A Vector for Measuring Obsolescence of Scientific Articles

Jianjun Sun[1], Chao Min[1] and Jiang Li[2]

[1] *sjj@nju.edu.cn*
School of Information Management, Nanjing University, Nanjing (China)

[1] *marlonmassine@yeah.net*
School of Information Management, Nanjing University, Nanjing (China)

[2] *li-jiang@zju.edu.cn*
Department of Information Resource Management, Zhejiang University, Hangzhou (China)

Abstract
Diachronous studies of obsolescence categorized articles into three general types: "flashes in the pan", "sleeping beauties" and "normal articles", by using quartiles to identify first 25% and last 75% articles reaching 50% of their total citations, or by using averages to define threshold values of sleeping and awakening periods. However, the average-based and quartile-based criteria, sometimes, less effectively distinguished "flashes in the pan" and "sleeping beauties" from normal articles. In this research, we proposed a vector for measuring obsolescence of scientific articles, as an alternative to these criteria. The obsolescence vector is designed as $O = (G_s, A^-, n)$, where n is the age of an article, G_s and A^- are parameters for revealing the shape of citation curves. Among Nobel laureates' 28,340 articles, each of which received over 20 citations, we identified 265 flashes in the pan (approximately 1%) and 40 sleeping beauties (approximately 0.1%) by the obsolescence vector. By a few case studies, it is verified that obsolescence vector yielded more reasonable classifications than did the average-based and quartile-based criteria.

Conference Topic:
Indicators

Introduction
In a previous study (Li et al., 2014), we introduced G_s index, an adjustment of Gini coefficient, for measuring the inequality of "heartbeat spectrum" of "sleeping beauties". "Sleeping beauty" in science was first proposed by van Raan (2004), in order to describe a phenomenon where papers did not achieve recognition in citations until many years after their original publication. As in the fairy tale, a princess (an article) sleeps (goes unnoticed) for a long time and then, almost suddenly, is awakened (receives a lot of citations) by a prince (another article). "Heartbeat spectrum" was defined as a vector of a sleeping beauty's annual citation(s) received in the sleeping period.

How to categorize recognition to a paper as "early", "delayed" or "normal"? Diachronous studies of obsolescence answered this question, by using quartiles to identify first 25% and last 75% articles reaching 50% of their total citations (Costas et al., 2010), or by using averages to define threshold values of sleeping and awakening periods (van Raan, 2004; van Dalen & Henkens, 2005). In this research, we propose an obsolescence vector based on the G_s index, as an alternative to both approaches.

Literature review
"Obsolescence" (or "ageing") studies, in the field of bibliometrics, attempt to answer the question how long does the information in a research paper remain current, by measuring the number of citations the paper received since publication (Cunningham & Bocock, 1995). There are two approaches to measure obsolescence: "synchronous" and "diachronous" distribution (Nakamoto, 1988). They are also referred to as "citations from" and "citations to" approaches (Redner, 2005), or "retrospective citation" and "prospective citation" approaches

(Burrell, 2002; Glänzel, 2004). The former considers the age distribution of references of a paper in a particular year, while the latter analyzes the distribution of citations over time.

A number of metrics has been proposed, from a synchronous perspective, to measure obsolescence of scientific literature. "Half-life" was described (Burton & Kebler, 1960) as "half the active life", which means the time during which one-half of the currently active literature was published. Price (1970) suggested the percentage of references (from all articles) up to five years old as an index to reveal obsolescence of scientific documents, which is also named "Price Index".

From a diachronous perspective, a citation curve (Garfield, 1989; Avramescu, 1979; Li et al., 2014) is the time distribution of citations a paper received. It is also referred to as "life-cycle" (Cunningham & Bocock, 1995), "citation patterns" (Li & Ye, 2014; Wang, Song, & Barabási, 2013; Guo & Suo, 2014; Redner, 2005), or "citation history" (Redner, 2005; ABT, 1981; Persson, 2005; Vlachý, 1985; Costas et al., 2010). A "typical citation curve" describes the history of an article which received a few citations in the first following years after publication, then rose to a citation peak, but afterwards was gradually less cited with time. It is identified that lognormal function best fits typical citation curves (Egghe & Rao, 1992). For most scientific papers, death (no longer being cited by other papers) comes within ten years after publication (Price, 1976). Nevertheless, the minority appears exponential increase in citations in a long time, whose citation curves fit exponential function (Li & Ye, 2014).

The peaking time of citations features the shape of citation curves, reflecting the immediacy of publications. Some articles were noticed immediately after publication but ignored very soon, and hence were named as "flashes in the pan" (van Dalen & Henkens 2005; Costas et al., 2010). Their citations peaked much earlier than typical citation curves. Some went unnoticed for a long time and then, almost suddenly, received a lot of citations, and hence were referred to as "sleeping beauties" (van Raan, 2004), "premature discoveries" (Stent, 1972; Wyatt, 1975), "resisted discoveries" (Barher, 1961) or "delayed recognition" (Cole, 1970). Their citations peaked much later than typical citation curves. Van Raan (2004) suggested three criteria for distinguishing sleeping beauties: (1) they deeply slept (receive at most 1 citation per year on average), or less deeply slept (between 1 and 2 citations per year on average) for a few years after publication; (2) they slept at least five years; and (3) they were awakened by over 20 citations during the four years following the sleeping period. However, the criteria are not always applicable to answer Garfield (1980)'s question how abrupt a citation boost must be to suggest delayed recognition. Moreover, the criteria ignored the citations received after the awakening period (Li, 2014; Li & Ye, 2012).

