Scientometrics (2018) 116:959–971
https://doi.org/10.1007/s11192-018-2780-0

A parameter-free index for identifying under-cited sleeping beauties in science

Jian Du[1] · Yishan Wu[2]

Received: 23 November 2017 / Published online: 28 May 2018

Abstract We developed a systematic methodology for identifying the under-cited (or not-so-highly cited) Sleeping Beauty (SB) publications and tried to figure out their key characteristics. Based on the identification framework of "beauty coefficient" (B) introduced by Ke et al. (Proc Natl Acad Sci USA 112:7426–7431, 2015), taking into account the whole citation history of the publications concerned, we substituted *yearly citations* in "beauty coefficient" with *yearly cumulative percentage of citations*, and eliminated the denominator in "beauty coefficient" since the curve of a given document's cumulative citations is always monotonically increasing if only the document is cited. The value of the modified beauty coefficient is denoted as Bcp. We also redefined the awakening year, sleeping length and sleeping depth within the Bcp framework with the intention of avoiding arbitrary thresholds as much as possible. We tested the new index using the data of SB articles identified from *Science* and *Nature*. The results showed that Bcp is more sensitive in identifying the "lower level SBs", which refers to the case when the total citations and the maximum annual citations of SBs are not so high in comparison with other typical SBs. Bcp works better than B in at least two aspects: (1) it "punishes" the situations when the SBs experienced early citations instead of continuous sleeping; (2) it allows for comparing the extent of delayed citation impact of publications in different fields with different citation patterns. We also figured out some key characteristics of such SB publications and pondered some policy implications about the associations of SB publications with transformative research, research front and research evaluation.

✉ Jian Du
windowsdujian@163.com

Yishan Wu
wuyishan1958@163.com

1 Institute of Medical Information and Library, Chinese Academy of Medical Sciences, Beijing, China

2 Chinese Academy of Science and Technology for Development, Beijing, China

Keywords Sleeping beauties · Citation analysis · Science policy · Identification methods

Introduction

A "Sleeping Beauty" (SB) in Science is a publication that goes unnoticed (or "sleeps") for a long time and then, almost suddenly, attracts a lot of attention, or "is awakened by a Prince" (van Raan 2004). This concept is actually a quantitative description of "delayed recognition of scientific achievements", a phenomenon widely discussed in sociology of science (Garfield 1989). "Premature discoveries" and "transformative innovations" are crucial for the development of science, but they are often initially neglected or resisted by the scientific community and thus are often subject to delayed recognition (Trapido 2015). In this paper, we try to propose a systematic identification method of SB publications, in order to extend the application scope of citation analysis, and discuss implications for identifying potential "ahead of time" discoveries or transformative research, and shorten the time lag for original research to get recognized.

There are three types of methods for identifying SBs in science, namely, citation curve fitting, arbitrary thresholds setting, and using parameter-free index, which were described at full length in our foregoing paper (Du and Wu 2016). Each of the three methods has both advantages and disadvantages.

Firstly, it is not precise and accurate to identify SB publications with citation curve fitting since it is inefficient to mine massive publications based on manual observation.

Secondly, the arbitrary thresholds methods include average-based and quartile-based criteria, as well as their combinations. Average-based criteria directly or indirectly referring to van Raan's definitions on sleeping period, sleeping depth, awakening period and awakening intensity (van Raan 2004), could better define the never cited or poorly cited papers in sleeping period, while quartile-based criteria can better reflect the extent of citation delay in the entire citation life time of an article. The former is based on a stricter threshold, and the latter on a more moderate threshold. It is, however, statistically difficult to apply the latter method since one paper's citation history must be compared with that of all papers across research fields (Costas et al. 2010). Both the average-based and quartile-based criteria are arbitrary definitions and do not take into account different citation patterns in various research fields. For example, van Raan defined four main variables, (1) length of the sleep in years after publication; (2) depth of sleep in terms of a maximum citation rate during the sleeping period (cs_{max}); (3) awake period in years after the sleeping period (a_{min} and a_{max}); and (4) awake intensity in terms of a minimum citation rate during the awake period (ca_{min}). The four main variables can be tuned. He denoted the SBs with these variables with [5, 1, 4, 5] in (van Raan 2004) and [10, 1, 10, 5] in (van Raan 2015, 2017). For example, he investigated a set of SBs with (1) sleeping period length $s = 10$ years (publication years starting in 1980), (2) deep sleep, $cs_{max} = 1$; (3) awake period of 10 years, $a_{min} = a_{max} = 10$; and (4) awake intensity $ca_{min} = 5$. The combination of average-based with quartile-based criteria enhanced the accuracy for identifying sleeping beauties, but the combined method increased complexity and reduced transparency.

