Modeling Citation Dynamics of “Atypical” Articles

Zhongyang He
Department of Energy and Mineral Engineering, Pennsylvania State University, University Park, PA, 16802, USA. E-mail: zhongyanghe@psu.edu

Zhen Lei
Department of Energy and Mineral Engineering, Pennsylvania State University, University Park, PA, 16802, USA. E-mail: zlei@psu.edu

Dashun Wang
Kellogg School of Management, Northwestern University, Evanston, IL 60208, USA

Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL 60208, USA. E-mail: dashunwang@gmail.com

Modeling and predicting citation dynamics of individual articles is important due to its critical role in a wide range of decisions in science. While the current modeling framework successfully captures citation dynamics of *typical* articles, there exists a nonnegligible, and perhaps most interesting, fraction of *atypical* articles whose citation trajectories do not follow the normal rise-and-fall pattern. Here we systematically study and classify citation patterns of atypical articles, finding that they can be characterized by awakened articles, second-acts, and a combination of both. We propose a second-act model that can accurately describe the citation dynamics of second-act articles. The model not only provides a mechanistic framework to understand citation patterns of atypical articles, separating factors that drive impact, but it also offers new capabilities to identify the time of exogenous events that influence citations.

Introduction

An improved ability to assess and foresee scientific impact of individual articles has important scientific and policy implications. As the most common proxy for impact, citations of an article have widely served as criteria in important decisions such as faculty hiring and promotions (Bornmann & Daniel, 2006), or grant awarding and budget appropriation to science policies (Li & Agha, 2015). They are also extensively used in the evaluation of faculty competence (Hirsch, 2005), impact of journals (Garfield, 2006), research performance of institutes (Kinney, 2007; Liu & Cheng, 2005), or nations (Moed, 2002). Thanks to the recent expanding available large-scale data sets that capture in great detail various activities in science, there is growing interest from multiple disciplines in quantitative understanding of science (Evans & Rzhetsky, 2010; Evans, 2008; Evans & Foster, 2011; Fortunato, et al., 2018; Sinatra, Deville, Szell, Wang, & Barabási, 2015). While there is increasing evidence showing that citations correlate well with perceived scientific impact (Ioannidis, 2014; Radicchi, Weissman, & Bollen, 2017), it is important to keep in mind that citations are merely a proxy for impact or quality (Hirsch, 2005; Moreira, Zeng, & Amaral, 2015). Citations of articles have been demonstrated to follow highly reproducible patterns (Glänzel & Schöpflin, 1994; Wang, Song, & Barabási, 2013). The citation model proposed by Wang et al. (2013, hereafter the WSB model), predicts that citations follow an endogenous process, characterized by a single citation peak, controlled by three parameters.

Yet while the WSB model and its derivatives (Shen, Wang, Song, & Barabási, 2014) successfully capture the citation dynamics of *typical* articles, there exists a nonnegligible, yet previously unknown, and perhaps most interesting, fraction of articles whose citations do not seem to follow typical rise-and-fall trajectories, suggesting that these articles, including many important works in history, may not

Additional Supporting Information may be found in the online version of this article.

Received February 13, 2017; revised January 30, 2018; accepted February 28, 2018

 • Published online May 6, 2018 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.24041

be recognized by the community in the same way as typical ones (Raan, 2004; Garfield, 1989; Redner, 2005; Ke, Ferrara, Radicchi, & Flammini, 2015). For simplicity, we call them *atypical* articles hereafter, referring to their atypical citation distributions.

The dynamic patterns underlying citation dynamics of atypical articles remain largely unknown, mainly because of difficulties in identifying atypical articles that could serve as input for systematic studies. This situation has changed, however, thanks to a recent study of awakened articles in science (Ke et al., 2015). Indeed, Ke et al. proposed an elegant, parameter-free methodology to calculate the beauty coefficient (*B* coefficient) based on citation patterns that can systematically characterize and identify awakened articles. With unusual citation distribution, awakened articles were assumed to fall out of existing modeling frameworks (Ke et al., 2015), a hypotheses that has not been tested directly. Other atypical articles, called "second-act" articles, which receive a second citation burst, have been also reported (Wang et al., 2013; Redner, 2005; Li, 2014), and their citation dynamics do not fit well with current frameworks either. It is unexplored whether these atypical citation patterns are driven by completely unknown mechanisms, or can be reproduced by improving existing models for typical articles. Note that "sleeping beauty" and "prince" have deep roots in the literature (Braun, Glänzel, & Schubert, 2010; J. Li & Fred, 2016; van Raan, 2004). In this article, at the request by the editors, we use awakened articles to denote sleeping beauty, and wakers for prince. The goal here is not to introduce new terminologies to an existing scientific consensus, but to fully respect the editorial guidelines on gender sensitivity in writing (Sugimoto & Mostafa, 2018).

In this article, we aim to fill in the gap between atypical articles and current citation models. Combining detected awakened articles and ones that cannot be fitted well with the WSB model, we gain access to a unique corpus of atypical articles for measuring and modeling their atypical citation dynamics. We first test to what degree awakened articles can be reproduced by typical citation models, such as the WSB model. An analytical connection between *B* coefficient and citation model is also discussed. For second-act articles (which partially overlap with awakened articles), we propose a second-act citation model, which substantially improves our ability to describe citation dynamics in second-act articles. Results suggest that atypical citation patterns are still largely driven by known mechanisms. We also find that there may exist two types of second-act, possibly due to different awakening mechanisms, suggesting the possibility of applying our framework to estimate awakening time and locate waker articles.

Related Work

There is a vast literature on citations that are related to our work. Here we classify them broadly into two categories (Modeling citations and Atypical articles) and discuss each of them below in detail.

Modeling Citations

Many factors affect the citation dynamics of an article, for example, individual competence (Sinatra, Wang, Deville, Song, & Barabási, 2016) and reputation (Petersen et al., 2014), nationality (Wardle, 1995), teamwork (Wuchty, Jones, & Uzzi, 2007; Jones, Wuchty, & Uzzi, 2008), topic (Jia, Wang, & Szymanski, 2017), year (Stringer, Sales-Pardo, & Amaral, 2010), fields (Garfield, 1979), and multidisciplinarity (Fay, Borrill, Amir, Haward, & West, 2006). Yet three main mechanisms have been identified in citation accumulation of an article in citation network: preferential attachment, aging, and fitness.

The concept of preferential attachment—"the rich get richer"—was first proposed by de Solla Price (1976) and referred to as "cumulative advantage," and later popularized by Barabasi and Albert in network contexts (1999) to explain power-law distributions found in real networks. It captures the process in which new citations are more likely to attach to heavily cited articles than their less-cited counterparts. Preferential attachment was validated in citation networks (Redner, 2005; Newman, 2009; Barabási et al., 2002) with a linear attachment rate, and its origin was explained by copying citation models (Krapivsky & Redner, 2001; Gabel & Redner, 2013; Oliveira & Spencer, 2005).

Aging in citations captures the fact that new articles are more likely to offer fresh ideas, but with time their attractiveness for new citations is expected to fade away, as their novelty becomes part of common knowledge. Several function forms have been proposed to approximate the aging process in citations, including power-law (Pollman, 2000; Hajra & Sen, 2005; Wen, Duan, Chen, & Geng, 2011; Medo, Cimini, & Gualdi, 2011), exponential (Glänzel & Schöpflin, 1994; Hajra & Sen, 2006; Wang, Yu, & Yu, 2009), log-normal (Wang et al., 2013; Shen et al., 2014; Yin & Wang, 2017), or simply polynomial (Bouabid, 2011).

Scientific ideas differ inherently in their novelty and communities, therefore it is impossible to objectively quantify them all. Instead, fitness parameters have been used mathematically for an article's ability to attract citations within a community (that is, in a citation network) and serve as a collective measure to the intrinsic value (Bianconi & Barabási, 2001; Caldarelli, Capocci, De Los Rios, & Munoz, 2002). Although an article's value could be reflected in multiple dimensions (such as the discovery or invention's commercial value) and go beyond its citation counts (Editorial, 2017; Van Noorden, 2017), the ability to attract citations still points to the importance of an article.

In this article, we apply a recent citation model (Wang et al., 2013), which incorporates all of the three mechanisms, that is, preferential attachment, aging, and fitness.

