A probe into the citation patterns of high-quality and high-impact publications

Jiang Li[1] and Fred Y. Ye [2]
[1]Department of Information Resource Management,
Zhejiang University, Hangzhou, CHINA
[2]School of Information Management,
Nanjing University, Nanjing, CHINA
e-mail: yye@nju.edu.cn (corresponding author)

ABSTRACT

Enlightened by Avramescu's studies, we classified the citation patterns of high-quality and high-impact publicationsinto five mathematical types: lognormal-type, exponential-type, polynomial-type, wave-type and sleeping-beauty. Based on practical datasets of high-quality and high-impact publications, the five types of citation patterns are fitted well at statistical level. The results reveal that high-quality publications usually appear irregular citation curves (wave-type and sleeping-beauty) and high-impact publications tend to show regular ones (lognormal-type, exponential-type and polynomial-type).

Keywords: Citation analysis; Citation pattern; Citation curve; Sleeping-beauty; Obsolescence.

INTRODUCTION

Citation analysis has become the central topic in bibliometrics, scientometrics as well as informetrics research since 1955, when Eugene Garfield introduced citation index (Garfield 1955) and citation indexing (Garfield 1979); and it has been further improved and used as a research evaluation tool (Moed 2005). Different citation patterns mark different citation history, which is especially meaningful in the studies of obsolescence (Line and Sandison 1974). There are two approaches in the measurement of obsolescence in citation studies: 'synchronous' and 'diachronous' distribution (Nakamoto 1988). These approaches are also referred to as 'citations from' and 'citations to' approaches (Redner 2004), or 'retrospective citation' and 'prospective citation' approaches (Burrell 2002; Glanzel 2004). The comparison of the two approaches has been made by Line and Sandison (1974), Griffith et al. (1979), Nakamoto (1988),Stinson and Lancaster (1987),Gupta (1990), Egghe (1993), and Redner (2004).

Citation pattern, also known as the shape of 'citation curves' (Garfield 1989), or 'citation histories' (Redner 2004), is the focus of diachronous approach. It is agreed that an article is likely to be cited in the first following years after publication, called 'immediacy effect', then rises to a citation peak, followed by a maximum level attained after a certain time and tends to be gradually less cited with time. Apart from this 'typical citation curve' or 'life cycle' (Cunningham 1995), this phenomenon called obsolescence is presented in five citation curves (Avramescu1979) as shown in Figure 1.

According to Avramescu, each curve has different meaning as depicted in Figure 1, where (1) refers to initially much praised articles, (2) basic recognized work, (3) scarcely reflected work, (4) well-received but later erroneous qualified work, and (5) genius work.Curves 1, 2 and 3 coincide with the typical citation curves, which belong to lognormal distribution. Curve 5 appears exponential. All five curves were reproduced by Avramescu's formula:

$$c(t) = C_0[\exp(-\alpha t) - \exp(-m\alpha t)],\ m>1, \qquad (1)$$

where $C_0$is the amplitude, and $\alpha$is the age decrement. However, some citation distributions cannot be described by Avramescu's formula, such as curves with two or more peaks or those with other irregular features.

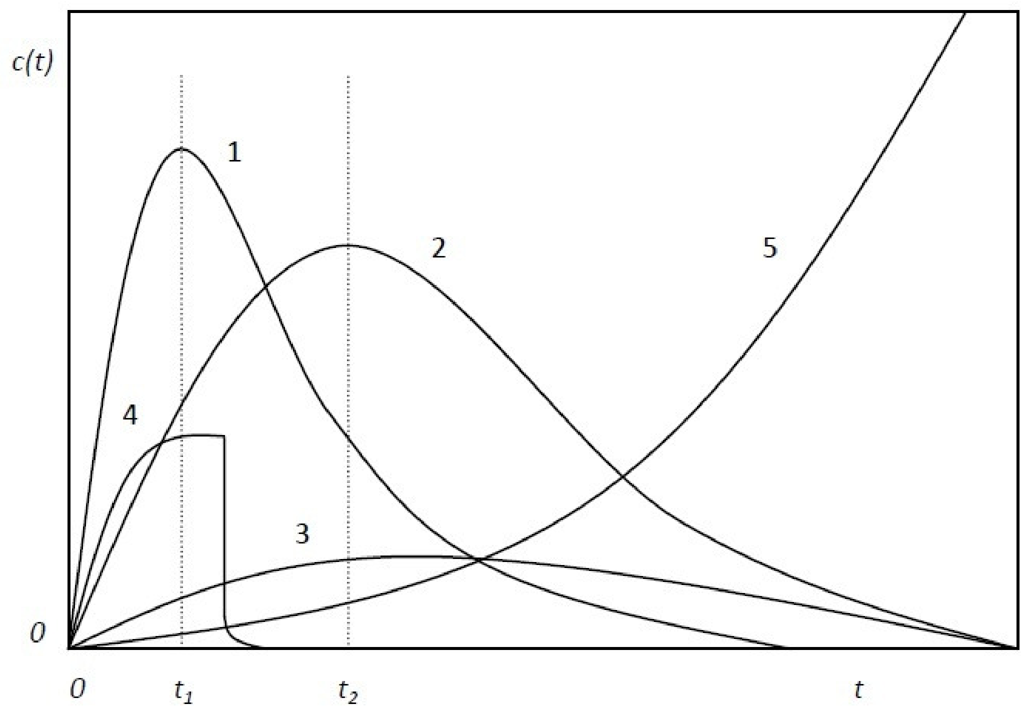


Figure 1: Avramescu's Citation Curves of Individual Articles

Other citation phenomena areknown as 'premature discoveries' (Stent 1972; Wyatt 1975), 'resisted discoveries' (Barher 1961), 'delayed recognition' (Cole 1970), or 'sleeping beauties' (Van Raan 2004; Glänzel and Garfield 2004; Burrell 2005; Braun et al. 2010; Egghe et al. 2011). They are publications that go unnoticed ('sleeps') for a long time and then, almost suddenly, attract a lot of attention ('is awakened by a prince'). Li and Ye (2012) revealed four special cases of sleeping beauties, in the phenomenon of 'all-elements-sleeping-beauties', which had a leaping before sleeping in citation curves.

The lognormal distribution is most advocated for typical citation curves, like curves 1 and 2 in Figure 1. It is proven by Egghe and Rao (1992) that lognormal model best fits the typical citation curve, other than Avramescu's model, negative binomial distribution (Mingers and Burrell 2006), and Weibull distribution. Not only does the lognormal distribution fit diachronous analysis, it is also used for synchronous analysis. Gupta (1997) stated that age of references cited is best modeled according to lognormal distribution, after analyzing the age of references cited in source papers of the theoretical population genetics specialty at different phases of its development. It has been corroborated by Rao and Meera (1991) that age distribution of cited journals follows lognormal distribution. In addition, some other models are also applied for describing the shape (or partial shape) of citation curves. Pollmann (2000) proposed that the decay from 'age 4' is best described by an inverse function. Multiple models were advocated and applied by Sangam (1999) for exploration of citation age distribution of psychology journals, including lognormal, negative binomial and exponential distribution. Bouabid (2011) proposed a model to represent the naturally observed citations distribution and citation aging in a diachronous-based method.

The peaking time of citations features the shape of citation curves, reflecting the immediacy of publications. A citation window of 3 to 5 years following the year of publication has proven to yield the most informative trend data (Moed 2005). In general, papers reach their citation peak two, three, or even four years after publication (http://sciencewatch.com/about/met/core-hp/). According to Amin and Mabe's(2000) statistics, 'citations to articles published in a given year rise sharply to a peak between two and six years after publication' (p. 2). Citation windows vary according to subject field, as well as journal type or article type. For letters, research articles and reviews, the citation peaks are farther and farther from the origin. For most scientific papers, death (no longer being cited by other papers) comes within ten years of the paper's publication (Price 1976).

In this article, we explore citation patterns of high-quality and high-impact publications, in order to construct a total framework for indicating high-quality or high-impact publications via characterized citation patterns, with clarifying the relations and differences between high-quality and high-impact ones.

