Journal of Informetrics 10 (2016) 1153–1165

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Measuring delayed recognition for papers: Uneven weighted summation and total citations

Chao Min[a,b], Jianjun Sun[a,*], Lei Pei[a], Ying Ding[b]

[a] School of Information Management, Nanjing University, Nanjing 210023, China
[b] School of Informatics and Computing, Indiana University, Bloomington, IN, USA

ARTICLE INFO

Article history:
Received 29 February 2016
Received in revised form
30 September 2016
Accepted 1 October 2016
Available online 20 October 2016

Keywords:
Delayed recognition
Sleeping beauty
Citation histories
Yearly citations
Citation delay
Weighted summation

ABSTRACT

This paper studies the so-called abnormal phenomenon of delayed recognition in bibliometrics and focuses on the first step in quantitatively measuring this phenomenon. As bibliometric analysis of a paper's recognition and influence is an uncertain and extended process, proper calculation of delayed recognition and "sleeping beauty" publications has limitations in current scientometric studies, such as restricted application indicators, scope, and complex calculation methods. This study suggests a solution for depicting the citation delay phenomenon of individual papers that avoids dividing them into different periods, is applicable to all papers with various types of citation curves, and is easy to calculate. Notably, this approach advocates using an uneven weighted summation based on earlier and later citation years when analyzing an individual paper's citation data. It demonstrates that the intrinsic relation between two independent indicators of citation delay and Gs index is based on the same logic of applying uneven weights to sum up yearly citations. This paper also recommends that simultaneous application of the new indicator D_a and final citation numbers can efficiently identify those delayed recognition papers, and that the criterion for selecting papers can be adjusted by the value of a.

1. Introduction

According to Kuhn's (1962) early model of scientific development, the sciences always face episodic anomalies that are influenced by sociological factors and illogical determinant procedures that may impact the reception of new discoveries. "Normal" progress in the sciences is thus viewed as a process of "development-by-accumulation" of accepted theories, but if new theories contradict or undermine normal science and its standard traditions, new questions may be asked of old data. This can eventually create radical paradigm shifts or even scientific revolutions. Research shows that the vast majority of cited papers reach a citation peak fairly quickly (Moed, Burger, Frankfort, & van Raan, 1985), generally two to six years after publication (Amin & Mabe, 2003). But some papers take a significant amount of time to reach their citation peaks, a phenomenon in scientific literature referred to in early studies as "resisted discoveries" (Barber, 1961), "scientific prematurity" (Stent, 1972), and the "Mendel syndrome" (Costas, van Leeuwen, & van Raan, 2011; Garfield, 1979), which refers to delayed recognition papers. More recently the term "sleeping beauty" was coined (van Raan, 2004, 2015) to describe this delayed recognition phenomenon where the significance of an important article is not recognized until a long time after its publication. Sleeping beauties often "wake up" in a gradual way and create milder impacts rather than sudden citation

* Corresponding author at: School of Information Management, Nanjing University, Nanjing 210023, China.
E-mail address: sjj@nju.edu.cn (J. Sun).

http://dx.doi.org/10.1016/j.joi.2016.10.001
1751-1577/

bursts that draw immediate attention (Li & Shi, 2016). While these various terms are used differently in the literature, for the purposes of this paper, a "delayed recognition" paper refers to a work that achieves influence and recognition long after its publication, in contrast to a "delayed citation" paper whose citations occur long after publication but that does not achieve distinction.

The unexpected citation distribution of delayed papers, which generally exhibits a dormancy period of a few years followed by a period of lifting citation counts, is an exception to the bibliometric rule of "cumulative advantage" as proposed by Price (1976). While papers with delayed recognition are considered to be rare, their degree of rarity has been debated based on how citations are analyzed. The actual reality of significant delayed recognition anomalies was identified by Glänzel (2008) to be the first of seven "myths" in bibliometrics because the particular choice of a standard citation window is often considered to be responsible for possible negative results of what would otherwise be seen as "correct" bibliometric evaluation.

The proper calculation of delayed recognition publications is generally out of reach of current scientometric indicators. Since article citations accumulate dynamically and continuously, just as the perceived value or influence of the articles varies over time, analysis of a paper's recognition and influence can also be an uncertain and extended process. This raises questions for scientometricians and informetricians as well as scientists concerning the basic characteristics of these special papers whose measurement approaches might be used to properly analyze or anticipate patterns within the delayed recognition phenomenon.

Many studies (Costas, van Leeuwen, & van Raan, 2010; Garfield, 1989; Glänzel, Schlemmer, & Thijs, 2003; Ke, Ferrara, & Radicchi, 2015; Li & Shi, 2016; van Raan, 2004) have paid special attention to the two periods of dormancy and awaking in delayed recognition papers, either by proposing several parameters with thresholds or trying to use a single standard to quantitatively measure this phenomenon. The outcomes of these approaches can be rough inspections of these two periods or black and white divisions between delayed papers and normal papers. In this study, we focus on the very first step of the quantitative measurement that relates to a paper's time of reception and level of recognition. We aim to make the following contributions to the literature by providing:

- A review of existing measures for the phenomenon of delayed recognition, with special focus on the intrinsic relation between two independent indicators;
- A convenient framework for analyzing the phenomenon of delayed recognition; and
- A simpler and more practical method for identifying delayed papers.

Using yearly citation data of articles by 629 Nobel Prize winners in science and economics, we propose that the two distinctive factors of a paper's (1) citation delay time table and (2) ultimate citation numbers should be considered simultaneously when identifying delayed recognition phenomenon. While previous studies have viewed delayed recognition from a primary angle, our two-factor solution for ranking papers provides a better measurement, where both are basic characteristics of all papers.

To quantify the extent to which a paper's citations are delayed, we propose a simple approach of assigning uneven weights to yearly citation counts based on which citations are to be emphasized and adding them up. This approach is useful for identifying nuances behind the phenomenon of delayed recognition, where the weights can be adjusted to emphasize the importance of certain citations of interest. Another important factor is determining how much recognition the paper has achieved on the whole, usually in the form of total citations. Instead of using thresholds to provide a definite boundary between delayed and normal papers, we only need to rank papers by the extent of delay and total citation counts in order to find potential "sleeping beauties." This allows for a better understanding of the salient features of the delayed recognition phenomenon, and can be applied to all papers with diverse types of citation curves.

