Journal of Informetrics 16 (2022) 101239

Contents lists available at ScienceDirect

Journal of Informetrics

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

Identifying keyword sleeping beauties: A perspective on the knowledge diffusion process

Jinqing Yang[a,b], Yi Bu[c], Wei Lu[a,b], Yong Huang[a,b,*], Jiming Hu[a,b], Shengzhi Huang[a,b], Li Zhang[a,b]

[a] *School of Information Management, Wuhan University, Wuhan 430072, China*
[b] *Information Retrieval and Knowledge Mining Laboratory, Wuhan University, Wuhan 430072, China*
[c] *Department of Information Management, Peking University, Beijing 100871, China*

A R T I C L E   I N F O

Keywords:
Sleeping beauty
Delayed recognition
Knowledge diffusion trajectory
Survival analysis

A B S T R A C T

Knowledge diffusion is a significant driving force behind discipline development and technological innovation. Keyword is a unique knowledge diffusion trajectory, in which the sleeping beauty phenomenon sometimes appears. In this paper, we first put forward the concept of Keyword Sleeping Beauties (KSBs) on the basis of the scientific literature phenomenon of sleeping beauties. Then, we construct a parameter-free identification method to distinguish KSBs based on beauty coefficient criteria. Furthermore, we analyze the intrinsic and extrinsic influencing factors to explore the awakening mechanism of KSBs. The experiment results show that sleeping beauty phenomena also exist in the keyword diffusion trajectory and 284 KSBs are identified. The depth of sleep has a positive correlation with awakening intensity, while the length of sleep has a negative correlation with awakening intensity. In the two years of pre-awakening, KSBs tend to appear in the journals with a higher impact factor. In addition, the adoption frequency and the number of KSBs both increase obviously in the one year of pre-awakening. The findings of this paper enrich the patterns of knowledge diffusion and extend academic thinking on the sleeping beauty in science.

1. Introduction

Knowledge diffusion is a significant driving force behind discipline development and technological innovation (Zheng, Pan, & Sun, 2019). It is of great theoretical and practical value to investigate the process of knowledge diffusion (Del Giudice, Della Peruta, & Maggioni, 2015; Liu, et al., 2020). To our knowledge, sleeping beauty in science is a specific pattern of knowledge diffusion, which signifies a unique knowledge diffusion trajectory after the publication of scientific literature (Hou & Yang, 2020). The citation network is widely adopted for investigating the diffusion trajectory of scientific knowledge (Braun, Glänzel, & Schubert., 2010; Lockett & McWilliams, 2005). However, knowledge diffusion is represented not only by the citation network but also by other means such as keyword adoption (Xu et al., 2018) and author collaboration (Wang, et al., 2015; Yang, Hu & Liu 2015).

Keywords have been used to uncover evolution laws of some research domains (Choi, Yi, & Lee, 2011; Ord, et al., 2005; Xu et al., 2018) since they were deliberately selected or created by authors to correctly express the subject matter of papers (Griffiths & Steyvers, 2004; Kim et al., 2013; Yang, Li, & Huang, 2018). Researchers have often regarded keywords as independent research topics in some studies of topic evolution (Duvvuru et al., 2013; Li & Yan, 2019; Osborne & Motta, 2015). Zhang et al. (2016a) and

* Corresponding author at: School of Information Management, Wuhan University, Wuhan 430072, China.
E-mail address: yonghuang1991@whu.edu.cn (Y. Huang).

https://doi.org/10.1016/j.joi.2021.101239
Received 21 April 2021; Received in revised form 28 November 2021; Accepted 1 December 2021
Available online 22 December 2021
1751-1577/

Zhang et al. (2016b) confirmed that author-keywords were more likely to represent research topics. More importantly, keywords have also been used as a medium for quantitatively (Callon, Law, & Rip, 1986) and flexibly (He, 1999) tracking the trajectory of knowledge diffusion. Keywords, as the smallest units of knowledge, have been used to express specific concepts intuitively and clearly, which is conducive to discovering the intricate patterns of sleeping beauty at the micro-level. González et al., (2018) found some keywords that had been considered to be dead appeared subsequently during their study of keyword survival analysis. In addition, it takes a while for a published article to be quoted (Gou et al., 2021; Mariani, Medo, & Zhang, 2016), which affects the accurate understanding of sleeping beauty from the perspective of the knowledge diffusion process. Thus, the citation relationship based on coarse-grained scientific literature was not conducive to revealing the fundamental rules of sleeping beauty in science. In contrast, in terms of time-sensitivity, keywords are timelier and more effective in the diffusion of knowledge. For example, when a keyword appears at different periods, it means the knowledge carried by the keyword is proliferating. Therefore, it is motivating to explore the sleeping beauty concept based on fine-grained keywords.

The "resurrection keywords" were also discovered during the keyword survival analysis. Peset et al. (2020) unexpectedly found the re-emergence of 3% of keywords judged to be dead and assumed that these might be keyword sleeping beauties (KSBs). With that in mind, this paper further explored sleeping beauty in science based on the keyword diffusion trajectory. Accordingly, we re-examined the scientific concept, identification method, and awakening factors of KSBs from the perspective of fine-grained keywords. In summary, this study extends academic thinking on the sleeping beauty in science at the fine-grained level. The main contributions are as follows:

(1) We propose a new idea for investigating sleeping beauties in science and redefine their concepts at the granularity level of keywords.
(2) Based on the survival analysis method, we compensate the defects of the beauty coefficient criterion by penalizing the early adoptions and propose an improved identification method for distinguishing keyword sleeping beauties.
(3) We explore the awakening mechanism of the keyword sleeping beauties based on intrinsic and extrinsic factors.

The remainder of this paper is structured as follows. Section 2 contains twofold surveys on a review of sleeping beauty criteria and its awakening mechanism as well as the novel perspectives of sleeping beauty. Section 3 introduces the data and methodology we used to identify KSBs. Section 4 presents the results of identifying KSBs and analyzing the factors which influence the awakening of KSBs. Section 5 mainly discusses the differences between KSBs and traditional sleeping beauty, the semantic problem of keywords whether it influences the experiment results, and the awakening mechanism and princes. Finally, we conclude with an overview of our findings and provide future research in Section 6.

2. Related work

2.1. A review of sleeping beauty criteria and its awakening mechanism

Van Raan (2004) described the phenomenon of delayed recognition as sleeping beauty in science and proposed three elements to identify sleeping beauties, namely depth of sleep, length of sleep and awakening intensity. On the basis of the three citation modes proposed by Van Dalen and Henkens (2005), Costas et al. (2010) further developed the quartile-based criteria, which classified the literature with the citation mode of the "slow increase and slow decrease" as sleeping beauties. By strictly defining the three elements of sleeping beauty, Li and Shi (2016) introduced a "boost effect" to devise rigorous formulaic identification criteria to avoid premature identification. On the other hand, researchers attempted to investigate the sleeping beauty criteria without parameters. For example, Li et al. (2014) adopted adjustment of the Gini coefficient to calculate the inequality of "heartbeat spectrum". After that, Sun, Min, and Li (2016) used an adjustment of the Gini coefficient for measuring the inequality of citation distribution to recognize sleeping beauties. Moreover, by using the ideas of mathematical linear programming optimality, Ke et al. (2015) proposed a "beauty coefficient" (B index) approach of the parameter-free measure to indicate the possibility for being a sleeping beauty. In reference to the "beauty coefficient" (B index), Du and Wu (2018) developed a novel "beauty coefficient percentage" (Bcp index) by improving yearly citations as a yearly accumulative percentage of citations. Sleeping beauty criteria could include the following twofold, namely the criteria with parameter and the parameter-free criteria.