Finally, the two main parameter-free indices are Citation Delay (Onodera 2016; Wang et al. 2015), an indicator opposite to Citation Speed (Wang 2013) and Beauty Coefficient (Ke et al. 2015). In our foregoing paper (Du and Wu 2015), these two parameter-free indices were used to identify SBs published between 1970 and 2005 in the four most prestigious clinical medicine journals, namely, *New England Journal of Medicine*, *The Lancet*, *Journal of the American Medical Association* and *British Medical Journal*. We found that because Citation Speed takes the citation curve of the whole citation window into account, it could identify the continuingly highly cited papers with "very long" life cycle but could not directly distinguish SB publications. The Beauty Coefficient was proved to be a good indicator to identify a SB, but it fails to cover the citation curve after the paper receives its maximum annual citations. Nevertheless, considering three factors namely: design, transparency in calculation and ease of application, the Beauty Coefficient is still the better measure.

Recently, Fang (2018) suggested that derivative analysis can be used alone to identify SBs and determine their awakening times or in combination with other methods of identifying SBs to improve their performance by assisting in the identification of abnormal SBs. In Ye and Bornmann (2018), the dynamic citation angle β is introduced as a simple way for identifying "smart girls" and "sleeping beauties" quantitatively. The new term smart girl (SG) is suggested to differentiate instant credit or "flashes in the pan" from SBs. Li and Ye (2016) proposed four rules that should be adhered to in distinguishing SBs in science: (1) early citations should be penalized; (2) the whole citation history should be taken into account; (3) the awakening time of a sleeping beauty should not vary over time; and (4) arbitrary thresholds on sleeping period or awakening intensity should be avoided. In this paper, we tried to combine the framework of Citation Delay and Beauty Coefficient to propose a revised parameter-free index for distinguishing under-cited sleeping beauties in science. Note that our definition of "under-cited" is in comparison with "highly-cited", from the perspective of the first generation citations.

Methods and materials

Determination of the beauty coefficient B

The Beauty Coefficient is based on the comparison between its citation history and a reference line, drawn from its publication year to the year of the peak of citations. Let's call t the time interval after publication and C_t the citation history of the paper. If C_{t_m} is the maximum of C_t, the straight line ℓ_{t} that connects the point $(0, C_0)$ and (t_m, C_{t_m}) is analytically described by the equation:

$$\ell_t = \frac{C_{t_m} - C_0}{t_m} \cdot t + C_0$$

$(C_{t_m} - C_0)/t_m$ is the slope of the line ℓ_t. For each $t < t_m$ we can compute the ratio between $\ell_t - C_t$ and $\max\{1, C_t\}$, and the definition of B is achieved by summing over these values.

$$\mathrm{B} = \sum_{t=0}^{t_m} \frac{\frac{C_{t_m} - C_0}{t_m} \cdot t + C_0 - C_t}{\max\{1, C_t\}}$$

The advantage of B is that it does not rely on arbitrary thresholds or on a certain percentage. By it we can investigate this phenomenon at a systematic level. But we do not agree with (Ke et al. 2015)'s argument that B can be calculated for any given paper that received at least one citation. For example, for a given paper, if the maximum number of citations received just in the publication year and the yearly citations decreased in the following years, namely, $C_0 = C_{t_m}$, then the reference line is C_0, the slope of the line is 1, and the value B could not be calculated.