2. Literature review

Quantitative measures of delayed recognition have been extensively addressed since Garfield's call in the 1980s for parameters to be set for what truly qualifies as delayed recognition (or "premature discovery") of scientific discoveries, which are affected by the scientists' own communication skills (Garfield, 1980; Garfield, 1989). But most of these quantitative measures were proposed 20 years later in separate components. The use of citation measurement based on averages (Glänzel et al., 2003; Li, Shi, Zhao, & Fred, 2014; van Raan, 2004) were first considered, where the citation history of a paper was divided into the standard dormancy and awakening periods. The dormancy period identified in this literature is usually three or five years long (Garfield, 1989; Glänzel et al, 2003; Li et al., 2014) while the awakening period is at least four years or longer (Glänzel et al., 2003; Li et al., 2014; van Raan, 2004). When a paper receives fewer citations per year than its lowest threshold (e.g. one or two citations) during the first period of dormancy, and more citations per year than a higher threshold (e.g. at least five citations) in the second period of awakening, it is viewed as a delayed recognition paper. These approaches can certainly identify some sleeping beauties, but may be unsatisfactory measures when considering how the division of different periods and adoption of citation thresholds is achieved. Notably, they may lead to the omission of some qualified papers, and as the citation histories of papers differ, neither dormancy nor awakening periods would appear at a fixed time point and last for a constant period after a paper's publication.

Costas et al. (2010) suggested using percentiles to differentiate "early," "delayed," or "normal" citation patterns of individual papers. Three parameters were calculated using yearly quartiles to identify the first 25% (P25) and the last 75% (P75) of articles that reached 50% (Year 50%) of their total citations. According to this approach, a paper is regarded as a delayed paper when its Year 50% >P75, and as a "flash in the pan" or short-lived renown when its Year 50% <P25. Obviously this approach requires a great deal of calculation, and is also a somewhat loose criterion in that it leads to a relatively large percentage of delayed recognition papers, nearly 20% of those found in the dataset. This deviates from the generally accepted notion that delayed recognition is a rather rare phenomenon in science (Lachance & Larivière, 2014).

A common approach in these studies is to differentiate delayed papers from normal papers based on citation distribution thresholds. For a given paper, however, this application can only answer whether a paper is delayed or not; it can't clarify how much a paper's recognition is delayed. Li et al. (2014) and Ke et al. (2015) took a different route in using citation curves, one that is relevant to this paper. Li et al. (2014) proposed an adjustment of the Gini Coefficient or Gini ratio (the most commonly used economic measure of income inequality), denoted as the *Gs* index, to investigate the inequality of the "heartbeat spectrum" or annual vector of sleeping beauties' annual citations during their dormancy period. They not only provided the graphical representation of the *Gs* index, but also deduced the calculation function of the new index based on its graph. Their conclusion is quite enlightening, that a paper with a *Gs* value existing in the interval (0.2, 0.6] is more likely to be "aroused" or awakened and achieve recognition status. Although the *Gs* index developed by Li et al. (2014) was applied only to the dormancy period of papers, it grasps an important feature of delayed recognition, that is, the uneven temporal distribution of a paper's citations. The *Gs* index thus shows a potential to expand its application range to the whole citation curve.

Another graph-based approach is Ke et al.'s (2015) "beauty coefficient" which was introduced to quantify the level to which a given paper could be considered as a sleeping beauty. They drew a reference line between the first point and the highest point in the citation curve of a paper, denoted as t_0 and t_m, respectively, to indicate a paper's publication year and the year when it received the maximum number of citations. The beauty coefficient was obtained by adding up the differentials between the reference line and the citation curve from its publication year to the year of t_m. Results found that the later the citations were accumulated, the higher the beauty coefficient value, ranging from −12.02 to 11,600 in this dataset. The awakening time was also defined by Ke et al. (2015) as the year before t_m that maximized the distance from the points in the citation curve to the reference line. On the one hand, Ke et al.'s results are in excellent agreement with Redner's (2005) study, and the method they use to identify a paper's awakening year is revelatory. On the other hand, their measure, as they recognized, only examines the citation curve before its peak and ignores any subsequent curve. This may result in an incomplete estimation of the paper's overall citation history and ultimate recognition.

Li and Shi (2016) just introduced an approach based on an exponential equation to investigate the awakening time of sleeping beauties in "genius" works. This approach works very well with these genius papers, whose citations can generally be expected to increase exponentially. But the awakening time of these papers is still dependent on an external parameter *k* that is open but indefinite. Furthermore, sample papers in the study's empirical analysis exhibit the same citation history pattern of smooth exponential curves. The applicability of this approach to other citation curves thus requires further investigation. In addition Wang (2013), constructed the indicator "citation speed" based on accumulative citation percentages of a single paper, and further defined the indicator "citation delay" to measure how fast citations accumulate for a single paper (Wang, Thijs, & Glänzel, 2015). These two indicators consider whole citation history and are applicable to all papers although they were not designed specifically for measuring delayed recognition.

3. Data

Following the practice of Li et al. (2014), the empirical analysis in this paper is based on a dataset of essays by Nobel Prize winners from Thomson Reuters' database Web of Science. During the period of 1901–2012, the Nobel Prizes in Chemistry, Physics, Physiology or Medicine, and Economic Sciences were rewarded to 163, 194, 201, and 71 scientists, respectively. Their publications from 1900 to 2000 in the Web of Science database were collected. Name disambiguation was done by manually scrutinizing the institution and research background information of each laureate on the official website of the Noble Prize (http://www.nobelprize.org/). Detailed citation data of these papers up to 2011 were also collected, thus giving each of these papers a citation window of at least 11 years. As a result, we obtained 58,963 papers and their citation data. To enhance the accuracy and reliability of the experiment results, we only retained papers with more than 19 citations[1] for empirical analysis in Section 5, that is, 28,769 papers.

Note that although Web of Science claims the coverage of its database dates back to 1900, it contains only 26 million pre-1996 records from among 90 million records.[2] One of the reasons for this might be due to the omission of publications being recorded for the first half of the 20th century. Although we tried to find the most comprehensive and complete citation database and chose Web of Science as our data source, the limited time coverage of this database may artificially reduce the citation counts of older publications during the first years after their appearance. But this limitation doesn't necessarily have

[1] 19 is the median value for papers in physics, and the median values for chemistry, physiology or medicine, and economic sciences are 21, 40, and 20, respectively.

[2] Thomson Reuters. Web of Science: The Complete Citation Connection. http://wokinfo.com/citationconnection/realfacts/.