Different from van Raan's average-based criteria, Costas et al. (2010) used quartiles. They identified the year after publication in which the document received for the first time at least 50% of its citations ("Year 50%"), then calculated, for all documents of the same year of publication in the same field, the percentiles 25 and 75 of the distribution function of the value of "Year 50%", and recorded them as "P25" and "P75". As a result, the articles were categorized into "flashes in the pan" ("Year 50%" <"P25"), "delayed recognition" ("Year 50%" >"P75") and the rest as "normal publications" ("P25"≤"Year 50%"≤"P75"). These criteria considered the whole citation history of articles rather than only sleeping and awakening periods, and avoided the deficiency of van Raan's definitions. However, the excessive percentages of early and delayed recognition identified by these criteria caused the originally rare phenomena normal.

Methodology

Design of the obsolescence vector

Suppose there are seven ten-year old articles whose citation curves are drawn in Figure 1. P_1 is a sleeping beauty who deeply slept for six years (received no citations) but was suddenly awakened by 40 citations in the following four years. P_2 is a flash in the pan, which immediately received 32 citations within the first two years after publication, but was ignored afterwards and rarely received citations. P_3 is a typical citation curve, which reached citation-peak in the fourth year. It was successfully fitted by the lognormal function in the program OriginPro 8 ($R^2 = 0.972$). P_4 is an article whose citations increase exponentially. Exponential function successfully fits the curve with $R^2 = 0.983$. Both P_5 and P_6 are waveform curves, but they have different initial values, hence have distinct normalized curves in Figure 1. P_7 is a horizontal line, and coincides with the 45 degree diagonal in the right side of Figure 1, which is called "the line of equality" and indicates absolutely even distribution.

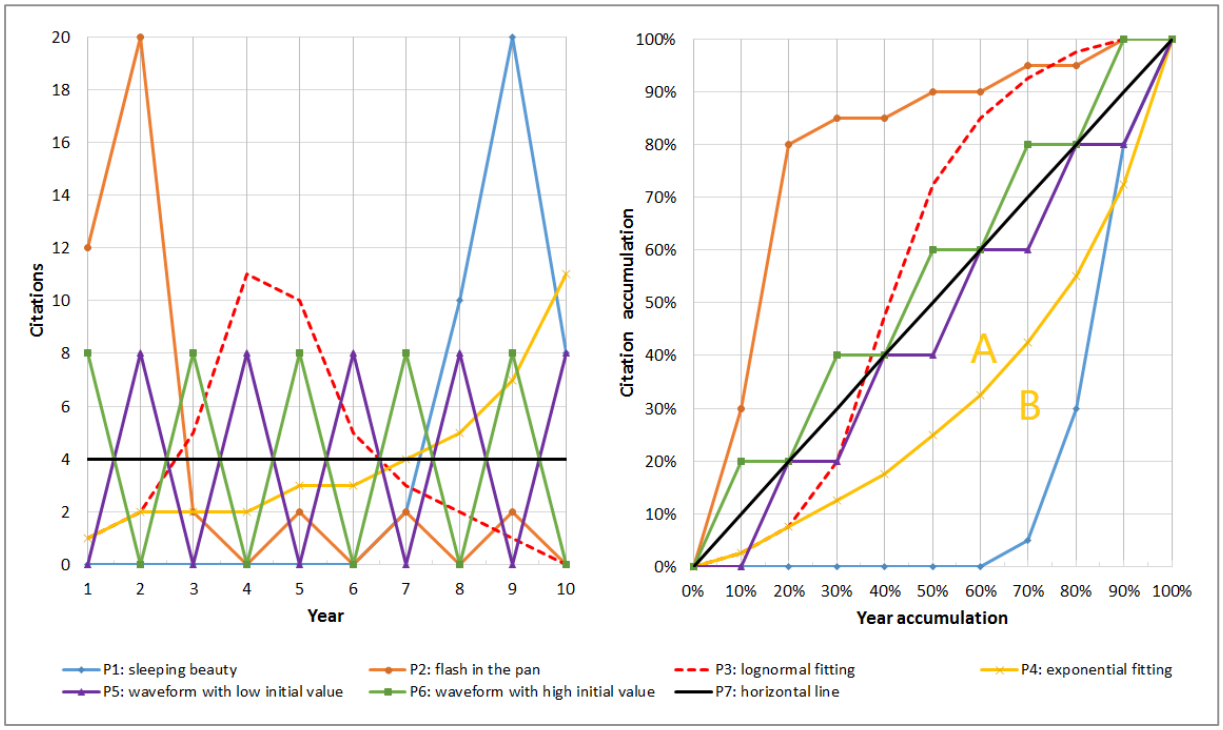

Figure 1. From citation curves to normalized cumulative citation curves of P1-P7 (left: citation curves; right: normalized cumulative citation curves).