One of the disadvantages of this definition is the high importance given to the peak. B works really well with top class SBs that after discovery have huge numbers of citations every year, but for lower level SBs with less total citations, it gives some unwanted results: many of the papers we had found as SBs had very low B values.

The other obvious disadvantage of B is that the denominator max$\{1, C_t\}$ does not penalize; instead, it may favor early citation accumulation. This runs against the intention of the authors (Ke et al. 2015). We hold that the role of the denominator is just to avoid division by zero, and calculating the ratio with a denominator could lead to the loss of the original information of the citation history.

Modifying beauty coefficient

So, to avoid the dependence of the Beauty Coefficient on just one year's citation history and to reduce the sensitivity of the measure to such extreme SBs, we propose a new measure in order to discover SBs among not-so-highly cited publications. Such under cited publications were often ignored in bibliometric related research.

Based on the identification framework of "beauty coefficient" (B) introduced by Ke et al. (2015), taking into account the whole citation history of the publications concerned, we substituted *yearly citations* in the formula for "beauty coefficient" with *yearly cumulative percentage of citations*, and eliminated the denominator since the curve of a given document's cumulative citations is always monotonically increasing if only the document is cited. The value of the modified beauty coefficient is denoted as Bcp.

Let's call t the time interval after publication and c_t the cumulative percentage of citation history of the paper, $c_t \in [0,1]$. If c_{t_m} is the maximum of c_t, $c_{t_m} = 1$, the straight line y_t that connects the point $(0, c_0)$ and $(t_m, 1)$ analytically is described by the equation:

$$y_t = \frac{1 - c_0}{t_m} \cdot t + c_0$$

$(c_{t_m} - c_0)/t_m$ is the slope of the line y_t. For each $t < t_m$ we can compute the value between y_t and c_t, and the definition of Bcp is achieved by summing over these values.

$$\text{Bcp} = \sum_{t=0}^{t_m} \frac{1 - c_0}{t_m} \cdot t + c_0 - c_t$$

The new beauty coefficient value Bcp for a given paper is based on the comparison between its annual cumulative percentage of citation history and a reference line that is determined by its publication year and the share of citations in this year, the maximum cumulative percentage of citations received in the last year (namely, 100% within a multiyear observation period), and the last year of the citation window.

Similar with Ke et al. (2015)'s definition, we give a plausible definition of awakening time—the year when the abrupt change in the accumulation of citations of SBs occurs. We defined the awakening time t_a as the time t at which the distance $d(t)$ between the point (t, c_t) and the reference line y_t reaches its maximum:

Within Bcp framework in Fig. 1,

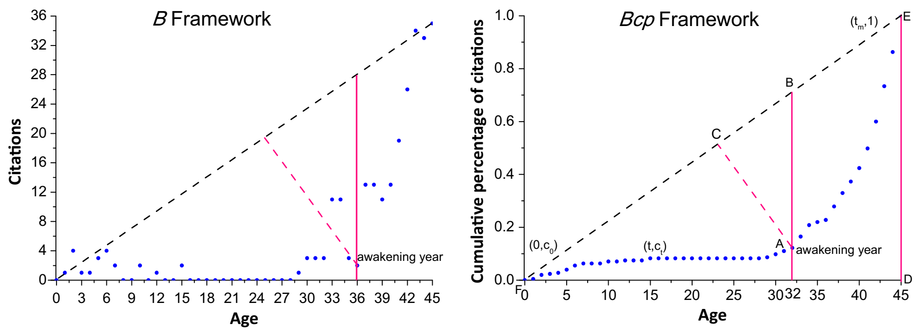

Fig. 1 Illustration of the definition of the new beauty coefficient and the awakening year of a given paper

$$\triangle ABC \sim \triangle FED$$
$$AC/FD = AB/FE$$

$AC = d(t)$, $d(t)$ is the distance between the point (t, c_t) and the reference line y_t.