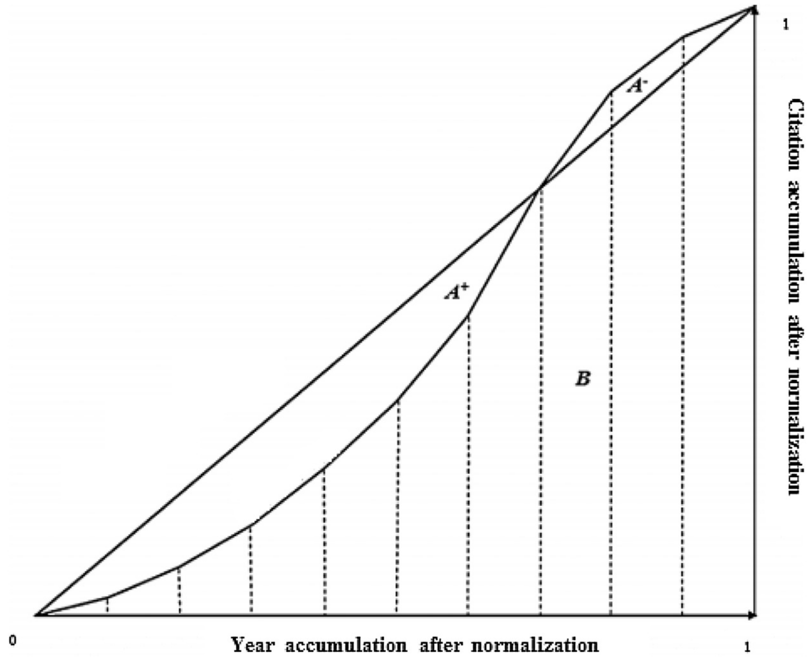

Fig. 1. Normalized citation accumulation curve for a single paper.

Adapted from Fig. 2 in Li et al. (2014).

much influence on the empirical analysis of this study, as noted in Section 5, where we aim to validate the effectiveness of our proposed method.

4. Comparison between the *Gs* index and *citation delay*

4.1. Correlation between the two indicators

As noted above, Li et al. (2014) proposed the *Gs* index[3] to investigate annual citation patterns of a paper during its dormancy period. Wang (2013) proposed an indicator *citation speed* by calculating the mean of the cumulative citation ratios of a single paper. Wang et al. (2015) then used the subtraction of *citation speed* to describe the process of citation accumulation for a single paper and developed a new indicator, *citation delay*.[4] Fig. 1 illustrates the design of the *Gs* index and the design of *citation delay*, both of which are based on a citation accumulation curve.

For the construction of the *Gs* index, Li et al. (2014) summed up the areas of the *n* trapeziums in the graph of year-citation accumulation. The original formula of *Gs* index is:

$$G_s = \frac{A}{A+B}, \tag{1}$$

where A and B, respectively, represent the areas of the two parts divided by the citation accumulation curve in the graph of year-citation accumulation. As area A can also be divided into two parts, Li et al. (2014) marked the one below the 45° line as A^+ and the one above is noted as A^-. The final formula of *Gs* index is expressed as a segment function:

$$G_s = \begin{cases} 1 - \dfrac{2 \times [n \times c_1 + (n-1) \times c_2 + \ldots + c_n] - C}{C \times n}, & C > 0 \\ 1, & C = 0 \end{cases} \tag{2}$$

where n is the total number of years after publication, C is the total number of citations the paper received during the n years, and $c_i \left(i \in \{1, 2, \cdots, n\}\right)$ is the number of citations the paper received in the i^{th} year after publication.

For the construction of *citation delay* Wang (2013), first proposed the indicator *citation speed*:

$$\textit{citation speed} = \frac{\sum_1^{n-1} C_i/C_n}{n-1} \tag{3}$$

[3] *Gs* index values range from −1 to 1. The more the recognition is delayed, the larger the value.
[4] *Citation delay* values range from 0 to 1. The more the recognition is delayed, the larger the value.

where n is the same as in Eq. (2), and C_i is the cumulative citation count by year i. Wang et al. (2015) later defined *citation delay* as 1–*citation speed*.

Although not aiming specifically to measure delayed papers, these two measures of *Gs* index and *citation delay* can both be applied to the whole citation curve of papers with all kinds of citation patterns, where results of this study find them to be highly correlated. If we expand the formula of *citation delay*, it could be expressed as:

$$\textit{citation delay} = 1 - \textit{citation speed} = 1 - \frac{\sum_1^{n-1} C_i/C_n}{n-1} = 1 - \frac{C_1 + C_2 + \ldots + C_{n-1}}{(n-1)\times C_n} = 1 - \frac{(n-1)\times c_1 + (n-2)\times c_2 + \ldots + c_{n-1}}{(n-1)\times C_n} \tag{4}$$

Eqs. (2) and (4) are similar to some extent. It is worth noting that Wang et al. (2015) left out the citation count of the last year when computing *citation delay*. If the omission is added to Eq. (4), a common term between the two equations becomes visible:

$$\textit{common term} = -\frac{n\times c_1 + (n-1)\times c_2 + \ldots + c_n}{n\times C_n} = -\frac{\sum_1^n (n+1-i)\times c_i}{n\times C_n} = -\sum\nolimits_1^n \frac{(n+1-i)}{n\times C_n}\times c_i \tag{5}$$

It can easily be inferred here that $-1 < \text{common term} < 0$, since:

$$0 < -\text{commonterm} = \frac{n\times c_1 + (n-1)\times c_2 + \ldots + c_n}{n\times C_n} < \frac{n\times c_1 + n\times c_2 + \ldots + n\times c_n}{n\times C_n} = \frac{n\times \sum_1^n c_i}{n\times C_n} = 1 \tag{6}$$

It is this common term, the kernel of *Gs* index and *citation delay*, which determines the high positive correlation between the two measures. In other words, this common term plays a crucial and equivalent role in measuring delayed citation of individual papers. Since two groups of researchers proposed two different indicators with a common term using different methods, the logic behind this common term may be worthy of further investigation.

4.2. Logic behind: weighted summation in time domain

From Eq. (2), we have:

$$G_s = 1 + 2\times \textit{common term} + \frac{1}{n} \tag{7}$$

Once we add in citation information of the last year for *citation speed*, then $\textit{citation speed} = \frac{\sum_1^n C_i/C_n}{n}$. So:

$$\textit{citation delay} = 1 - \textit{citation speed} = 1 - \frac{\sum_1^n C_i/C_n}{n} = 1 + \textit{common term} \tag{8}$$

Using 2 × Eq. (8) – Eq. (7), we have

$$\textit{citation delay} = \frac{1+G_s}{2} - \frac{1}{2n} \tag{9}$$

The *Gs* index and the *citation delay* can thus be converted to each other via the common factor. To better understand how this common term works, we examine its numerator and denominator in Eq. (5). For one thing, its numerator can be reckoned as a weighted summation of yearly citations of a paper, with the weights being $(-n+i-1)$ $(i \in \{1, 2, \cdots, n\})$ and n being the total number of years after the paper's publication. Its denominator functions as a normalization factor, which makes the value of common term fall between −1 and 0. The entire common term is also a weighted summation of yearly citations, with the weights being $-\frac{(n+1-i)}{n\times C_n}$ $(i \in \{1, 2, \cdots, n\})$.

