



天鹅展翅:高品质论文的引文模式探析

□曾继城 张家榕 叶鹰*

摘要 “天鹅”现象是高品质论文中出现的一种特殊引文模式,定义为具有关联意义的成对出现的高被引论文,其中前期发表的具有科学发现意义的是“白天鹅”而后来发表的象征科学突破并引起前期论文引用下降的是“黑天鹅”。该文以诺贝尔物理学奖、化学奖、生理学或医学奖得主的关键论文为样本集,用“天鹅指数”测量了每类中选出的“黑、白天鹅”样例。“天鹅”引文模式结合了科学发现实质和科学引文数据,可望为突破性科学论文提供一个可量化测评的思路和方法。

关键词 高品质论文 引文分析 天鹅

分类号 G250.252

DOI 10.16603/j.issn1002-1027.2019.02.014

1 引言

学术论文的质性判断和量化评价之间的关联一直困扰着学术界^[1],我们知道质性判断和量化评价皆优的论文一般是高品质论文^[2],但我们不知道高品质论文在引文分析中究竟有何可测度性质。最近,本课题组合作提出了把高品质论文集和科学突破论文关联起来的特殊引文模式并命名为“天鹅”^[3],用“天鹅”比喻高品质论文集并把一般高品质论文通称为“白天鹅”,而以“黑天鹅”作为具有科学突破性质的论文且引起了前期发表的“白天鹅”引文曲线改变。

通常,“黑天鹅”代指高度不可能事件^[4],在经济和社会事件中多为贬义词,而在引文分析^[5]中,我们赋予其褒义用法:“黑天鹅”和“白天鹅”被定义为特别的引文组合,“白天鹅”本身即属于高被引论文,而“黑天鹅”的出现会改变“白天鹅”的引文形态,从而把质性的高品质论文与量化的高被引论文关联起来,用引文曲线^[6]与引用计量^[7]揭示这一特殊的引文现象。

“天鹅展翅”,既在科学史背景上展开了高品质科学论文的量化特征,也为纯粹定量的引文分析研究增添了新的乐趣。兹述如下。

2 方法与数据

以“天鹅”代称科学史上质性判断为高品质的论文,因此,“天鹅”首先是具有科学意义的优质论文。为避免争议,我们把自然科学类诺贝尔奖得主的关键论文视为“天鹅”源并作为研究基础。

在高被引的高品质论文中,有的论文发表后引用高涨并导致原有高引论文引用下降,这种高品质论文我们称为“黑天鹅”,而把原有高品质论文称为“白天鹅”。“黑天鹅”和“白天鹅”形成一对特别的引文组合,反映的是高品质论文的特殊引文现象。“黑天鹅”不仅具备高引量化测度,而且具备突破性科学发现性质,因其具有重要科学意义而被授予诺贝尔奖。

为测度黑白“天鹅”,我们选取具有代表性的两只“白天鹅”与独特的“黑天鹅”进行定量比较,建立图1模型,其中 T 为时间轴, C 为引文量, W_1 , W_2 分别表示具有代表性的两只“白天鹅”,而 B 代表“黑天鹅”。

设 T_s 为“黑天鹅”发表年, T_w 为两只“白天鹅”发表年中的极小值 ($TW = \min\{T_1, T_2\}$), C_b 为“黑天鹅”发表5年内的被引总量, C_{wi} ($i=1, 2, \dots$) 是“白天鹅” i 在 T_s 前5年的被引总量,则定义天鹅指数:

$$S_w = \frac{C_b}{C_{w_1} + C_{w_2}} \quad (1)$$

* 通讯作者:叶鹰, ORCID: 0000-0001-9426-934X, 邮箱: yye@nju.edu.cn.