METHODOLOGY

Referring to Avramescu's five types of curves, by using practical citation data, we found that there were different five types of curves for high-quality and high-impact publications. Mathematically, the curves can be classified as lognormal-type, exponential-type, polynomial-type, wave-type and sleeping-beauty. The method of Curve Fitting is used to describe citation curves, using the computer program OriginPro 8. Prior to curve fitting, citation curves need to be classified according to their shape. Lognormal function covers Avramescu's curves of 1, 2 and 3 (Figure 1). Exponential and polynomial functions are applied for curve 5 and others.

As a result of curve fitting, whether 'failed' or 'succeeded' is reported, as well as the goodness of fit, R^2,for successful fitting. The lognormal function is

$$y = y_0 + \frac{A}{\sqrt{2\pi}wx} e^{\frac{-\left[ln\frac{x}{x_c}\right]^2}{2w^2}} , \qquad (2)$$

wherey_0 = offset, x_c = center, w = width, A = amplitude. The exponential function is

$$y = y_0 + Ae^{R_0 x} , \qquad (3)$$

where y_0 = offset, A = initial value, R_0 = rate. The polynomial function with order of three is

$$y=c+b_1x+b_2x^2+b_3x^3 , \qquad (4)$$

where b_1, b_2 and b_3 are constant coefficients. In this research, it is required that the value of R^2 must be no less than 0.700 for a successful lognormal, exponential or polynomial fitting.

Meanwhile, there are irregular or waveform curves. As any function y=f(x) can become Fourier expansion, as follows,

$$y = f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty}(a_n \cos nx + b_n \sin nx) \quad (5)$$

and cosine and sine functions looks like true waves, we apply the wave-type to cover all other curves except regular lognormal, exponential, polynomial ones, and sleeping-beauties.

It is defined in this research that a sleeping beauty in science is an article cited less than 2.5 times on average per year (depth of sleep) in continuous eight years or longer (length of sleep), and no less than 20 times in the four years (awake intensity) following the sleeping period, furthermore, the value of annual citations in awakening period must be at least four times that in sleeping period. It is also defined in this research that an all-elements-sleeping-beauty in science is an article cited over ten times in the publication year or the next year, and then it quickly falls into sleep and becomes a sleeping beauty as defined above.

Data

The data consist of five datasets: (1) high-quality publications (N_{21}); (2) near high-quality publications (N_{82}); (3) high-impact publications in SCI (SCI_{100}); (4) high-impact publications in SSCI ($SSCI_{100}$); and (5) high-impact publications in A&HCI ($A\&HCI_{100}$).

N_{21} comes from twenty-one discoveries that changes science and the world. It includes twenty-one papers which are chosen from the book *A Century of 'Nature': Twenty-One Discoveries that Changed Science and the Worldby Laura Garwin and Tim Lincoln, University of Chicago Press, 2003*,(Garwin and Lincoln 2003) as shown in Table 1, most of which were awarded Nobel prizes so that they belong to high-quality publications.

N_{82} contains 82 papers from *Nature*, which are determined by the following procedures: First, select top five most cited *Nature* publications in each publication year in Table 1; second, eliminate papers appeared in Table 1 from the collection; and third, delete papers appeared in SCI_{100}, $SSCI_{100}$ or $AHCI_{100}$. It is remarkable that high-quality publications are not necessarily of low-influence, and high-impact papers are not necessarily of low-quality. Articles in N_{82} are of high-impact in corresponding years and their quality is not as high as that of N_{21}, although there are 12 out of 21 among the top five most cited publications in corresponding years, and six ranking the first, as shown in Table 1.

SCI_{100}, $SSCI_{100}$, $A\&HCI_{100}$ are defined as high-impact articles involving three groups of top 100 most cited papers selected from SCI, SSCI, and A&HCI databases of the Thomson Reuters, respectively. The influence of articles in SCI_{100} is significantly higher than N_{21}, and both of them are high science articles, therefore the comparisons characterize citation curves of high science articles. In addition, the selection of $SSCI_{100}$ and $AHCI_{100}$ covers social science, arts and humanities, and it extends characteristics of citation curves of high-impact articles. The influence of $SSCI_{100}$ articles is only inferior to SCI_{100}, as shown in Table 2. Although $AHCI_{100}$ has the lowest citation rate, it characterizes the citation curves of high-impact articles of arts and humanities.

Table 1: High-quality Publication Group: Twenty-One Discoveries that Changed Science and the World

Code	Author(s)	PY	Vol: page	C	R
N_{21}-01	Dart, R.A.	1925	115: 195-199	211	1
N_{21}-02	Davisson, C.&Germer, L.H.	1927	119: 558 - 560	100	1
N_{21}-03	Chadwick, J.	1932	129: 312	122	1
N_{21}-04	Kapitza, P.	1938	141: 74	162	6
N_{21}-05	Meitner, L. & Frisch, O.R.	1939	143: 239-240	172	3
N_{21}-06	Watson, J. D. & Crick, F. H. C.	1953	171: 737 - 738	3760	1
N_{21}-07	Kendrew, J. C., Phillips, D. C., et al.	1958	181: 662 - 666	383	6
N_{21}-08	Maiman, T.H.	1960	187: 493-494	1006	1
N_{21}-09	Schmidt, M.	1963	197:1040	351	9
N_{21}-10	Vine, F.J. & Matthews, D.H.	1963	199: 947-949	660	3
N_{21}-11	Hewish, A., Collins, R.A., et al	1968	217: 709 - 713	540	7
N_{21}-12	Baltimore, D.	1970	226: 1209 -1211	1509	4
N_{21}-13	Lauterbur, P. C.	1973	242: 190 - 191	1330	2
N_{21}-14	Zinkernagel, R. M. & Doherty, P. C.	1974	248: 701 - 702	1539	4
N_{21}-15	Neher, E &Sakmann, B.	1976	260: 799 - 802	872	8
N_{21}-16	Sanger, F., Smith, M., et al.	1977	265: 687 - 695	1042	9
N_{21}-17	Nüsslein-Volhard, C. &Wieschaus, E.	1980	287: 795 - 801	1848	3
N_{21}-18	Kroto, H. W., Smalley, R.E., et al.	1985	318: 162 - 163	7090	1
N_{21}-19	Farman, J.C., Gardiner, B.G., and Shanklin, J.D.	1985	315: 207-210	1585	7
N_{21}-20	Mayor, M. &Queloz, D.	1995	378: 355-359	1328	20
N_{21}-21	Wilmut, I., Schnieke, A.E., McWhir, J., et al.	1997	385: 810-813	1745	17
Notes: PY = Publication Year; C = Citations; R = Ranking of the year					

Table 2: Citations Distribution of the High-quality and High-impact Publications

	Age			Citations			Annual citations		
	Min	Max	Avg	Min	Max	Avg	Min	Max	Avg
N_{21}	14	86	47.3	100	7090	1302.6	1.2	272.7	43.1
N_{82}	14	86	47.6	24	8706	1076.0	0.3	327.7	23.6
SCI_{100}	3	86	34.5	8915	294484	13594.5	150.6	4913.6	504.9
$SSCI_{100}$	8	55	32.0	2759	34634	3915.5	57.7	677.7	127.4
$AHCI_{100}$	8	36	22.0	272	2414	353.5	8.3	92.1	17.6

After selecting the five groups of articles datasets, citations each year to each of the publications are searched in the Web of Science database, and their citation curves are hence drawn. The latest paper in SCI_{100} was published in 2008 and it is too young to be curved by citations.

RESULTS

As there are five groups of citation curves, we also apply five mathematical types for fittings: lognormal-type, exponential-type, polynomial-type, wave-type and sleeping-beauty. Appendices 1 and 2 show the age and citation distribution of the fivegroups of citation curves with their respective mathematical fittings.

Lognormal-type

a) Lognormal

A lognormal fitting of a citation curve for a selected paper is shown in Figure 2. Its shape is similar to curve 2 in Figure 1, which is basically a recognized work.Among the high-impact articles in SCI_{100}, $SSCI_{100}$ and $AHCI_{100}$, higher citations indicate more lognormal shape and better fitting. The articles in SCI_{100} are more cited than those in $SSCI_{100}$and those in $AHCI_{100}$. Its percentage of lognormal shape is much higher than that of $SSCI_{100}$ and $AHCI_{100}$, i.e. 9.0% and 5.0% respectively. Its goodness of fit of lognormal function is also higher (Appendix 2).