The logic behind the common term of the two measures is now clear. In considering the weighted summation of a paper's yearly citation counts, the selection of the weights depends on the yearly citations to be focused on. The more the yearly citations are to be emphasized, the larger the corresponding weights should be. For instance, the weights in Eq. (5) are: $-\frac{(n+1-i)}{n\times C_n} = \frac{i-n-1}{n\times C_n}$ $(i \in \{1, 2, \cdots, n\})$, showing a monotonic increasing order with time, as (i-n-1) increases with an increasing i. This approach to citation summation agrees very well with the phenomenon of delayed recognition, in that the weights of late yearly citations should be larger than those of early yearly citations to quantify this time lag. The reasonability of the *Gs* index and the *citation delay* is thus confirmed here.

5. New understandings of delayed recognition

5.1. Uneven weights for yearly citations

Since the *Gs* index and the *citation delay* formulas have a common term that plays a decisive role in gauging delayed citation, and are in essence the same measure for that purpose, these two measures can be converted to each other by the use of linear transformation. The implicit logic behind this approach is the assignation of uneven weights to citations according to different measurement requirements of yearly citation counts and adding them up. It is important to note that, this method for citation summation may seem somewhat unusual, as citations in different years are usually viewed as equal when calculating a paper's total citation counts, but it is rather effective for measuring delayed citation. This method for summarizing yearly citations of a paper is only concerned with citation distribution for a paper in a time domain, and has nothing to do with the subjective impact or "quality" of that paper.

The key issue here is the determination of the weight for every yearly citation count. For example, when every yearly citation is viewed as equally important, and where the weight for every yearly citation is 1, the summation is the *total* citations of a paper. This is how the impact of an individual paper is generally measured. In another example, if every yearly citation is viewed as equally important, while the weight is $1/n$ (n being the total number of years after publication of a paper), the summation is the average number of the paper's annual citations. This is the usual approach to "normalizing" the impact of time when evaluating an individual paper's influence. Yet both of these examples are cases where yearly citations of a paper are treated equally. The question addressed herein is, could there be an advantage to allocating different weights to the citations?

When other dimensions such as time are involved, we suggest that yearly citations of a paper may assume unequal importance in cases such as delayed citation, in which later citations should be stressed, based on the paper's more immediate relevance under potentially changed circumstances and in light of new research. In using weighted summation, new indicators can be designed to measure the extent to which a paper's citations are delayed. If we assign weight $\frac{i}{n \times C_n}\left(i \in \left\{1, 2, \cdots, n\right\}\right)$, which increases with time to the i^{th} yearly citation count, we could obtain a single indicator D_1 that unifies the way in which both the *Gs* index and the *citation delay* functions to measure delayed citation:

$$D_1 = \frac{1}{n \times C_n} \times c_1 + \frac{2}{n \times C_n} \times c_2 + \cdots + \frac{n}{n \times C_n} \times c_n = \sum\nolimits_1^n \frac{i}{n \times C_n} c_i = \frac{\sum_1^n i \times c_i}{n \times C_n} \tag{10}$$

where c_i is the citation count of year i, and C_i is the cumulative citation count by year i. In Eq. (10), $n \times C_n$ is also a normalization factor that makes $0 < D_1 < 1$.

In fact, a generalized weight $\frac{i^a}{n^a \times C_n}$ could be assigned to the i^{th} yearly citation count in order to develop a more generalized indicator. With this generalized weight, we could amplify or lessen the relative weights among yearly citations by changing the value of parameter a.

For example, if we want to further strengthen late citations and weaken early citations to put more emphasis on the former, we could amplify the relative difference between weights of the early years and later years with the use of a new indicator D_2 by setting $a = 2$:

$$D_2 = \frac{1^2}{n^2 \times C_n} \times c_1 + \frac{2^2}{n^2 \times C_n} \times c_2 + \cdots + \frac{n^2}{n^2 \times C_n} \times c_n = \sum\nolimits_1^n \frac{i^2}{n^2 \times C_n} c_i = \frac{\sum_1^n i^2 \times c_i}{n^2 \times C_n} \tag{11}$$

where $\frac{i^2}{n^2 \times C_n}$ is the weight assigned to the i^{th} yearly citation count c_i. It can also be easily demonstrated that $0 < D_2 < 1$, making it a generalized D_a. Just as the logic used to construct the indicators implies, the later a paper's citations are accumulated, the closer the values of D_a to 1, and the earlier a paper's citations are accumulated, the closer the values of D_a to 0. The feature that D_a ranges from 0 to 1 normalizes the measurement of citation delay, thus providing a convenient measure for future studies on this phenomenon.

5.2. *Using both D_a and total citations to identify delayed papers*

To measure how much a paper's citations converge in the later years of its lifetime, we use D_1 as an example. As indicators with other a values emphasize late citations of a paper, their performance for measuring citation convergence in later years can also be reflected in the following analysis. We calculate the value of D_1 for papers in our dataset and confirm that later citation convergence doesn't necessarily ensure recognition in the form of large total citations. Papers with top 15 D_1 values are selected in Table 1. Because D_1 might not be very reliable when the denominators in Eqs. (10) and (11) are too small, and papers with very small numbers of citations might not fit into the phenomenon of delayed recognition we focus on, we only retained papers with more than 19 citations, leading to 28,769 papers used for calculation.

A significant proportion of papers have very high D_1 values but not very large total citation numbers (Table 1). This outcome supports the argument that delayed recognition is a rare phenomenon in science (Lachance & Larivière, 2014). These papers in general received hardly any citations for a long time after publication, although their citations began to increase in later years. Their long-lasting period of dormancy (see Fig. 2, for example) makes the weights assigned to yearly citations of the late years very large, leading to high values of D_1. On the other hand, typically delayed recognition papers also have relatively large citations as well as high D_1 values (see Fig. 3, for example), such as Lippmann's paper in 1908, Rayleigh's paper in 1916, Staudinger's paper in 1907, Zernike's paper in 1955, and Einstein's paper in 1935 (Einstein, Podolsky & Rosen, 1935, the pioneering essay on quantum mechanics) that had gained 4937 citations by the time we collected the data, a classic sleeping beauty.