Atypical Articles

As mentioned earlier, we refer to "atypical articles" as articles with atypical citation dynamics, unlike the typical rise-and-fall pattern right after publication. The current literature has primarily focused on two types of atypical

DOI: 10.1002/asi

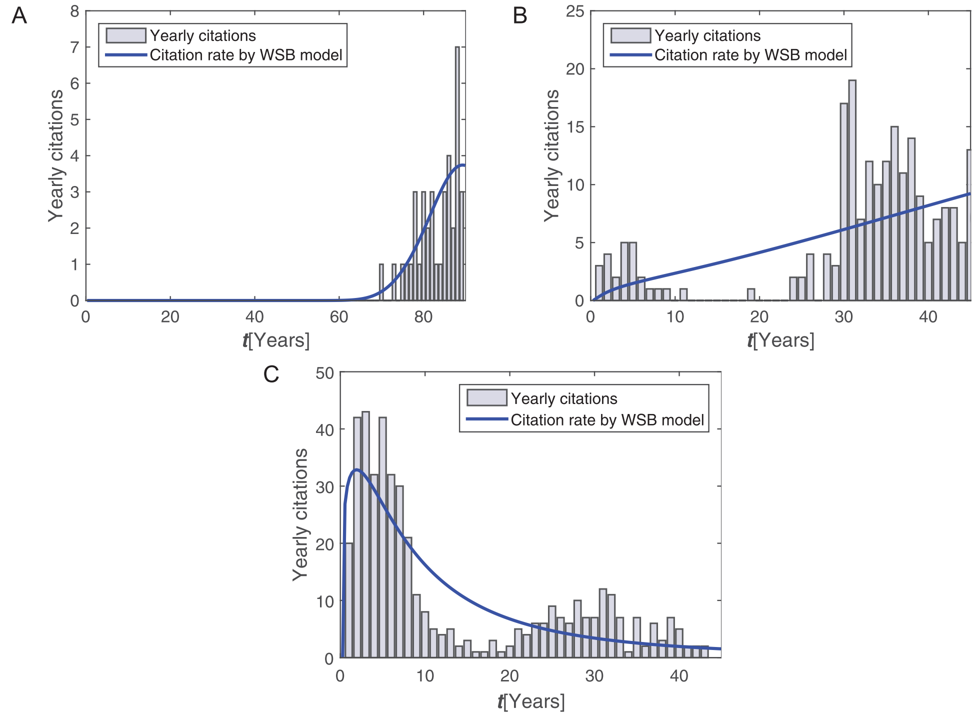

FIG. 1. Examples of atypical articles. In each example, yearly citations data and fitting result of the WSB model are compared. K-S test is used to evaluate the fitting performance. Higher p value suggests better fitting. (**A**) Example of single-peak awakened article, $p = .978$. (**B**) Example of second-act awakened article, $p < .001$. (C) Example of second-act non-awakened article, $p < .001$. [Color figure can be viewed at wileyonlinelibrary.com]

articles. The first one is *awakened articles*, whose scientific impact receives delayed recognition and only starts to emerge after a long hibernation period after publications (Figure 1A,B) (Raan, 2004; Garfield, 1989; Burrell, 2005). Awakened articles in science could be due to various reasons, such as nonideal technological or social conditions that forbid immediate follow-up research when the discovery was made (Garfield, 1980; Raan, 2015; Wang, Ma, Chen, & Rao, 2012), resistance to new findings that contradict traditional wisdom (Cole, 1970), or late recognition from other fields to their value (Ke et al., 2015). Many definitions of awakened articles have been proposed; for instance, by calculating the length of hibernation, the intensity of awakening, and publication date (Raan, 2004; Redner, 2005; Huang, Hsu, & Ciou, 2015). Here we follow the most recent work by using the B coefficient (Ke et al., 2015) to detect awakened articles. There are many advantages of this definition, one of which is that it involves no arbitrary threshold (Li & Fred, 2016), and hence provides a universal metric to all articles. Researchers have also studied "waker" articles, which help awakened articles reveal their true value to scientific communities (Braun, Glänzel, & Schubert, 2010; Li, Yu, Zhang, & Zhang, 2014). However, current studies have not focused on the mechanistic modeling framework of awakened articles.

The second type of atypical articles is *second-act* articles (Figure 1C) (Redner, 2005). One example is the BCS article (Bardeen, Cooper, & Schrieffer, 1957), whose citation trajectories experienced a second burst following the discovery of high-temperature superconductivity. The same phenomenon is also found in some awakened articles, where these articles experienced "awakening-sleeping-awakened" cycles (Li & Ye, 2012; Li, 2014). For these articles, their citation dynamics are characterized by more than one citation peak; hence, are not captured by existing single-peak citation models.

Research Questions

Previous studies have documented the existence of atypical articles whose citation histories do not seem to follow typical rise-and-fall patterns. Yet it remains unclear whether there is any modeling framework that can reproduce their citation dynamics. In this article, we build specifically on two recent advances in science of science—theoretical modeling of typical articles' citation dynamics (Wang et al., 2013; Shen et al., 2014) and empirical identifications of atypical articles (Ke et al., 2015)—allowing us to shed light on the following research questions:

- To what extent can citation models of typical citation patterns reproduce atypical citation patterns?
- What are the differences between atypical and typical patterns?
- Can we extend the current modeling framework to reproduce atypical citation patterns?

Data and Methods

Data

In this study we used both the American Physics Society (APS) citation data set and the Web of Science (WOS) data

DOI: 10.1002/asi

set. The APS data set includes articles published in *Physical Review* (1893–1970), *Physical Review* A/B/C/D (1970–present), *Physical Review Letters* (1958–present), etc. We focus on articles published in the top three journals in terms of publication volumes: *Physical Review (*PR*), Physical Review B (*PRB*)*, and *Physical Review Letters* (PRL). We include articles with at least 10 citations (Wang et al., 2013; Shen et al., 2014), resulting in 14,788 articles in PR, 33,203 in PRB, and 40,292 in PRL. The percentages of articles selected are 37.58%, 29.08%, and 47.65% for PR, PRB, and PRL respectively. To enrich our available data and address the limitation that the APS data set contains only articles and citations in physics, we gather articles (including both journal articles and conference articles) from the WOS data set (1900–2015). For our study, we select about 8 million articles that were published before 2000 (since recent articles have too short citation histories to show atypical citation patterns) and have at least 10 citations, which are 32.79% of all articles in the same time.

WSB Citation Model

The citation model we start with is the WSB model proposed by prior work (Wang et al., 2013; Shen et al., 2014). The model assumes an endogenous process of an article receiving citations and predicts a single citation peak. Broadly speaking, the model can be defined as follows,

$$c_i(t) = \lambda_i \left(c_i^t + m\right) f(t|\theta_i), \tag{1}$$

where $c_i(t)$ is the citation rate of an article i, that is, the expected number of new citations that this article receives at time t. In the model, λ_i is the article's fitness; c_i^t is the cumulative number of citations the article has received by time t, assuming a linear rate of preferential attachment in citation networks; m is the initial attractiveness of articles, that they may receive some citations regardless of their existing ones; $f(t|\theta_i)$ is the aging function that describes the temporal change of the impact of the article in the scientific community, in the following form of log-normal distribution:

$$f_i(t|\theta_i) = \frac{1}{\sqrt{2\pi}\sigma_i t} \exp\left(-\frac{1}{2}\left(\frac{\ln t - \mu_i}{\sigma_i}\right)^2\right), \quad t > 0, \tag{2}$$

where μ_i and σ_i are parameters for log-normal distribution. μ_i is the immediacy, controlling the time of maximum in the aging function, and σ_i is the longevity, governing the span of the aging function. Since the integration of the aging function from zero to infinity is always 1, smaller σ_i narrows the nonnegligible part of the distribution to a short time range and results in a greater value for the maximum point, while greater σ_i spreads out the distribution to a long time range and lowers the maximum point.

Based on Equation (1), suppose we observe n_i citations of article i within time T after its publication, the cumulative number of citations collected by article i at time t is:

$$c_i^t = (m + n_i) \exp\left[\lambda_i \left(F(t|\theta_i) - F(T|\theta_i)\right)\right] - m, \tag{3}$$

where F is the cumulative density function of log-normal distribution. The parameters λ_i, μ_i, and σ_i of article i can be estimated by maximum likelihood method.

