Sleeping Beauties in Meme Diffusion

Leihang Zhang\textsuperscript{1}, Jichang Zhao\textsuperscript{2,*} and Ke Xu\textsuperscript{1}

\textsuperscript{1} State Key Lab of Software Development Environment, Beihang University
\textsuperscript{1} School of Economics and Management, Beihang University

\textsuperscript{*}Corresponding author: jichang@buaa.edu.cn

(Dated: April 27, 2016)

Abstract

A sleeping beauty in diffusion indicates that the information, can be ideas or innovations, will experience a hibernation before a sudden spike of popularity and it is widely found in citation history of scientific publications. However, in this study, we demonstrate that the sleeping beauty is an universal phenomenon in information diffusion and even more inspiring, there exist two consecutive sleeping beauties in the entire lifetime of propagation, suggesting that the information, including trending topics, search queries or Wikipedia views, which we call memes, will go unnoticed for a period and suddenly attracts some attention, and then it falls asleep again and later wakes up with another unexpected popularity peak. Further explorations on this phenomenon show that intervals between two wake ups follow an exponential distribution and the second awakening stage generally reaches its peak at a higher velocity and will bring a wider dissemination. Taking these findings into consideration, the upgraded Bass model can lead to promising predictions for the meme diffusion on different media. Our results shed lights on disclosing the common mechanism behind different memes and can help locate the tipping point in marketing like realistic scenarios.

PACS numbers: 89.65.-s, 89.75.Fb
I. INTRODUCTION

Meme is usually defined as the simplest cultural unit that spreads between different individuals and may gain collective attention within a community or culture \[2, 11\]. Dawkin even postulates meme as a cultural analogy of genes in order to explain how innovations, ideas, catchphrases, melodies, rumors, or fashion trends disseminate through a population \[11\]. In recent decades, Internet and its various applications, like pubMeds, Wikipedia or online social media, provide massive digital fossils of meme diffusion, which offer us a decent proxy to disclose the mechanism beyond the propagation. The insights from these investigations can help us understand the rules that produce the dynamics and establish prediction models that estimate the future trends.

Basic dynamics of the meme diffusion within the same media has been comprehensively studied from different perspectives. For example, mathematical epidemiology as well as simple log-normal distributions are suggested to profile the growth and decline of diffusion \[2, 21, 32, 34\], how competition, homophily and network cooperatively affect the spread is discussed \[7, 8, 14, 15, 35\], different roles played by different individuals are revealed \[3, 43\] and even simulation models are established to replicate the meme diffusion in Twitter \[38–41\]. However, except disclosing the common features of successful memes in different online social networks \[9, 10, 31, 33\], the universal mechanism that essentially drives the propagation of memes in different media still remains unclear. In this study, we argue that the sleeping beauty existing in the lifetime of different memes can be a path to unravel the common knowledge behind diffusions of different media.

Sleeping beauty, exhibiting a hibernation before an unexpected popularity peak, is pervasively found and studied in the diffusion of memes like ideas or innovations in scientific publications. Garfield first provide examples of articles with delayed recognition \[12, 13\], which can be identified through the citation history \[16, 37\]. Van then coined the term “sleeping beauty” in reference to the delayed recognition \[37\] and several basic features, including sleep length, sleep depth and awake intensity are proposed to measure the sleeping beauty. Later, finding general features of the sleeping beauty \[24, 26, 28, 30\] and explaining the awakening reasons in paper citations \[4–6, 19, 23\] attract most of the attention. Indeed, understanding the sleeping beauty in scientific development will help improve the citation impact and mine the surprising innovation \[36\], however, most of the previous studies only
focus on the scientific publications and ignore the possibility that in other memes, like trending topics in social media, hot queries in Google or popular items in Wikipedia, might also experience the similar sleeping beauty, which in fact greatly motivates the present work.

Actually the existing evidence already implicitly implies the latent connection between different memes in terms of sleeping beautifies. For scientific papers, Li et al. find that there are some papers appearing “flash in the pan” first and then “delay recognition”\(^{22}\), i.e., these papers experience two sleeping beautifies in their citation history. More uplifting is that the Internet slang words also demonstrate the same phenomenon in Weibo\(^{43}\). This similarity suggests that for a meme, no matter it is an idea in scientific citations or a slang word in social media, its diffusion might be produced by a common rule. Hence in this study, by assuming that different memes in different media share the same diffusive mechanism, we try to disclose the possible common rule through the in-depth investigation on the phenomenon of sleeping beauties. And the knowledge of this common rule can indeed help upgrade the existing prediction approaches and extend them to many different domains.