It is also observed that a substantial number of papers with high D_1 values (Table 1) received quite a few citations over the last five years compared with prior years, including papers published decades or even more than a century ago. This finding

Table 1
Papers with top 15 D_1 values (Citation counts greater than 100 are in bold).

Discipline	Publication year	Author	Title	Journal, Volume, Issues	D_1	Citations by 2011	Citations by 2015	Citation increase in recent five years (%)
Physics	1907	Rayleigh, Lord	Note on the remarkable case of diffraction spectra described by Prof. Wood	PHILOSOPHICAL MAGAZINE, 14, 79–84	0.9436	23	64	178
Physics	1908	Lippmann, G	Reversible test prints. Integral photographies	COMPTES RENDUS HEBDOMADAIRES DES SEANCES DE L ACADEMIE DES SCIENCES, 146	0.9326	**403**	**630**	56
Physics	1916	Rayleigh, Lord	On convection currents in a horizontal layer of fluid, when the higher temperature is on the under side	PHILOSOPHICAL MAGAZINE, 32, 187-92	0.8935	**139**	**248**	78
Economics	1980	Diamond P	Income taxation with fixed hours of work	JOURNAL OF PUBLIC ECONOMICS, 13, 1	0.8882	30	47	57
Physics	1957	TAMM, I	Ribonucleic acid synthesis and influenza virus multiplication	SCIENCE, 126, 3285	0.888	63	102	62
Economics	1968	Stiglitz J	Note on technical choice under full employment in a socialist economy	ECONOMIC JOURNAL, 78, 311	0.8859	21	23	10
Chemistry	1907	Staudinger, H	Announcements from the Chemical Institute at Strasbourg University in Alsace, France – Ketene	JUSTUS LIEBIGS ANNALEN DER CHEMIE, 356, 1/3	0.8856	**251**	**325**	29
Physics	1931	Einstein, A	Unified theory of gravitation and electricity	SITZUNGSBERICHTE DER PREUSSICHEN AKADEMIE DER WISSENSCHAFTEN PHYSIKALISCH-MATHEMATISCHE KLASSE	0.8851	26	36	38
Physics	1955	Zernike, F	How I discovered phase contrast	SCIENCE, 121, 3141	0.8811	**189**	**289**	53
Physics	1926	Einstein, A	The cause of meander formation of river paths and the so-called Baer law	NATURWISSENSCHAFTEN, 14	0.8788	20	40	100
Physics	1915	Einstein, A	On the general theory of relativity	SITZUNGSBERICHTE DER KONIGLICH PREUSSISCHEN AKADEMIE DER WISSENSCHAFTEN	0.8735	68	84	24
Physics	1935	Einstein, A	Can quantum-mechanical description of physical reality be considered complete?	PHYSICAL REVIEW, 47	0.8711	**4937**	**6775**	37
Physics	1946	Purcell, EM	Spontaneous emission probabilities at radio frequencies	PHYSICAL REVIEW,69, 11-1	0.871	**1631**	**2363**	45
Physics	1915	Einstein, A	The field equations of gravity	SITZUNGSBERICHTE DER KONIGLICH PREUSSISCHEN AKADEMIE DER WISSENSCHAFTEN	0.8701	70	88	26
Chemistry	1911	Sabatier, P	Announcement: Hydrogenation and dehydrogenation for catalysis	BERICHTE DER DEUTSCHEN CHEMISCHEN GESELLSCHAFT, 44	0.8693	57	137	140

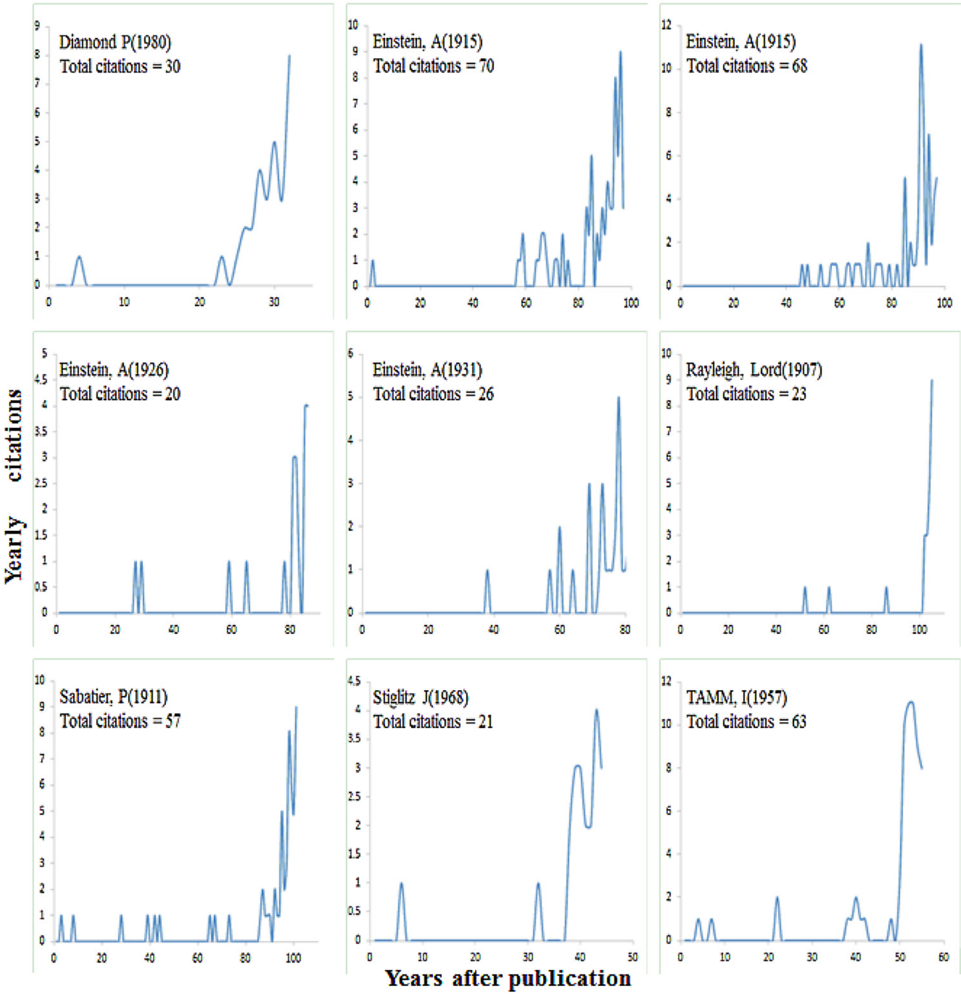

Fig. 2. Citation histories of papers with less than 100 citations in Table 1.