B Coefficient

To identify awakened articles, we employ the B coefficient, proposed by Ke et al. (2015). Suppose for an article i, in its yearly citation history, the number of citations in the publication year is $c_{0,i}$ and the peak yearly citation occurs $c_{m,i}$ at year t_m. A line connecting two points $\left(0, c_{0,i}\right)$ and $\left(t_{m,i}, c_{m,i}\right)$ is $l_{t,i}$. The coefficient of the article, B_i, is defined as follows (see Supporting Figure S1 for illustration):

$$B_i = \sum_{t=0}^{t_m} \frac{l_{t,i} - c_{t,i}}{\max\left(1, c_{t,i}\right)}, \tag{4}$$

where $c_{t,i}$ is the number of citations received in year t. Intuitively, B increases with the length of hibernation (t_m) and the intensity of awakening (c_m), but is penalized by citations before the citation peak.

Connecting the WSB Model and the B Coefficient

If an awakened article is captured by the WSB model (for example, the case in Figure 1A, later called a single-peak awakened article), we can build an analytical connection between the model and B coefficient. To begin with, we examine the definition of B in Equation (4). The denominator penalizes early citations and deals with zero yearly citations, but it also introduces discontinuity in the definition. Neglecting the denominator, it is easy to see that the sum of l_t is nearly half of the product of the difference between the first year citation c_0 and peak year citation c_m multiplied by the peak year t_m, and the sum of c_t is cumulative citations by year t_m. In other words, the B coefficient without the denominator for a given article i is:

$$B_{nd,i} = \sum_{t=0}^{t_m} \left(l_{t,i} - c_{t,i}\right) \approx \frac{1}{2}\left(c_{m,i} - c_{0,i}\right) t_{m,i} + c_{0,i} t_{m,i} - \sum_{t=0}^{t_m} c_{t,i}. \tag{5}$$

Our test shows that B_{nd} and B are highly correlated. For APS data, for example, the overall rank correlation between B_{nd} and B is over 0.8. For the top 1% of awakened articles identified by B, 80% of them, including nearly all top 0.5%, can be recovered by the top 1% using B_{nd}. So we can use B_{nd} as our starting point to connect the WSB model and B coefficient without much loss.

The WSB model uses time intervals between publication date and forward citation date; therefore, we calculate yearly citations starting from the date of publication (rather than using calendar years). Thus $c_0 = 0$, and B_{nd} can be simplified to:

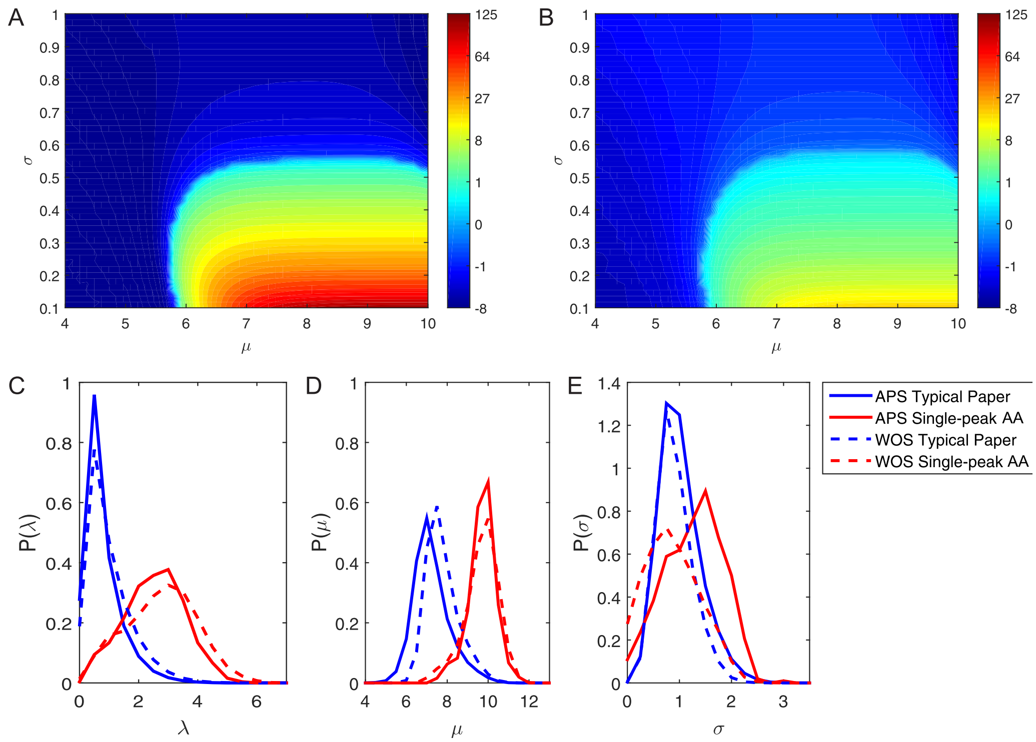

FIG. 2. Analytical and empirical results of single-peak awakened articles. (**A,B**) Contour of B_m on $\mu-\sigma$ plane. The color scale corresponds to the value of B_m. λ is constant in each figure. (A) λ=1.5. Awakened articles ($B_m > 45$) should have low σ and high μ. (B) λ=0.5. Although the same region as in (A) has the highest B_m, the maximum is much lowered, roughly from 120 to 20. No awakened article is found in this plot. (**C–E**) Distributions of WSB model parameters of typical articles and single-peak awakened articles in APS and WOS. (C) Distribution of λ. Awakened articles are likely to have higher fitness. (D) Distribution of μ. The citations of awakened articles are likely to arrive later. (E) Distribution of σ. There's less significant difference between awakened articles and typical ones. However, some of awakened articles have higher σ due to their long citation spans, and others have smaller sigma due to very late citations and observed short citation spans. [Color figure can be viewed at wileyonlinelibrary.com]

$$B_{nd,i} = \frac{1}{2} c_{m,i} t_{m,i} - \sum_{t=1}^{t_m} c_{t,i}, \tag{6}$$

where B_{nd} has the same intuition as B: it increases with the duration of hibernation (t_m) and the intensity of awakening (c_m), but declines with cumulative citations by the citation peak$(\sum_{t=1}^{t_m} c_t)$.

With this simplified form, we can express the B coefficient B_{nd} in Equation (6) in terms of parameters in the citation model, if the model fits the data well. Specifically, we propose a similar coefficient B_m that is based on model parameters:

$$B_{m,i} = \frac{1}{2} c_i\left(t'_{m,i}\right) t'_{m,i} - c_i^{t'_{m,i}}, \tag{7}$$

where $c_i(t'_{m,i}), t'_{m,i}, c_i^{t'_{m,i}}$ are the peak citation rate, the peak time, and the number of cumulative citations at peak time for article i, all derived from the citation model. The difference between Equations (6) and (7) is that (6) is calculated based on an article's yearly citation data while (7) is derived from citation model, using estimated or given parameters.

To compute $B_{m,i}$, we first obtain the expression of citation rate of the WSB model, the first-order derivative of c_i^t, and set to zero to solve for the peaking time $t'_{m,i}$. After numerically solving for $t'_{m,i}$, $c_i^{t'_{m,i}}$ and $c_i(t'_{m,i})$ can be calculated based on Equation 3 and its first-order derivative, respectively.

To understand how the model can produce the citation growth distribution of a single-peak awakened article, we further investigate the impact of model parameters λ_i, σ_i, and μ_i on its $B_{m,i}$. We plot a contour of B_m with respect to μ &σ to show variations of B_m to different $\mu-\sigma$ combinations (Figure 2A,B). In Figure 2A, we assume that the article has a relatively large fitness (that is, the article is inherently a high quality one). For $\sigma > 0.6$, B_m stays at a relative low level (less than zero), and its variation is insensitive to μ. A larger σ implies a smaller peak in aging function; without a reasonably sized citation peak, a large B_m can never be achieved regardless of the time of citation peak. For $\mu < 6$, B_m is also small even with a small σ, suggesting that an early citation peak goes against a large B_m. For $\mu > 6$ and $\sigma < 0.6$, as B increases, σ decreases and μ increases; a smaller σ corresponds to a narrow citation span and a higher citation peak, and a larger μ fits a longer delay in citations. In Figure 2B, we show the variation of B_m given a relatively small λ. B_m tends to be small regardless of μ and σ, suggesting an awakened article is hardly a low-quality one.

To sum up, parameters of a single-peak awakened article are likely to be a combination of a high fitness, which reflects the value of the article, a relative large immediacy

DOI: 10.1002/asi

parameter μ, which is determined by the long hibernation period, and a small longevity parameter σ, which is required by the short citation span.