In order to systematically explore the sleeping beauties in different memes’ diffusion, we investigate three kinds of datasets, including utilization volume of \( n \)-grams in scientific publications titles during a long period, Wikipedia item view count, and search queries in Google. That’s to say, the memes we focus come from different backgrounds and guarantee the universality of our following findings. Note that the time granularity in meme diffusion is also diverse, including year, week or month for different data sets and it further ensure that our study can discuss the entire lifetime of the propagation from a long-term perspective. As shown in Fig. 1, six typical meme are randomly sampled to demonstrate their diffusion dynamics with different time granularities. As can be seen, all the memes experience two obvious sleeping beauties, which means that each of them goes unnoticed for a period and suddenly attracts considerable attention, and then falls asleep again, following with another unexpected popularity spike even higher than the previous one. This interesting phenomenon of two sleeping beauties, which to our best knowledge not discussed in the previous work, is independent to the media and time granularity. And it raises many unsolved but natural questions like how to identify these two beautifies without parameter settings, how they distribute in lifetime of the diffusion and can they be used to predict the future trends.

Inspired by the index of beauty coefficient\(^{17}\), we introduce a parameter-free framework to identify two sleeping beauties in each meme diffusion. We demonstrate that the phe-
Figure 1. The popularity dynamics of six memes from different data sets. For convenience, we use popularity to denote the utilization volume, which is generally considered as the degree of public concern on a particular meme. (A) The yearly popularity of 2-grams “swine flu” from 1809 to 2013. (B and C) The relative search volume of two queries, “Oki Matsumoto” and “Iwai Yukiko” in Google and the time granularity are respectively week and month. (D, E and F) The Wikipedia page views for three items, including “Sano Gaku”, “Shield of Straw” and “Kurebayashi AsaTakeshi”, and the time granularity is respectively set to day, week, and month.

The phenomenon of two sleeping beauties is pervasively existing in diffusion of different memes and time intervals between the two awakening stages follow an exponential distribution. Besides, the second awakening stage generally rises to peak at a higher velocity and thus brings a wider dissemination. Based on these findings, we model the two sleeping beauties through the classical Bass diffusion model and produce promising predictions for the meme diffusion. We disclose a common mechanism beyond meme diffusion in terms of two sleeping beauties and our results will shed lights on revealing the essential rule that drives the
popularity dynamics of memes in different media.

II. RESULTS

Inspired by approaches presented in [29] and [17], we develop methods to detect and measure the phenomenon of two sleeping beauties in diffusion of different memes (see Methods). In addition, the measurement solution is parameter-free and can be easily extended to different domains. The probe on different data sets demonstrates that the two sleeping beauties are pervasively existing in different media and can be convincing proxy to explore the common mechanism beyond different meme diffusion. Hence we observe several key characteristics of the sleeping beauties first, including the time interval between two beauties, the velocity of rise after awakening and the length of the wake up and then we try to establish a prediction model to replicate the popularity dynamics of memes from different media.

A. Exponential intervals between beauties

For each meme with two sleeping beauties in diffusion, we can obtain the first awakening time $t_{a1}$, the time of obtaining the first peak popularity $t_1$, the first falling asleep time $t_{f1}$, the second awakening time $t_{a2}$, the time of obtaining the second peak popularity $t_2$ and the second falling asleep time $t_{f2}$ (see Methods). We define the time interval between different beauties as $t_{a2} - t_{f1}$, which reflect the length of the second sleeping. Assuming the first sleeping beauty is observed, then this time interval can be employed to predict the second awakening time, i.e., when the meme will start to experience a new spike of popularity. And from [43] it is also demonstrated that generally the spike in the second sleeping beauty will be much higher than the first one, so given the first sleeping and awakening information, successful prediction of the second awakening time will be a sharp break.