The value of G_s, taking P4 as an example, equals to the ratio of the area that lies between the line of equality and the normalized cumulative citation curve (marked A in Figure 1) over the total area under the line of equality (sum of A and B), i.e.,

$$G_s = \frac{A}{A+B}. \quad (1)$$

The normalized cumulative citation curve (hereafter "normalized curve") of P4 is a "Lorenz curve", because the sequence of citations is in an ascending order. Since the areas A and B form an isosceles right triangle, we have

$$A + B = \frac{1}{2}. \quad (2)$$

Thus, putting Eq. (2) into Eq. (1), we have

$$G_s = 2A. \quad (3)$$

The calculation of G_s is determined by the calculation of the area B which can be divided into several trapeziums and a triangle. In this study, we remain the expression of the segment function of G_s in our previous study (Li et al., 2014),

$$G_s = \begin{cases} 1 - \frac{2\times[n\times c_1+(n-1)\times c_2+\cdots+c_n]-C}{C\times n}, & C > 0 \\ 1, & C = 0 \end{cases} \quad (4)$$

but redefine the parameters. In the new definition, n is the age of a paper, C is the total number of citations the paper received during the n years, and $c_i (i \in \{1,2,\cdots,n\})$ is the number of citations the paper received in the i^{th} year after publication. Here, $Gs \in (-1,1]$ and depends on the age (n) of articles. The value of G_s gradually approaches to -1, if the article no longer receives citations.

The value of G_s, to certain extent, characterizes the shape of citation curves:

(1) large G_s indicates delayed recognition, while small G_s denotes early recognition, as P_1 and P_2 shown in Table 1;

(2) $G_s < 0$ implies that there exists leaping early in citation curves, for example, both P_2 and P_6 received a large number of citations immediately after publication, while P_3 has a fast rising period although it does not have immediacy; and

(3) $G_s = 0$ suggests a horizontal citation curve (as P_7), or a citation curve including at least one high-citation period (to guarantee $A^- < 0$) which is offset by at least one low-citation period.

The value of A is not always positive. For P_2, $A<0$, since its normalized curve in Figure 1 is above the line of equality. Since

$$A = A^+ + A^- , \quad (5)$$

putting Eq. (5) into Eq. (3), we have

$$A^- = \frac{1}{2}G_s - A^+. \quad (6)$$

A^+ is the area between the line of equality and the normalized curve under the line of equality. Similar to the calculation of G_s, we calculate A^+, and accordingly have the value of A^-. In case of P_3, the intersection of the normalized curve and the line of equality in Figure 1 exists in between the accumulation year 30% and 40%. Therefore, there is a minor error (a difference) between the output and target of A^+ values of P_3. In cases of P_1, P_4 and P_5, there is no error in the calculation of A^+.

The fast rising period of a citation curve is hidden from the value of G_s if $A^- < 0 < A^+$. In case of $A^+ = 0$, we have

$$A^- = A = \frac{1}{2}G_s. \quad (7)$$

Hence, the value of A^- provides complementary explanation to the shape of citation curves:

(1) recognition to the article is normal or delayed rather than early if $A^-=0$;

(2) there exists leaping in the citation curve of the article if $A^-<0$; and

(3) citation leaping appears early if $A^- = \frac{1}{2}G_s$.

We propose a vector for measuring obsolescence of scientific articles: $O=(G_s, A^-, n)$, where G_s is an index revealing the history of citations, A^- is a parameter uncovering citation leaping and age n is an adjusting parameter. We calculated the obsolescence vectors for P_1-P_7 as shown in Table 1.

Table 1. Obsolescence vectors for P1-P7.

Article	*Citation curve*	*Citations*	*A*	A^+	*Obsolescence vector*		
					G_s	A^-	*n*
P1	Sleeping beauty	40	0.335	0.335	0.670	0.000	10
P2	Flash in the pan	40	-0.300	0.000	-0.600	-0.300	10
P3	Lognormal fitting	40	-0.075	0.028	-0.150	-0.103	10
P4	Exponential fitting	40	0.183	0.183	0.365	0.000	10
P5	Waveform with low initial value	40	0.050	0.050	0.100	0.000	10
P6	Waveform with high initial value	40	-0.050	0.000	-0.100	-0.050	10
P7	Horizontal line	40	0.000	0.000	0.000	0.000	10

Criteria for categorizing the patterns of obsolescence

In this research, we use the terms "flashes in the pan", "sleeping beauties" and "normal articles" as the patterns of obsolescence, but provide three different approaches for measurement, in order to characterize obsolescence vector. We remain van Raan's average-based criteria in the first approach. By following the criteria, we define variables for "flashes in the pan": "noticed" (van Dalen and Henkens, 2005) as receiving over 10 citations, "ignored" as receiving less than two citations per year on average and "immediately" as within two years since publication. We also define the duration of light disappearing for at least five years, since a flash is likely to reappear. Then, we suggest average-based criteria as follows:

flashes in the pan (F_1): articles which received more than 10 citations in the first two years since publication, and then in the next five years received no more than 2 citations per year on average;

sleeping beauties (S_1): articles which received no more than 2 citations per year on average in the first five years since publication, and then in the next four years received more than 20 citations; and

normal articles (N_1): which neither satisfy the criteria for F_1 nor for S_1.