As components of the sleeping beauty research, it is essential to analyze and understand how sleeping beauties are awakened. The concept of "Prince" was introduced at the same time as sleeping beauties (Van Raan, 2004). According to the definition of "Prince" literature, Braun, Glänzel, and Schubert (2010) significantly found the average impact factor of journals was higher for "Prince" literature during analyzing the awakening mechanism. Ohba and Nakao (2012) further made a rigorous definition of "Prince" literature and found self-citation might be an important impact factor for awakening the sleeping beauties. Since scientific literature of sleeping beauty might be likely to be cited by patent literature, the scope of "Prince" has been extended in the term of document types (Van Raan, 2015, 2017; Van Raan, & Winnink, 2018). The "Prince" of sleeping beauties also involved some patents besides academic literature.

There are merits and demerits to each kind of sleeping beauty criteria. Although the criteria with parameters require a fix-threshold value which is subjective, we could receive detailed effects of three elements of sleeping beauties, i.e., the characteristics of sleeping beauties for different depths of sleep, lengths of sleep, and awakening intensities. The parameter-free criteria involve few subjective threshold values so that it provided a common model to identify sleeping beauties. The awakening of sleeping beauties is influenced by several factors. The journal impact factor and document type both had a significant influence on the awakening of sleeping beauties.

Our study, therefore, integrates two types of identification methods and analyses the awakening mechanisms of sleeping beauties from a journal perspective.

2.2. The novel perspectives of sleeping beauty

The sleeping beauty phenomenon is widely presented in different disciplines. Specifically, researchers have explored this phenomenon in different domains, such as medicine (El Aichouchi & Gorry, 2018), biology (Ke, 2018), computational science (Dey et al., 2017), and internet business (Teixeira, Fonseca, & Vieira, 2020). Although sleeping beauty was first discovered in journal papers, researchers have continued to explore this topic in different types of documents. Teixeira, Vieira, and Abreu (2017) selected innovative journals to explore the sleeping beauty literature therein and attempted to explain their genesis and inherent mechanisms of being awakened. Van Raan (2017) investigated sleeping beauties of scientific literature that are cited in patents to find potential sleeping innovations. Further, Hou and Yang (2019) constructed a quantitative model to identify the sleeping beauties on the concept of “all-elements -sleeping-beauties” (Li, 2014; Li &Ye, 2012) from the perspective of patents. Du, et al. (2020) predicted early indications of the sleeping beauty journal literature based on the citation relationship between patents and journal literature. Hou and Yang (2020) defined sleeping beauties in science based on a new diffusion trajectory by using social media metrics and investigated an awakening mechanism for sleeping beauties from the perspective of social media (Hou, Li, & Zhang, 2020).

The patent text has demonstrated the “sleeping beauties” phenomenon. Likewise, sleeping beauties have been identified at the fine-grained level. Delayed recognition has already appeared in early studies of information diffusion, and Wang et al. (2015) used a novel concept of “information sleep” in the context of the sleeping beauty study from the information diffusion perspective. The sleeping beauty phenomenon was also exhibited in the process of online information diffusion (Zhang, Xu, & Zhao, 2017; Zheng, Pan, & Sun, 2019). Du, et al. (2020) investigated the object of knowledge delayed recognition research at the fine-grained level. In the concept level, Van Raan (2015) analyzed the cognitive environment of sleep beauties by measuring all co-occurrences of any possible pair of concepts. In the subsequent research, concept maps and citation-based maps have strengths for discovering the awakening mechanism and other important application-oriented research directly related to sleeping beauty (Van Raan, 2017). From this, we could see that the concept maps were leveraged for making an in-depth analysis of sleeping beauties and their awakening mechanism. In addition, it was found that fine-grained units such as information, knowledge entities, and discipline topics commonly undergo a period of hibernation before a sudden outbreak of popularity. In particular, the concept maps were constructed to analyze the sleeping beauties by measuring all co-occurrences of any possible pair of concepts, which inspires the idea of understanding the sleeping beauty at a fine-grained level.

3. Data & Methodology

3.1. Data collection and processing

The data for this study was gained from an open dataset in the field of LIS submitted by Peset et al. (2018). To ensure data reliability, the first step is to automatically convert to lowercase letters and remove the spaces in front of and behind the keywords. The second step is to solve the abbreviation, morphology, and synonym of some keywords, for example, we unified the “part of speech blocks” and “POS blocks” into “part of speech blocks”, “distortion rate” and “rate distortion” into “distortion rate”. Then, new keyword data was first extracted in the period of 1990-1999, which consists of four columns, namely keywords, year of publication, source title, and ISSN. A total of 7,136 new keywords were collected, which appeared for the first time in the initial year. Then, the history data of all new keywords were used for identifying KSBs between 1990 and 2014.

In terms of morphology, since keywords generally end in nouns, we only unified the singular with the plural for the keywords. In general, stemming is one of the more popular methods to unify keyword morphology, but the lexical diversity of keywords is also destroyed. For example, “indexing” and “index” have the same etyma, but they represent two different concepts in the information retrieval domain.

3.2. Defining concepts of KSBs

Van Raan (2004) defined the concepts of sleeping beauty in science, including depth of sleep, length of sleep, and awakening intensity. To explore sleeping beauties from a fine-grained perspective, we first defined the keyword sleeping beauties (KSBs) on the basis of the scientific literature phenomenon of sleeping beauties and redefined characteristics of KSBs, and the corresponding metrics are shown in Fig. 1.

(1) Definition of KSBs

At the granularity level of keywords, we employed a keyword diffusion trajectory to describe the KSBs and quantify the approval degree through the keyword adoption frequency.

(1) Depth of sleep

Depth of sleep refers to how low is the frequency of keyword adoption. Specifically, it can be expressed as the average value of adoption frequency ($\overline{AF}$) divided by the maximum value of adoption frequency during the sleeping period ($\underset{t_a-t_0}{max}(AF)$), namely $\overline{AF}/\underset{t_a-t_0}{max}(AF)$.

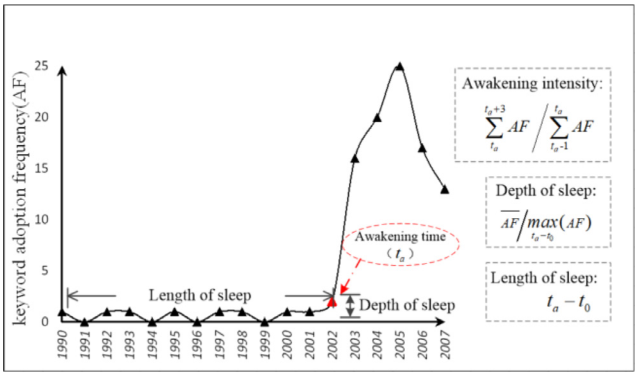


Fig. 1. The characteristics of keyword sleeping beauties.

(1) Length of sleep

Length of sleep is the period experienced from the appearance of the keyword to the rapid acquisition of a large amount of attention, and it can also be described as the duration experienced from the first appearance of the keyword to awakening, namely $(t_a - t_0)$.

(1) Awakening intensity

Awakening intensity refers to the degree of state shift from sleeping to awakening, which could be formulated as the cumulative frequency of keyword adoption in the three years of post-awakening ($\sum_{t_a}^{t_a+3} AF$) divided by that in the one year of pre-awakening ($\sum_{t_a-1}^{t_a} AF$), namely $\sum_{t_a}^{t_a+3} AF / \sum_{t_a-1}^{t_a} AF$.