Second-Act Citation Model

As shown in Figure 1B,C, the WSB model failed to explain citation patterns of second-act articles. To capture this pattern, we propose a novel second-act citation model, by incorporating a second aging function to account for the second citation peak, hence enriching existing modeling frameworks while preserving the three mechanisms driving citations.

As illustrated in Figure 4A, we modify the aging function by adding another log-normal distribution,

$$f'(t)=f_1(t)+rf_2(t)=f_1(t|\mu_1,\sigma_1)+rf_2(t|\tau,\mu_2,\sigma_2), \quad (8)$$

where f_1 is the aging function characterizing the first period of citations, and f_2 is the one for citations induced by second-act.

More important, we introduce two key parameters, the delay of second-act, τ, and the fitness multiplier for the second-act (relative to the first act), r. τ captures the time when the second-act takes place, that is, the article starts to be recognized by the scientific community again. Note it is possible that the second-act occurs before citations of the second-act become apparent, since log-normal distribution alone can induce some hibernation period. r captures the fitness change of the article in the second-act. Naturally, we assume r is positive: when the second-act happens, it is usually the case that the article has been rediscovered by researchers who might find the inherent value to be different (greater or smaller) than that in the first act.

More specifically, f_2 is a shifted log-normal distribution to characterize the temporal change of citation rate in the second-act, which is given as:

$$f_2(t|\tau,\mu_2,\sigma_2)=\begin{cases} 0, & t\le\tau \\ \dfrac{1}{\sqrt{2\pi}\sigma_2(t-\tau)}\exp\left[\dfrac{1}{2}\left(\dfrac{\ln(t-\tau)-\mu_2}{\sigma_2}\right)^2\right], & t>\tau \end{cases} \quad (9)$$

where σ_2 and μ_2 are the immediacy of citation peak and the longevity of citation span of the second-act, respectively, the same as σ_1 and μ_1 in the first act.

Substituting the aging function in the WSB model by Equation 8, we obtain the number of expected cumulative citations of a second-act article i at time t,

$$c_i^t=(m+n_i)\exp\left[\lambda_i(F'(t|\theta_i)-F'(T|\theta_i))\right]-m, \quad (10)$$

where $F'(t|\theta_i)=\int_0^t f'(t)dt$ and θ_i denotes all parameters in these two distributions.

Using citation data, we can estimate model parameters of a second-act article by maximizing the log-likelihood function,

$$\ln L=n_i\ln\lambda_i+\sum_{k=1}^{n_i}\ln(k+m-1)+\sum_{k=1}^{n_i}\ln(f'(t_k|\theta_i)) -\lambda_i(n_i+m)F'(T|\theta_i)+\lambda_i\sum_{k=1}^{n_i}F'(t_k|\theta_i). \quad (11)$$

Although we classify the second-act articles into second-act awakened articles and second-act non-awakened articles, the model is able to reproduce both of them. The difference between the two groups will be reflected in the distribution of estimated parameters (see Figure S7).

The MatLab (MathWorks, Natick, MA) codes used to implement discussed models and estimate parameters of articles are available on FigShare.[1]

Results and Discussion

Identified Atypical Articles

After obtaining B of all articles, we take articles within the top 1% of B in APS data set ($B>45$, see Figure S2A; here we select a lower threshold of B due to smaller size of this data set), gathering 594, 107, and 211 awakened articles in PR, PRB, and PRL, respectively. Since PR is the oldest of these three journals, it is reasonable that it has the most awakened articles. In the WOS data set, we collect articles $B>100$, resulting in 24,870 awakened articles (about top 0.3%, or 0.1% if articles with less than 10 citations are included; also see Ke et al., 2015). These articles constitute the set of awakened articles.

To identify second-act articles, we apply the WSB model to all articles in both data sets and measure the fitting performance of the model using the two-sample Kolmogorov–Smirnov test (Massey, 1951). For a given article, we compare the expected arrival time of each citation calculated from the WSB model vs. its actual time in citation history. If the p-value of test is lower than .1, then the citation history deviates significantly from the pattern predicted by the model, and we take it as a second-act candidate. We identify 708 articles from APS (0.8% of data set, Figure S2B), involving 376, 82, and 250 articles from PR, PRB, and PRL, respectively. From the WOS data set, we identify 54,388 second-act articles. After examining their citation patterns, we find they mostly experience two citation peaks, and for this reason we call this phenomenon the second-act in citation history.

While the two types of atypical articles, awakened articles and second-act articles, are largely different from each other, the boundary that separates them is often blurry. To systematically identify and categorize atypical articles, we divide them into four quadrants based on B and p (Figure 3A,B).

Typical articles ($B\le 45$ in APS, $B\le 100$ in WOS, and $p\ge .1$) are located in the second quadrant. The WSB model can fit them well (mean p-value .907 of APS, .861 in PR,

[1]https://figshare.com/projects/Modeling_citation_dynamics_of_atypical_articles/26350

DOI: 10.1002/asi

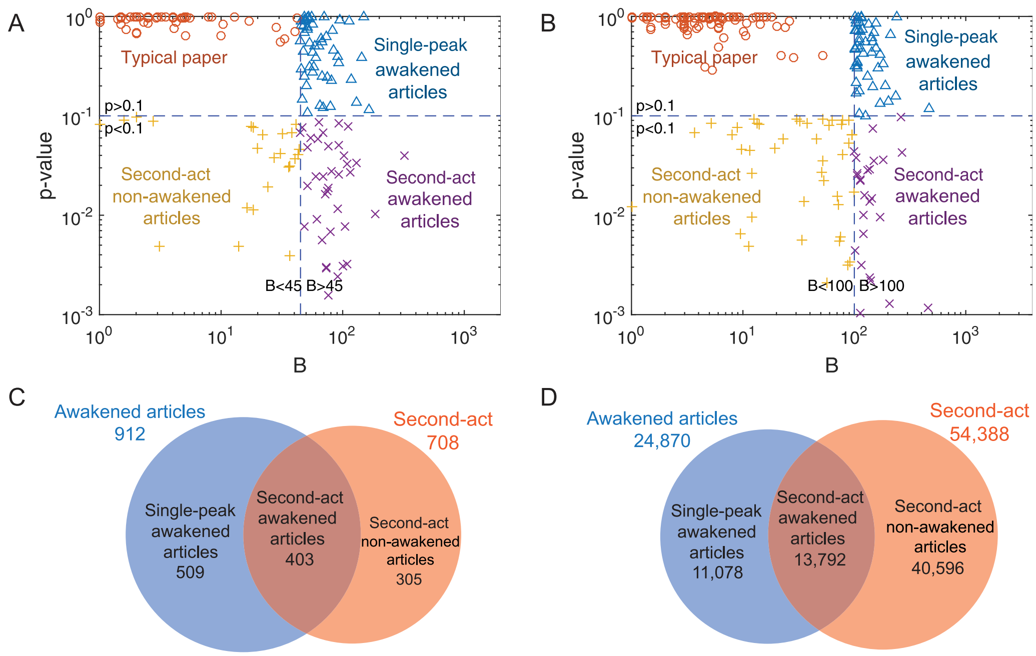

FIG. 3. (**A**) Four quadrants of APS articles. The vertical dash line is the threshold of awakened articles ($B > 45$). The horizontal dash line is the threshold of second-act articles ($p < .1$). These two thresholds divide all articles into four quadrants. Percentages of articles in each quadrant are: 98.6% (typical article), 0.58% (single-peak awakened article), 0.35% (second-act non-awakened article), and 0.46% (second-act awakened article). The data points are randomly drawn from each quadrant. (**B**) Four quadrants of WOS articles. The threshold of awakened articles is $B > 100$. The percentages are: 99.2% (typical article), 0.14% (single-peak awakened article), 0.50% (second-act non-awakened article), and 0.17% (second-act awakened article). (**C**) The number of atypical articles in each group in APS data set. We identify 509 single-peak awakened articles (343 in PR, 63 in PRB, and 103 in PRL), 403 second-act awakened articles (125 in PR, 38 in PRB, and 142 in PRL), and 305 second-act non-awakened articles (251 in PR, 44 in PRB, and 108 in PRL). (**D**) The number of atypical articles in each group in WOS data set. [Color figure can be viewed at wileyonlinelibrary.com]

.924 in PRB, and .909 in PRL, .892 in WOS), and they are not awakened articles.