The second approach uses quartiles. We adjust "relative ranking in a field" in Costas et al. (2010) to "relative age", since the former requires the population of articles in a filed which involves a huge dataset. Thus, for a single article, we record the percentiles 25 and 75 of its age as "A25" and "A75". Then, we define quartile-based criteria for the patterns of obsolescence as follows:

flashes in the pan (F_2): articles that reached "Year 50%" within 25% of its age, i.e., "Year 50%" <"A25";

sleeping beauties (S_2): articles that reached "Year 50%" with the time exceeding 75% of its age, i.e., "Year 50%" >"A75"; and

normal articles (N_2): which neither satisfy the criteria for F_1 nor for S_1, i.e., "A25"≤"Year 50%"≤ "A75".

Based on the obsolescence vectors of the seven cases in Table 1, we propose new criteria for categorizing the patterns of obsolescence as follows,

flashes in the pan (F_3): $G_s \leq$ -0.6 and $A^- = \frac{1}{2} G_s$;

sleeping beauties (S_3): $G_s \geq 0.6$ and $A^- = 0$; and

normal articles (N_3): which neither satisfy the criteria for F_3 nor for S_3.

Data

A dataset was prepared to make comparisons of the above three sets of criteria, and to verify the efficiency of the proposed obsolescence vector. From the Web of Science, we collected 58,963 articles of 629 Nobel Prize winners during the period of 1901-2012, in the fields of Chemistry, Physics, Physiology or Medicine, and Economic Sciences. The definition S_2 requires that a sleeping beauty should have more than 20 citations. For the purpose of comparisons, we eliminated articles, which received no more than 20 citations, and remained a collection of 28,340 articles published between 1900 and 2000. Then, we searched the number of annual citations to these articles up to 2011 in the Web of Science. Thus, every article in this collection aged at least eleven, which is sufficient for a sleeping beauty with the shortest sleeping period to be awakened.

Results

Obsolescence vector as an alternative to average-based and quartile-based criteria

The life-cycles of most articles in the dataset have already drawn to their close. As shown in Table 2, the peak of G_s distribution appears in the interval (-0.4,-0.2] and the values of G_s for 84.3% articles are negative. Moreover, 95.0% of the articles have $A^-<0$. Small G_s values (minus) indicate the end of cife-cycles, as shown by article P_2 in Figure 1. It is calculated that 68.4% of the articles with $G_s > 0$ have $A^- < 0$. Thus, there are only a small fraction of citation curves having the shape of P_1, P_4 and P_5 in Figure 1. What they have in common is that there is no citation rise and fall in the initial stage of citation curves. The rise and fall of citations must be a citation leaping or like a lognormal shape. Articles with the largest and smallest G_s values are categorized into sleeping beauties (S_3) and flashes in the pan (F_3), respectively. The obsolescence vector for the former (Rayleigh, 1914) is O = (0.892, 0, 98). Although published as early as in 1914, it received no citations until 1992. It does not satisfy S1, since it was not awakened by more than 20 citations within four years after sleeping period. However, it satisfies S_2, since recognition to it was delayed to the last four years of its age. This example reveals the deficiency of S_1. The latter (Ryle & Bailey, 1968) has an obsolescence vector O = (-0.960, -0.480, 44). The article received 26 citations immediately in the publication year, but the number rapidly fell to zero four years later and it was never cited till the end. It satisfies both F_1 and F_2.

Table 2. Comparisons of the three approaches to measuring obsolescence.

G_s	N	$N(A^-<0)$	F_1	S_1	F_2	S_2	F_3	S_3	$F_1\&F_3$	$F_2\&F_3$	$S_1\&S_3$	$S_2\&S_3$
(-1,-0.8]	494	494	41	0	489	0	265	0	34	262	0	0
(-0.8,-0.6]	3,897	3,897	62	6	3,856	0	1,734	0	57	1,704	0	0
(-0.6,-0.4]	6,808	6,808	30	16	5,250	0	0	0	0	0	0	0
(-0.4,-0.2]	7,213	7,213	21	22	985	0	0	0	0	0	0	0
(-0.2,0]	5,477	5,477	7	25	25	0	0	0	0	0	0	0
(0,0.2]	2,894	2,344	7	27	0	15	0	0	0	0	0	0
(0.2,0.4]	1,140	543	5	26	0	228	0	0	0	0	0	0
(0.4,0.6]	348	141	2	7	0	304	0	0	0	0	0	0
(0.6,0.8]	65	17	1	1	0	65	0	37	0	0	1	37
(0.8, 1)	4	0	0	0	0	4	0	3	0	0	0	3
Total	**28,340**	**26,934**	**176**	**130**	**10,605**	**616**	**1,999**	**40**	**91**	**1,966**	**1**	**40**