3.3. KSB identification

As mentioned in related work, extant sleeping beauty criteria were divided into criteria with parameters and parameter-free criteria. However, criteria with parameters were rules of thumb that summarized by citation count the papers received, and are not suitable for distinguishing fine-grained KSBs. Therefore, this paper integrated the survival analysis method based on the beauty coefficient parameter-free criteria to design a suitable KSB identification method.

3.3.1. Keyword survival analysis

Survival analysis is a branch of statistics for investigating the expected duration of time until one or more incidents happen which is widely utilized in medicine, economics, engineering science, social science, and other domain-specific fields (Singer & Willett, 1993). Peset et al. (2020) applied a survival analysis method to calculate the survival probability of new keywords over the following 10 years. To adapt the analysis of the survival time of the keywords, the survival analysis function proposed by Kaplan-Meier was improved and re-designed (González et al, 2018); i.e., after a specific time point t, the status of some keywords changed from 1 to 0. The proportion of keywords whose status remained 1 after time t was calculated using the following formula (1):

$$S(t) = \prod_{i: t_i \le t} \left(1 - \frac{d_i}{n_i}\right) \tag{1}$$

Where t_i is the interval time, d_i represents number of keyword status changes, and n_i indicates the element whose state remains equal to the original value 1.

Keyword survival analysis utilizes the whole history data of a keyword, which could compensate for the defects of the beauty coefficient criterion to penalize premature accumulation. The KSBs are special keywords beyond the routine, just as many sleeping papers are not awakened (Van Raan, 2004) so that they barely become sleeping beauty papers. The keywords that have not been awakened are regarded as dead keywords in survival analysis. Therefore, this study used the sleeping duration over the average survival time of all keywords as a penalty condition for premature accumulation to distinguish KSBs.

3.3.2. The identification method for KSBs

The keywords collected were first scrubbed according to the rules in part 3.1, and then the beauty coefficient criterion was utilized to find out the linear equation (2) of the primary function from the initial year (t_0) to the year with a maximum value of adoption frequency (t_m). After that, we calculated the B value, which represents the cumulative difference between the equation of

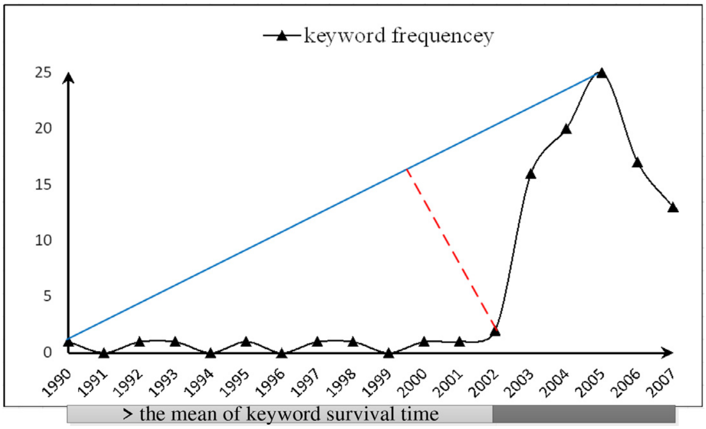


Fig. 2. Diagram of improved beauty coefficient criteria.

the line and the actual value, using equation (3), i.e., the integral area formed by cutting the real line to the curve in Fig. 2, and used equations (4) and (5) to determine the point in time when the sleeping keyword is awakened; $t_a - t_0$is length of sleep time:

$$\ell_t = \frac{c_{t_m} - c_0}{t_m} \cdot t + c_0 \tag{2}$$

$$B = \sum_{t=0}^{t_m} \frac{\frac{c_{t_m} - c_0}{t_m} \cdot t + c_0 - c_t}{\max\{1, c_t\}} = \sum_{t=0}^{t_m} \frac{\ell_t - c_t}{\max\{1, c_t\}} \tag{3}$$

$$d_t = \frac{\left|(c_{t_m} - c_0)t - t_m c_t + t_m c_0\right|}{\sqrt{(c_{t_m} - c_0)^2 + t_m^2}} \tag{4}$$

$$t_a = \arg\left\{\max_{t \le t_m} d_t\right\} \tag{5}$$

Where $d_t$is the distance from the point on the curve to the line, t_a indicates awakening time point, t_m is the year with a maximum value of adoption frequency, $c_{t_m}$is the maximum value of adoption frequency, c_0 is the adoption frequency in the initial year, t is the age of the keywords, and c_t represents adoption frequency of keywords in the t th year.

Finally, we found the distribution of B values calculated by the formula (3) follows power-law distribution and decided to adopt this rule for identifying KSBs based on long-tail theory. The method for identifying KSBs is shown in Fig. 3.

4. Results of KBS identification

In order to identify KSBs, we first need to sort out the history data of each keyword, so that the new keywords from 1990 to 1999 can be identified from the keywords appearing in 1990 to 2014.

4.1. Beauty coefficient

According to the parameters of the beauty coefficient criterion, the maximum and initial value of adoption frequency are used to construct the linear equation. Then the B value can be calculated by the formula (3). Some of the results are shown in Table 1.

By analyzing the distribution of B values, it was found that the keyword B values excluding the zero follow a power-law. Furthermore, to verify the power-law distribution, we employed four representative fitting functions from the statistical analysis software Origin 8.5 (Stevenson, 2011) to fit the B values that appeared in 1990-1999, respectively. Taking the B values in 1990 as an example, the fitting results are shown in Table 2.

According to the goodness-of-fit condition that larger R^2 and F values indicate better curve fitting, it can be seen that the power function model (R^2=0.8939, F=1010.8042) has the best fit, but the quadratic function also has a good fit when comparing the fitting results of the four models in Table 2. To further determine the distribution function of B values, we fit power and quadratic functions to the 1990-1999 data, respectively. The distributions of R^2 are shown in Fig. 4.

The mean value of the power function R^2 is greater than that of the quadratic function (0.88>0.87) and the standard deviation of the R^2 is less than that of the quadratic function (0.036<0.051). In addition, as shown in Table 3, the mean values of the quadratic coefficients a, b and the constant c are 73.38, -1.68, and 0.01 respectively. Additionally, the mean of a values is nearly 43 times that of b values.

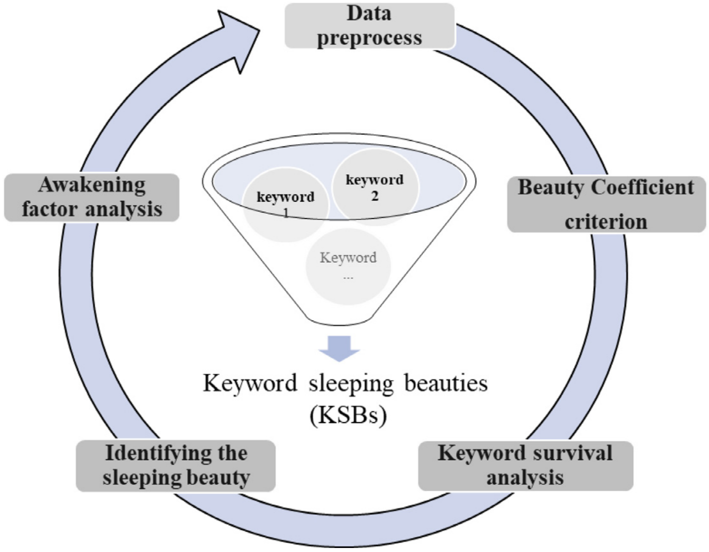


Fig. 3. Identification method of KSBs.