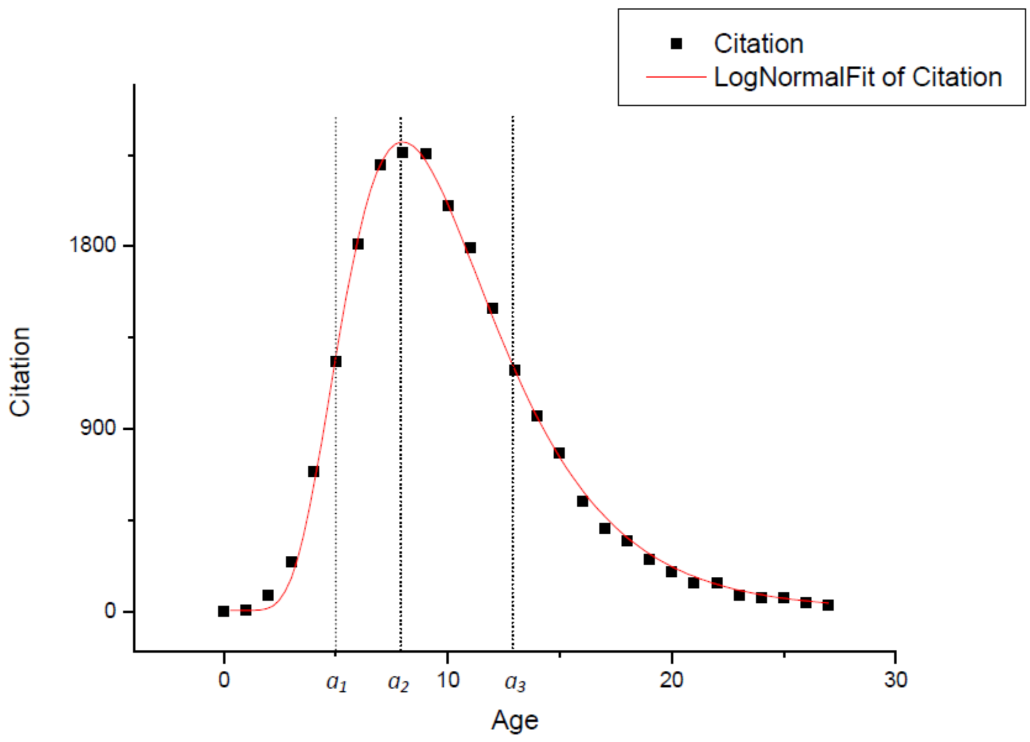


Figure 2: Lognormal Fitting for the Curve Peaking in the Ninth Year for Paper SCI_{100}-29: Feinberg, A.P, & Vogelstein, B. 1983, *Analytical Biochemistry*, 132(1):6-13. (R^2=0.998)

b) Incomplete Lognormal

The curve in Figure 3 is experiencing the stage of decreasing after peaking. The curve 'failed' to be fitted with lognormal or growing functions by OriginPro 8, but its shape is similar to the curve depicted in Figure 2. It is therefore regarded as an incomplete lognormal curve, e.g. a curve between age=0 and age=a_3 in Figure 2.

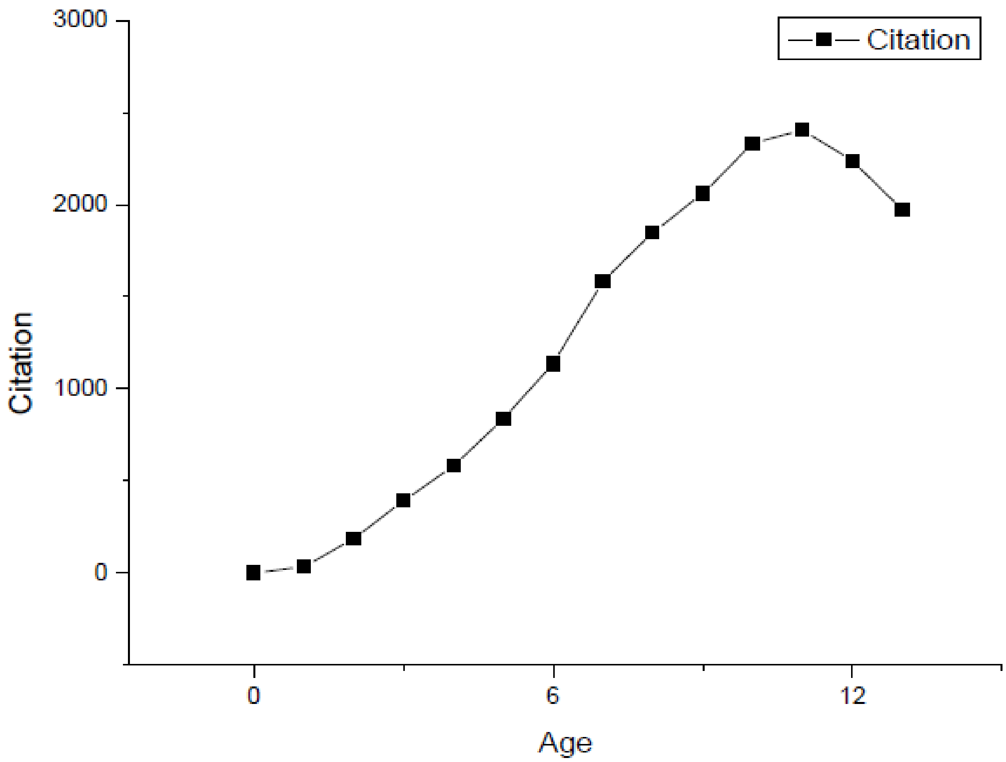


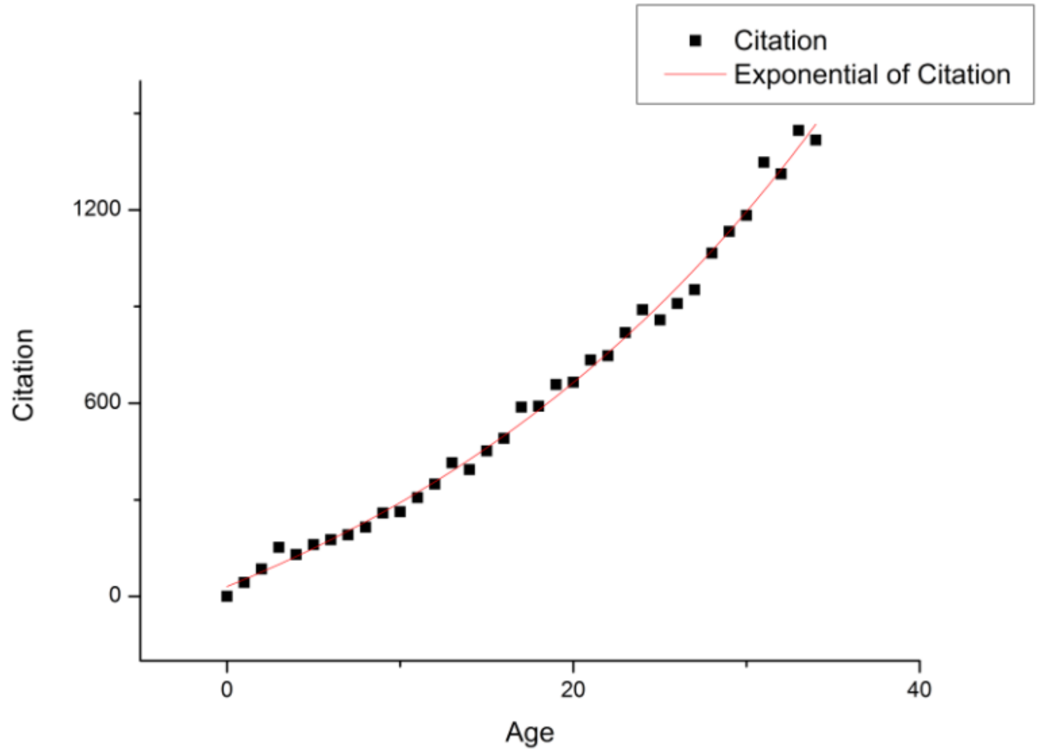
Figure 3: Incomplete Lognormal Curve for Paper SCI_{100}-33: Thompson, J.D., Gibson, T.J., Plewniak, F., et al. 1997. *Nucleic Acids Research*, 25(24): 4876-4882.