It seems that the condition $G_s \leq$ -0.6 and $A^- = \frac{1}{2} G_s$ for flashes in the pan is a loose condition, since it yields 1,999 flashes in the pan in the dataset. If it is intensified to be $G_s \leq$ -0.8 and $A^- = \frac{1}{2} G_s$, the number of flashes in the pan shrinks to 262, closer to the result of criterion F_1. Considering that 81.6% of the articles aged over 20, we suggest the criterion for flashes in the pan be $G_s \leq$ -0.8 and $A^- = \frac{1}{2} G_s$ on condition that n≥20.

The criterion S_3 for sleeping beauties is more stringent than S_1 and S_2, and selected only 40 qualified articles from the dataset. The 40 articles is a subset of the collection by S_2, but covers 39 articles out of the collection by S_1. In Table 2, there are six articles satisfying S_1 whose G_s values exist in the interval (-0.8, -0.6]. For example, the article in Figure 2 received only nine citations within the first five years after publication, but suddenly received 25 citation in the following four years. It also satisfies S_2, since it reached "Year 50%" within ten years (13.9% of its age) after publication. Nevertheless, this article is more like a "typical citation curve" which spent seven years to gradually reach citation-peak and slowly declined to death afterwards. The obsolescence vector of this article is O = (-0.648, -0.324, 72) which does not satisfy S_3. Moreover, we identified 3,897 articles of its kind, which have $G_s \in$ (-0.8, -0.6]. Therefore, it is more reasonable to categorize it as a "normal article" rather than a "sleeping beauty".

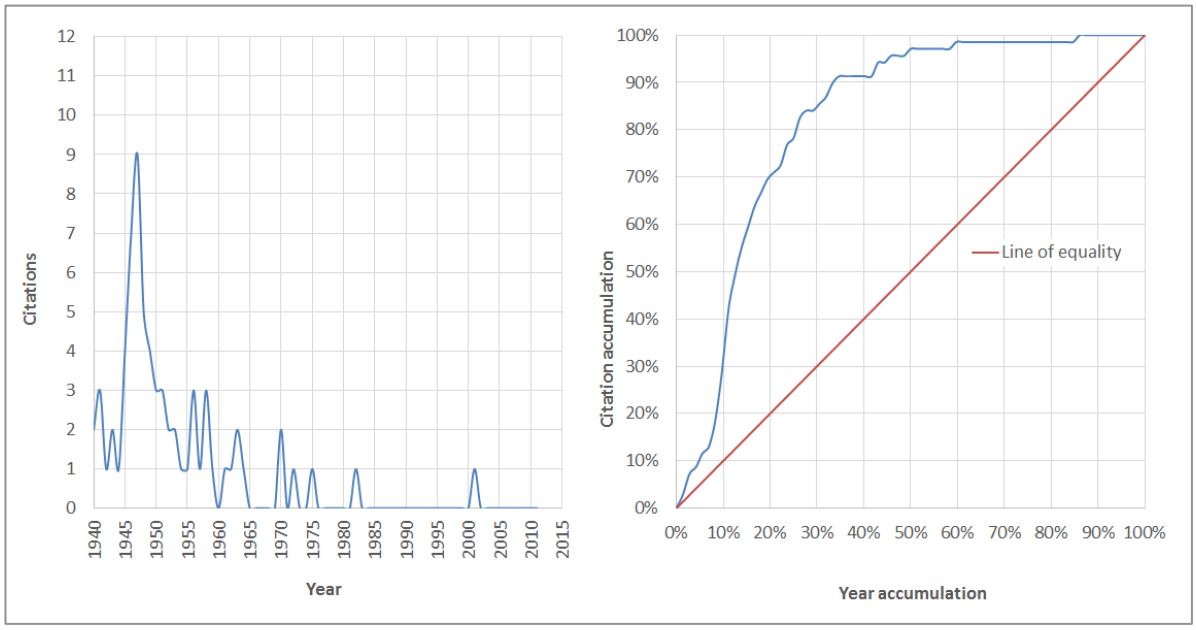

Figure 2. A sleeping beauty by average-based and quartile-based criteria, but a normal article by obsolescence vector (Landsteiner, 1940).

Citation-curve differences of obsolescence

The calculation of G_s values, sometimes, remains citation leaping under cover. As shown in Figures 3, Zewail's and Corey's articles were published in the same year of 2000, and have the same G_s values 0.083. However, they received different citations and have different citation curves. The obsolescence vector of the two articles are O=(0.083, 0, 12) and O=(0.083, -0.004, 12), respectively. Due to the citation leaping since 2007, the normalized curve of Corey's article in Figure 3 surpassed the line of equality in 2010 and yielded $A^- < 0$ which does not appear in the normalized curve of Zewail's article. Therefore, it is a sign of citation leaping to have A^-<0.