$$FD = t_m$$
$$AB = \left|\frac{1 - c_0}{t_m} \cdot t + c_0 - c_t\right|$$

The absolute sign "||" is needed because a cumulative citation curve can be convex as well as concave, or a mixture of both (Liu and Rousseau 2014). When it is convex:

$$FE = \sqrt{(1 - c_0)^2 + t_m^2}, \quad so$$
$$d(t) = \frac{\left|\frac{1-c_0}{t_m} \cdot t_m \cdot t + t_m \cdot c_0 - t_m \cdot c_t\right|}{\sqrt{(1 - c_0)^2 + t_m^2}} = \frac{|(1 - c_0) \cdot t - t_m(c_t - c_0)|}{\sqrt{(1 - c_0)^2 + t_m^2}}$$

The reference line y_t is connecting the point $(0, c_0)$ and $(t_m, 1)$. It is easy to check that the points with coordinates $(0, c_0)$ and $(t_m, 1)$ have a distance zero to the reference line as it must be. In this paper, t_m is defined as the last year of the citation window, namely the year that the percentage of citations accumulated to 100%.

The awakening time reflects the year when the abrupt change in the accumulation of citations of SBs occurs, not the abrupt change from one year's citations to other year's citations. Being able to pinpoint the awakening time may help identify possible general trigger mechanisms behind the said change.

In the Bcp framework of Fig. 1, the reference line coincides with the case that a given paper's annual citations are equal. In general, when a given paper's cumulative citation curve is concave, then Bcp > 0 denotes delayed citations and Bcp < 0 early citations (the cumulative citation curve is convex). The lager the Bcp value, the more extent a given article is delayed recognized in terms of the citation curve. And the maximum value of Bcp index is (n-1)/2, n is the age of a given paper, when the total number of citations received are all in the last year and the yearly citations after publication till the last year is zero.

In the framework of the Bcp index, the paper shown in Fig. 1 was awakened in the 32nd year after its publication, going from 4 citations in the 32nd year to 12 in the 33rd year (Fig. 1). In the framework of B index, the citation burst occurs in the 36th year, from 3 times to 13 times (Fig. 1). However, looking at the annual citation curve, it is clearly appropriate to define the wakening time as the 32nd year. The Bcp framework can reveal an accumulate citation burst in the whole life cycle, while the B framework only reflects an annual citation burst.

In should be noted that the framework of Bcp allows comparing the extent of delayed citation impact of publications in different disciplines with different citation patterns. Bcp depends on the relative shape of the graph shown in Fig. 1 but not on the total number of citations.

Dataset

Two parameter-free indices, namely the Beauty Coefficient (B) and the modified Beauty Coefficient (Bcp) were used to identify sleeping beauties articles published between 1970 and 2005 in *Science* and *Nature*. The total period in which the SBs and their citation data are searched for is 1970–2015. Thus, 2005 is the last year for publications having in total a ten year time span until 2015. Articles with at least 200 citations, in total 20,000 publications were included in the following analysis.

Results

Bcp works better than B

A case study from (Li and Ye 2016)

As for the example shown in (Li and Ye 2016), two papers P1 and P2 were published in the same year, received the same number of citations, but had different citation curves as

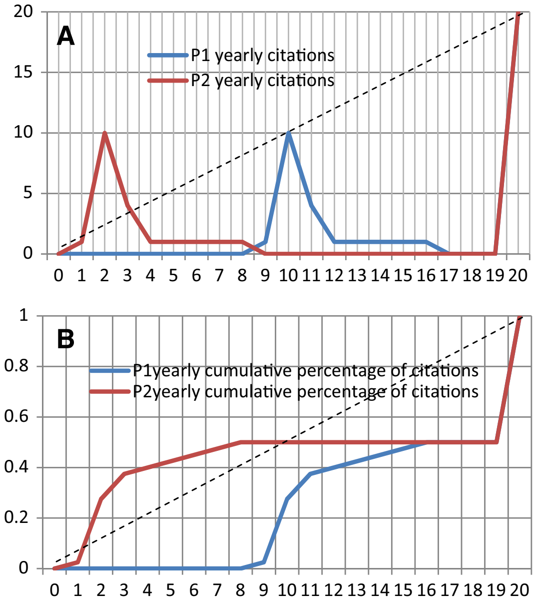

Fig. 2 Two fictitious articles P1 and P2. The upper graph illustrates the B framework (yearly citations) while the bottom one illustrates the Bcp framework (yearly cumulative percentage of citations)

shown in Fig. 2. Both of them received no citation in the publication year, and reached a citation peak at the age of twenty. Citations of P2 were received earlier than those of P1.