High-impact articles are more inclined to show lognormal-type citation curves, especially in high-science. There are 58.0% SCI_{100} articles appearing lognormal-type, but only 19.0% accounted for the high-quality articles. As to the science articles near high-quality, e.g. N_{82}, the percentage is 39.0%. However, high-impact articles in social science, arts and humanities tend to present exponential-type citation curves.

Exponential-type Citations

Exponential growth of citations is shown in Figures 4 and 5. The functions of the depicted are concave and convex, and the equations are $y = -583.227+613.517e^{0.035x}$ and $y = 561.826-586.327e^{-0.040x}$, respectively.

A growing curve resembling Exponential-a or Exponential-b is a partial lognormal curve, e.g. the part before peaking at time a_1 shown in Figure 2. An initial stage of a life cycle indicates a work that is not fully recognized yet. Therefore, exponential growth does not necessarily indicate 'genius work', as Avramescu's curve 5 in Figure 1. High-impact articles in social science are significantly more inclined to show exponential-type curve than any other articles in this research. There are 45.0% exponential-type citation curves in $SSCI_{100}$.



1976. *ActaCrystallographica Section A*, 32, Sept: 751-767.(R^2=0.994)

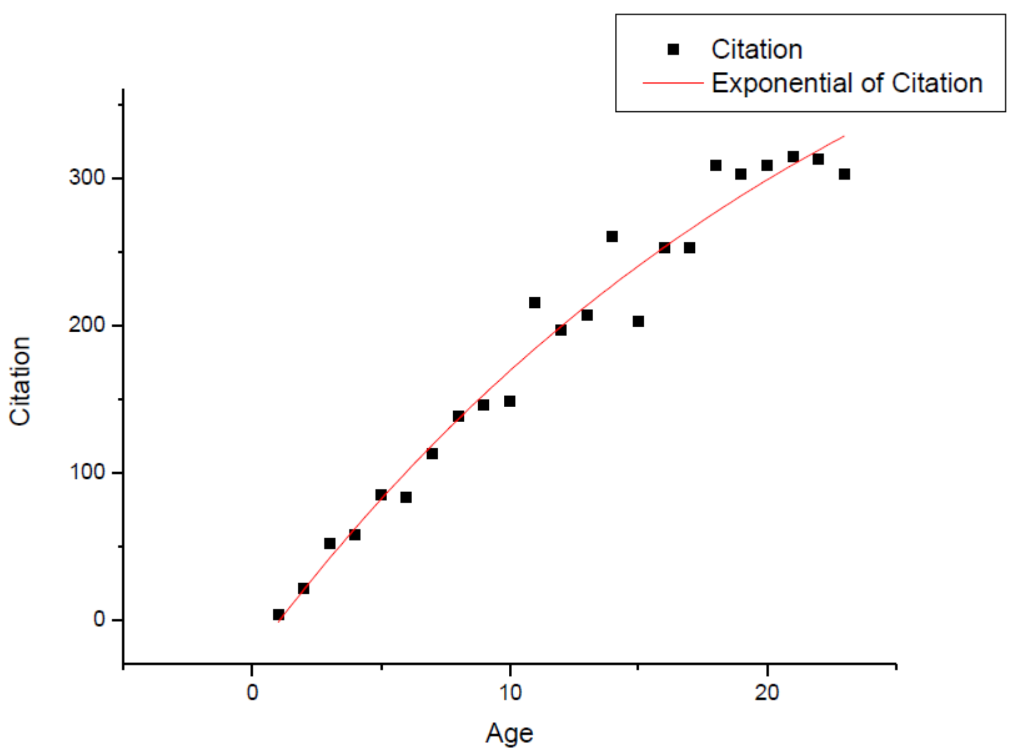


Figure 5: Exponential-b (Convex Exponential) Fitting for Paper $SSCI_{100}$-40: DiMaggio, P.J., & Powell, W.W. 1983. *American Sociological Association*, 48:147-160. (R^2=0.969)

Polynomial-type

A polynomial curve indicates a growth which means that an article is not fully recognized. Polynomial fitting with order three is shown in Figures 6 and 7. Citations in Figure 6 increase slowly at its beginning phase and ending phase but rapidly in the middle part. On the contrary in Figure 7, citations grow rapidly at its beginning phase and ending phase but slowly in the middle part. Their functions of polynomial fitting are $y=-64.316+106.990x_1+22.138x_2-1.070x_3$ and $y=-34.081+14.532x_1-0.612x_2+0.008x_3$, respectively.

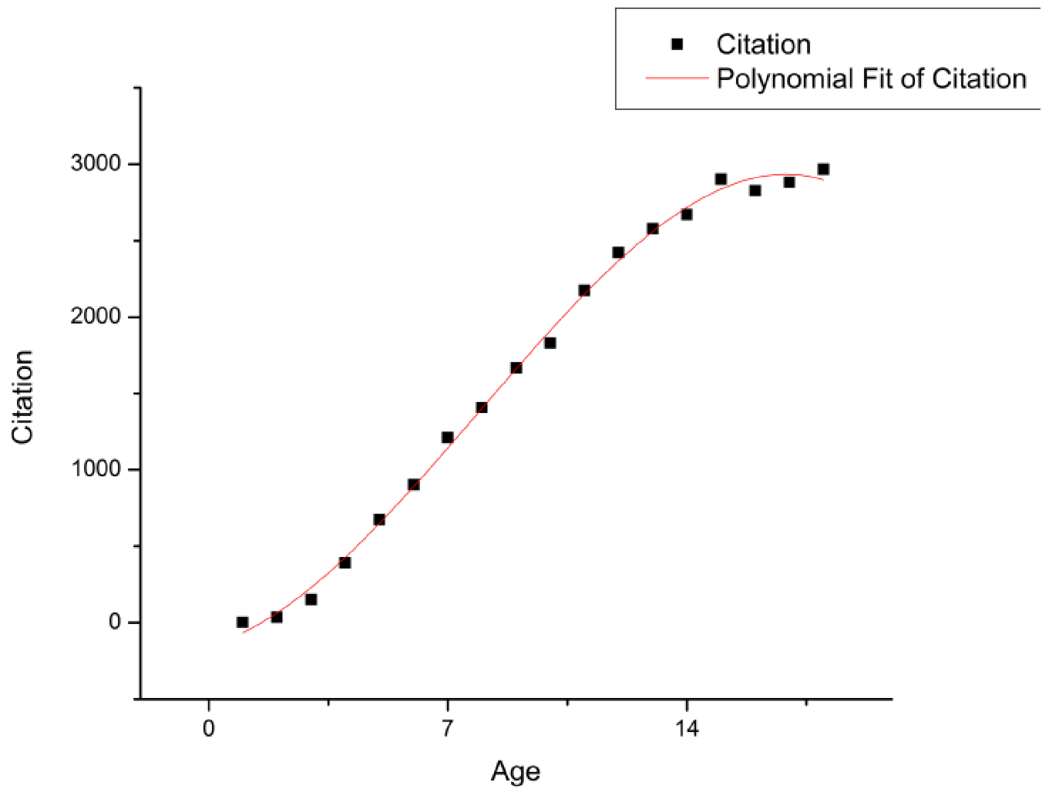


Figure 6: Polynomial-a Fitting for Paper SCI_{100}-11: Becke, A.D. 1993. *Journal of Chemical Physics*, 1993, 98(7): 5648-5652. (R^2=0.997)

The status of Polynomial-a is between the exponential in Figure 4 and the incomplete lognormal curve in Figure 3. It is near the peaking of its life cycle. The curve of Polynomial-a is not found among the high-quality or the near high-quality publications, but found in the high-impact articles. Polynomial-b is unusual in the slow down at the middle of the curve.Nineteen percent of the curves show that Polynomial-b in $SSCI_{100}$, are rarely in other high-impact groups, and none in the high-quality or near high-quality groups. It demonstrates that a period of stagnation by citations usually appears in high-impact social science articles. In general, high-impact articles are significantly more inclined to show polynomial-type curves than high-quality ones in this research.