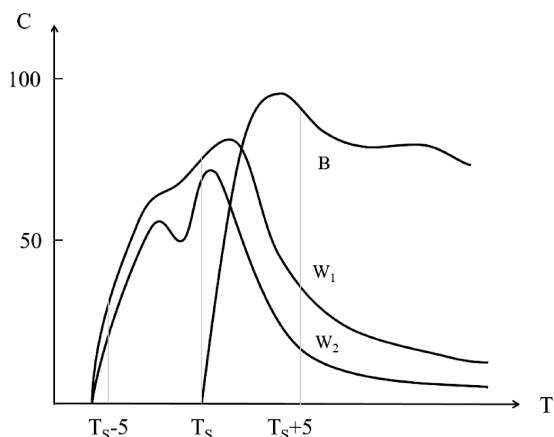


图1 黑白“天鹅”交互作用模型

由此构成的各类天鹅测度参数如表1所示,其中“灰天鹅”是“黑天鹅”的一种“瘦弱型”或非典型形态。

表1 黑白“天鹅”测度参数

类型	Cb	Cw	$T_s - T_w$	S_w
黑天鹅	>100		≥ 5	>0.5
白天鹅		>50		(Cw follows Cb)
灰天鹅			<5	<0.5

为检测上述多变量指标的适用性,以下选取三个不同层面的数据集进行实证研究:

样本一:1950—2010年诺贝尔物理学奖得主关键论文;

样本二:1950—2010年诺贝尔化学奖得主关键论文;

样本三:1950—2010年诺贝尔生理学或医学奖得主关键论文。

诺奖得主的论文向来是科学史和科学计量学研究的重点对象^[8],猜想与数据的契合是立论的重要依据^[9],原始数据来自诺贝尔物理学奖列表(http://www.nobelprize.org/nobel_prizes/physics/laureates/)、诺贝尔化学奖列表(http://www.nobelprize.org/nobel_prizes/chemistry/laureates/)、诺贝尔生理学或医学奖列

表(http://www.nobelprize.org/nobel_prizes/medicine/laureates/),从中分析诺贝尔奖得主获奖演讲中提及的关键论文,筛选并识别出白天鹅、黑天鹅(以及灰天鹅)。在Web of Knowledge平台中检索对应文献,并获取其参考文献。

3 结果与讨论

通过表1测度判断所得自然科学类诺贝尔奖关键论文中的“黑(灰)天鹅”文献总量如表2所示。

表2 自然科学类诺贝尔奖关键论文中的“天鹅”

	关键论文	黑天鹅文献	灰天鹅文献	非天鹅文献	黑-灰天鹅文献占比
物理学	118	16	8	94	20%
化学奖	102	23	18	61	40%
生理或医学奖	120	25	25	70	42%

由此可见“黑(灰)天鹅”文献在自然科学类诺贝尔奖关键论文中的比例不低,故有一定分析意义。

下面选取诺贝尔物理学奖、化学奖、生理或医学奖中比较典型的“黑白天鹅”论文的实例,为同时展示“黑天鹅”和“灰天鹅”,每个学科选“黑天鹅”和“灰天鹅”各一。

3.1 诺贝尔物理学得主关键论文中的天鹅实例

图2左侧的黑天鹅是 Anderson, MH; Ensher, JR; Matthews, MR; Wieman, CE; CORNELL, EA. Observation of Bose-Einstein Condensation in a Dilute Atomic Vapor. Science, 1995, 269(5221):198.。具有代表性的两只白天鹅是 W1: Monroe C, Swann W, Robinson H, et al. Very cold trapped atoms in a vapor cell. Physical Review Letters, 1990, 65(13):1571; W2: Sesko D, Walker T, Monroe C, et al. Collisional losses from a light-force atom trap. Physical Review Letters, 1989, 63(63):961-964.。其天鹅指数为 $S_w = 1065 / (162 + 83) = 4.35 > 1$, $T_s - T_w = 8$ 。这是一只典型的黑天鹅。

表3 天鹅实例数据

学科	天鹅论文	$T_w - T_s$	Cb	Cw1	Cw2	Swan-index	天鹅类型
物理学	2001Cornell	8	1065	162	83	4.35	黑天鹅
物理学	2008Kobayashi	3	341	343	191	0.64	灰天鹅
化学	S2009Ramakrishnan	9	519	95	345	1.17	黑天鹅
化学	S1989Cech	7	177	472	366	0.21	灰天鹅
生理或医学	S2007Capecchi	9	295	351	654	0.29	灰天鹅
生理或医学	S1982Vane2	7	779	229	153	2.04	黑天鹅