We measure the time intervals between different sleeping beauties for different memes in our data sets and surprisingly find that they follow stable exponential distributions in different media and the coefficients are very close to each other. As can be seen in Fig. 2, $\lambda$, the exponential coefficient, is respectively 0.0792, 0.0246, 0.0461, 0.0263, 0.0218 and 0.0511 for different data sets with different time granularities. Considering the fact that expo-
Figure 2. Distribution of time intervals between beauties (A) Distribution of time interval in n-grams of publication titles. (B and C) Distribution of time interval in search queries of Google Trends. Note that here the popularity is defined as the normalized search frequency in Google. (D, E and F) Distribution of time interval in Wikipedia page views. Note that \( R \) denotes Pearson’s correlation and higher values stand for better fittings.

The exponential distribution has the key property of being memoryless, the above finding suggests that only temporal patterns may not be enough to predict the second awakening instant. Meanwhile, note that \( \lambda \) in different data sets mainly locates in the narrow range of \([0.02, 0.08]\), indicating that even for different media, the distribution is almost the same and it further support our hypothesis that there exists a common and media-independent mechanism beyond the meme diffusion.

B. The comparison between two wake ups

During a wake up, the meme obtains collective attention and demonstrates popularity spikes. Being consistent with the previous finding in [43], the wake up of the second beauty
**Figure 3.** Correlation between the total popularity of two wake ups. $m_1$ is the total popularity during the first wake up and $m_2$ is the total popularity in the second wake up. (A) N-grams in publication titles. (B and C) Search queries from Google Trends with time granularity of hour and week. Popularity denotes the normalized search frequency in Google. (D, E and F) Wikipedia page views with time granularity of day, week and month.

will get much more popularity than the first one (see Fig 4). Moreover, the comparison of the total popularity, i.e., the total search volume in Google, Wikipedia page views or occurrences in paper titles, shows that more attention in the first wake up will lead to even more popularity in the second wake up, as can be seen in Fig 3. Meanwhile, we also compare the length of the rising stage in the two waking periods, which can be intuitively defined as $t_i - t_{a_i}$ ($i = 1, 2$), and as shown in Fig 4 it can be seen that independent to the media, memes in different data sets generally reach their peak popularity within a shorter rising time in the second waking period. Another metric to reflect the form of the popularity spike is the rising velocity, which can be directly defined as $v_i = (p_{t_i} - f_{t_{a_i}})/(t_i - t_{a_i})$. As can be seen in Fig. 5 the rising velocity in the second wake up is 3-5 times faster than the one in
Figure 4. Comparison of rising stage length in awakening period. $G_1$ and $G_2$ denote the first and second wake up. (A) $N$-grams of publication titles. (B and C) Search queries from Google Trends with time granularity of week and month. (D, E and F) Wikipedia views with with time granularity of day, week and month.

The rising stage length

with the first wake up. In addition, we also compare the length of falling stage in the two wake ups in Fig. E which are almost equal to each other. With significantly faster rising velocity and approximately equal rising stage length, the latter beauty accordingly produces more popularity than the first.

C. Prediction of the popularity

With the aim of predicting future popularity of different memes in different media, we have to neglect many detailed and domain-dependent factors like community structure, homogeneity or competition and focus on establishing a general framework only based on the statistics from the two sleeping beauties. Because of this, we upgrade the classical diffusion model named Bass model, which was developed by Bass [1] and Norton [27], to
Figure 5. **Comparison of rising velocity in wake up.** $\mu_1$ and $\mu_2$ denote the average rising velocity in $G_1$ and $G_2$. (A) $N$-grams of publication titles. (B and C) Search queries from Google Trends with time granularity of week and month. (D, E and F) Wikipedia page views with time granularity of day, week and month.

predict the memes’ popularity dynamics.

Bass model originally depicts the diffusion of innovation and imitation. Specifically, innovators create innovation and the other individuals in the social system might adopt the innovation at different time. Considering the pervasive existing of two sleeping beauties in the lifetime of memes, here we correspondingly separate the entire diffusion into two generations $G_i (i = 1, 2)$. Let $S_i(t)$ represent the popularity of the meme at time $t$, which can be obtained from

$$S_1(t) = m_1 F_1(t) - m_1 F_1(t) F_2(t - ta_2) = m_1 F_1(t)[1 - F_2(t - ta_2)],$$  \hspace{1cm} (1)

and

$$S_2(t) = m_2 F_2(t - ta_2) + m_1 F_1(t) F_2(t - ta_2) = F_2(t - ta_2)[m_2 + m_1 F_1(t)],$$  \hspace{1cm} (2)
$m_i$ represents the diffusion potential for the $i_{th}$ diffusion of the meme while $F_i(t)$ is the diffusion rate of the $i_{th}$ diffusion at time $t$ and can be evaluated by