High-quality publications are significantly less inclined to present regular citation curves, including lognormal-type, exponential-type and polynomial-type curves. In the high-quality group N_{21}, 23.8% of the curves demonstrate regular growth. However, the percentages are 85%, 88% and 51% in SCI_{100}, $SSCI_{100}$ and $AHCI_{100}$, respectively. The irregular citation curves involve wave-type and sleeping-beauty.

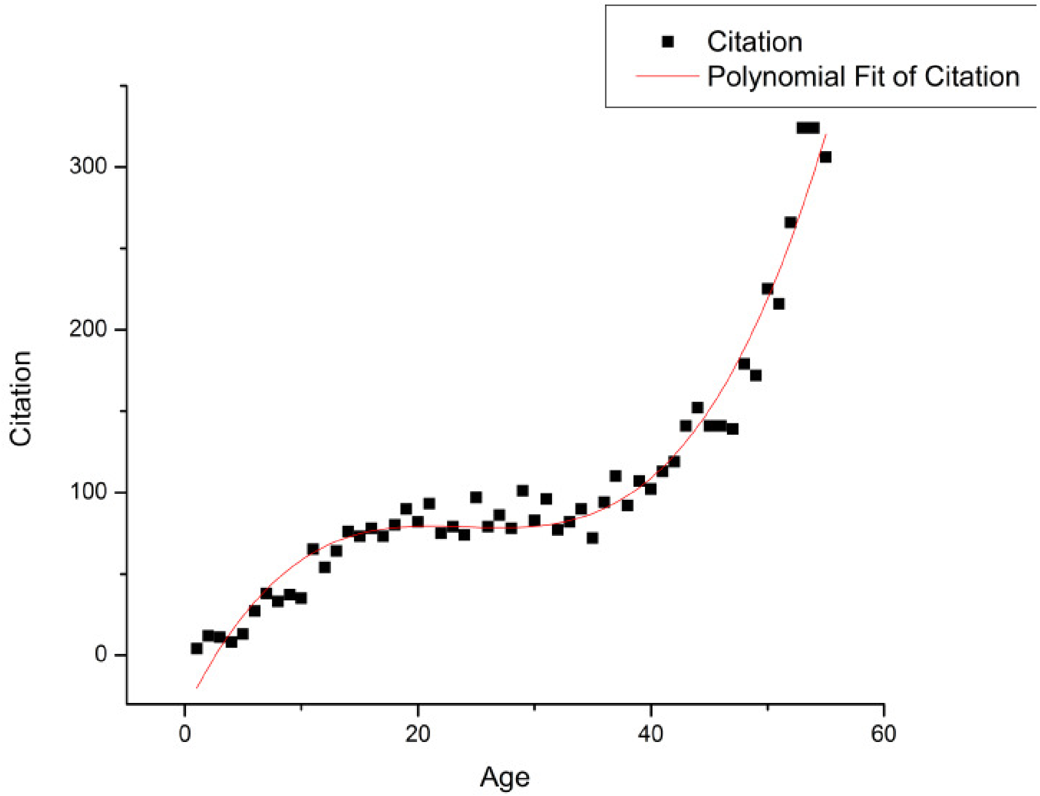


Figure 7: Polynomial-b fitting for Paper $SSCI_{100}$-23: Shrout, P.E. & Fleiss, J.L. 1979. *Psychological Bulletin*, 86(2): 420-428. (R^2=0.958)

Wave-type

a) Waveform-a: curves with signs of second life cycles

The curve in Figure 8 comprises two parts: the first part is a shape of a lognormal curve which shows signs of life cycle, and the second part is a curve which causes the lognormal fitting failed and makes the whole curve fluctuated in shape. The time a_0 in Figure 8 is a turning point, implying the start of the second life cycle of the publication. The curve in Figure 9 shows that the number of citations fluctuates after publication, but appears two peaks. Similar to the curve in Figure 8, it also begins with a lognormal shape but suffers a turning point which shows sign of a second life. Therefore, these curves are named 'Waveform-a' and belong to 'Wave-type' in shape.

The 'Waveform-a' shape exists in both high-quality and high-impact groups, but more in the former. In N_{21}, 33.3% of the articles appear 'Waveform-a' curves, and the percentages in SCI_{100}, $SSCI_{100}$ and $AHCI_{100}$ are 15.0%, 8.0% and 9.0%, respectively.

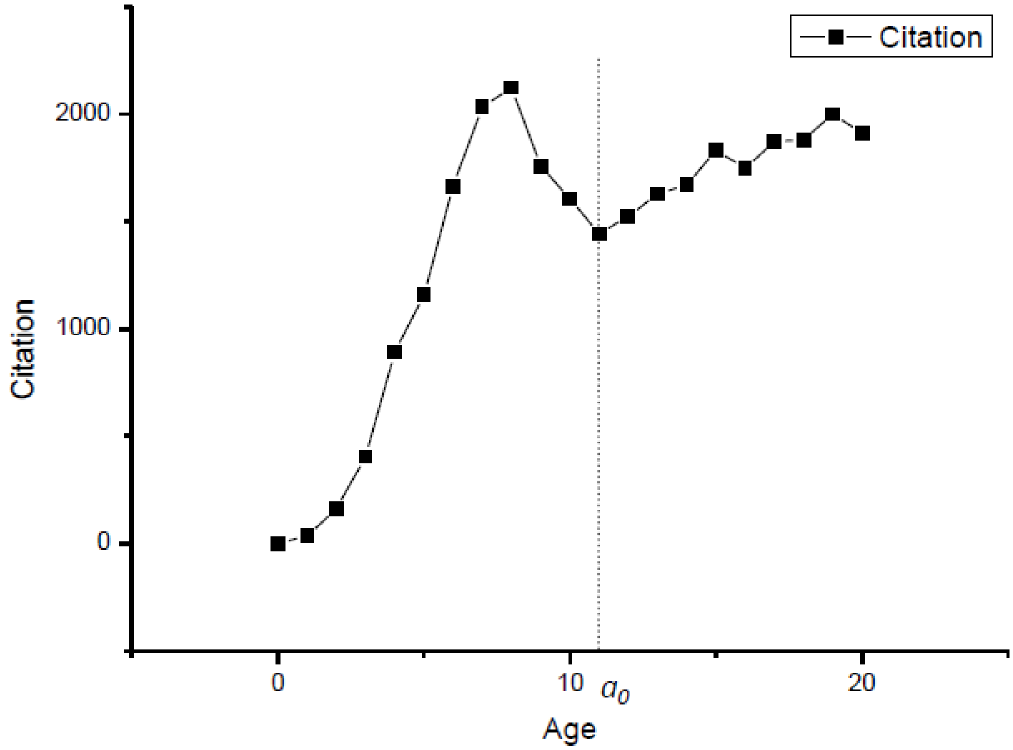


Figure 8: 'Waveform-a' curve for Paper SCI_{100}-14: Altschul, S.F., Gish, W., Miller, W., et al. 1990.*Journal of Molecular Biology*, 215(3): 403-410.