Table 1
The beauty coefficient of new keywords in 1990.

keyword	B value	keyword	B value
information management	168.14	management	69.71
electronic medium	142.70	cooperation	66.73
leadership	130.07	computational modeling	63.00
budget	125.70	philosophy	62.78
machine learning	96.62	index	61.04
critical thinking	93.99	writing	60.67
information security	93.55	principal component analysis	58.46
creativity	90.19	bibliography	52.37
user	88.11	game theory	50.79
productivity	83.87	linguistics	50.69
competition	72.28	uncertainty	50.62

Note: See all results at https://doi.org/10.6084/m9.figshare.14393249.v1.

Table 2
The model functions and parameter values.

Model name	equation	R-square	F value
Allometric1	$y = ax^b$	0.8939	1010.8042
Exp1p2Md	$y = B^x$	-1.0350	7.7435
Parabola	$y = c + bx + ax^2$	0.8900	649.2300
Logarithm	$y = \ln(x - A)$	-0.1610	91.1145

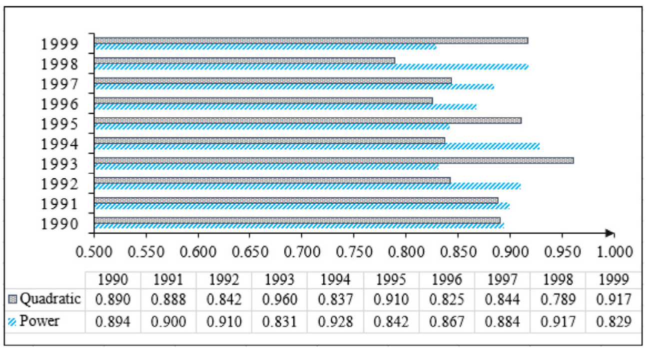


Fig. 4. R^2values distribution of power and quadratic functions.

Table 3
Fitting parameter values of the quadratic function.

Parameter	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
a	110.19	94.41	84.02	67.69	94.72	78.16	53.55	50.38	46.09	34.56
b	-2.55	-3.13	-2.88	-2.64	-2.50	-1.50	-0.36	-0.42	-0.51	-0.27
c	0.02	0.03	0.03	0.03	0.02	0.01	0.00	0.00	0.00	0.00

Table 4
Average survival time of keywords for each group.

Groups	Means	Std. Error	Lower Bound	Upper Bound
1990	11.59	0.561	10.49	12.691
1991	9.996	0.609	8.803	11.189
1992	9.996	0.592	8.836	11.156
1993	9.079	0.57	7.963	10.196
1994	9.656	0.499	8.678	10.633
1995	8.407	0.384	7.655	9.159
1996	8.136	0.204	7.736	8.536
1997	6.621	0.184	6.26	6.982
1998	5.809	0.176	5.465	6.154
1999	6.281	0.178	5.931	6.631

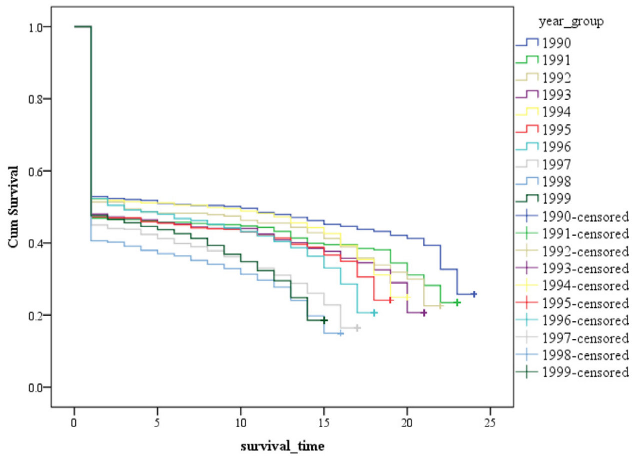


Fig. 5. Plot of keyword survival function between 1990 and 1999.

According to the above analysis, it can be seen that B values distribute like a power law, following the formula $y = ax^b$,where y indicates the B values; x is the inverse serial number; a, b are the constants estimated from the domain data set. It means that there are only a few keywords with a high B value and mounts of keywords with a small B value, which alludes to the fact that sleeping beauties are scarce. Since the B value is used to measure the possibility of sleeping beauties, the top 20% of keywords are considered to be the candidate KSBs considering the power-law distribution of B values. It is worth noting that the top-n could be adjusted as required.

4.2. The results of keyword survival analysis

To compensate for the drawback that the beauty coefficient criterion has failed to penalize premature growth, we treated sleeping duration exceeding the average survival time of all keywords as a penalty condition for early accumulation. Firstly, we extracted the keywords that appeared for the first time in their history data from 1990 to 1999. Secondly, the frequency of new keywords was calculated in each year, and the death year is the time when the adoption frequency is not zero for the first time in descending order. Finally, the survival time was calculated by using the Kaplan-Meier analysis in SPSS 22 with the initial year (1990-1999) as the “treatment group” and the survival time and survival status (disappearance as “1” and presence as “0”) as the variables. The average survival time for each group is shown in Table 4, and the overall survival time distribution is shown in Fig. 5.

In Table 4, the mean survival time of new keywords was different in each group within the 95% confidence interval, and the p-values for the A log-rank test and Wilcoxon test were less than 0.001. The survival time of keywords is correlated with the age of the grouping data. Therefore, the grouping calculation should be performed on the average survival time to penalize premature accumulation.

Table 5
KSBs and their characteristic values.

Keyword	year	Depth of sleep	Length of sleep	Awakening intensity
machine learning	1990	7.91	22	9.71
competition	1990	22.22	22	15.00
information security	1990	14.29	21	12.33
creativity	1990	14.47	21	23.00
user	1990	25.00	20	13.00
...	...	...	...	...
bias	1999	12.44	9	12.00
journal	1999	12.14	9	8.00
advertising	1999	25.60	8	8.00
scientific journal	1999	24.00	8	14.00
text mining	1999	33.45	7	12.50
knowledge creation	1999	26.67	7	19.00
scholarly communication	1999	20.57	7	13.00

Note: For all KSBs see https://doi.org/10.6084/m9.figshare.14393249.v1.

Table 6
Correlations among the depth of sleep, awakening intensity, and length of sleep.

Data item	Correlation parameter	Depth of sleep	Awakening intensity	Length of sleep
Depth of sleep	Pearson Correlation	1	0.228**	- 0.147*
	Sig. (2-tailed)	—	0.000	0.013
Awakening intensity	Pearson Correlation	0.228**	1	
	Sig. (2-tailed)	0.000	—	- 0.181**
Length of sleep	Pearson Correlation	- 0.147*	- 0.181**	1
	Sig. (2-tailed)	0.013	0.002	—

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

4.3. The findings of awakening mechanism analysis

According to the candidate KSBs, the average survival time of the keywords in each group was used as the penalty condition to eliminate the candidate ones with early fast accumulating frequency, and finally, a total of 284 KSBs were identified. The length of sleep, depth of sleep, and awakening intensity of the KSBs are shown in Table 5.

The awakening mechanism of sleeping beauty is one of the crucial tasks in the study of the delayed recognition phenomenon. How to detect the KSBs as early as possible is of great interest to researchers. In general, awakening intensity is influenced by intrinsic factors, i.e., depth of sleep and length of sleep, and external factors, i.e., journal impact and keyword distribution in the journal. As a result, we provide the following three findings:

(1) A correlation was observed among the depth of sleep, awakening intensity, and length of sleep.