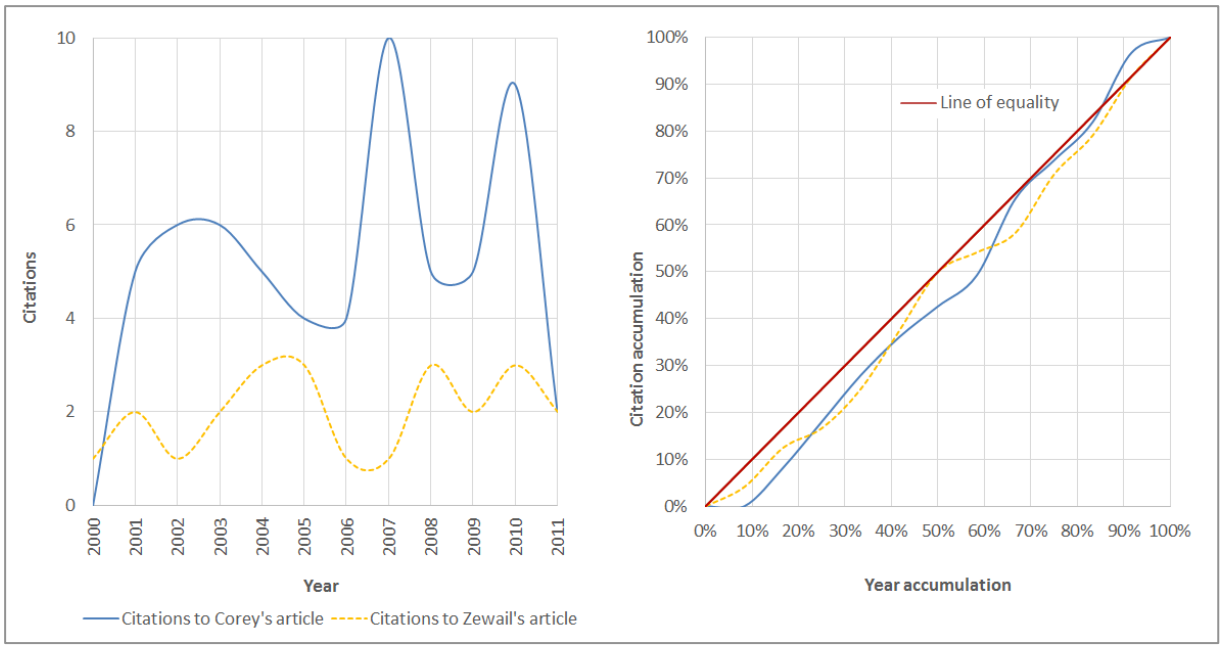

Figure 3. Zewail's article with O = (12, 0.083, 0) and Corey's article with O = (12, 0.083, -0.004).

Age differences of obsolescence

The years of 1950, 1990 and 2000 were selected for the publication years for sampling articles, in order to explore age differences of obsolescence. They were aged 62, 22 and 12, respectively. It appears that older articles have smaller G_s values while younger ones have larger G_s values. It is clear in Table 3 that the peak of G_s distribution among the intervals shifted from (-0.6, -0.4] in 1950, to (-0.4,-0.2] in 1990, even to (-0.2, 0] in 2000. Most of the old articles have been ignored and receive rare or no citations after recognition, similar to the example in Figure 2. Therefore, their G_s values gradually decline. It is hence identified that age exerts significant influence on the values of G_s.

Table 3. Age differences of obsolescence.

G_s	*Year* 1950		*Year* 1990		*Year* 2000	
	N	*N*(A^-<0)	*N*	*N*(A^-<0)	*N*	*N*(A^-<0)
[-1,-0.8]	11	11	12	12	0	0
(-0.8,-0.6]	65	65	45	45	8	8
(-0.6,-0.4]	66	66	190	190	31	31
(-0.4,-0.2]	42	42	250	250	81	81
(-0.2,0]	28	28	148	148	216	216
(0,0.2]	22	16	80	68	173	117
(0.2,0.4]	8	3	27	9	46	10
(0.4,0.6]	6	0	5	2	8	1
(0.6,0.8]	0	0	0	0	0	0
(0.8, 1]	0	0	0	0	0	0
Total	**248**	**231**	**757**	**724**	**563**	**464**

Disciplinary differences of obsolescence

The obsolescence of economic sciences is slower than that of fundamental sciences, including chemistry, physics and physiology & medicine. It is a sign of slow obsolescence to have more positive G_s values and less $A^- < 0$. In Table 4, the distribution of G_s values of economic sciences peaked in the interval (0, 0.2], while in other disciplines, it peaked in the interval (-0.4,-0.2] or (-0.6,-0.4]. The percentage of $A^- < 0$ in positive G_s values is only 50.4%, far less

than 69.8-75.8% in fundamental sciences. Moreover, older articles tend to have higher absolute G_s values, in each of the four disciplines.