In the beauty coefficient framework the two articles have the same reference line. B-values are such that B1 = 164.75 is smaller than B2 = 177.95, although P1 accumulated citations later than P2. This is a counterintuitive result! Within the Bcp framework, however, the Bcp value for P1 is 5.075, while the Bcp-value for P2 is 1.075. Obviously, P1 has a larger and delayed citation impact than P2. From this case, we can see that the new beauty coefficient is better than B. Next, we will validate the new index through a large-scale bibliographic dataset.

A large-scale SBs investigation of articles published in science and nature

We selected the top 1% ($N = 200$) sleeping beauty publications using B and Bcp, respectively. There are 133 papers that occur in the two top 1% lists and 134 ((200–133) × 2) that occurred in just one. So, we analyze the 134 discrepant papers. Within our Bcp framework, we redefined two variables: (1) length of the sleep: years in the sleeping period, namely, the time interval between publication year and awakening year, (2) depth of sleep: the accumulated percentage of citations in the year before the awakening year, that is the accumulated percentage of citations at the end of sleeping period. The difference between B and Bcp using an independent samples T test is shown in Table 1.

The average length of sleep of sleeping beauty publications identified by Bcp is significantly smaller than B. In other words, the awaking year identified with the B framework comes later than the one determined by Bcp because of the computing mechanism. The average depth of sleep of sleeping beauty publications identified by Bcp is significantly smaller than B. That is to say, Bcp works better than B in penalizing the early citations. Although not statically significant, the total citations and annual maximum citations of sleeping beauty publications identified by Bcp is also smaller than B. The average age of sleeping beauty publications identified by Bcp is also significantly smaller than B. Bcp is more sensitive for identifying "the new and the lower level SBs", which refers to the case when the total citations and the maximum annual citations of SBs are not so high in comparison with other typical SBs.

We then selected the top 0.1% ($N = 20$) sleeping beauty publications with B and Bcp, respectively. There are 12 papers occurring in both top 0.1% lists and 16 papers which

Table 1 The difference test between B and Bcp

	Index	N	Mean	Sig. (2-tailed)
Length of the sleep	B	67	19.13	0.028 (*)
	Bcp	67	21.73	
Depth of sleep	B	67	42.3%	0.000 (**)
	Bcp	67	28%	
Total citations	B	67	3908.01	0.507
	Bcp	67	1561.94	
Annual maximum citations	B	67	183.43	0.646
	Bcp	67	112.6	
Age	B	67	35.39	0.006 (**)
	Bcp	67	38.13	

Significant at the 0.05 level (2-tailed); **Significant at the 0.01 level (2-tailed)