$$F_i(t) = \begin{cases} 0, & t < 0 \\ 1 - e^{-(p_i+q_i)t}, & t \geq 0 \\ \frac{(q_i/p_i) e^{-(p_i+q_i)t} + 1}, & t \geq 0 \\ \end{cases}$$

in which $p_i$ is the innovation coefficient and $q_i$ denotes the imitation coefficient. For the case of two sleeping beauties, the innovation coefficient can be calculated through

$$p_i = \frac{S_i(ta_i + 1)}{m_i}.$$  

Note that the approach of sleeping beauty identification (see Methods) tends to pick the instant with lowest and nearest popularity before $t_1$ as the awakening point, indicating that the popularity at $ta_i$ is generally close or equal to zero. In order avoid the this, it is
reasonable to choose the popularity at the next point, i.e. \( ta_i + 1 \), to estimate the innovation coefficient. And the imitation coefficient \( q_i \) can be calculated from

\[
t_i - ta_i = \frac{1}{pi + qi} \ln(q_i/pi).
\]

(5)

![Histograms](image)

Figure 7. Distribution of \( p_i \) from different datasets. \( \sigma_i \) and \( \mu_i \) are standard deviation and mean of innovation coefficient in \( G_i \). (A) N-grams of publication titles. (B and C) Search queries of Google Trends with time granularity of week and month. (D, E and F) Wikipedia page views with time granularity of day, week and month.

In meme diffusion, \( p_i \) can model the influence brought by external factors while \( q_i \) in fact represents the internal factors (e.g. social networks) that might also shape the popularity dynamics. With respect to \( p_i \), as can be seen in Fig. 7, it is surprising to find that all the memes’ \( p_i \) follows a Gaussian-like distribution and the first beauty possesses a higher averaged value than the second. The Gaussian-like distribution, especially the mean locating in a very narrow range, further testify our assumption that different memes from different media are driven by a common mechanism beyond to attract collective attention. And higher \( p_i \)
for the first beauty also suggests that for each meme, the external factors functions more significantly in the first beauty than in the second. Besides, it should be noted that the variance of $p_i$, reflected by $\sigma_i$, is lower in the second beauty, indicating that in diffusion of the second beauty, the external factors function more homogeneously for different memes than that in the first one. Regarding to imitation coefficient $q_i$, as seen in Fig. 8, the average value in the first beauty is about two times higher than the second one. However, higher imitation coefficient cannot guarantee broader diffusion, because the imitation pressure at $t_n$ from the internal network can be quantified as $(q_i/m_i) \sum_{t=1}^{t_n} S_i(t)$ while $\sum_{t=1}^{t_n} S_i(t)$ is greatly determined by the rising velocity in the awakening period and as shown in Fig. 5 the rising velocity in the second beauty is usually about five time higher than the first one.

In the conventional parameter estimation, Bass model infers the innovation coefficient,
Figure 9. Evaluation of the prediction. The innovation coefficient is set to the mean of all memes in each data set. \(R\) denotes Pearson’s correlation between the curves of average actual and simulation popularity. (A) N-grams of publications titles. (B and C) Search queries of Google Trends with time granularity of week and month. Popularity denotes the normalized search frequency in Google. (D, E and F) Wikipedia page views with time granularity of day, week and month.

imitation coefficient and total diffusion amount from the first three observations of the diffusion. However, because many memes reach their peak popularity in no more than three time units, the previous method is not applicable here and we have to present new estimation approach. We set the first awaken instant as the start timing for each meme and \(ta_1, t_1, tf_1, ta_2, t_2\) and \(tf_2\) can be obtained from each sleeping beauty’s popularity history. Gaussian-like distribution of \(p_i\) implies that the average value from historical observations can be the estimation for the innovation coefficient and accordingly the imitation coefficient can be obtained through Eq. 5. Then the total popularity \(m_i\) for \(G_i\) can be estimated by \(\sum_{t=ta_i}^{tf_i} S_i(t)\).
Figure 10. **Evaluation of the prediction.** The innovation coefficient is set to the realistic value of each meme. (A) N-grams of publications titles. (B and C) Search queries of Google Trends with time granularity of hour, week and month. Popularity denotes the normalized search frequency in Google. (D, E and F) Wikipedia page views with time granularity of day, week and month.