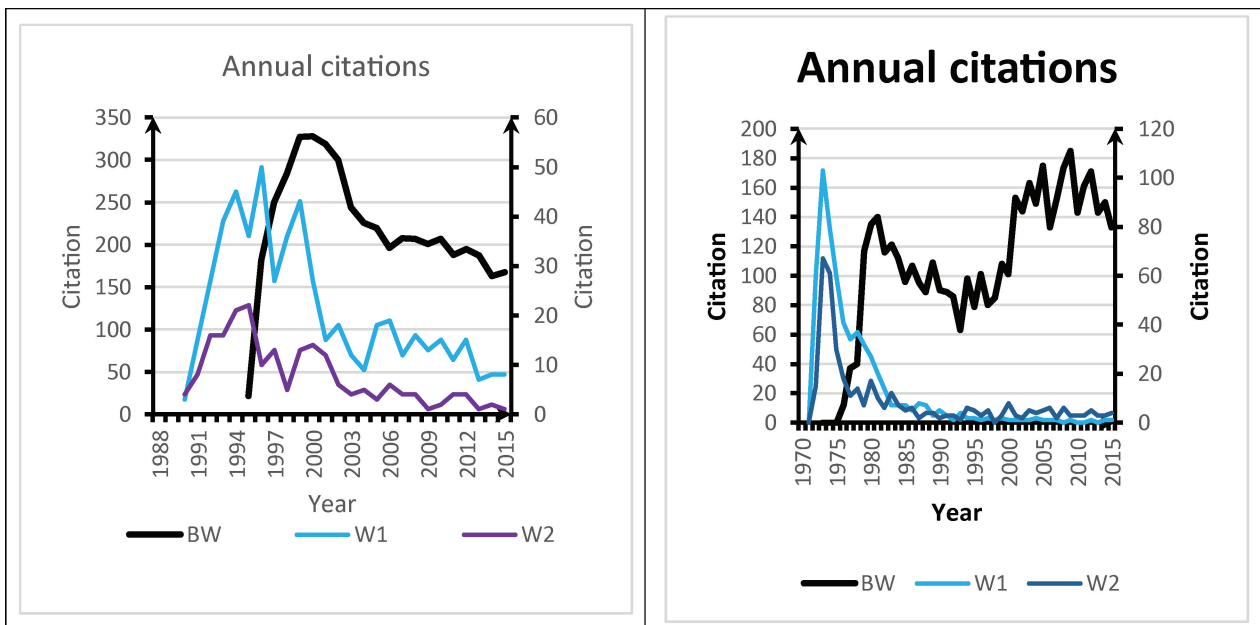


图2 诺贝尔物理学奖中的两例天鹅

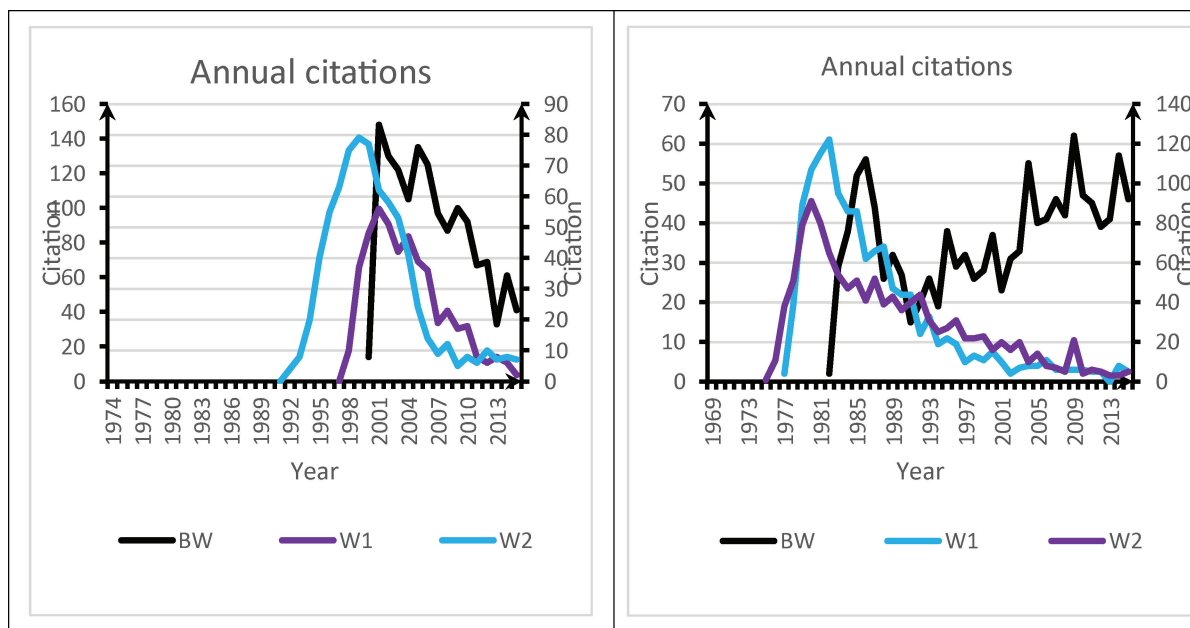


图3 诺贝尔化学奖中的两例天鹅

图2右侧的灰天鹅是 Kobayashi, M., &Maskawa, T. (1989). Cp-violation in the renormalizable theory of weak interaction. Cp Violation, 49(2), 199-204.。具有代表性的两只白天鹅是 W1: Dicus, D. A. (1972). Stellar energy-loss rates in a convergent theory of weak and electromagnetic interactions. Physical Review D, 6 (4), 941-949; W2: Georgi, H., & Glashow, S. L. (1972). Unified weak and electromagnetic interactions without neutral currents. Physical Review Letters, 28

(22), 1494-1497.。其天鹅指数为 $341/(343+191)=0.64 > 0.5$, $T_s - T_w = 3$ 。这是一只灰天鹅。

3.2 诺贝尔化学奖得主关键论文中的天鹅实例

图3左侧的灰天鹅是 Wimberly, BT; Brodersen, DE; Clemons, WM; Morgan-Warren, RJ; Carter, AP; Vonrhein, C; Hartsch, T; Ramakrishnan, V. (2000). Structure of the 30s ribosomal subunit. Nature, 407 (6802), 327.。具有代表性的两只白天鹅是 W1: Jones, T. A., & Kjeldgaard, M. (1997). electron-density map



interpretation. *Methods in Enzymology*, 277, 173–208.; W2: Carson M. RIBBONS 2 0, *Journal of Applied Crystallography*, 1991, 24(5): 958。其天鹅指数为 $Sw = 519 / (95 + 349) = 1.17 > 0.5$, $Ts - Tw = 4$ 。这是一只灰天鹅。