The first quadrant contains awakened articles. These are ones that have a large B value ($B > 45$ in APS and $B > 100$ in WOS), yet they do not reject the hypothesis that their citation dynamics deviates significantly from the WSB model ($p \geq .1$). Therefore, they are referred to as *single-peak awakened articles*. Figure 1A shows an example of such articles. Indeed, it received no citation during its early periods but obtained a large number of citations after nearly 70 years. Although its citation pattern appears different from typical articles (for which most citations are several years after publication), we find the WSB model fits appropriately with the delayed impact pattern, suggesting that this subset of awakened articles follows the same citation mechanisms as typical articles.

Articles in the third quadrant ($B \leq 45$ in APS, $B \leq 100$ in WOS, and $p < .1$) are second-act articles but not awakened ones. For example, the article in Figure 1C received the majority of its citations during the first decade after publication and collected much less in the second decade. This is most certainly a typical citation pattern. Then it was actively cited again for about another 20 years. Its citation rate is not captured by the WSB model (Figure 1C, blue line), showing large deviations from the actual citation trajectory ($p < .001$), thus we call articles like this *second-act non-awakened articles*.

Finally, in the fourth quadrant, articles have larger B ($B > 45$ in APS, $B > 100$ in WOS) and $p < .1$. We call them *second-act awakened articles*. Figure 1B shows an example of such articles. According to the B parameter, this article is an awakened one. Yet it also has two separated short-term citation peaks that single-peak citation models failed to capture (the first one in 5 years after publication and the second one in 31 years).

Note that the definition of article quadrants is not unique. The threshold of B or p can be adjusted to obtain a different classification of articles as long as the insight of classification remains unchanged. Our goal here is not to seek an optimal classification, but to use the article quadrants as a way to understand observed differences in citation dynamics. Given selected thresholds, we identify 509 single-peak awakened articles, 403 second-act awakened articles, and 305 second-act non-awakened articles from the APS data set (Figure 3C). Similarly, we collect 11,078 single-peak awakened articles, 13,792 second-act awakened articles, and 40,596 second-act non-awakened articles from the WOS data set (Figure 3D). The numbers show that there is considerable overlap between awakened articles and second-act ones, and every type is nontrivial in the data set. In following sections, we provide empirical analyses for each type of atypical article.

DOI: 10.1002/asi

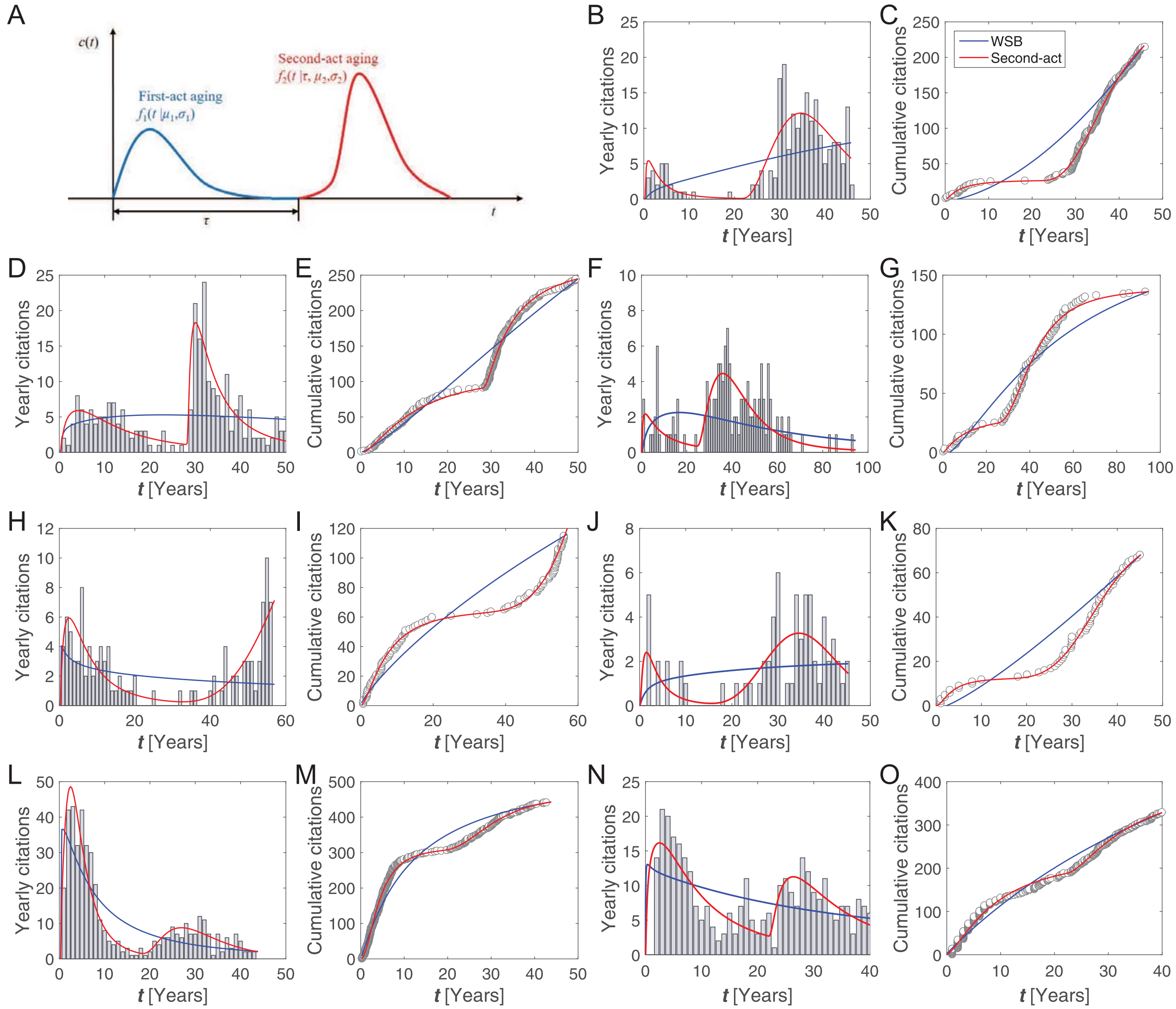

FIG. 4. (**A**) Illustration of the second-act model. The blue line represents the period in which the citation rate is dominated by the first-act aging function f_1. The red line represents the period dominated by the second-act aging function f_2. τ represents the delay of second-act. (**B–O**) Fitting examples of second-act articles. (B,C) Fitting example of second-act awakened article. (B: yearly citations, C: cumulative citations. The same layout for other examples.) $p = 1.000$ for K-S test of second-act model fitting. (D,E) Fitting example of second-act awakened article. $p = .996$. (F,G) Fitting example of second-act awakened article. $p = .968$, from WOS. (H,I) Fitting example of second-act awakened article. $p = .980$. Different from previous examples, this article has a nearly zero delay and exhibits slow increase of citation rate in the second-act. (J,K) Fitting example of second-act awakened article similar to (H). $p = 1.000$, from WOS. (L,M) Fitting example of second-act non-awakened article. $p = .894$. The second citation peak is much lower than the first one. (N,O) Fitting example of second-act non-awakened article. $p = .822$, from WOS. [Color figure can be viewed at wileyonlinelibrary.com]

Empirical Analysis of Single-Peak Awakened Articles

We first test the WSB model to all single-peak awakened articles. The mean p-value of K-S test to them is .523, which, although lower than the mean p-value of all articles (around .9), still suggests that these single-peak awakened articles, while being atypical articles, can be approximated by the WSB model.

To illustrate the difference between single-peak awakened articles and typical ones, we compare distributions of their model parameters (Figure 2C–E). Awakened articles are characterized by higher fitness, corresponding to higher overall impact than typical articles. The distinct difference in μ also confirms that awakened articles experience longer hibernation in citations. The difference in σ is less significant. Some awakened articles have higher σ due to their long citation spans, and others have smaller σ due to very late citations and short citation spans.

Empirical Analysis of Second-Act Articles

Applying the second-act model to second-act articles, we obtain significant improvement on fitting performance (see Figure 4B–O for examples).

DOI: 10.1002/asi

TABLE 1. Summary of fitting performance for second-act awakened articles.