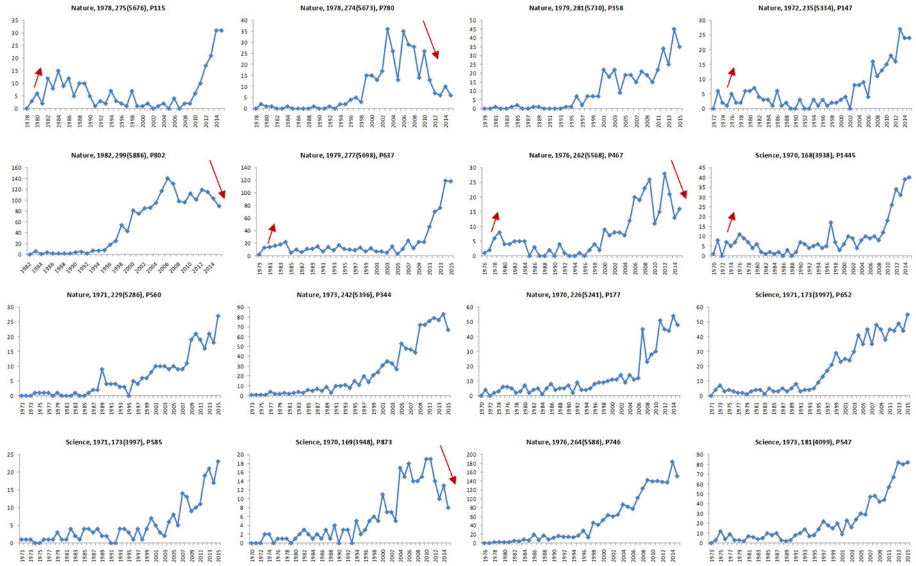

Fig. 3 The citation cures of extreme SBs identified only by B (the upper) and Bcp (the bottom), respectively

only occurred in one top 0.1% list. So, we analyze the citation curve of the 16 discrepant papers. As shown in Fig. 3, we found that in contrast with B, Bcp works well in penalizing early citations, so that, at parity of the total citations received, the later such citations accumulate, the higher the value of the proposed measure. Bcp takes into account the entire, rather than the partial, citation history of an article, so extreme sleeping beauty publications identified with Bcp show a continually increasing trend in terms of annual citations. Such articles have a delayed but durable citation impact.

The extreme SBs identified by Bcp tend to be landmark publications of a specific research field and the SBs are often technique and application-oriented work

As is shown in Table 2, in terms of content analysis, we can see that the top 10 SBs identified by Bcp were all landmark publications of a specific research field, such as “the first report on …”, or “the classic theory about …”. Three papers were Nobel laureate’s publications. It appears that high quality publications tend to encounter delayed recognition and thus showed delayed citation impact. For example, John Maynard Smith’s concept of protein space proposed in 1970 was assessed by *Nature* in 2004 writing that “1970 Foreshadows concepts now widely applied in studies of molecular evolution, such as genotype–phenotype mapping” (http://www.nature.com/nature/focus/maynardsmith/).

The SBs are often technique and application-oriented work

One is perhaps more inclined to believe that Sleeping Beauties relate to more fundamental and basic, and less to application-oriented work. But a surprising finding is that half of the SBs are application oriented and significantly more cited in patents than ‘normal’ papers

Table 2 Top 10 SBs identified by Bcp and their citations by patents

Bcp	Title	Authors	Source	Fields	Landmark publications	Citations	Awakening year	Citations by patent families	Priority date (the earliest application date)	Patent Number (the earliest application date
15.024	Electrochemical Photolysis of Water at a Semiconductor Electrode	Fujishima, A; Honda, K	Nature, 1972, 238(5358), P37	Chemistry	The landmark publication in photocatalyst research field	10776	2004	49	Nov 17, 1975	US 63255775
14.792	Structural Aspects of Interatomic Charge-Transfer Bonding	Hassel, O	Science, 1970, 170(3957), P497	Chemistry	1969 Nobel laureate's publication	255	2002	–		
14.724	Reduction of Lens Reflection by Moth Eye Principle	Clapham, PB; Hutley, MC	Nature, 1973, 244(5414), P281	Physics	The first report on the reduction of lens reflection	415	2005	6	Sep 27, 1982	GB 8227500
14.167	Impact of Population Growth	Ehrlich, PR; Holdren, JP	Science, 1971, 171(3977), P1212	Social Science	The first report on IPAT model. I = PAT, I- environmental impact; P- number of people; A- Affluence; T-technology	553	2002	–		
13.942	Central Dogma of Molecular Biology	Crick, F	Nature, 1970, 227(5258), P561	Life Science	1962 Nobel laureate's publication, the classic theory in molecular biology	602	2003	31	Sep 10, 1998	US 9976398
13.697	Natural Selection and Concept of a Protein Space	Smith, JM	Nature, 1970, 225(5232), P563	Life Science	1970 foreshadows concepts now widely applied in studies of molecular evolution	241	2005	2	Jul 17, 1986	US 88707086

