Accounts by Nobel Laureates[J]. Scientometrics, 2009, 81(2): 549-565.

李江 浙江大学信息资源管理系讲师。通讯地址:杭州市余杭塘路866号。邮编:310058。

姜明利 李玥婷 浙江大学信息资源管理系本科生。通讯地址同上。

(收稿日期:2013-06-24;修回日期:2013-09-02)