We performed correlation analysis among the length of sleep, depth of sleep, and awakening intensity for the 284 KSBs. It can be seen in Table 6 that the correlation between depth of sleep and awakening intensity was 0.228 ($p<0.01$), with a positive correlation between the two, while the length of sleep was negatively correlated with the depth of sleep (-0.147, $p<0.05$) and awakening intensity (-0.181, $p<0.01$), respectively. It can be noticed that the depth of sleep has a positive correlation with awakening intensity, while the length of sleep has a negative correlation with awakening intensity.

(2) After being published by high-impact journals, these KSBs obtained more attention and were then awakened.

To verify this assumption, we first obtained information on journals in the field of LIS for 1990 to 2014 from the SJR (Scimago Journal & Country Rank), which includes SJR for evaluating journal impact (Guerrero-Bote & Moya-Anegón, 2012), H index, Cites / Doc. (2years), etc. Then, the journal impact was calculated before and after the awakening of the KSBs. Since SJR, H index and cites are different perspectives for measuring journal impact, we performed the normalization for the three impact indicators and took the average value of the three as a criterion in this experiment. We selected the Z-score normalization method to statistically calculate the journal impact values for six years of pre-awakening and three years of post-awakening, as shown in Fig. 6.

Fig. 6 shows that the impact of the journals in which KSBs were published increased significantly in the two years of pre-awakening, which indicates that KSBs were adopted by the journals with high impact before they were awakened. And the impact of the journals decreases after awakening in a short time.

(3) The KSBs were not universally adopted before the awakening, and the spread trajectory in the journals was single. After KSBs awakened, they were widely adopted, and the number of distributed journals was enhanced.

To verify the distribution of journals during the proliferation of KSBs, we counted the adoption frequency (AF) of KSBs, and the number of KSBs, the number of journals in which KSBs were published, etc., as shown in Table 7.

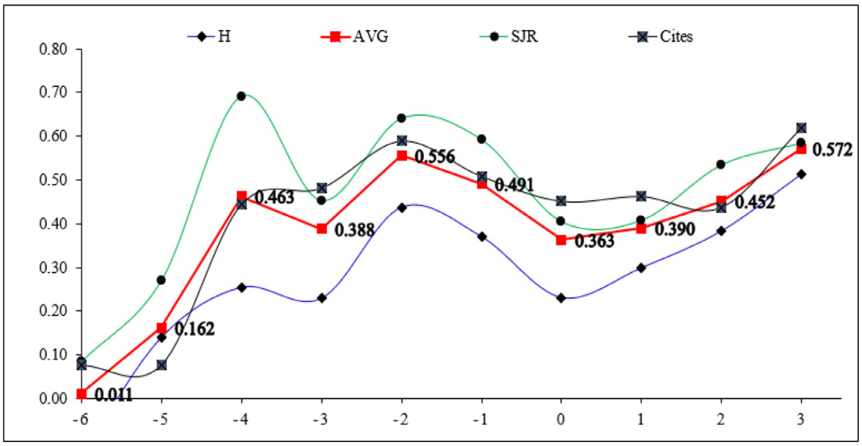


Fig. 6. The calculation of the journal impact before and after awakening.
Note: The red curve with decimals is the average value of z-score normalization.
"-1" represents the 1 year of pre-awakening, "1" shows the 1 year.
of post-awakening, and "0" annotates the awakening year in the x-axis.

Table 7
Distribution of journals and adoption frequency before and after awakening.

Time	Adoption frequency (AF) of KSBs	Number of KSBs	Number of journals	Average AF in each journal	Average AF of each KSB
-6	253	105	48	5.271	2.410
-5	415	151	71	5.845	2.748
-4	521	167	78	6.680	3.120
-3	596	182	86	6.930	3.275
-2	723	212	96	7.531	3.410
-1	982	246	105	9.352	3.992
0	411	139	87	4.724	2.957
1	1912	266	121	15.802	7.188
2	2072	258	121	17.124	8.031
3	2341	232	135	17.341	10.091

Note: "1" represents the 1 year of post-awakening, "-1" shows the 1 year of pre-awakening, "0" annotates the awakening year in the "Time" column.

As can be seen in Table 7, the number of journals in which KSBs were published increased significantly in the one year of pre-awakening, and journal number increased significantly. The average adoption frequency of each KSB also showed a significant increase in the one year of pre-awakening, while more KSBs were adopted, up to 86.6%. The average adoption frequency of KSBs in each journal also showed a significant increase in the one year of the pre-awakening. The increase of journal numbers was accompanied by the growth of keyword adoption frequency. Therefore, the growth of KSBs started quietly before awakening and it was more stable after awakening.

5. Discussion

Knowledge diffusion is one kind of scientific development pattern. When the delayed recognition in science was first announced, it was investigated from the perspective of the knowledge diffusion trajectory of the literature. Although citation relationships between literatures are the most direct way of characterizing knowledge diffusion trajectories, they are not the only way. Moreover, the citation relationship is established by the selection of papers to support the ideas in the text, but the variety of citation intentions, the possible non-existence of topic relevance between the citing and the cited literature, and the phenomenon of self-citation all have impacts on the real value of the cited frequency. Especially, the cumulative effect of the literatures being cited over time might cause an awakening time lag. Keywords from one literature to another also signal the diffusion trajectory of knowledge. Authors select a limited number of keywords to indicate the research topics in conjunction with the research content of their papers. In other studies (Van Raan, 2015; 2017), domain terms and concepts are also used to define the research topics, which brought the idea of concept-based sleeping beauties. To the best of our knowledge, a concept generally includes multiple terms or keywords, thus keywords have narrower semantics than their concepts. If research topics were represented by concepts, the keywords natively correspond to sub-topics. Thus, the identification of the keyword sleeping beauty not only contributes to a deeper understanding of knowledge diffusion patterns but also expands the researches of sleeping beauty in science from a more fine-grained perspective.

Although keywords refine the granularity of knowledge, they also bring about the problem of polysemy and synonymy. In the polysemy issue, there will be the following doubts: multi-semantic keywords and mono-semantic keywords are adopted with relatively different adoption frequencies. Multi-semantic KSBs are easier to awaken while mono-semantic ones are more difficult, which also has an influence on the results of sleeping beauty identification. To address this issue, we designed the experiment with the hypothesis that the sleeping depth of multi-semantic keywords is shorter than that of mono-semantic keywords. We grouped the keywords according to the number of their sub-word units and calculated the sleeping depth of the identified KSBs separately. The results are shown in Fig. A.1. of Appendix A. The KSBs are mostly single words or two words combined, and the average difference between the two sleeping depths is only less than 0.2%. In addition, the correlation test between the number of keyword combinations and sleeping depth was $p>0.05$. Thus, the semantics of keywords within a given domain are more specific and have little effect on the experimental results. In the synonymy issue, to further verify whether the keywords that mean almost the same thing have existed in the reprocessed keywords, we calculated the semantic similarity scores between reprocessed keywords. The calculation strategy and results are shown in Fig. A.2. of Appendix A. By observing and analyzing the results, we find that although some keywords have similar semantics, they still represent two different topics. There are few synonyms in our studies after reprocessing keywords.