图3右侧的黑天鹅是 Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., & Cech, T. R. (1982). Self-splicing RNA: autoexcision and auto-cyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell*, 31(1), 147–57。具有代表性的两只白天鹅是 W1: Donis-Keller, H., Maxam, A. M., & Gilbert, W. (1977). Mapping adenines, guanines, and pyrimidines in RNA. *Nucleic Acids Research*, 4(8), 2527–38; W2: Burgess, R. R., & Jendrisak, J. J. (1975). A procedure for the rapid, large-scale purification of *Escherichia coli* DNA-dependent RNA polymerase involving Polymin P precipitation and DNA-cellulose chromatography. *Biochemistry*, 14(21), 4634。其天鹅指数为 $Sw = 177 / (472 + 366) = 0.21 < 0.5$, $Ts - Tw = 9$ 。这是一只黑天鹅。

3.3 诺贝尔生理或医学奖得主关键论文中的天鹅实例

图4左侧的灰天鹅是 Thomas, K. R., & Capecchi, M. R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell*, 51(3), 503–12。具有代表性的两只白天鹅是 W1: Orr-Weaver, T. L., Szostak, J. W., & Rothstein, R. J. (1981). Yeast transformation: a model system for the

study of recombination. *Proceedings of the National Academy of Sciences of the United States of America*, 78(10), 6354–8; W2: Hinnen, A., Hicks, J. B., & Fink, G. R. (1978). Transformation of yeast. *Proceedings of the National Academy of Sciences of the United States of America*, 75(4), 1929–33。其天鹅指数为 $Sw = 295 / (351 + 654) = 0.29 < 0.5$, $Ts - Tw = 9$ 。这是一只灰天鹅。