Data set	APS	WOS
Number of articles	403	13,792
Mean of *p*-value, WSB model	.025	.018
Mean of *p*-value, second-act model	.898	.800
Ratio of $p < .1$, second-act model	2.2% (9 in 403)	7.8% (1,074 in 13,792)

The mean *p*-values of K-S test to the fitting from both the WSB model and the second-act model are listed in the table.

TABLE 2. Summary of fitting performance for second-act non-awakened articles.

Data set	APS	WOS
Number of articles	305	40,596
Mean of *p*-value, WSB model	.044	.0437
Mean of *p*-value, second-act model	.798	.752
Ratio of $p < .1$, second-act model	12.1% (37 in 305)	10.7% (4,351 in 40,596)

The *p*-values of K-S test to the fitting from both the WSB model and the second-act model are listed in the table.

First, we examine the results in second-act awakened articles. Table 1 summarizes overall fitting performance. Although as a simple extension to the single-peak model, the second-act model significantly improves in its predictive power. The mean *p*-value of K-S test increases from .025 to .898 in the APS data set, the same level when we apply the WSB model to typical articles. The improvement is similar for the WOS data set, from .018 to .800. Only 2.2% and 7.8% of second-act awakened articles in APS and WOS respectively reject the second-act model.

The result of second-act non-awakened articles is also convincing. As shown in Table 2, the mean of the *p*-value increases from .044 to .798 in APS, and from .044 to .752 in WOS, again a significant enhancement over the single-peak model. Indeed, only 12% and 10.7% of second-act non-awakened articles in APS and WOS, respectively, reject the second-act model. Overall, the result suggests that the majority of second-act articles, regardless being an awakened one or not, are well captured by the second-act model.

By the improved fitting performance of the second-act model, we are able to study second-act articles and find some interesting facts.

To start with, we examine the second-act awakened article illustrated in Figure 1B, noted as article A. The fitting result with the second-act model is given in Figure 4B,C. The model is able to capture the change of citation rate in two periods and gives convincing fitting to actual trajectories. The estimated delay for the second-act is about 20 years, just before actual citations in the second-act become noticeable. The relative fitness r is 2.08, reflecting the fact that the article receives more citations in the second-act. (See similar example from APS in Figure 4D,E, and example from WOS in Figure 4F,G.)

Meanwhile, we find another second-act awakened article (Figure 4H,I), noted as article B, whose citation dynamic is different from that of article A. Article B has a significant portion of citations that are received right after its publication. If it was not for the peak near the end of data, the article would not be considered an awakened article. More interestingly, for this article the second-act model gives $\tau = 0y$, even though the second-act only becomes visible after 40 years. By comparing citation histories, we see that the citation rate in the second-act of article B rose gradually over a long time (from 35–55 years), while that of article A rose quickly (from 28–30 years). The difference suggests a possibility that the second-act is induced by different sources. For article A, it may be due to a sudden exogenous factor outside the article's topic, but for article B, it could be more endogenous and lie in the article's own field, that is, the value of the article has had different latent within-field impacts since its publication. (Figure 4J,K provides another example from WOS as article B.)

Thus, there might be two different types of second-act awakened articles. One has a long τ in the second-act, likely caused by exogenous shocks. The other has a small τ and the second-act actually starts near $t = 0$ but peaks very late, suggesting a likely endogenous and inherent reason for the second-act. For the latter, parameters of the second-act aging function will be similar to ones in single-peak awakened articles, that is, large μ_2 and small σ_2.

To further investigate the above difference within second-act awakened articles we plot values of B and τ over $\mu_2 - \sigma_2$ plane using APS results (see Figure 5A) (see Figure S7 for the result of the WOS data set). We find that awakened articles can be roughly divided into two kinds: articles on the left with similar $\mu_2 \& \sigma_2$ close to typical articles but with large τ, in which the arrival of the second-act is mainly reflected in the delay (similar to Figure 4B–G); and articles on the right with small τ, in which case larger μ_2 and smaller σ_2 are required to reproduce the hibernation before the second-act (similar to Figure 4H–K). To illustrate the role of τ, we pick two examples (the red cross) from each side of Figure 5A. The article shown in Figure 5B belongs to the first kind, where τ is large. As such, its citation rate increased rapidly at year 39, when it experienced its second-act. Whereas in Figure 5C, the citation rate slowly increased from year 40 till 70, our model suggests its second-act began much earlier. Thus, consistent with insights from previous discussions, a second-act awakened article either has a long delay in the second-act aging function or has a relatively short delay, but the second-act involves a late and more noticeable peak.

Interestingly, the distribution of τ (Figure S6G) shows that for second-act articles, no matter awakened article or

DOI: 10.1002/asi

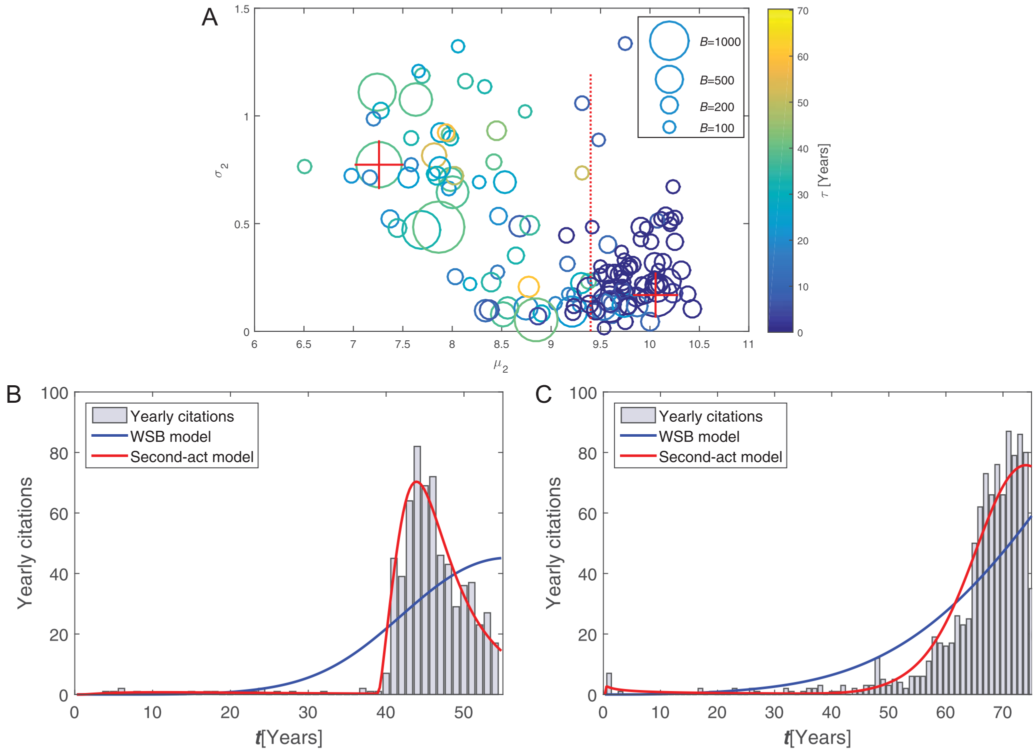

FIG. 5. (**A**) Scatterplot of key second-act model parameters for top 0.2% second-act awakened articles in APS. We select fewer articles to make the figure clearer while maintaining the message. Higher B corresponds to larger circle and the value of τ (in years) is illustrated by the color of circles. Articles can be roughly divided by the red line. Articles on the left have longer delay, lower immediacy, and higher longevity, contrary to articles on the right. The two red crosses are selected examples shown in B,C. (**B**) Example on the left side of (A), in which $\tau=39y$. After that, its citation rate increased rapidly. (**C**) Example on the right side of (A), in which $\tau=0y$. The citation rate of the second-act only became visible around year 40 and slowly increased till year 70. [Color figure can be viewed at wileyonlinelibrary.com]

non-awakened ones, about 30% have a second-act with a delay of less than 1 year after publication. Thus, for all second-act articles, reasons for the second-act might not be all exogenous: for some articles, endogenous factors could give rise to the second-act.

Finally, for second-act non-awakened articles, the example in Figure 1C is also reestimated (Figure 4 L,M). The obtained relative fitness, r, is only 0.13, indicating that the second-act is much weaker than the first-act, yet the second-act still deviates the article's citation trajectory from a typical single-peak pattern. (See Figure 4N,O for similar example from WOS.)