As a novel perspective of sleeping beauty in science, there are some consistencies and differences between the keyword sleeping beauties and traditional sleeping beauties (sleeping beauties-based citation). In the views of internal characteristics, comparing with van Raan's summaries about the length of sleep, depth of sleep, and awakening intensity (Van Raan, 2015): (1) The probability of awakening declines for longer length of sleep (2) Length of sleep matters less for the probability of awakening for less deep sleep, we found that KSBs has the same finding that the depth of sleep has a positive correlation with awakening intensity, whereas the correlation analysis result that the length of sleep has a negative correlation with awakening intensity is different from that of van Raan's sleeping beauties-based citation. Besides, the length of sleep is not less than 5 years in van Raan's definition of sleeping beauties, whereas the minimum of the sleeping duration of KSBs is 6 years.

In the terms of the awakening mechanism, since the author keywords generally appear once a time in the keyword list, we identified keywords as the princes of the KSBs within 3 years before or after in which KSBs awakens (Dey et al., 2017). The detailed process of identifying the princes of the KSBs is shown in Table A.1 of Appendix A. We found that each KSB could have about 3 princes who awaken her. Moreover, from the journal perspective, we found after being published by high-impact journals, the KSBs obtained more attention and were then awakened. Namely, the KSBs adopted in a journal of lower impact may be spotlighted by getting adopted in a higher impact journal, which is consistent with that of the traditional sleeping beauties (Braun et al., 2010). We also found the KSBs were not universally adopted before the awakening, and the spread trajectory in the journals was single. After KSBs awakened, they were widely adopted, and the number of distributed journals was enhanced, which also implies that the KSBs were paid more attention and then awakened. In order to verify the robustness of the results at the journal level, we conducted the additional experiment in an interdisciplinary field. The results are shown in Appendix B.

6. Conclusion

Based on the knowledge diffusion trajectory, this study refined the description of knowledge diffusion from a fine-grained perspective and deeply explored the diffusion pattern of the sleeping beauty phenomenon at the keyword level. Accordingly, the scientific concept, identification method, and awakening factors of sleeping beauty are re-examined in this study. We proposed the KSB identification method based on the previous studies of sleeping beauty. Firstly, keywords in the field of LIS were obtained and pre-processed, and the keywords newly appearing from 1990 to 1999 were extracted as the target of the experiments. Then the beauty coefficient criterion was implemented to calculate the B values of these keywords. Since it is a lack of punishment for the early cumulative frequency that gives rise to prematurely identifying some keywords as KSBs, the average survival time of domain keywords was inserted as one of the adjudication conditions of the KSBs. Finally, we selected the top 20% as the candidate KSBs based on the power-law distribution of B values and combined this with the average survival time of the keywords to identify KSBs.

For identified KSBs, we explored the factors that influenced their awakening time. In terms of internal factors, according to the definitions of the length of sleep, depth of sleep, and awakening intensity, we calculated the values of the three attributes respectively. And then the correlation degree of sleeping length—sleeping depth, sleeping length—awakening intensity, and sleeping depth—awakening intensity was statistically analyzed in groups, and the results showed that length of sleep was negatively correlated with the depth of sleep, length of sleep was negatively correlated with awakening intensity, and depth of sleep was positively correlated with awakening intensity. In terms of external factors, we verified that, before the awakening, KSBs tend to appear in the journals with a higher impact factor, and the adoption frequency and the number of KSBs both increased significantly in the journals. In future research, these research directions are worth investigating, e.g., keywords mobility across disciplinary fields whether has an effect on sleeping beauties identification, the characteristics differences of keyword sleeping beauties and concept sleeping beauties, and other awakening factors.

CRediT authorship contribution statement

Jinqing Yang: Conceived and designed the analysis, Collected the data, ontributed data or analysis tools, Performed the analysis, Wrote the paper. **Yi Bu:** Conceived and designed the analysis, Collected the data, Contributed data or analysis tools, Performed the analysis, Wrote the paper. **Wei Lu:** Conceived and designed the analysis, Collected the data, Contributed data or analysis tools, Performed the analysis, Wrote the paper. **Yong Huang:** Conceived and designed the analysis, Contributed data or analysis tools, Performed the analysis, Wrote the paper. **Jiming Hu:** Performed the analysis, Wrote the paper. **Shengzhi Huang:** Collected the data, Contributed data or analysis tools. **Li Zhang:** Collected the data, Contributed data or analysis tools.

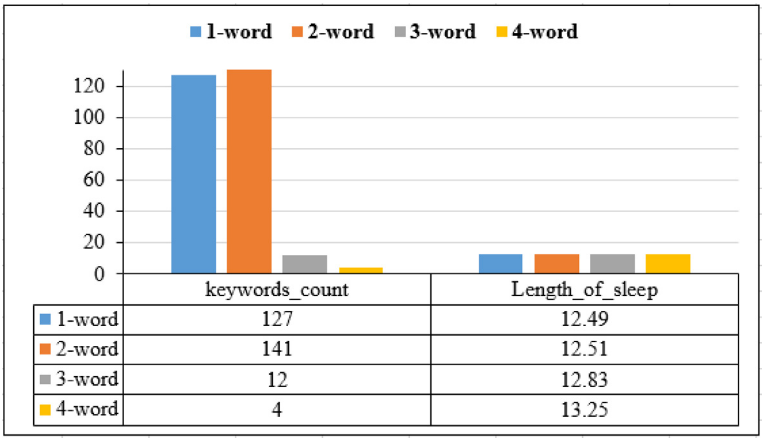
Acknowledgments

This work is supported by Major Projects of the National Social Science Foundation of China (no. 17ZDA292). We are also grateful to editors and anonymous reviewers for their helpful and valuable comments on our work.

Appendix A

1. We designed the experiment with the hypothesis that the sleeping depth of multi-semantic keywords is shorter than that of mono-semantic keywords and grouped the keywords according to the number of their sub-word units and calculated the sleeping depth of the identified KSBs separately.
2. To further verify whether the keywords that mean almost the same thing have existed in our study, we conducted the semantic analysis in the two levels. Specifically, we firstly considered a normalization index ($S_{ij} = c_{ij}/s_i \times s_j$), defined by Eck and Waltman (2009) to calculated the similarity values (S_{ij}) between two keywords i and j. c_{ij} indicates the count of co-occurrences, $s_i$and s_j denotes the occurrences of keywords i and j, respectively. Then, since there are few synonyms in the same keyword list, we need to verify the synonyms from another perspective. The basic idea is that two keywords that frequently co-occur with the same keyword may be synonyms.

The analysis results show that the co-occurrence count of the 676, 477 pairs of keywords was calculated and the keywords with co-occurrence greater than 2 are selected for semantic similarity calculation, as shown in Fig. A.2. And, we could extract all triads of keywords from the semantic network and gained 1508 pairs of keywords, which might be synonyms. By observing and analyzing keyword pairs with high similarity scores, we find that although some keywords are semantically similar, they still



	keywords_count	Length_of_sleep
1-word	127	12.49
2-word	141	12.51
3-word	12	12.83
4-word	4	13.25

Fig. A.1. Relationship between the keyword length and sleeping depth of KSBs.
Note: 1-word means the keywords consisted of one word.