To facilitate the model evaluation, the realistic trend of the popularity is averaged over different memes and the average popularity at $t$ is defined as $\langle S_i(t) \rangle = 1/n \sum_{j=1}^{m} S_{ij}(t)$, where $n$ stands for the number of memes whose $S_i(t) \neq 0$ and $m$ is the total number of sleeping beauties. Then its similarity (e.g., Person correlation) to the averaged prediction will be used to vividly demonstrate the model performance.

As can be seen in Fig. 9, we obtain promising predictions for the locations of different beauties from different data sets and the Pearson correlation between the simulated and realistic popularity dynamics in monthly Google trends and Wikipedia page views are particularly high. Fig. 10 shows the replication by setting each sleeping beauty’s innovation coefficient to its observed value instead of the average (same to the imitation coefficient),
the minor improvement further justifies the validity of using the mean of distribution to estimate the innovation coefficient for different memes, implying that our prediction approach is robust to memes, domains and time granularities.

Note that we may not get the precise estimation of the absolute value of the popularity at certain time, but we aim to give a promising prediction of where two sleeping beauties will locate. What’s more, the volume difference between the prediction and the realistic is partly caused by the fact that the popularity in sleeping period is assumed to zero, which actually is a nonzero number. The above results indicate that based on the statistics of two sleeping beauties in different memes, we can use the upgraded Bass model to replicate the popularity trend for memes of different media, especially the location of two beauties, which greatly help to identify key points in the entire diffusion.

III. DISCUSSION

By providing detailed tracks of diffusion, Internet indeed offers us an unprecedented opportunity to compressively understand the dynamics of information. However, the previous study tends to only focus on one particular domain and neglect the possibility that there might be a common mechanism beyond the propagation of different information. Motivated by this, in this study we try to disclose the common mechanism from the perspective of the sleeping beauty, which is formerly studied in citation history of scientific publications.

Surprisingly, we find the sleeping beauty, especially the phenomenon of two consecutive sleeping beauties, is pervasively existing in diffusion of different information, i.e., memes, from different media. We systematically explore the basic features of two beauties and establish a Bass-model-based approach to predict the future dynamics. The promising result of our prediction model solidly demonstrates that the mechanism behind different memes exists and its function in driving the formation of popularity dynamics can be reflected by the two sleeping beauties we reveal.

Even more inspiring, our investigations also show that for all memes with two sleeping beauties, the second beautify generally produces much higher popularity spike than the first one, i.e., it leads to the highest peak in the lifetime of diffusion. Recall that the highest peak is conventionally thought to be caused by the tipping point of the meme, hence our prediction approach can be employed directly to locate the timing of the tipping point in
specific media, which is intuitively the awakening time of the second beauty. So from this view, our findings and solutions can be important to realistic applications like marketing business.

In summary, the mechanism beyond the diffusion dynamics of memes might be complicated and hard to be tracked, however, we argue that the phenomenon of two sleeping beauties could provide a new and insightful view to understand it. And in the future work, we would like to develop more detailed models to capture the fine-granular dynamics of memes and provide the prediction both in volume and timing.

IV. MATERIALS AND METHODS

A. Datasets

The experiment is conducted on three kinds of datasets. The first dataset is metadata for the complete set of all PubMed records from 1809 to 2012 (with part of 2013 available as well), including title, authors, and year of publication [20, 25]. We segment the title into $n$-grams ($n=1, 2, 3$) and finally get 86784 uni-gram, 190992 bi-grams, and 47646 tri-grams. From the total 325422 $n$-grams, we obtained 734 memes with two obvious sleeping beauties in the diffusion.

The other two datasets are obtained from Wikipedia and Google Trends, which are publicly available from [42]. The frequency of an item in Wikipedia or a Google search query generally reflects the collective attention on a particular subject in the real world [18]. So both of Wikipedia page views and Google Trends are valuable resources for information diffusion study. The search query set contains 3231 keywords about Cartoon, 7251 keywords about comic, 10000 keywords about movie and 10000 celebrity names. These queries were accessed most frequently from 2008 to 2014 and their daily, weekly and monthly search volumes in Google Trends and page views in Wikipedia are all collected [42]. It should be noted that the number of queries in Google Trends might be a little different from that in Wikipedia, because Google Trends does not provide search statistics for queries with very low frequencies. What’s more, Google Trends only offers relative search volume of queries and the maximum value is 100.