Table 4. Disciplinary differences of obsolescence

G_s	*Chemistry*			*Physics*			*Physiology & Medicine*			*Economic sciences*		
	N	*N(A-<0)*	*Age*	*N*	*N(A-<0)*	*Age*	*N*	*N(A-<0)*	*Age*	*N*	*N(A-<0)*	*Age*
[-1,-0.8]	34	34	56.1	124	124	36.4	336	336	51.0	0	0	0.0
(-0.8,-0.6]	625	625	49.8	653	653	35.1	2,615	2,615	45.9	4	4	38.3
(-0.6,-0.4]	1,727	1,727	41.4	1,185	1,185	33.2	3,850	3,850	41.0	44	44	36.2
(-0.4,-0.2]	2,690	2,690	37.5	1,212	1,212	35.0	3,193	3,193	36.2	118	118	36.8
(-0.2,0]	2,236	2,236	35.3	1,008	1,008	34.6	1,972	1,972	30.7	263	263	35.6
(0,0.2]	1,099	926	39.3	576	483	42.2	730	594	34.5	489	341	30.0
(0.2,0.4]	307	161	53.9	289	180	58.9	155	78	49.8	389	124	28.2
(0.4,0.6]	67	34	71.1	147	63	71.9	33	13	60.4	101	31	37.2
(0.6,0.8]	10	3	90.5	38	10	86.9	5	0	47.2	12	4	52.3
(0.8, 1]	0	0	0.0	4	0	90.0	0	0	0.0	0	0	0.0
Total	**8,795**	**8,436**		**5,236**	**4,918**		**12,889**	**12,651**		**1,420**	**929**	

Discussion

Further discussion on $A^- < 0$

Significant citation leaping is likely to result in recurring appearance of $A^-<0$ area. For example of Hsu et al.'s article (1997), citation leaping appeared twice in the citation curve. The first citation peak appeared in 1998, the second year after publication, which led the normalized curve to reach the line of equality. In 1999, the article received six citations. The normalized curve hence surpassed the line of equality. However, the citation leaping disappeared afterwards, and the normalized curve dropped under the line of equality. Nevertheless, the second citation peak, higher than the first one, appeared in 2005 and boosted the normalized curve above the line of equality again. Comparing this example with the supposed waveform citation curves, i.e., P_5 and P_6 in Figure 1, it is identified that the appearance of $A^-<0$ area is originated by citation leaping. Furthermore, double appearance of $A^-<0$ area indicates double citation leaping in which the first one happened immediately after publication and the second one is higher. However, the characteristics of double or multiple appearance of $A^-<0$ area are not in consideration of the new designed obsolescence vector, since the number of this kind is rare.

Limitations

The obsolescence vector cannot differentiate two citation curves if there is multiplier relationship between their annual citations. For example, both (0, 8, 0, 8, 0, 8, 0, 8, 0, 8) and (0, 4, 0, 4, 0, 4, 0, 4, 0, 4) have the same obsolescence vector O=(0.1, 0, 10). The obsolescence vector is applicable to categorize articles into "flashes in the pan", "sleeping beauties" or "normal articles", by distinguishing citation leaping in citation curves. It does not characterize citation history of "normal" articles, which account for a large percent. As normal articles, P_3-P_6 in Figure 1 have entirely different obsolescence patterns. However, they cannot be uncovered by obsolescence vector.

It is controversial whether someone who won a major prize has received increased citations on all his/her work (Hugget, 2013; Mazloumian et al., 2011). However, the results are generalized from articles of Nobel laureates rather than randomly sampled authors, and hence are potentially biased. In addition, "recognition" is referred to as a large number of citations,

e.g., 20. Thus, whether the obsolescence vector is applicable to articles receiving less than 20 citations requires further research.

Conclusions

We proposed a vector for measuring obsolescence of scientific articles, $O = (G_s, A^-, n)$, where n is the age of an article, G_s and A^- are parameters for the shape of the article's citation curves. By distinguishing inequality of citation distribution, obsolescence vector is applicable to categorize articles into three general types:

flashes in the pan: $G_s \leq$ -0.8 and $A^- = \frac{1}{2} G_s$ for n≥20 or $G_s \leq$ -0.6 and $A^- = \frac{1}{2} G_s$ for n<20;
sleeping beauties: $G_s \geq 0.6$ and $A^- = 0$; and
normal articles: which neither satisfy the criteria for F_3 nor for S_3.

The age, subject category and citation curve of articles exert significant influence on G_s values. Older articles tend to have higher absolute G_s values. The criterion for "flashes in the pan" is adjustable in terms of the age of articles. In case of articles younger than, e.g., ten years old, as shown in Figure 1, it is feasible to mildly adjust the criterion as $G_s \leq$ -0.6. Disciplinary differences exist in the proposed obsolescence vector. Articles in economic sciences appear higher G_s values than those in fundamental sciences, including chemistry, physics and physiology & medicine. In case of articles receiving no more citations, their G_s values tend to decline, till to -1.