图4右侧的黑天鹅是 Moncada, S., Gryglewski, R., Bunting, S., & Vane, J. R. (1976). An enzyme isolated from arteries transforms prostaglandin endoperoxides to an unstable substance that inhibits platelet aggregation. *Nature*, 263(5579), 663。具有代表性的两只白天鹅是 W1: Piper P J, Vane J R. Release of additional factors in anaphylaxis and its antagonism by anti-inflammatory drugs. *Nature*, 1969, 223(5201): 29–35; W2: Gilmore N, Vane J R, Wyllie J H. Prostaglandins released by the spleen. *Nature*, 1968, 218(218): 1135–1140。其天鹅指数为 $Sw = 779 / (229 + 153) = 2.04 > 1$, $Ts - Tw = 7$ 。这是一只黑天鹅。

黑白天鹅的隐喻为我们从定性与定量角度考察科学研究历史中具有突破性的科学发现提供了分析可能。部分科学研究成果在发表后受到很多抨击与批评, 反映在引文数据上就是获得了大量的负面引用, 但随着时间的推移和科学的进步, 那些超越时代的科学发现终会被学术界所认可, 甚或被授予权威的

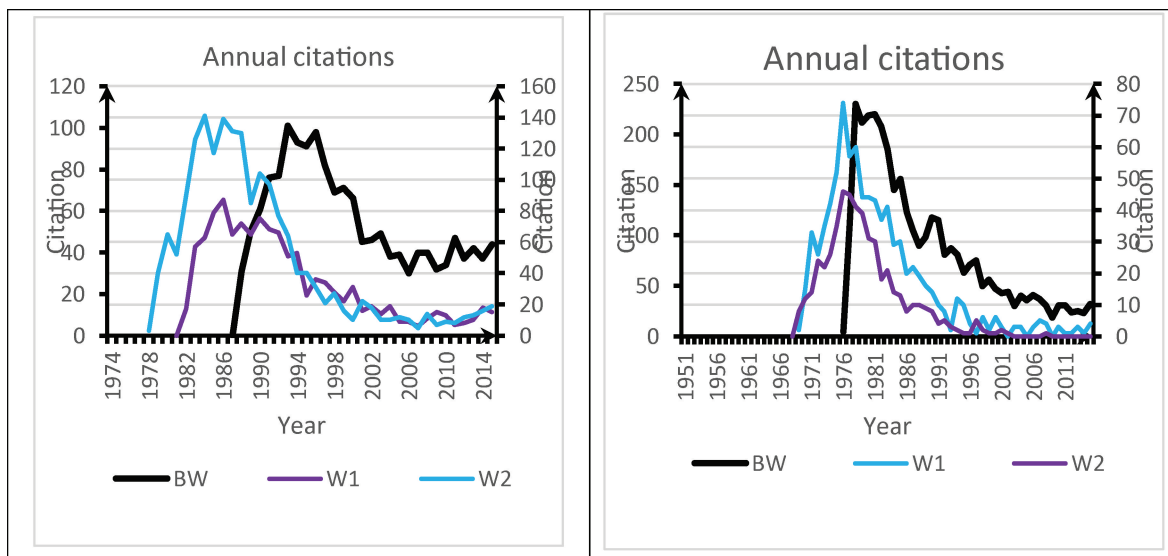


图4 诺贝尔生理或医学奖中的两例天鹅



科研奖项(比如诺贝尔奖),并开辟了新的研究方向,论文因而受到学术界的正面关注和引用。例如,获得2011年诺贝尔化学奖的谢赫特曼(Daniel Shechtman):1982年4月,作为以色列理工学院材料科学家的谢赫特曼休假期间从事航空用高强度合金研究时,从快速冷却的铝锰合金中发现一种特殊的“晶体”及准晶。他借助电子显微镜获得一幅电子衍射图,发现它不具有平移对称性却拥有长程有序性。而按照当时的理论,具有此种原子排列方式的固体物质是不存在的。因此,谢赫特曼的发现在当时引起极大争议。继续深入研究后,历经艰苦努力,谢赫特曼的题为《一种长程有序但不具有平移对称的金属相》的论文得以在《物理评论通讯》期刊上发表,引发了化学界的轩然大波。甚至名声显赫的获得诺贝尔奖的化学家鲍林也公开质疑他的研究成果。然而,后来事情发生了转变,1987年,法国和日本科学家制造出了足够大的准晶,并可以由X射线和电子显微镜直接观察这种晶体。至此,他的发现才得到了科学界的认可^[10]。