Table 2 continued

Bcp	Title	Authors	Source	Fields	Landmark publications	Citations	Awakening year	Citations by patent families	Priority date (the earliest application date)	Patent Number (the earliest application date
13.053	More Is Different—Broken Symmetry and Nature of Hierarchical Structure of Science	Anderson, PW	Science, 1972, 177(4047), P393	Physics	1977 Nobel laureate's publication	888	2002	–		
12.360	Selection and Covariance	Price, GR	Nature, 1970, 227(5257), P520	Biology, Life Science	In the theory of evolution and natural selection, the Price equation (also known as Price's equation or Price's theorem) describes how a trait or gene changes in frequency over time.	736	2003	–		
11.979	Role of Pericorneal Papillary Structure in Renewal of Corneal Epithelium	Davanger, M; Evensen, A	Nature, 1971, 229(5286), P560	Medical Science	The first report on the role of pericorneal papillary structure in renewal of corneal epithelium	286	1999	1	Sep 27, 2000	US 23604500
11.695	Silicon. A Possible Factor In Bone Calcification	Carlisle, EM	Science, 1970, 167(3916), P279	Medical Science	The first report on the silicon as a possible factor in bone calcification	423	2004	6	Dec 27, 1988	US 29030788

Citations are counted through Web of Science between the publication year and 2015. Citations by patent families are searched for through lens.org from the publication year till 14 Feb, 2017 (van Raan 2015, 2017). The scientific non-patent references (SNPRs) represent a bridge between science and technology although they do not necessarily indicate the direct scientific basis of the invention described in the patent. In this study we focus on a particular phenomenon, namely the extent to which the extreme SBs show up as SNPRs. Patent publications were gathered by searching the lens.org, created by Cambia (a non-profit organization in Australia dedicated to facilitating innovation) and Queensland University of Technology. The platform lens.org has linked the world's patent information to most of the scholarly literature with collaborations with CrossRef and National Library of Medicine. We group patent publications describing the same invention in 'patent families' to prevent double counting. As is shown in Table 2, six of the top 10 SBs identified by Bcp have been cited by patents, and the SB's first citation in terms of priority date (the earliest application date) in a patent usually appears to be earlier than the awakening year.

Concluding remarks

In this paper, we developed a systematic methodology for identifying SB publications and figured out their key characteristics. We tested the new index using the data of SB articles identified from *Science* and *Nature*. The results showed that both B and Bcp worked well with top class SBs, but Bcp is more sensitive for the lower level SBs, namely, the total citations and the maximum number of citations received in a year is moderate. Bcp works better than B in at least two aspects: (1) it "punishes" the situations when the SBs experienced early citations instead of continuous sleeping; (2) it allows for comparing the extent of delayed citation impact of publications in different fields with different citation patterns. We also figured out some key characteristics of SB publications. The extreme SBs are application oriented and significantly more cited by patents than 'normal' papers at the initial stage. The SB's first citation in a patent usually appears to be earlier than the awakening year. These findings demonstrated the potential technical-research and applied-research properties of Sleeping Beauties, rather than pure fundamental or basic-research. Based on the above mentioned analysis, we pondered some policy implications associated with SB publications.

The first is SB publications and transformative research. According to Thomas Kuhn's "paradigm" concept, there should be two types of innovative research, namely, cumulative processes and revolutionary breakthroughs. The latter is also called transformative research, which refers to research that shifts or disrupts established scientific paradigms. Identifying potential transformative research early and accurately is important for funding agencies to maximize the impact of their investments. It also helps scientists identify promising emerging works and focus their attention on them. Therefore, it is imperative for us to spare no effort to avoid delayed recognition and to detect SBs as early as possible, in order to promote potentially valuable but not readily accepted innovative research. Transformative research tends to be neglected or resisted by the scientific community initially and this neglect or resistance could be regarded as the key clue for the early prediction of sleeping beauty literature. Publications belonging to so-called transformative research, even when less frequently cited than others, should be given special attention as early as possible, because they may suddenly attract many citations after a period of sleep. We hold that scholars in both scientometrics and library and information science (LIS) should initiate the research for identifying transformative ideas. One could identify transformative research through some text terms (such as "disagree", "overcome", "break", "dispute"…). In order to discern such potential transformative research, one could observe whether the relevant documents get early citation from patents or not, because many Sleeping Beauty documents tend to be cited by patents and thus have more influence on technology and invention.