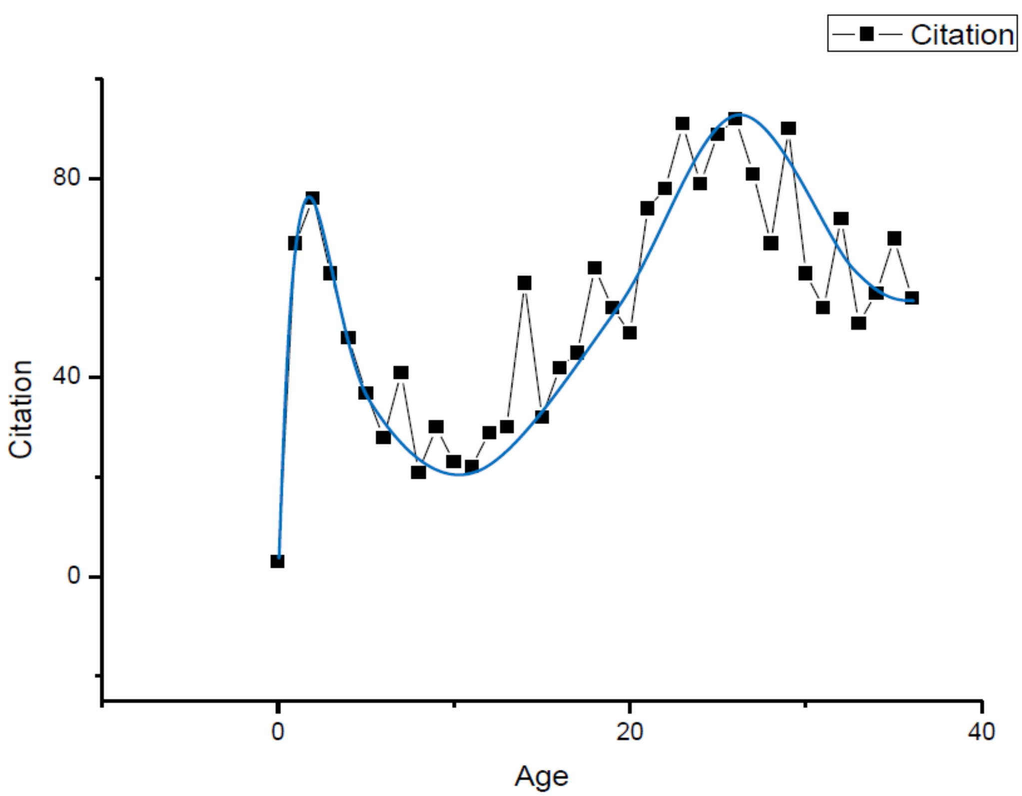


Figure 9: 'Waveform-a' curve for Paper N_{82}-52: Hawking, S.W. 1974. *Nature,* 248:30-31

b) Waveform-b: curves without signs of life cycle

A 'Waveform-b curve' is a curve whose citations fluctuate without signs of life cycles, and its shape does not satisfy any of the other curve types. Figure 10 shows a waveform-b curve. Waveform-b citation curves are easily found in articles of low annual citations. For example, 81.2% of the Wave-type articles in $AHCI_{100}$ have annual citations lower than 20, whereas no waveform-b curve is found in SCI_{100} or $SSCI_{100}$, whose lowest annual citations are 150.6 and 57.7, respectively. Waveform-b curves are of significantly lower annual citations than waveform-a curves, as shown in Appendices1 and 2. Wave-type citation curves mainly appear in publications of low annual citations.

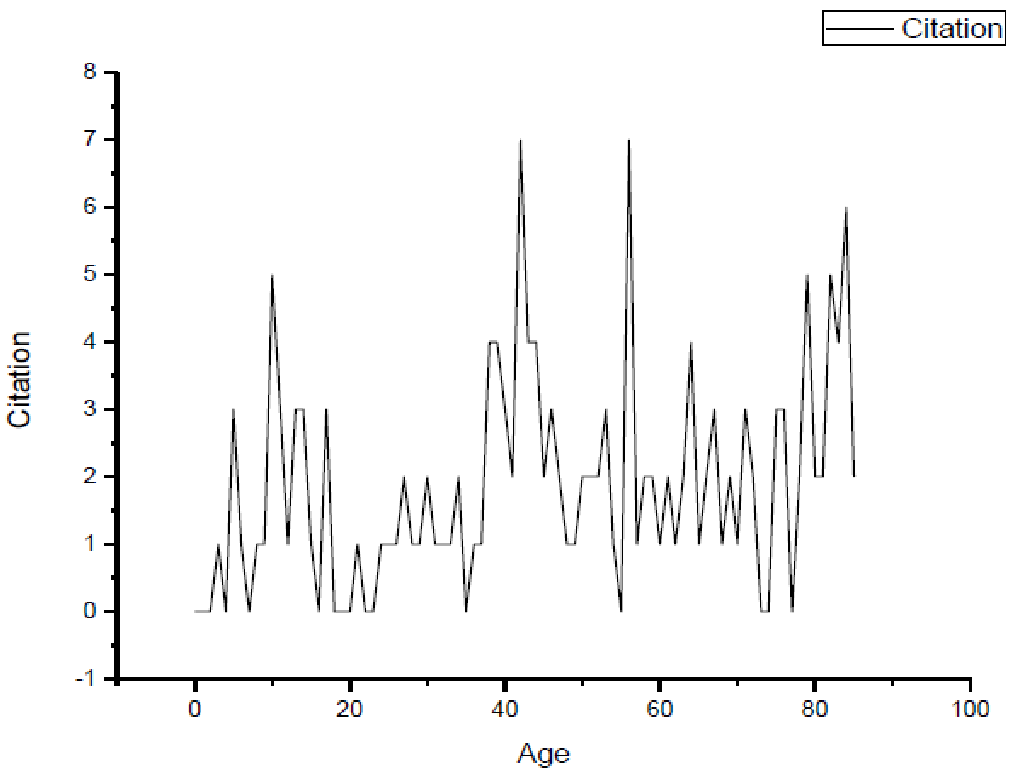


Figure 10: Waveform-b4 Curve for Paper N_{82}-1: Rowan, W. 1925. *Nature*, 115:494-495

Sleeping-beauty

A citation curve of a sleeping beauty, e.g. Sb-a, is shown in Figure 11. It had been sleeping for nine years since publication and had attracted 20 citations in total during this period. Then, it was "awakened by a prince in all of a sudden", attracting 114 citations in the four subsequent years. In general, the curve of a sleeping-beauty is irregular, but reveals completely different citation history compared to wave-type curves.

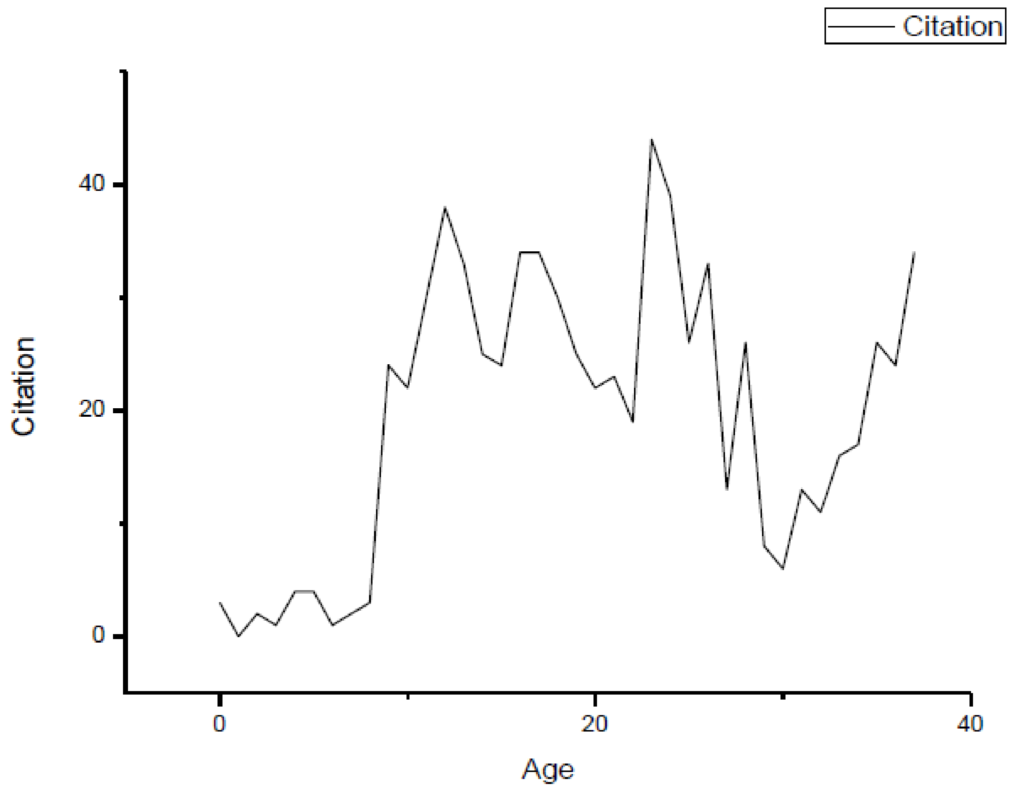


Figure 11: Sb-a Curve for Paper N_{82}-48: Smith, J.M., & Exponential, G.R. 1973, *Nature*, 246:15-18