“黑天鹅”论文可能革新人们的科学认识,开辟新的研究方向,为科学进步作出巨大贡献。“黑白天鹅”的隐喻也为卡尔·波普尔的证伪主义、库恩在《科学革命的结构》中提出的范式转移两大影响深远的科技哲学理论提供了量化解释。

3.4 “天鹅”隐喻的意义

“黑白天鹅”隐喻使我们能够确定重要的科学发现,并定性和定量地衡量科学进步的影响。一篇“黑天鹅”文献的出现引起科学变革,可能改变旧的范式,尤其是“强壮黑天鹅”的效果较强,而“弱小黑天鹅”或“灰天鹅”的效果则相对较弱,但“灰天鹅”也反映出在主流科学中取得了进展。“黑白天鹅”隐喻兼顾了论文质量和引用数量,这种均衡使得该方法的价值凸显出来。在“黑白天鹅”的研究中,科学发现(科学史)的记载提供基础,而科学计量数据(引文曲线)给出了量化补充。当“白天鹅”遇到“黑天鹅”,科学出现突破,“白天鹅”的引用模式发生变化,这为识别和验证科学发现的重要性提供了量化方法。

“黑白天鹅”之间的关系是:没有“白天鹅”就谈不上“黑天鹅”,“黑天鹅”论文的出现会导致“白天鹅”论文的引用频次显著下降,“黑、白天鹅”是相对应而存在的,因此,“黑、白天鹅”之间存在互动和相互依存关系。

4 小结

综上所述,天鹅是一种在高品质论文中出现的特殊引文模式,黑天鹅是指具有科学突破性的论文,白天鹅是指为数众多的高品质论文,黑天鹅的出现会导致白天鹅的引用曲线改变。天鹅指数是测量黑天鹅影响力的重要指标,根据天鹅指数的高低,可以将天鹅分为黑天鹅和灰天鹅。天鹅引文模式的判定需要建立在对高品质文献进行识别和判定的基础上,因而需要结合科学史和具体学科的发展脉络,才能够更加科学和有效地评判高品质论文特别是科学突破论文。

总之,天鹅这一特殊引文模式的提出,可望为判别具有科学突破性质的成果提供一种鉴别思路和方法。

参考文献

- 1 叶鹰. 国际学术评价指标研究现状及发展综述[J]. 情报学报, 2014, 33(2): 215-224.
- 2 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, 35(1): 100-103.
- 3 Zeng, C.J., Qi, E. P., Li, S.S., Stanley, H. E. & Ye, F. Y. Statistical characteristics of breakthrough discoveries in science using the metaphor of black and white swans[J]. Physica A, 2017, 487: 40-46
- 4 Taleb N.N. Black Swans: the impact of the highly improbable [M]. New York: Penguin Books, 2007.
- 5 Garfield E. Citation indexing: Its theory and application in science, technology and humanities [M]. New York: Wiley, 1979.
- 6 Avramescu A. Actuality and obsolescence of scientific literature [J]. Journal of the American Society for Information Science, 1979, 30(5): 296-303.
- 7 Redner S. Citation statistics for more than a century of physical review[J]. Physics Today, 2004, 58(6): 49-54.
- 8 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.
- 9 Glänzel W. Seven myths in bibliometrics: about facts and fiction in quantitative science studies [J]. Collnet Journal of Scientometrics and Information Management, 2008, 2(1): 9-17.
- 10 陈其荣, 廖文武. 科学精英是如何造就的: 从 STS 的观点看诺贝尔自然科学奖[M]. 上海: 复旦大学出版社, 2011: 74.

作者单位: 南京大学信息管理学院, 南京, 2100232
南京大学-UIUC 国际联合信息学实验室, 南京, 210023

收稿日期: 2018年2月8日

(转第 112 页)