In total, we get a large number of memes with two sleeping beauties from the three
TABLE I. Statistics of the Data sets

| Time Granularity | Google Trends | Wikipedia | Paper Titles |
|------------------|---------------|-----------|--------------|
| Day              | –             | –         | 29072 1859   | – –          |
| Week             | 13195 1402    | 29072 3815| – –          |
| Moth             | 22535 1641    | 29072 1946| – –          |
| Year             | –             | –         | 325422 734   |

For each dataset, the left column denotes the number of all memes and the right column denotes the number of memes with significant phenomenon of two sleeping beauties.

datasets and the detailed statistics are summarized in Table I. The fraction of memes with two sleeping beauties ranges from 5% to 10% in Wikipedia page views and Google search queries, which agrees well with the statistics in [38]. Note that because of noisy grams in paper titles, the ratio of memes with two sleeping beauties in data set of n-grams is relatively low.

B. Identification of the two sleeping beauties

The first step of identifying sleeping beauties is to locate two significant peaks in the meme’s popularity dynamics by successfully filtering out the nosily cascades with many but trivial popularity peaks. Palshikar [29] proposed a simple but effective algorithm for peak detection from a number of noisy cascades. By using this method with $k = 5$ and $h = 0.5$ (the parameter selection is introduced in [29]), we first locate the highest peak ($P_2$) from the popularity dynamics and then target another meaningful peak that happens before $P_2$. If there exists another peak $P_1$, the meme will be labelled as the one with two sleeping beauties.

According to three dimensions of describing sleeping beauty, i.e., the length of sleep, the depth of sleep, and the awake intensity presented in [37], Ke et al. proposed a index named Beauty coefficient ($B$) and then introduced a parameter-free framework to measure the sleep beauty [17]. Inspired by their approach, as illustrated in Fig. 11, we further establish a method to measure the phenomenon of two sleeping beauties. Given a meme, we define $p_t$ as the popularity at time $t$ after its appearance, i.e., $t$ stands for the age of the meme and
Figure 11. Illustration of the identification of two awakening instants $t_{a1}$ and $t_{a2}$, and two falling asleep instants $t_{f1}$ and $t_{f2}$. The blue curve represents the popularity dynamics of a meme at age $t$ (i.e., $t$ represents the time since its appearance). Dotted lines represent the reference lines.

the diffusion starts at $t_0 = 0$ and ends at time $T$.

Assume that the meme receives its maximum popularity $p_{t_2}$ at time $t_2$, then the straight line that connects points $(0, p_0)$ and $(t_2, p_{t_2})$ in the time-popularity plane, denoted as $l_{a_t}$, can be depicted as

$$l_{a_t} = \frac{p_{t_2} - p_0}{t_2} t + p_0. \quad (6)$$

The awakening time for peak $P_2$, denoted as $t_{a2}$, can be defined as the time $t$ at which the distance ($d_{a_t}$) between point $(t, p_t)$ and the reference line $l_{a_t}$ reaches the maximum. Specifically, we have

$$t_{a2} = \arg\left\{ \max_{t < t_{2}} d_{a_t} \right\}, \quad (7)$$

where

$$d_{a_t} = \frac{|(p_{t_2} - p_0)t - t_2p_t + t_2p_0|}{\sqrt{(p_{t_2} - c_0)^2 + t_2^2}}. \quad (8)$$
Similarly, the straight line that connects the points \((T, p_T)\) and \((t_2, p_{t_2})\) in the time-popularity plane, denoted as \(lf_t\), can be formulated as

\[
lf_t = \frac{p_{t_2} - p_T}{t_2 - T} (t - T) + p_T. \tag{9}
\]

Then the time of falling asleep, denoted as \(tf_2\), can be taken as the time \(t\) at which the distance \(df_t\) between the point \((t, p_t)\) and the reference line \(lf_t\) reaches the maximum. Accordingly, we have

\[
 tf_2 = \arg \left\{ \max_{t > t_2} df_t \right\}, \tag{10}
\]

where

\[
df_t = \frac{|(p_{t_2} - p_T)(t - T) - Tp_T + t_2p_T|}{\sqrt{(p_{t_2} - p_T)^2 + t_2^2}}. \tag{11}
\]

Again, considering the straight line that connects points \((0, p_0)\) and \((t_1, p_{t_1})\) and the straight line that connects points \((T, p_T)\) and \((t_1, p_{t_1})\) in the time-popularity plane, the awakening time \(ta_1\) and the falling asleep time \(tf_1\) can be similarly located.