Locating the "Waker"

It is well known that an awakened article is often revived after a waker article (likely an important one) cites it, bringing a new burst of citations. However, identifying such wakers is still challenging. Many studies resort to cocitation networks between wakers and awakened articles (Gorry & Ragouet, 2016; Li et al., 2014; Braun et al., 2010; Li, 2014; Ohba & Nakao, 2012; Du & Wu, 2016). This method requires a thorough analysis of the citation network of the awakened article and all citing articles and, consequently, current work focuses only on case studies, not on systematic methods for a large corpus of awakened articles.

Based on our work, the delay of an awakened article can pin down the time of waker candidates using only citation data of the awakened article. This method can be applied to a large number of awakened articles. If an awakened article has a noticeable delay, then we can just focus our attention on citing articles around the delay time.

Here we give two examples where waker articles have been identified in prior works (Li, 2014; Ohba & Nakao, 2012), and test if our model can find the time of wakers. We show their citation histories, and plot the fitted citation rate by the second-act model (Figure 6). The delay of second-acts in two articles is 26.6 years and 56 years, respectively, which matches the gaps between awakened articles and identified wakers, suggesting the potential of our model to locate wakers.

Due to limited data the model requires, the result of delay cannot identify an exact waker article when there are multiple candidates. However, it can narrow down the scope of investigation when the time of second-act is not easy to determine from yearly citation data.

Conclusion

This article aims to improve our understanding of atypical citation patterns, by developing mathematical models that can capture citation dynamics of awakened articles and

DOI: 10.1002/asi

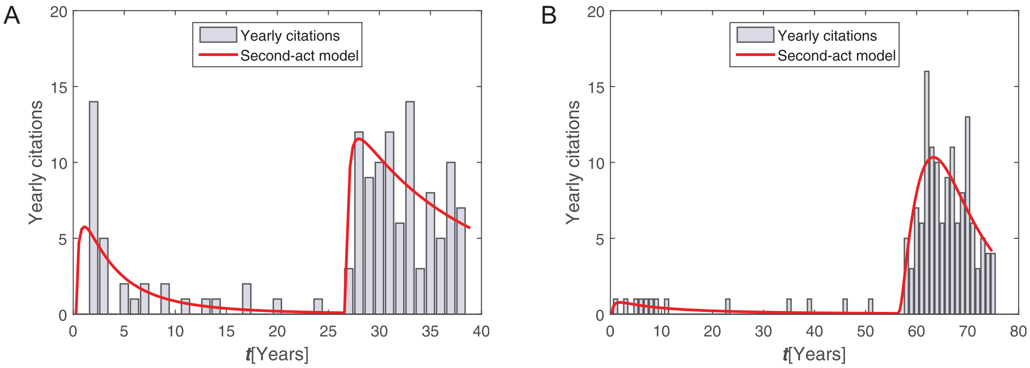

FIG. 6. Examples of locating waker using second-act model. (**A**) The awakened article (De Rujula, Georgi, & Glashow, 1977) is published in 1977 and the waker (Swanson, 2004), identified by Li (2014), is published in 2004, matching the delay of 26.6 years estimated from our model. (**B**) The awakened article (De Rotth, 1940) is published in 1940 and the waker (Tsubota et al., 1996), identified by Ohba and Nakao (2012), is published in 1996, matching the delay of 56 years from our model. [Color figure can be viewed at wileyonlinelibrary.com]

second-act articles. We first identified awakened articles and second-act articles from two large-scale corpora by applying existing methods. We then analyzed their citation patterns, classifying them into three categories: single-peak awakened articles, second-act awakened articles, and second-act non-awakened articles, based on their *B*s and *p*-values in fitting to citation models of typical articles. The percentages of the three groups are: 0.58% (single-peak awakened articles), 0.47% (second-act awakened articles), and 0.35% (second-act non-awakened articles) in APS; 0.14%,0.17%, and 0.50% in WOS (0.21%,0.17%, and 0.13% in APS, and 0.04%, 0.06%, and 0.16% in WOS, if articles less than 10 citations are included).

We find that about half of awakened articles can be described by typical citation trajectories, suggesting an endogenous process for these awakened articles to emerge. We derive the *B* coefficient analytically in terms of parameters of citation models, offering further insights on the citation dynamics of this type of awakened articles in comparison with typical articles.

To understand citation patterns of second-act articles, we extend the existing single-peak citation model to include a second peak, allowing us to develop a second-act citation model. We find the second-act model substantially improves our ability to characterize atypical citation patterns of both second-act awakened articles and second-act non-awakened articles, suggesting that while citation patterns of these articles appear atypical and citation peaks may come early or late, they may follow the same mechanisms that drive citations of typical articles. The second-act model may have applications in estimating the time of potential waker articles for second-act awakened articles.

One limitation of our model is that it captures only two peaks in citation history. In some cases, we observed three or more citation bursts, which may explain why there are $\sim$10% second-act articles rejecting the second-act model. Another limitation lies in the identification of second-act articles. We resort to the WSB model to detect them and choose the significance level of .1, yet there are some, albeit much fewer, similar second-act articles when $p \in [0.1, 0.3]$.

Taken together, we contribute a modeling framework for the analysis of atypical citation patterns in articles. As more works now are focused on predictions of future citations, our results indicate that there are second-act articles whose future citations may be inherently unpredictable (Garfield, 1980), as their delays in impact can be modeled by exogenous origins. Yet for some second-act articles, our results offer a possibility of predicting future citations once the second-act emerges. What remains an open question is to what degree additional information on authors, topics, or other metrics in citation networks may shed light on the predictability of second-acts. Lastly, it is worth noting that, although this article focuses on citation dynamics in academic articles, the awakened and second-act phenomena are also observed in other networked systems such as patent citation networks and online or social networks (Zhao, Erdogdu, He, Rajaraman, & Leskovec, 2015; Zhang, Xu, & Zhao, 2017). Applying and adapting our framework to these settings outside science could be an interesting future direction.

Acknowledgments

This work is supported by the Air Force Office of Scientific Research under award number FA9550-15-1-0162 and FA9550-17-1-0089, Northwestern University's Data Science Initiative, and U.S. National Science Foundation Grant SMA-1360205. We thank the reviewers for their insightful comments that greatly improved our work.

Author Contributions

All authors designed the experiments. Z.H. conducted experiments and prepared the figures. All authors analyzed the results and wrote the article.

Competing Financial Interests

The authors declare no competing financial interest.

References

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. Science, 286(5439), 509–512.

Barabási, A.-L., Jeong, H., Ravasz, R., Néda, Z., Vicsek, T., & Schubert, A. (2002). On the topology of the scientific collaboration networks. Physica A, 311(590–614), 4–6.

Bardeen, J., Cooper, L.N., & Schrieffer, J.R. (1957). Theory of superconductivity. Physical Review, 108(5), 1175.

Bianconi, G., & Barabási, A.-L. (2001). Competition and multiscaling in evolving networks. EPL (Europhysics Letters), 54(4), 436.

Bornmann, L., & Daniel, H.-D. (2006). Selecting scientific excellence through committee peer review-a citation analysis of publications previously published to approval or rejection of post-doctoral research fellowship applicants. Scientometrics, 68(3), 427–440.

Bouabid, H. (2011). Revisiting citation aging: a model for citation distribution and life-cycle prediction. Scientometrics, 88(1), 199–211.

Braun, T., Glänzel, W., & Schubert, A. (2010). On sleeping beauties, princes and other tales of citation distributions. Research Evaluation, 19(3), 195–202.

Burrell, Q.L. (2005). Are "sleeping beauties" to be expected? Scientometrics, 65(3), 381–389.

Caldarelli, G., Capocci, A., De Los Rios, P., & Munoz, M.A. (2002). Scale-free networks from varying vertex intrinsic fitness. Physical Review Letters, 89(25), 258702.

Cole, S. (1970). Professional standing and the reception of scientific discoveries. American Journal of Sociology, 76(2), 286–306.

De Rotth, A. (1940). Plastic repair of conjunctival defects with fetal membranes. Archives of Ophthalmology, 23(3), 522.

De Rujula, A., Georgi, H., & Glashow, S. (1977). Molecular charmonium: A new spectroscopy? Physical Review Letters, 38(7), 317.

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.

Editorial. (2017). Citation is not the only impact. Nature. Retrieved from https://www.nature.com/news/citation-is-not-the-only-impact-1.22917

Evans, J.A. (2008). Electronic publication and the narrowing of science and scholarship. science, 321(5887), 395–399.