The second concerns SB publications and research fronts. The research front(s) studies based on citation analysis and visualization methods were one of most important topics in informetrics. However, research front and research frontier are different terms, which translated into same Chinese Word. The former usually took the "fast-highly cited papers" as basic data, and latter tend to be revealed by the delay-highly cited papers from the perspective of scientists, as is shown in terms of the landmark characteristic of the top ten sleeping beauty publications identified with Bcp. The Bcp index proposed in this paper may be used to identify the papers at the sleeping-awakening interface, which provided new tools for arousing attention of the science community to previously overlooked but important research.

The third is SB publications and research evaluation. Since the major achievements often encounter delayed recognition than the hot tracking research, it is recommended that citation delay reflected by Bcp can be used as an important index to evaluate the academic quality of papers. In the same vain, research evaluators should moderately extend the evaluation cycle. In bibliometric-based research assessment, one should give special attention to papers with a higher value of Bcp, because such papers accumulate their citations slowly but may show a longer and durable citation impact.

Acknowledgements This study was supported by the National Natural Science Foundation of China (Grant Nos. 71603280, 71373252). The present study is an extended version of an article presented at the 16th International Conference on Scientometrics and Informetrics, Wuhan (China), 16–20 October 2017. We thank Prof. Ronald Rousseau for his detailed review of our language and enlightening suggestions for the calculation of the Bcp index.

References

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,* 329–339.

Du, J., & Wu, Y. (2015). Defining and identifying the Sleeping Beauty and her Princes in Science (in Chinese). *Library and Information Service, 59,* 6–11.

Du, J., & Wu, Y. (2016). A bibliometric framework for identifying "Princes" who wake up the "Sleeping Beauty" in challenge-type scientific discoveries. *Journal of Data and Information Science, 9,* 50–68.

Fang, H. (2018). Analysing the variation tendencies of the numbers of yearly citations for sleeping beauties in science by using derivative analysis. *Scientometrics, 115,* 1051–1070.

Garfield, E. (1989). Delayed recognition in scientific discovery: Citation frequency analysis aids the search for case histories. *Current Contents, 23,* 3–9.

Ke, Q., Ferrara, E., Radicchi, F., et al. (2015). Defining and identifying Sleeping Beauties in science. *Proceedings of the National Academy of Sciences of the United States of America, 112,* 7426–7431.

Li, J., & Ye, F. Y. (2016). Distinguishing Sleeping Beauties in science. *Scientometrics, 108,* 821–828.

Liu, Y., & Rousseau, R. (2014). Citation analysis and the development of science: A case study using articles by some Nobel Prize winners. *Journal of the American Society for Information Science and Technology, 65,* 281–289.

Onodera, N. (2016). Properties of an index of citation durability of an article. *Journal of Informetrics, 10,* 981–1004.

Trapido, D. (2015). How novelty in knowledge earns recognition: The role of consistent identities. *Research Policy, 44,* 1488–1500.

Van Raan, A. F. J. (2004). Sleeping Beauties in science. *Scientometrics, 59,* 467–472.

Van Raan, A. F. J. (2015). Dormitory of physical and engineering sciences: Sleeping Beauties may be sleeping innovations. *PLoS ONE, 10,* e0139786.

Van Raan, A. F. J. (2017). Sleeping beauties cited in patents: Is there also a dormitory of inventions? *Scientometrics, 110,* 1123–1156.

Wang, J. (2013). Citation time window choice for research impact evaluation. *Scientometrics, 94,* 851–872.

Wang, J., Thijs, B., & Glanzel, W. (2015). Interdisciplinarity and impact: Distinct effects of variety, balance, and disparity. *PLoS ONE, 10,* 18.

Ye, F. Y., & Bornmann, L. (2018). "Smart girls" versus "sleeping beauties" in the sciences: The identification of instant and delayed recognition by using the citation angle. *Journal of the Association for Information Science and Technology, 69,* 359–367.