The datasets and code used in this study are both publicly available to the research community and the details can be found in S1 Datasets and S1 Code, respectively.

**ACKNOWLEDGMENT**

This work was supported by the National Natural Science Foundation of China (Grant Nos. 71501005 and 71531001) and the fund of the State Key Lab of Software Development Environment (Grant Nos. SKLSDE-2015ZX-05 and SKLSDE-2015ZX-28).

[1] Frank M. Bass. A new product growth for model consumer durables. *Manage. Sci.*, 15(5):215–227, 1969.

[2] Christian Bauckhage. Insights into internet memes. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, ICWSM-11, pages 42–49, 2011.

[3] R. Alexander Bentley, Philip Garnett, Michael J. O’Brien, and William A. Brock. Word diffusion and climate science. *PLoS ONE*, 7(11):1–9, 11 2012.

[4] Lutz Bornmann, Loet Leydesdorff, and Jian Wang. How to improve the prediction based on citation impact percentiles for years shortly after the publication date? *Journal of Informetrics*, 8(1):175–180, 2014.
[5] Tibor Braun, Wolfgang Glänzel, and Andrs Schubert. On sleeping beauties, princes and other tales of citation distributions. *Research Evaluation*, 19(3):195–202, 2010.

[6] L. Quentin Burrell. Are “sleeping beauties” to be expected? *Scientometrics*, 65(3):381–389, 2005.

[7] Damon Centola. The spread of behavior in an online social network experiment. *Science*, 329(5996):1194–1197, 2010.

[8] Damon Centola. An experimental study of homophily in the adoption of health behavior. *Science*, 334(6060):1269–1272, 2011.

[9] Michele Coscia. Competition and success in the meme pool: a case study on quickmeme.com. In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media*, AAAI 2013, pages 100–109, 2013.

[10] Michele Coscia. Average is boring: How similarity kills a meme’s success. *Scientific Reports*, 4(6477), 2014.

[11] R Dawkins. *The Selfish Gene*. Oxford University Press, Oxford, UK, 1976.

[12] Eugene Garfield. Premature discovery or delayed recognition—why. *Current Contents*, 21:5–10, 1980.

[13] Eugene Garfield. Essays of an information scientist: Creativity, delayed recognition, and other essays. *Current Contents*, 23:3–9, 1989.

[14] James P. Gleeson, Jonathan A. Ward, Kevin P. O’Sullivan, and William T. Lee. Competition-induced criticality in a model of meme popularity. *Phys. Rev. Lett.*, 112:048701, 2014.

[15] Kevin P. and Baos Raquel A. and Moreno Yamir Gleeson, James P. and O’Sullivan. Determinants of meme popularity. arXiv:1501.05956, 2015.

[16] Wolfgang Glänzel, Balzs Schlemmer, and Bart Thijs. Better late than never? on the chance to become highly cited only beyond the standard bibliometric time horizon. *Scientometrics*, 58(3):571–586, 2005.

[17] Qing Ke, Emilio Ferrara, Filippo Radicchi, and Alessandro Flammini. Defining and identifying sleeping beauties in science. *Proceedings of the National Academy of Sciences*, 112(24):7426–7431, 2015.

[18] Ladislav Kristoufek. Bitcoin meets google trends and wikipedia: Quantifying the relationship between phenomena of the internet era. *Scientific Reports*, 3(3415), 2013.
[19] Christian Lachance and Vincent Larivire. On the citation lifecycle of papers with delayed recognition. *Journal of Informetrics*, 8(4):863 – 872, 2014.

[20] Gavin LaRowe, Sumeet Ambre, John Burgoon, Weimao Ke, and Katy Brner. The scholarly database and its utility for scientometrics research. *Scientometrics*, 79(2):219–234, 2008.

[21] Jure Leskovec, Lars Backstrom, and Jon Kleinberg. Meme-tracking and the dynamics of the news cycle. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pages 497–506, New York, NY, USA, 2009. ACM.

[22] Jiang Li. Citation curves of “all-elements-sleeping-beauties”: “flash in the pan” first and then “delayed recognition”. *Scientometrics*, 100(2):595–601, 2013.