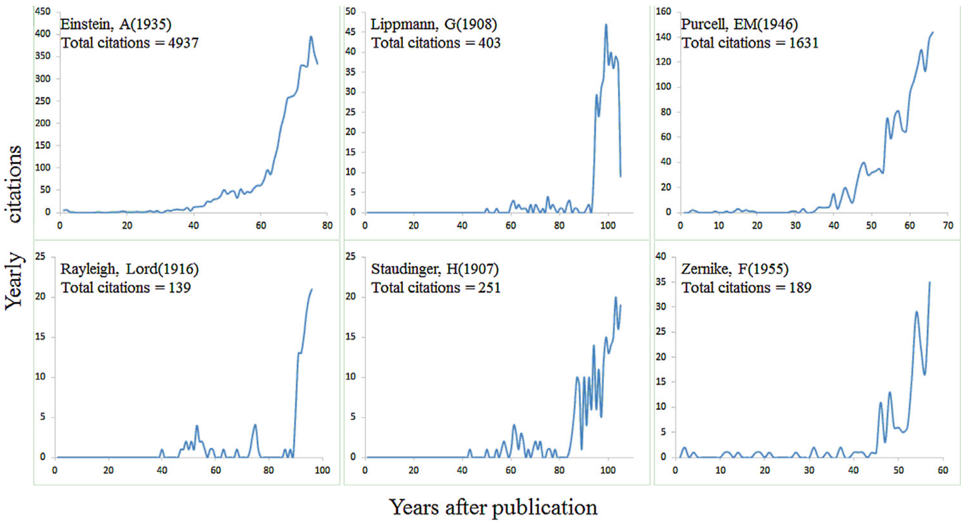

Fig. 3. Citation histories of papers with more than 100 citations in Table 1.

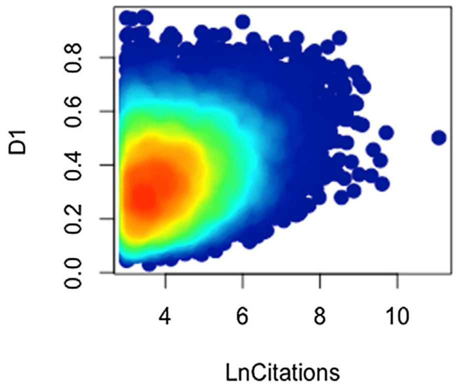

Fig. 4. Scatter plot of natural logarithm of total citations and D_1 (the lighter the colors, the denser the observations).

to some extent reflects D_1's potential in identifying papers that will continue to get a considerable amount of citations long after their original publication.

The papers in Table 1 are divided into two groups, with one group having fewer than 100 citations, and the other group having more than 100 citations. Their citation histories are shown in Figs. 2 and 3, respectively. All these citation curves take on the similar pattern of a dormancy period followed by a lifting of citation counts, in cases of both sudden and gradual increases. This visualization shows D_1's effectiveness in identifying papers with citation delay. But the total citations of these papers are different in number and potential recognition. Among papers whose citations are delayed, there are "sleeping beauties" with considerable total citations as well as "sleeping mediocrities" with small total citations. These results suggest that ultimate recognition doesn't necessarily co-occur with citation lag, although in some cases it does. We have to also consider, therefore, the ultimate recognition of a paper in addition to its citation lag.

Figs. 2 and 3 suggest that citation delay and total citations are quite different dimensions in describing a paper's citation pattern. To learn more about their associations, we explore the correlations between them. For example, as citations can be very large in number, we take their natural logarithm. Correlation analysis suggests a significant positive, although relatively low correlation, between the natural logarithm of citations and the values of D_1, showing a Spearman correlation coefficient of 0.21.

In Fig. 4, a scatter plot between the natural logarithm of citations and D_1 reflects again their positive correlation as well as the density of observations. But it is not easy to describe the exact association that exists between them. The above results show that citation delay and total citations measure different dimensions of a paper, and that although D_1 is good at measuring how much a paper's citations are delayed, it alone can't efficiently identify delayed recognition, let alone the existence of sleeping beauties. We therefore argue that neither citation delay nor total citations should be ignored when considering delayed recognition phenomenon.

5.3. Measuring delayed recognition with different a values

To further emphasize the importance of late citations, we introduce D_2 as an indicator when $a = 2$ for comparison with D_1. We also calculate the value of D_2 for papers in our dataset and list papers with top 15 D_2 values in Table 2. On the one hand, we see in Table 2 that D_2 can also effectively identify papers that still receive a large number of citations long after their original publication. On the other hand, when we again make a correlation analysis between D_2 and total citations, we find that the correlation coefficient is 0.20, similar to the coefficient 0.21 between D_1 and total citations.

To test the effect of changing relative weights among yearly citations, we sort the papers in our dataset in a descending order by their total citation number, and then label each paper with its rank by descending D_1 values and also label each paper with its rank by descending D_2 values, respectively. Table 3 lists 10 papers from our dataset as an example. It is found in the whole dataset that for papers that are ranked as the top 1000 by total citations, 727 papers' D_2 rank is lower than their D_1 rank. Furthermore, correlation analysis suggests a very small negative correlation between the rank of total citations and the rank difference between D_2 and D_1. This suggests that the ranking results for D_2 and D_1 are different: with the decreasing of citation rank (and increasing of total citations), the D_2 rank tends to become lower than the D_1 rank for the whole dataset. An obvious example is Einstein's paper in 1935, which now has 4937 citations that appear in Table 1 but disappear in Table 2. It is confirmed that D_2 can to some extent reduce the impact of large total citations on the measurement of delayed citation.

By setting different values of a, we can further change relative weights among yearly citations. We calculate D_a values for each paper when $a = 1/3$, 1/2, 2/3, 1, 2, 3 and 4, respectively. Replicating the previous ranking process, we then rank all the papers by $D_{1/3}$, $D_{1/2}$, $D_{2/3}$, D_1, D_2, D_3, D_4 and total citations separately. Obviously, we have more reason to view a paper as a delayed recognition paper if it has both higher D_a rank and higher total citation rank than other papers in the dataset. According to this principle, we conduct 8 experiments to test the effect of choosing different a values when identifying delayed recognition papers. In each experiment, we count the number of papers that are both within a certain range of D_a

Table 2
Papers with top 15 D_2 values (Citation counts greater than 100 are in bold).