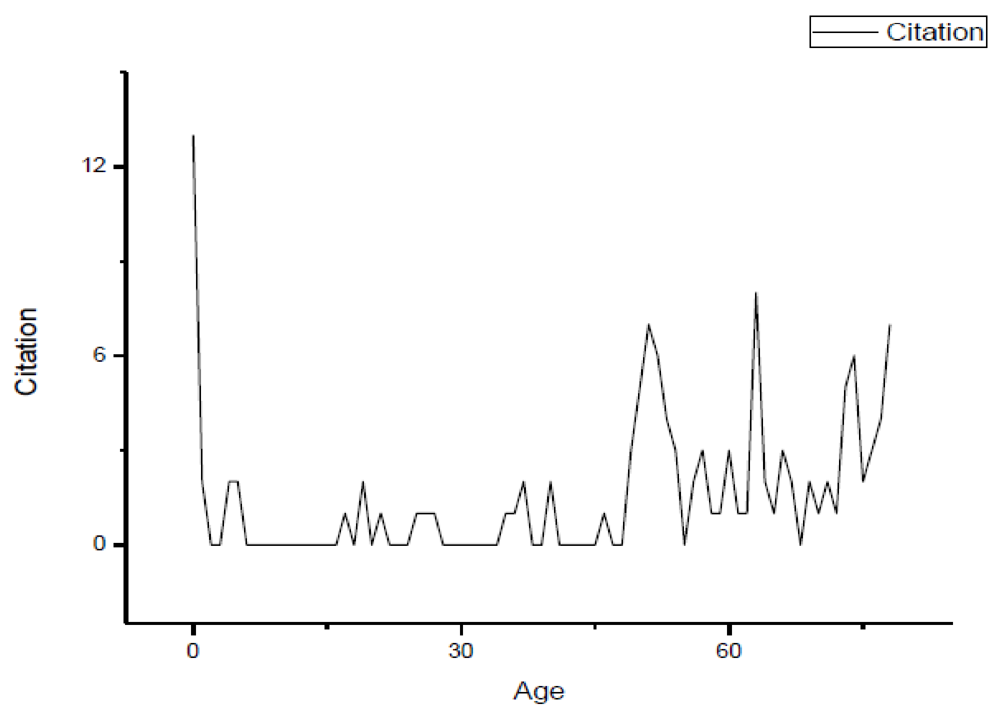


Figure 12: Sb-b curve for Paper N_{21}-3: Chadwick, J. 1932, *Nature*, 129: 312

A citation curve of an all-element-sleeping-beauty is shown in Figure 12, e.g. Sb-b. The publication N_{21}-3 was cited 13 times as soon as it is published, but had been suddenly ignored in the following 48 years as if the "princess pricked her fingers on the spindle and then fell into sleep". During the period of 48 years, it attracted only 0.417 citations per year. However, "the princess" was awakened afterwards since the publication was cited 21 times in the four years following the sleeping period.

Sleeping beauties in science are easily found in publications with low citation rate. There is no sleeping beauty found in the SCI_{100}, but five, six, seven and five in $SSCI_{100}$, $A\&HCI_{100}$, N_{82} and N_{21}, respectively, accounting for 5.0%, 6.0%, 8.5% and 23.8%. It reveals that the high-quality articles are more inclined to show sleeping-beauty curves. Furthermore, all of the four all-elements-sleeping-beauties in science are found in the high-quality articles, accounting for 19.0%, which suggests potential correlation between sleeping-beauty curves and high-quality articles, especially all-elements-sleeping-beauty.

DISCUSSION AND CONCLUSION

Similar to Avramescu's studies, we classified citation patterns into five types mathematically, i.e. lognormal-type, exponential-type, polynomial-type, wave-type and sleeping-beauty. Lognormal-type and sleeping-beauty citation curves link with high-impact and high-quality publications, respectively. Among the high-impact articles, SCI articles attract significantly more citations than SSCI articles and A&HCI articles. They are more inclined to appear regular citation curves, e.g. lognormal-type, exponential-type and polynomial-type than SSCI articles, and A&HCI articles. Exponential-type curves is a sign of increasing recognition which rarely exists in high-quality articles but extensively in high-impact publications, especially in SSCI articles. The irregular curves, e.g. wave-type and sleeping-beauty, largely emerge in high-quality articles rather than high-impact ones. The life cycle changed unexpectedly in waveform-a. It is more inclined to appear in the high-quality articles, compared to waveform-b which has no signs of life cycles. Irregular curves without signs of life cycles are easily found in publications with low annual citations.

The age of regular citation curves, e.g. lognormal-type, exponential-type and polynomial-type, are significantly younger than that of irregular ones among all the high-quality and high-impact articles collected in this study. The oldest is sleeping-beauty among all of the citation curves. The peaking age of high-impact publications is older than that of high-quality publications. Among the high-impact articles, the peaking age of SCI articles is significantly older than that of SSCI and A&HCI articles. Among all the articles chosen in this research, very few articles peaked before five years since their publication. Among all the high-impact articles in SCI, SSCI and A&HCI, very few articles peaked within fifteen year since their publication. In this case, the two-year citation window of journal impact factor is too short for these high-impact publications to make the greatest contribution to the corresponding journals.

As the datasets N_{21}, N_{82}, SCI_{100}, $SSCI_{100}$, and $A\&HCI_{100}$ belong to high-quality or high-impact publications, we ignore the low-impact or low-citation publications, so that our findings are confined to high-quality or high-impact publications. It is not suggested that the quality of publications is judged in terms of citation curves, but citation curves do give clues of recognition. It is suggested that publications for citation curve analysis be aged ten or older, to make the curve meaningful and curve fitting practical.

The five types of citation patterns construct a total framework for indicating the quality and impact of publications. As the framework is mainly derived from high-quality and high-impact publications, it may not be always suitable for fitting general publications, particularly average articles with low citations.

The findingsreveal that high-quality publications usually appear irregular citation curves (wave-type and sleeping-beauty) and high-impact publications tend to show regular ones (lognormal-type, exponential-type, polynomial-type). Meanwhile, among the high-impact publications, the citation curves of high-science, social science, and arts and humanities publications are principally lognormal-type, exponential-type and irregular, respectively. Our studies suggest that the quality and the impact of publications are not inevitably associated.

ACKNOWLEDGEMENTS

The authors thank the National Natural Science Foundation of China (NSFC Grant Nos. 71203193 and 71173187) and the Department of Education of Zhejiang Province (Y201223174) for financial support. The authors also thank Professor Peter Willett and Mr. Xing Zhao for their helpful comments.

REFERENCES

Avramescu, A. 1979. Actuality and Obsolescence of Scientific Literature. *Journal of the American Society for Information Science*, Vol. 30, no. 5: 296-303.

Amin, M. and Mabe, M. 2000. Impact factors: use and abuse. *Perspectives in Publishing*, no. 1: 1-6.

Barher, B. 1961. Resistance by scientists to scientific discovery. *Science,* Vol. 134: 596–602.

Bouabid, H. 2011. Revisiting citation ageing: a model for citation distribution and life cycle prediction. *Scientometrics*, Vol. 88, no. 1: 199-211.

Braun, T., Glanzel, W., and Schubert, A. 2010. On sleeping beauties, princes and other tales of citation distributions*. Research Evalua*tion, Vol. 19, no. 3: 195–202.