[23] Jiang Li, Dongbo Shi, Star X. Zhao, and Fred Y. Ye. A study of the heartbeat spectra for sleeping beauties. *Journal of Informetrics*, 8(3):493 – 502, 2014.

[24] Jiang Li and Fred Y. Ye. The phenomenon of all-elements-sleeping-beauties in scientific literature. *Scientometrics*, 92(3):795–799, 2012.

[25] Robert P. Light and Katy Polley, David E.and Brner. Open data and open code for big science of science studies. *Scientometrics*, 101(2):1535–1551, 2014.

[26] Werner Marx. The shockley-queisser paper - a notable example of a scientific sleeping beauty. *Annalen der Physik*, 526(5-6):A41–A45, 2014.

[27] John A. Norton and Frank M. Bass. A diffusion theory model of adoption and substitution for successive generations of high-technology products. *Manage. Sci.*, 33(9):1069–1086, September 1987.

[28] Norio Ohba and Kumiko Nakao. Sleeping beauties in ophthalmology. *Scientometrics*, 93(2):253–264, 2012.

[29] G Palshikar. Simple algorithms for peak detection in time-series. In *Proc. 1st Int. Conf. Advanced Data Analysis, Business Analytics and Intelligence*, pages 1–13, 2009.

[30] S. Redner. Citation Statistics from 110 Years of Physical Review. *Physics Today*, 58(6):49–54, jun 2005.

[31] Ceyda Sanl and Renaud Lambiotte. Local variation of hashtag spike trains and popularity in twitter. *PloS one*, 10(7):e0131704, 2015.

[32] Kazutoshi Sasahara, Yoshito Hirata, Masashi Toyoda, Masaru Kitsuregawa, and Kazuyuki Aihara. Quantifying collective attention from tweet stream. *PLoS ONE*, 8(4):1–10, 04 2013.

[33] Limor Shifman. An anatomy of a youtube meme. *New Media & Society*, 14(2):187–203, 2012.
[34] Brian H. Spitzberg. Toward a model of meme diffusion (m3d). *Communication Theory*, 24(3):311–339, 2014.

[35] Carlos Senz-Royo, Carlos Gracia-Lázaro, and Yamir Moreno. The role of the organization structure in the diffusion of innovations. *PLoS ONE*, 10(5):1–13, 05 2015.

[36] Anthony F. J. van Raan. Dormitory of physical and engineering sciences: Sleeping beauties may be sleeping innovations. *PLoS ONE*, 10(10):e0139786, 10 2015.

[37] Anthony FJ Van Raan. Sleeping beauties in science. *Scientometrics*, 59(3):461–466, 2004.

[38] Senzhang Wang, Zhao Yan, Xia Hu, Philip S. Yu, and Zhoujun Li. Burst time prediction in cascades. In *Proceedings of the Twenty-Ninth Conference on Artificial Intelligence*, AAAI 2015, pages 325–331, 2015.

[39] Lilian Weng, Filippo Menczer, and Yong-Yeol Ahn. Virality prediction and community structure in social networks. *Scientific reports*, 3(2522), 2013.

[40] Lilian Weng, Filippo Menczer, and Yong-Yeol Ahn. Predicting successful memes using network and community structure. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, ICWSM-14, pages 535–544, Ann Arbor, MI, June 2014.

[41] Lillian Weng, Alessandro Flammini, Alessandro Vespignani, and Fillipo Menczer. Competition among memes in a world with limited attention. *Scientific Reports*, 2(335), 2012.

[42] M Yoshida, Y Arase, T Tsunoda, and M & Yamamoto. Wikipedia page view reflects web search trend. In *Proceedings of the WebSci15*, Web Science 2015, pages 1–2, 2015.

[43] Leihan Zhang, Jichang Zhao, and Ke Xu. Who creates trends in online social media: The crowd or opinion leaders? *Journal of Computer-Mediated Communication*, 21(1):1–16, 2016.

**SUPPORTING INFORMATION**

**S1 Datasets.** The datasets download location. All the datasets collected can be freely downloaded from the permanent location in figshare.com: [https://figshare.com/articles/Meme_popularity_and_d](https://figshare.com/articles/Meme_popularity_and_d)

**S1 Code.** The code downloads location. All the code used can be freely downloaded from the permanent location in figshare.com: [https://figshare.com/articles/Meme_popularity_and_d](https://figshare.com/articles/Meme_popularity_and_d)