As an alternative to average-based and quartile-based criteria, the obsolescence vector avoided overlooking the period after sleeping beauties being awakened, and tightened the loose conditions by using quartiles. By obsolescence vectors, we identified 265 flashes in the pan (approximately 1%) and 40 sleeping beauties (approximately 0.1%), among 28,340 articles of Nobel laureates, which receive more than 20 citations by the year of 2011. The low percentages of flashes in the pan and sleeping beauties remained them rare phenomena.

Acknowledgement

This research was financially supported by the National Natural Science Foundation of China (NSFC No. 71203193 and 71273125).

References

ABT, H. A. (1981). Long-term citation histories of astronomical papers. *Publications of the Astronomical Society of the Pacific*, *93*, 207-210.

Avramescu, A. (1979). Actuality and obsolescence of scientific literature. *Journal of the American Society for Information Science, 30*(5), 296-303.

Burrell, Q. L. (2002). The nth-citation distribution and obsolescence. *Scientometrics*, *53*(3), 309-323.

Burton, R. E., & Kebler, R. W. (1960). The "half-life" of some scientific and technical literatures. *American Documentation*, *11*(1), 18-22.

Cole, S. (1970). Professional standing and the reception of scientific discoveries. *American Journal of Sociology*, *76*, 286–306.

Costas, R., van Leeuwen, T. N., & van Raan, A. F. J. (2010). Is scientific literature subject to a "sell-by-date"? A general methodology to analyze the "durability" of scientific documents. *Journal of the American Society for Information Science and Technology*, *61*(2), 329–339.

Cunningham, S. J., & Bocock, D. (1995). Obsolescence of computing literature. *Scientometrics*, *34*(2), 255-262.

Egghe, L.,, & Rao, I. K. R. (1992). Citation age data and the obsolescence function: Fits and explanations. *Information and Processing Management, 28*(2), 201-217.

Garfield, E. (1980). Premature discovery or delayed recognition-why? *Current Contents*, *4*, 488-493.

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*, *38*, 3-8.

Guo, J. L., & Suo, Q. (2014). Comment on" Quantifying Long-term Scientific Impact". *Science*, *345*(6193), 149.

Hugget, S. (2010). Does a Nobel Prize lead to more citations. *Research Trends*, Retrieved April 7, 2015 from http: //www.researchtrends.com/issue20-november-2010/does-a-nobel-prize-lead-to-more-citations.

Hsu, C. P., Song, X.,, & Marcus, R. A. (1997). Time-dependent Stokes shift and its calculation from solvent dielectric dispersion data. T*he Journal of Physical Chemistry B*, *101*(14), 2546-2551.

Li, J. (2014). Citation Curves of "All-elements-sleeping-beauties": "Flash in the Pan" first and then "Delayed Recognition". *Scientometrics*, *100*(2), 595-601.

Li, J., & Ye, F. Y. (2012). The phenomenon of all-elements-sleeping-beauties in scientific literature. *Scientometrics*, *92*(3), 795–799.

Li, J., & Ye, F. Y. (2014). A Probe into the Citation Patterns of High-quality and High-impact Publications. *Malaysian Journal of Library and Information Science*, *19*(2), 31-47.

Li, J., Shi, D., Zhao, S. X., & Ye, F. Y. (2014). A study of the "heartbeat spectra" for "sleeping beauties". *Journal of Informetrics*, *8*(3), 493-502.

Mazloumian, A., Eom, Y., Helbing, D., Lozano, S., & Fortunato, S. (2011). How citation boosts promote scientific paradigm shifts and nobel prizes. *Plos One*, *6*(5), e18975.

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

Persson, O. (2005). "Citation Indexes for Science"-A 50 year citation history. *Current Science*, *89*(9), 1503-1504.

Price, D. (1970). Citation measures of hard science, soft science, technology, and non-science. In C. E. Nelson & D. K. Pollock (Eds.), *Communication among Scientists and Engineers* (pp. 3-22). Lexington, MA: Heath.

Price, D. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, *27*(5), 292–306.

Rayleigh, L. (1914). On the theory of long waves and bores. Proceedings of the royal society of London series A- Containing papers of a mathematical and physical character, 90(619), 324-328.

Redner, S. (2005). Citation statistics from more than a century of physical review. *Physics Today*, *58*(1), 49–54.

Stent, G. S. (1972). Prematurity and uniqueness in scientific discovery. *Scientific American*, *227*(6), 84–93.

van Dalen, H. P., & Henkens, K. (2005). Signals in science – On the importance of signaling in gaining attention in Science. *Scientometrics*, *64*(2), 209–233.

van Raan, A. F. J. (2004). Sleeping beauties in science. *Scientometrics*, *59*(3), 467–472.

Vlachý, J. (1985). Citation histories of scientific publications. The data sources. *Scientometrics*, *7*(3), 505-528.

Wang, D., Song, C., & Barabási, A. L. (2013). Quantifying long-term scientific impact. *Science*, *342*(6154), 127-132.

Wyatt, H. V. (1961). Knowledge and prematurity-journey from transformation to DNA. *Perspectives in Biology and Medicine, 18*(2), 149-156.