Discipline	Publication year	Author	Title	Journal, Volume, Issues	D_2	Citations by 2011	Citations by 2015	Citation increase in recent five years (%)
Physics	1907	Rayleigh, Lord	Note on the remarkable case of diffraction spectra described by Prof. Wood	PHILOSOPHICAL MAGAZINE, 14, 79–84	0.9082	23	64	178
Physics	1908	Lippmann, G	Reversible test prints. Integral photographies	COMPTES RENDUS HEBDOMADAIRES DES SEANCES DE L ACADEMIE DES SCIENCES, 146	0.8784	**403**	**630**	56
Physics	1957	TAMM, I	Ribonucleic acid synthesis and influenza virus multiplication	SCIENCE, 126, 3285	0.8267	63	102	62
Physics	1916	Rayleigh, Lord	On convection currents in a horizontal layer of fluid, when the higher temperature is on the under side	PHILOSOPHICAL MAGAZINE, 32, 187-92	0.8253	**139**	**248**	78
Economics	1968	Stiglitz J	Note on technical choice under full employment in a socialist economy	ECONOMIC JOURNAL, 78, 311	0.8216	21	23	10
Economics	1980	Diamond P	Income taxation with fixed hours of work	JOURNAL OF PUBLIC ECONOMICS, 13, 1	0.8208	30	47	57
Physics	1926	Einstein, A	The cause of meander formation of river paths and the so-called Baer law	NATURWISSENSCHAFTEN, 14	0.8152	20	40	100
Physics	1955	Zernike, F	How I discovered phase contrast	SCIENCE, 121, 3141	0.8122	**189**	**289**	53
Chemistry	1911	Sabatier, P	Announcement: hydrogenation and dehydrogenation for catalysis	BERICHTE DER DEUTSCHEN CHEMISCHEN GESELLSCHAFT, 44	0.8084	57	137	140
Medicine	1952	HUXLEY, AF	Propagation of electrical signals along giant nerve fibres	PROCEEDINGS OF THE ROYAL SOCIETY SERIES B-BIOLOGICAL SCIENCES, 140, 899	0.808	31	65	110
Chemistry	1907	Staudinger, H	Announcements from the Chemical Institute at Strasbourg University in Alsace, France – Ketene	JUSTUS LIEBIGS ANNALEN DER CHEMIE, 356, 1/3	0.8008	**251**	**325**	29
Physics	1931	Einstein, A	Unified theory of gravitation and electricity	SITZUNGSBERICHTE DER PREUSSICHEN AKADEMIE DER WISSENSCHAFTEN PHYSIKALISCH-MATHEMATISCHE KLASSE	0.7994	26	36	38
Chemistry	1900	Ostwald, W	On the assumed isomerism of red and yellow mercury oxide and the surface-tension of solid bodies	ZEITSCHRIFT FUR PHYSIKALISCHE CHEMIE–STOCHIOMETRIE UND VERWANDTSCHAFTSLEHRE, 34	0.7912	**534**	**685**	28
Economics	1969	Schelling T	Models of segregation	AMERICAN ECONOMIC REVIEW, 59, 2	0.7859	**165**	**302**	83
Physics	1915	Einstein, A	The field equations of gravity	SITZUNGSBERICHTE DER KONIGLICH PREUSSISCHEN AKADEMIE DER WISSENSCHAFTEN	0.784	70	88	26

Table 3
Illustration of ranking and labelling papers (N = 28,769).

Paper	Citations	Citation rank	D_1 rank	D_2 rank	D_2 rank $-D_1$ rank
Paper #1	64519	1	5448	8303	2855
Paper #2	16519	2	4584	5973	1389
Paper #3	14845	3	16885	18114	1229
Paper #4	14188	4	10567	14427	3860
Paper #5	11822	5	7906	9286	1380
Paper #6	11228	6	14592	16977	2385
Paper #7	9190	7	568	650	82
Paper #8	8887	8	3423	4510	1087
Paper #9	8292	9	416	520	104
Paper #10	8178	10	14287	13138	−1149

Table 4
The numbers of top-ranked papers based on total citations and D_a (N = 28,769).

Experiment #	Selection criterion		$D_{1/3}$	$D_{1/2}$	$D_{2/3}$	D_1	D_2	D_3	D_4
	D_a rank	Citation rank							
#1	Top 200	Top 200	16	16	16	14	14	13	13
#2	Top 400	Top 400	48	47	48	45	42	42	39
#3	Top 600	Top 600	76	76	76	74	70	67	62
#4	Top 800	Top 800	115	115	112	107	105	97	95
#5	Top 1000	Top 100	26	25	25	25	23	22	22
#6	Top 1000	Top 200	52	51	51	51	47	44	43
#7	Top 1000	Top 500	92	90	90	91	84	80	78
#8	Top 1000	Top 1000	153	150	149	153	142	136	132

Note: In each experiment, we first select papers with top D_a rank, and then among these papers we count the number of papers that are also in a top ranking list of total citations. For example, the value in column 4 in Experiment #1 means that there are 16 papers who simultaneously rank top 200 by $D_{1/3}$ and top 200 by total citations.

rank and total citation rank. For example, in Experiment #1 of Table 4, the numbers are counted for papers who rank top 200 by D_a values and simultaneously rank top 200 by total citations.

Since the experimental results don't show much difference, D_a is generally a robust indicator for identifying papers with delayed citations. This is especially true when we use narrow ranking ranges (columns 2 & 3 in Table 4). However, when expanding ranking ranges, the differences in the number of papers identified by different D_a increase. In experiment #4, $D_{1/3}$ identifies 20 more papers than D_4 does; while in Experiment #8, $D_{1/3}$ identifies 21 more papers than D_4 does, indicating the effect of choosing different a values on identifying delayed recognition papers. This effect can also be seen by comparing the results of Experiments #5, #6, #7 and #8. What's more, it is observed in Table 4 that lower a values tend to result in more identified papers with both large D_a values and total citations. There are always more papers identified when $a < 1$ than when $a > 1$. Therefore, decreasing the values of a usually means loosening the criterion for selecting delayed recognition papers. Increasing the values of a usually means tightening the criterion. In addition, changing a values within the interval $[1,\infty)$ seems to have an effect more significant than changing a values within $(0, 1]$ on the number of identified papers.

6. Discussion and conclusion

This paper studies the less common phenomenon of delayed recognition in bibliometrics and suggests an approach for understanding and measuring this phenomenon. The literature review shows limitations in terms of applicability, reasonability, and complexity in existing measures for delayed recognition. Formula derivation uncovers a decisive common term between the two measures of *citation delay* and *Gs* index in the literature. The use of uneven weighted summation for yearly citations in the time domain is proposed herein to better understand the nuanced phenomenon of citation delay, as well as to estimate how much a single paper's citations are lagged. We recommend the extent to which citations are delayed as a basic attribute for a single paper, along with its other basic characteristics such as age and total citation numbers. This approach is promising in its ability to depict delayed recognition in a simple way for all individual papers. Overall recognition, usually in the form of total citations, should also be considered at the same time, as delayed citation does not reflect a paper's ultimate recognition as a significant contribution or a sleeping beauty.