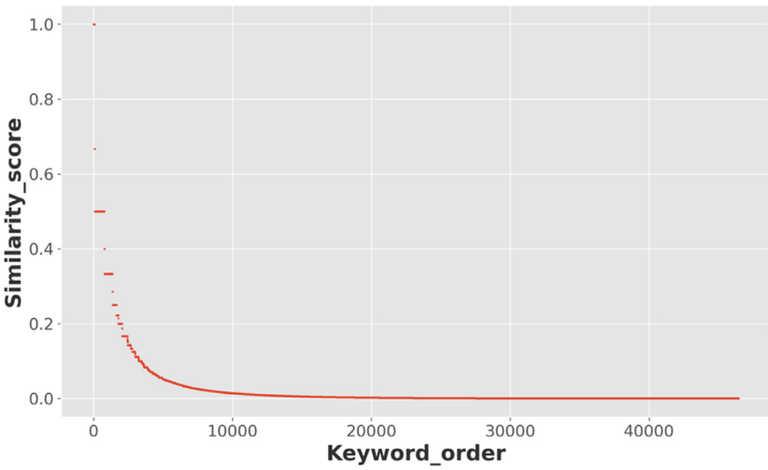


Fig. A.2. Distribution of all similarity scores.

Table A.1
The prince and retinues of the KSBs.

Keyword	year	Number of prince and its retinues
machine learning	1990	7
competition	1990	1
information security	1990	3
creativity	1990	1
user	1990	1
...	...	...
bias	1999	1
journal	1999	3
advertising	1999	1
scientific journal	1999	3
text mining	1999	2
knowledge creation	1999	1
scholarly communication	1999	2

Note: For all KSBs see https://doi.org/10.6084/m9.figshare.15172296.v2.

represent two different topics. Therefore, there are few synonyms after reprocessing keywords. The all results of similarity calculation see https://doi.org/10.6084/m9.figshare.15172296.v2

1. Since the author keywords generally appear once a time in the keyword list, we identified literature in which KSBs awakens as the princes that adopted KSBs within 3 years before or after (Dey et al., 2017). The prince and retinues (Du & Wu, 2016) of the KSBs is shown in Table A.1 of Appendix A.

Appendix B

In order to verify the robustness of the results of influencing factors at the journal level, we first selected the data sources of interdisciplinary field, medical informatics, an interdisciplinary field. We used the query "SU = Medical informatics" to obtain 58,986 papers and 75,721 keywords in the field of medical informatics from the Web of Science (WOS) (Duration until 31 December 2019).

Secondly, since medical informatics is an emerging interdisciplinary field and the history span of publication is from 1990 to 2019, we obtained 14,013 new keywords from 1995 to 2004 to identify the KSBs. 406 KSBs were then obtained. Thirdly, we selected SJR and Cites / Doc (2 years) for evaluating journal impact to verify the result "After being published by high-impact journals, these KSBs obtained more attention and were then awakened." We still selected the z-score normalization method to statistically calculate the journal impact values for six years of pre-awakening and three years of post-awakening, as shown in Fig. B.1.

Additionally, to verify the results of the distribution of journals during the proliferation of KSBs, namely "The KSBs were not universally adopted before the awakening, and the spread trajectory in the journals was single. After KSBs awakened, they were widely adopted, and the number of distributed journals was enhanced.". We counted the adoption frequency (AF) of KSBs, and the number of KSBs, the number of journals in which KSBs were published, etc., as shown in Table B.1.

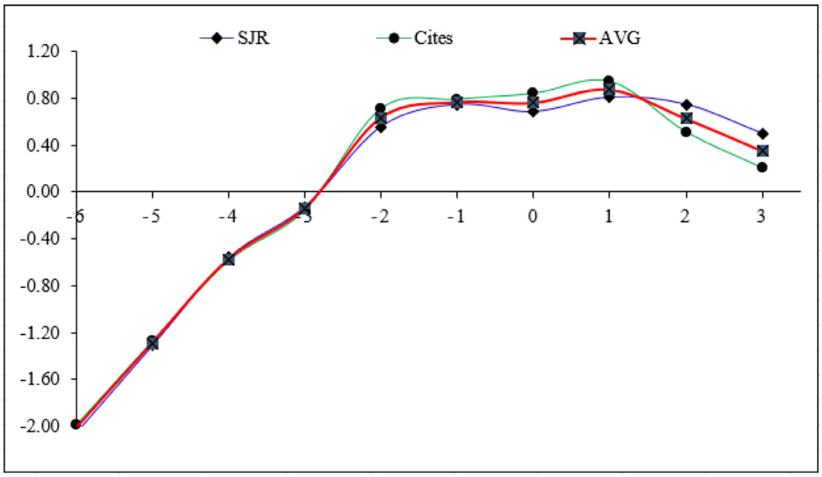


Fig. B.1. The calculation of the journal impact before and after awakening in medical informatics.
Note: The red curve with decimals is the average value of z-score normalization.
"-1" represents the 1 year of pre-awakening, "1" shows the 1 year.
of post-awakening, and "0" annotates the awakening year in the x-axis.

Table B.1
Distribution of journals and adoption frequencies before and after awakening.
in the field of medical informatics.

Time	Adoption frequency (AF) of KSBs	Number of KSBs	Number of journals	Average AF in each journal	Average AF of each KSB
-6	374	204	34	11.000	1.833
-5	414	211	39	10.615	1.962
-4	460	222	42	10.952	2.072
-3	546	244	42	13.000	2.238
-2	698	264	43	16.233	2.644
-1	772	306	46	16.783	2.523
0	221	95	31	7.1293	2.326
1	1173	300	41	28.610	3.910
2	1329	319	42	31.643	4.166
3	1458	322	44	33.136	4.528

Note: “1” represents the 1 year of post-awakening, “-1” shows the 1 year of pre-awakening, “0” annotates the awakening year in the “Time” column.

Therefore, the results of influencing factors at the journal level look robust. After we delve into our thoughts, the results seem to be a common phenomenon in the diffusion of knowledge. Specifically, the journals with high impact would get more attention from researchers because of their high-quality papers; the keyword contained also obtained more attention. Yet, this phenomenon deserves further study in the future.

References

Braun, T., Glänzel, W., & Schubert, A. (2010). On Sleeping Beauties, Princes and other tales of citation distribution. *Research Evaluation, 19*(3), 195–202.

Callon, M., Law, J., & Rip, A. (1986). Qualitative scientometrics. In M. Callon, J. Law, & A. Rip (Eds.), *Mapping the dynamics of science and technology: Sociology of science in the real world* (Eds.). London: The Macmillan Press Ltd.

Choi, J., Yi, S., & Lee, K. C. (2011). Analysis of keyword networks in MIS research and implications for predicting knowledge evolution. *Information & Management, 48*(8), 371–381.

Costas, R., van Leeuwen, T. N., & van Raan, A. F. (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.

Del Giudice, M., Della Peruta, M. R., & Maggioni, V. (2015). A model for the diffusion of knowledge sharing technologies inside private transport companies. *Journal of Knowledge Management, 19*(3), 611–625.

Dey, R., Roy, A., Chakraborty, T., & Ghosh, S. (2017). Sleeping beauties in Computer Science: characterization and early identification. *Scientometrics, 113*(3), 1645–1663.

Du, J., Li, P., Haunschild, R., Sun, Y., & Tang, X. (2020). patent citation linkages as early signs for predicting delayed recognized knowledge: Macro and micro evidence. *Journal of Informetrics, 14*(2), Article 101017.

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, 1*(1), 50–68.

Du, J., & Wu, Y. (2018). A parameter-free index for identifying under-cited sleeping beauties in science. *Scientometrics, 116*(2), 959–971.

Duvvuru, A., Radhakrishnan, S., More, D., Kamarthi, S., & Sultornsanee, S. (2013). Analyzing structural & temporal characteristics of keyword system in academic research articles. *Procedia Computer Science, 20*, 439–445.

Eck, N. J. V., & Waltman, L. (2009). How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American society for information science and technology, 60*(8), 1635–1651.