Evans, J.A., & Foster, J. G. (2011). Metaknowledge. Science, 331(6018), 721–725.

Evans, J., & Rzhetsky, A. (2010). Philosophy of science: Machine science. Science, 329(5990), 399.

Fay, D., Borrill, C., Amir, Z., Haward, R., & West, M.A. (2006). Getting the most out of multidisciplinary teams: A multi-sample study of team innovation in health care. Journal of Occupational and Organizational Psychology, 79(4), 553–567.

Fortunato, S., Bergstrom, C. T., Börner, K., Evans, J. A., Helbing, D., Milojević, S., … Vespignani, A. (2018). Science of science. Science, 359(6379), eaao0185.

Gabel, A., & Redner, S. (2013). Sublinear but never superlinear preferential attachment by local network growth. Journal of Statistical Mechanics: Theory and Experiment, 2013(02), P02043.

Garfield, E. (1979). Is citation analysis a legitimate evaluation tool? Scientometrics, 1(4), 359–375.

Garfield, E. (1980). Premature discovery or delayed recognition-why. Current Contents, 21, 5–10.

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

Garfield, E. (2006). The history and meaning of the journal impact factor. JAMA, 295(1), 90–93.

Glänzel, W., & Schöpflin, U. (1994). A stochastic model for the ageing of scientific literature. Scientometrics, 30(1), 49–64.

Gorry, P., & Ragouet, P. (2016). "Sleeping beauty" and her restless sleep: Charles Dotter and the birth of interventional radiology. Scientometrics, 107(2), 773–784.

Hajra, K.B., & Sen, P. (2005). Aging in citation networks. Physica A: Statistical Mechanics and its Applications, 346(1), 44–48.

Hajra, K.B., & Sen, P. (2006). Modelling aging characteristics in citation networks. Physica A: Statistical Mechanics and Its Applications, 368(2), 575–582.

Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of the United States of America, 102(46), 16569–16572.

Huang, T.-C., Hsu, C., & Ciou, Z.-J. (2015). Systematic methodology for excavating sleeping beauty publications and their princes from medical and biological engineering studies. Journal of Medical and Biological Engineering, 35(6), 749–758.

Ioannidis, J.P. (2014). Is your most cited work your best? Nature, 514(7524), 561.

Jia, T., Wang, D., & Szymanski, B.K. (2017). Quantifying patterns of research-interest evolution. Nature Human Behaviour, 1, 0078.

Jones, B.F., Wuchty, S., & Uzzi, B. (2008). Multi-university research teams: Shifting impact, geography, and stratification in science. science, 322(5905), 1259–1262.

Ke, Q., Ferrara, E., Radicchi, F., & Flammini, A. (2015). Defining and identifying sleeping beauties in science. Proceedings of the National Academy of Sciences, 201424329.

Kinney, A.L. (2007). National scientific facilities and their science impact on nonbiomedical research. Proceedings of the National Academy of Sciences, 104(46), 17943–17947.

Krapivsky, P.L., & Redner, S. (2001). Organization of growing random networks. Physical Review E, 63(6), 066123.

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, D., & Agha, L. (2015). Big names or big ideas: Do peer-review panels select the best science proposals? Science, 348(6233), 434–438.

Li, J., & Fred, Y.Y. (2016). Distinguishing sleeping beauties in science. Scientometrics, 108(2), 821–828.

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

Li, S., Yu, G., Zhang, X., & Zhang, W.-F. (2014). Identifying princes of sleeping beauty-knowledge mapping in discovering princes. In 2014 International Conference on Management Science & Engineering 21th Annual Conference Proceedings (pp. 912–918).

Liu, N.C., & Cheng, Y. (2005). The academic ranking of world universities. Higher Education in Europe, 30(2), 127–136.

Massey Jr, F.J. (1951). The Kolmogorov-Smirnov test for goodness of fit. Journal of the American Statistical Association, 46(253), 68–78.

Medo, M., Cimini, G., & Gualdi, S. (2011). Temporal effects in the growth of networks. Physical Review Letters, 107(23), 238701.

Moed, H. (2002). Measuring china's research performance using the science citation index. Scientometrics, 53(3), 281–296.

Moreira, J.A., Zeng, X.H.T., & Amaral, L.A.N. (2015). The distribution of the asymptotic number of citations to sets of publications by a researcher or from an academic department are consistent with a discrete lognormal model. PLoS One, 10(11), e0143108.

Newman, M. (2009). The first-mover advantage in scientific publication. EPL (Europhysics Letters), 86(6), 68001.

Ohba, N., & Nakao, K. (2012). Sleeping beauties in ophthalmology. Scientometrics, 93(2), 253–264.

Oliveira, R., & Spencer, J. (2005). Connectivity transitions in networks with super-linear preferential attachment. Internet Mathematics, 2(2), 121–163.

Petersen, A.M., et al. (2014). Reputation and impact in academic careers. Proceedings of the National Academy of Sciences, 111(43), 15316–15321.

Pollman, T. (2000). Forgetting and the ageing of scientific publications. Scientometrics, 47(1), 43–54.

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

Raan, A.F. van. (2015). Dormitory of physical and engineering sciences: Sleeping beauties may be sleeping innovations. PLoS One, 10(10), e0139786.

Radicchi, F., Weissman, A., & Bollen, J. (2017). Quantifying perceived impact of scientific publications. Journal of Informetrics, 11(3), 704–712.

DOI: 10.1002/asi

Redner, S. (2005). Citation statistics from 110 years of physical review. Physics Today, 6(58), 49–54.

Shen, H. W., Wang, D., Song, C., & Barabási, A. L. (2014). Modeling and Predicting Popularity Dynamics via Reinforced Poisson Processes. AAAI, 14, 291–297.

Sinatra, R., Deville, P., Szell, M., Wang, D., & Barabási, A.-L. (2015). A century of physics. Nature Physics, 11(10), 791–796.

Sinatra, R., Wang, D., Deville, P., Song, C., & Barabási, A.-L. (2016). Quantifying the evolution of individual scientific impact. Science, 354(6312), aaf5239.

Solla Price, D. de. (1976). A general theory of bibliometric and other cumulative advantage processes. Journal of the American society for Information science, 293.

Stringer, M.J., Sales-Pardo, M., & Amaral, L.A.N. (2010). Statistical validation of a global model for the distribution of the ultimate number of citations accrued by articles published in a scientific journal. Journal of the American Society for Information Science and Technology, 61(7), 1377–1385.

Sugimoto, C. R., & Mostafa, J. (2018). A note of concern and context: On careful use of terminologies. Journal of the Association for Information Science and Technology, 69(3), 347–348.

Swanson, E.S. (2004). Short range structure in the x (3872). Physics Letters B, 588(3), 189–195.

Tsubota, K., et al. (1996). Surgical reconstruction of the ocular surface in advanced ocular cicatricial pemphigoid and stevens-johnson syndrome. American Journal of Ophthalmology, 122(1), 38–52.

Van Noorden, R. (2017). The science that's never been cited. Nature. Retrieved from https://www.nature.com/articles/d41586-017-08404-0

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

Wang, J., Ma, F., Chen, M., & Rao, Y. (2012). Why and how can "sleeping beauties" be awakened? The Electronic Library, 30(1), 5–18.

Wang, M., Yu, G., & Yu, D. (2009). Effect of the age of articles on the preferential attachment in citation networks. Physica A: Statistical Mechanics and Its Applications, 388(19), 4273–4276.

Wardle, D.A. (1995). Journal citation impact factors and parochial citation practices. Bulletin of the Ecological Society of America, 76(2), 102–104.

Wen, G., Duan, Z., Chen, G., & Geng, X. (2011). A weighted local-world evolving network model with aging nodes. Physica A: Statistical Mechanics and Its Applications, 390(21), 4012–4026.

Wuchty, S., Jones, B.F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. Science, 316(5827), 1036–1039.

Yin, Y., & Wang, D. (2017). The time dimension of science: Connecting the past to the future. Journal of Informetrics, 11(2), 608–621.

Zhang, L., Xu, K., & Zhao, J. (2017). Sleeping beauties in meme diffusion. Scientometrics, 1–20.

Zhao, Q., Erdogdu, M. A., He, H. Y., Rajaraman, A., & Leskovec, J. (2015). Seismic: A self-exciting point process model for predicting tweet popularity. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1513–1522.

DOI: 10.1002/asi