This paper contributes to the delayed recognition literature based on three perspectives it provides. First, it demonstrates that the intrinsic relation between two independent indicators of *citation delay* and *Gs* index is based on the same logic of applying uneven weights to sum up yearly citations, where larger weights are given to later paper citations. Previous studies usually treat yearly citations as equal in weight in the time domain at the level of single papers, where such measures as total citation and average citation per year come into use when an individual paper's overall impact is evaluated. The idea of assigning larger weights to later citations to a paper nicely accords with the essence, practice, and manifestation of citation delay, and can thus provide a quantitative description of this phenomenon. The weights given to yearly citations can also be determined according to various measurement requirements and thresholds, leaving an open space for future studies on scientific evaluation of citation phenomenon.

Second, this paper suggests a significant and convenient framework for analyzing delayed recognition. Different from previous studies that view delayed recognition from an integral or singular angle, we suggest that the extent of citation delay and the extent of recognition in terms of contribution should be measured separately. On the one hand, citation delay and total citations serve as quite different dimensions in describing a paper's citation patterns, despite their relatively low correlation. A paper with large D_a value usually experiences citation delay, and a paper with both large D_a value and large total citations is often a "sleeping beauty." In light of this idea, it is in practice very easy to discover papers most likely to achieve sleeping beauty status from a given set of papers, such as the whole database of Web of Science. We only need to rank all papers by their D_a values in a descending order, and then select the most cited papers from those with top D_a ranks. On the other hand, papers with high citation delay but medium or low total citations are also worthy of attention. Although they don't achieve much immediate recognition, their academic values may nonetheless be realized after undergoing a long dormancy period, where some of them might have the potential to become sleeping beauties. These papers with fewer citations might deserve more attention from scientists since they may contain hidden values to be exploited. This is particularly relevant due to increasing trends of big data and cross-disciplinary work in the sciences, where research shows that for about 80% of citations of top Sleeping Beauties, up to three quarters are interdisciplinary in nature (Ke et al., 2015).

Third, this paper suggests a simple indicator D_a for depicting delayed citation of individual papers. This indicator allows us to adjust the criterion for selecting delayed recognition papers by changing the values of a. Existing measures for delayed recognition have limitations in terms of restricted application scope, division of different periods, and complex calculation methods. Our new method avoids these limitations and allows for a simpler measure in relation to the extent of citation delay. We recommend D_a as a basic attribute of a single paper to show the degree to which citations are delayed, because: (1) it makes full use of citation data of a paper without important citation information being omitted; (2) it avoids the division of different citation time periods; (3) it is applicable to all papers with various types of citation curves; (4) it is simple and easy to calculate; and (5) its values range from 0 to 1, a convenient measure for further study on citation delay phenomenon.

Author contributions

Chao Min: Conceived and designed the analysis; performed the analysis; wrote the paper.
Jianjun Sun: Conceived and designed the analysis.
Lei Pei: Performed the analysis.
Ying Ding: wrote the paper.

Acknowledgements

This research was supported by the National Natural Science Foundation of China (NSFC No. 71273125) and the China Scholarship Council. The authors are grateful to Dr. Filippo Radicchi, Dr. Yuxiang Zhao, Mr. Alexander T J Barron, Miss Yizhi jing, and Ms. Kathleen Connors. The authors thank Dr. Jiang Li for Nobel Laureate citation data. We would also like to express our special thanks to the editor and three reviewers for their very constructive comments and suggestions.

References

Amin, M., & Mabe, M. A. (2003). Impact factors: Use and abuse. *Medicina*, *63*(4), 347–354.
Barber, B. (1961). Resistance by scientists to scientific discovery. *Science*, *134*, 596–602.
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.
Costas, R., van Leeuwen, T. N., & van Raan, A. F. (2011). The Mendel syndrome in science: Durability of scientific literature and its effects on bibliometric analysis of individual scientists. *Scientometrics*, *89*(1), 177–205.
Einstein, A., Podolsky, B., & Rosen, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, *47*(10), 777.
Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics*, *1*(4), 359–375.
Garfield, E. (1980). Premature discovery or delayed recognition-why? *Current Contents*, *4*, 488–493.
Garfield, E. (1989). Delayed recognition in scientific discovery-citation frequency-analysis aids the search for case-histories. *Current Contents*, *23*, 3–9.
Glänzel, W., Schlemmer, B., & Thijs, B. (2003). Better late than never? On the chance to become highly cited only beyond the standard bibliometric time horizon. *Scientometrics*, *58*(3), 571–586.
Glänzel, W. (2008). Seven myths in bibliometrics about facts and fiction in quantitative science studies. *Collnet Journal of Scientometrics and Information Management*, *2*(1), 9–17.
Ke, Q., Ferrara, E., Radicchi, F., & Flammini, A. (2015). Defining and identifying sleeping beauties in science. *Proceedings of the National Academy of Sciences of the United States*, *112*(24), 7426–7431.
Kuhn, T. (1962). *The structure of scientific revolutions*. Chicago, IL: University of Chicago Press.
Lachance, C., & Larivière, V. (2014). On the citation lifecycle of papers with delayed recognition. *Journal of Informetrics*, *8*(4), 863–872.
Li, J., & Shi, D. (2016). Sleeping beauties in genius work: When were they awakened? *Journal of the Association for Information Science and Technology*, *67*(2), 432–440.
Li, J., Shi, D., Zhao, S. X., & Fred, Y. Y. (2014). A study of the heartbeat spectra for sleeping beauties. *Journal of Informetrics*, *8*(3), 493–502.
Moed, H. F., Burger, W. J. M., Frankfort, J. G., & van Raan, A. F. (1985). The use of bibliometric data for the measurement of university research performance. *Research Policy*, *14*(3), 131–149.
Price, D. D. S. (1976). A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, *27*(5), 292–306.

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*, 84–93.
van Raan, A. F. (2004). Sleeping beauties in science. *Scientometrics*, *59*(3), 467–472.
van Raan, A. F. (2015). Dormitory of physical and engineering sciences: Sleeping beauties may Be sleeping innovations. *PloS One*, *10*(10), e0139786.
Wang, J., Thijs, B., & Glänzel, W. (2015). Interdisciplinarity and impact: Distinct effects of variety, balance, and disparity. *PLoS One*, *10*(5), e0127298.
Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics*, *94*(3), 851–872.