Burrell, Q. L. 2002. The nth-citation distribution and obsolescence. *Scientometrics*, Vol. 53, no. 3: 309–323.

Burrell, Q. L. 2005. Are 'Sleeping Beauties' to be expected? Scientometrics, Vol. 65, no. 3: 381–389.

Cole, S. 1970. Professional standing and the reception of scientific discoveries. *American Journal of Sociolog*y, Vol. 76, no. 2: 286–306.

Cunningham, S. J. 1995.An empirical investigation of the obsolescence rate for information systems literature. (Working paper 95/16). Hamilton, New Zealand: University of Waikato, Department of Computer Science.

Egghe, L. 1993. On the influence of growth on obsolescence. *Scientometric*s, Vol. 27, no. 2: 195-214.

Egghe, L., Guns, R., and Rousseau, R. 2011. Thoughts on uncitedness: Nobel laureates and fields medalists as case studies. *Journal of the American Society for Information Science and Technolog*y, Vol. 62, no. 8: 1637–1644.

Egghe, L., and Rao, I. K. R. 1992. Citation age data and the obsolescence function: Fits and explanations. *Information and Processing Management*, Vol. 28, no. 2: 201–217.

Garfield, E. 1955. Citation indexes for science – new dimension in documentation through association of ideas. *Science*, Vol. 122, no. 3159: 108-111

Garfield, E. 1979. *Citation indexing: its theory and applications in science, technology, and humanities*. New York: Wiley

Garfield, E. 1989. More delayed recognition. Part 1. Examples from the genetics of color blindness, the entropy of short-term memory, phosphoinositides, and polymer Rheology. *Current Contents*, Vol. 38: 3-8.

Garwin, L., and Lincoln, T. 2003. *A century of 'Nature': Twenty-One discoveries that changed science and the world*. University of Chicago Press.

Glanzel, W. 2004. Towards a model for diachronous and synchronous citation analyses. *Scientometrics*, Vol. 60, no. 3: 511–522.

Glänzel, W. and Garfield, E. 2004. The myth of delayed recognition. *Scientist,* Vol. 18, no. 11: 8–9.

Griffith, B.C., Servi, P. N., Anker, A., and Drott, M. C. 1979. The aging of scientific literature: a citation analysis, *Journal of Documentation*, Vol. 35, no. 3: 179-196.

Gupta, B.M. 1997. Analysis of distribution of the age of citations in theoretical population genetics. *Scientometrics*, Vol. 40, no. 1: 139-162.

Gupta, U. 1990. Obsolescence of physics literature: Exponential decrease of the density of citations to Physical Review articles with age. *Journal of the American Society for Information Science*, Vol.41, no. 282–287.

Li, J., and Ye, F.Y. 2012. The phenomenon of all-elements-sleeping-beauties in scientific literature. *Scientometrics,*Vol. 92, no. 3: 795-799.

Line, M. B. and Sandison, A. 1974. 'Obsolescence' and changes in the use of literature with time. *Journal of Documentation*, Vol. 30, no. 3: 283-350.

Mingers, J., and Burrell, Q. L. 2006. Modeling citation behavior in management science journals. *Information Processing and Management*, Vol. 42, no. 6: 1451–1464.

Moed, H.F. 2005. *Citation analysis in research evaluation.* Dordrecht: Springer.

Nakamoto, H. 1988. Synchronous and dyachronous citation distributions. In L. Egghe& R. Rousseau (Eds.), *Informetrics*87/88. Amsterdam: Elsevier Science Publishers. pp. 157–163.

Pollmann, T. 2000. Forgetting and the ageing of scientific publications. *Scientometrics,*Vol. 47, no. 1: 43-54.

Price, D. J. de Solla 1976. A general theory of bibliometrics and other cumulative advantage distribution. *Journal of the American Society for Information Science*, Vol. 27, no. 5: 292–306.

Price, D. J. d. S. 1965. Networks of scientific papers. *Science*, Vol. 149, no. 3683: 510-515.

Rao, I. K. R., &Meera, B. M. 1991. Growth and obsolescence of literature: An empirical study. Third International Conference on Informetrics, Aug. 9-12, Bangalore India.

Redner, S. 2004. Citation statistics for more than a century of physical review. *Physics Today,*Vol. 58, no. 6: 49–54.

Sangam, S. L. 1999. Obsolescence of literature in the field of psychology. *Scientometrics,*Vol. 44, no. 1: 33-46.

Stent, G. S. 1972. Prematurity and uniqueness in scientific discovery. *Scientific American*, Vol. 227, no. 6: 84–93.

Van Raan, A. F. J. 2004. Sleeping Beauties in science. *Scientometrics*, Vol. 59, no. X: 467–472.

Wyatt, H. V. 1975. Knowledge and prematurity-journey from transformation to DNA. *Perspectives in Biology and Medicine,*Vol. 18, no. 2: 149–156.

Appendix 1: Distribution of Citation Patterns of N_{21} and N_{82}

		N_{21}						N_{82}					
		N	**Age**	**TC**	**AC**	**R^2**	**Peak**	**N**	**Age**	**TC**	**AC**	**R^2**	**PA**
Lognormal-type	Lognormal	3	41.0	1070.3	27.1	0.941	4.3	28	34.3	2150.1	78.2	0.917	8.1
	Incomplete lognormal	1	14.0	1745.0	124.6		6.0	4	14.5	3136.5	219.0		9.8
Exponential-type	Exponential-a							5	42.6	1005.6	25.5	0.897	
	Exponential-b	1	16.0	1328.0	83.0	0.810		1	14.0	2738.0	195.6	0.772	
Polynomial-type	Polynomial-a												
	Polynomial-b												
Wave-type	Waveform-a	7	44.1	2113.1	61.7		6.9	9	43.7	1336.4	40.3		5.4
	Waveform-b	4	34.3	1378.0	43.7			28	67.7	580.0	12.8		
Sleeping-beauty	Sb-a	1	86.0	211.0	2.5			7	52.9	787.3	17.0		
	Sb-b	4	77.0	139.0	1.8								

N = Number of publications; TC = Total citations; AC = Annual citations; PA= Peaking age.

Appendix 2: Distribution of Citation Patterns of SCI_{100}, $SSCI_{100}$ and $AHCI_{100}$

		SCI_{100}						$SSCI_{100}$						$A\&HCI_{100}$					
		N	**Age**	**TC**	**AC**	**R^2**	**PA**	**N**	**Age**	**TC**	**AC**	**R^2**	**PA**	**N**	**Age**	**TC**	**AC**	**R^2**	**PA**
Lognormal-type	Lognormal	48	38.3	29711.7	794.6	0.952	15.8	9	36.7	4493.3	128.5	0.910	15.6	5	23.8	411.0	18.8	0.842	5.0
	Incomplete lognormal	10	30.4	18285.0	812.0		23.0	14	26.9	4215.6	187.2		21.9	9	15.3	527.2	33.2		12.8
Exponential-type	Exponential-a	20	33.4	14739.5	525.4	0.980		42	31.7	5155.6	177.3	0.970		23	23.5	507.6	23.7	0.861	
	Exponential-b	1	19.0	10193.0	536.5	0.991		3	28.3	3340.0	124.1	0.914		11	17.9	425.3	24.8	0.811	
Polynomial-type	Polynomial-a	4	27.8	19992.0	844.9	0.987		1	44.0	4756.0	108.1	0.978		1	32.0	307.0	9.6	0.833	
	Polynomial-b	2	27.5	15170.0	550.8	0.965		19	35.9	4843.7	144.7	0.928		2	18.0	506.5	27.5	0.910	
Waveform-type	Waveform-a	14	41.4	18410.2	622.6		24.3	7	34.4	6664.1	201.4		19.0	11	23.9	659.8	27.0		6.7
	Waveform-b													32	24.9	381.9	16.1		
Sleeping-beauty-type	Sb-a							5	47.8	10622.0	211.3			6	27.2	319.0	12.0		
	Sb-b																		

Notes: N = Number of publications; TC = Total citations; AC = Annual citations; PA= Peaking age.