El Aichouchi, A., & Gorry, P. (2018). Delayed recognition of judah folkman’s hypothesis on tumor angiogenesis: When a prince awakens a sleeping beauty by self-citation. *Scientometrics, 116*(1), 385–399.

González, L. M., García-Massó, X., Pardo-Ibañez, A., Peset, F., & Devís-Devís, J. (2018). An author keyword analysis for mapping Sport Sciences. *PloS one, 13*(8), Article e0201435.

Gou, Z., Meng, F., Chinchilla-Rodríguez, Z., & Bu, Y. (2021). Revisiting the obsolescence process of individual scientific publications: Operationalisation and a preliminary cross-discipline exploration. In *Proceedings of the 18th International Conference on Scientometrics and Informetrics (ISSI 2021)* (pp. 477–488). July 12-15, 2021.

Griffiths, T. L., & Steyvers, M. (2004). Finding scientific topics. In *Proceedings of the National Academy of Sciences of the United States of America: 101* (pp. 5228–5235).

Guerrero-Bote, V. P., & Moya-Anegón, F. (2012). A further step forward in measuring journals’ scientific prestige: The SJR2 indicator. *Journal of informetrics, 6*(4), 674–688.

He, Q. (1999). Knowledge discovery through co-word analysis. *Library Trends, 48*(1), 133–159.

Hou, J., Li, H., & Zhang, Y. (2020). Identifying the princes base on Altmetrics: An awakening mechanism of sleeping beauties from the perspective of social media. *Plos one, 15*(11), Article e0241772.

Hou, J., & Yang, X. (2019). Patent sleeping beauties: evolutionary trajectories and identification methods. *Scientometrics, 120*(1), 187–215.

Hou, J., & Yang, X. (2020). Social media-based sleeping beauties: Defining, identifying and features. *Journal of Informetrics, 14*(2), Article 101012.

Ke, Q. (2018). Comparing scientific and technological impact of biomedical research. *Journal of Informetrics, 12*(3), 706–717.

Ke, Q., Ferrara, E., Radicchi, F., & Flammini, A. (2015). Defining and identifying sleeping beauties in science. In *Proceedings of the National Academy of Sciences of the United States of America: 112* (pp. 7426–7431).

Kim, S. N., Medelyan, O., Kan, M. Y., & Baldwin, T. (2013). Automatic keyphrase extraction from scientific articles. *Language resources and evaluation, 47*(3), 723–742.

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., & 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., & Ye, F. Y. (2014). A study of the “heartbeat spectra” for “sleeping beauties. *Journal of informetrics, 8*(3), 493–502.

Li, K., & Yan, E. (2019). Are NIH-funded publications fulfilling the proposed research? An examination of concept-matchedness between NIH research grants and their supported publications. *Journal of Informetrics, 13*(1), 226–237.

Liu, W., Tan, R., Li, Z., Cao, G., & Yu, F. (2020). A patent-based method for monitoring the development of technological innovations based on knowledge diffusion. *Journal of Knowledge Management, 25*(2), 380–401.

Lockett, A., & McWilliams, A. (2005). The balance of trade between disciplines: Do we effectively manage knowledge? *Journal of Management Inquiry, 14*(2), 139–150.

Mariani, M. S., Medo, M., & Zhang, Y. C. (2016). Identification of milestone papers through time-balanced network centrality. *Journal of Informetrics, 10*(4), 1207–1223.

Ohba, N., & Nakao, K. (2012). Sleeping beauties in ophthalmology. *Scientometrics, 93*(2), 253–264.
Ord, T. J., Martins, E. P., Thakur, S., Mane, K. K., & Börner, K. (2005). Trends in animal behavior research (1968–2002): Ethoinformatics and the mining of library databases. *Animal Behavior, 69*(6), 1399–1413.
Osborne, F., & Motta, E. (2015). Klink-2: integrating multiple web sources to generate semantic topic networks. In *International Semantic Web Conference* (pp. 408–424). Springer. pages.
Peset, F., Garzón-Farinós, F., González, L. M., García-Massó, X., Ferrer-Sapena, A., Toca-Herrera, J. L., & Sánchez-Pérez, E. A. (2020). Survival analysis of author keywords: An application to the library and information sciences area. *Journal of the Association for Information Science and Technology, 71*(4), 462–473.
Peset, F., Garzón-Farinos, F., González, L., García-Massó, X., Ferrer-Sapena, A., Toca-Herrera, J., & Sánchez-Perez, E. (2018). Dataset. *Figshare*. Retrieved from https://figshare.com/s/5313334d203f800e869c.
Singer, J. D., & Willett, J. B. (1993). It's about time: Using discrete-time survival analysis to study duration and the timing of events. *Journal of Educational Statistics, 18*(2), 155–195.
Stevenson, K. J. (2011). Review of originpro 8.5. *Journal of the American Chemical Society, 133*(14), 5621.
Sun, J., Min, C., & Li, J. (2016). A vector for measuring obsolescence of scientific articles. *Scientometrics, 107*(2), 745–757.
Teixeira, A. A., Fonseca, Â., & Vieira, P. C. (2020). Sleeping beauties and their princes in international business. *Journal of Business & Finance Librarianship, 25*(1-2), 44–72.
Teixeira, A. A., Vieira, P. C., & Abreu, A. P. (2017). Sleeping Beauties and their princes in innovation studies. *Scientometrics, 110*(2), 541–580.
Van Dalen, H. P., & Henkens, K. N. (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.
Van Raan, A. F. (2015). Dormitory of physical and engineering sciences: Sleeping beauties may be sleeping innovations. *PloS one, 10*(10), Article e0139786.
Van Raan, A. F. (2017). Sleeping beauties cited in patents: Is there also a dormitory of inventions? *Scientometrics, 110*(3), 1123–1156.
Van Raan, A. F., & Winnink, J. J. (2018). Do younger Sleeping Beauties prefer a technological prince? *Scientometrics, 114*(2), 701–717.
Wang, J. P., Guo, Q., Yang, G. Y., & Liu, J. G. (2015). Improved knowledge diffusion model based on the collaboration hypernetwork. *Physica A: Statistical Mechanics and its Applications, 428*, 250–256.
Xu, J., Bu, Y., Ding, Y., Yang, S., Zhang, H., Yu, C., & Sun, L. (2018). Understanding the formation of interdisciplinary research from the perspective of keyword evolution: A case study on joint attention. *Scientometrics, 117*(2), 973–995.
Yang, G. Y., Hu, Z. L., & Liu, J. G. (2015). Knowledge diffusion in the collaboration hypernetwork. *Physica A: Statistical Mechanics and its Applications, 419*, 429–436.
Yang, L., Li, KP., & Huang, HF. (2018). A new network model for extracting text keywords. *Scientometrics, 116*(1), 339–361.
Zhang, J., Yu, Q., Zheng, F., Long, C., Lu, Z., & Duan, Z. (2016a). Comparing keywords plus of WOS and author keywords: A case study of patient adherence research. *Journal of the Association for Information Science and Technology, 67*(4), 967–972.
Zhang, L., Xu, K., & Zhao, J. (2017). Sleeping beauties in meme diffusion. *Scientometrics, 112*(1), 383–402.
Zhang, Y., Shang, L., Huang, L., Porter, A. L., Zhang, G., Lu, J., & Zhu, D. (2016b). A hybrid similarity measure method for patent portfolio analysis. *Journal of Informetrics, 10*(4), 1108–1130.
Zheng, W., Pan, H., & Sun, C. (2019). A friendship-based altruistic incentive knowledge diffusion model in social networks. *Information Sciences, 491*, 138–150.