

Article

Collective Dynamics in the Awakening of Sleeping Beauty Patents: A BERTopic Approach

Hee Jin Mun  and Sanghoon Lee * 

Technology Commercialization Division, Electronics and Telecommunications Research Institute, 218 Gajeong-ro, Yuseong-gu, Daejeon 34129, Republic of Korea; hjmun99@etri.re.kr

* Correspondence: leeshe@etri.re.kr

Abstract

Prior research has emphasized individual patent characteristics in identifying the awakening of sleeping beauty patents (SBPs) that remain unnoticed for long periods before suddenly attracting substantial attention. However, less attention has been paid to how collective dynamics shape these awakenings. This study examines whether field-level topic patterns—observable manifestations of collective perceptions and choices—are associated with SBP awakenings. We derived two indicators from U.S. patent abstracts by using BERTopic: Jensen–Shannon Divergence (JSD), which reflects temporal shifts in topic distributions, and topic entropy, which captures the breadth of technological exploration across topics. The logistic regression results showed that JSD is negatively associated with SBP awakenings, whereas entropy is positively associated with them. These findings suggest that SBPs are more likely to reemerge when technological exploration spans a broader range of topics while the topic structure remains relatively stable. In this way, the study contributes by demonstrating how outputs of collective dynamics are linked to the delayed recognition of SBPs.

Keywords: BERTopic; collective dynamics; sleeping beauty patents; topic modeling



Academic Editor: Jose María Alvarez Rodríguez

Received: 1 September 2025

Revised: 22 September 2025

Accepted: 23 September 2025

Published: 25 September 2025

Citation: Mun, H.J.; Lee, S. Collective Dynamics in the Awakening of Sleeping Beauty Patents: A BERTopic Approach. *Appl. Sci.* **2025**, *15*, 10395. <https://doi.org/10.3390/app151910395>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Charles Theodore Dotter, widely recognized as the father of interventional radiology, introduced the first percutaneous transluminal angioplasty in his landmark paper published in 1964 [1]. Although this pioneering technique marked the advent of a new era in vascular treatment, the article received few citations until 1979 [2]. Publications that remain unnoticed for an extended period (“sleeping”) and then suddenly attract substantial attention (“awakening”) are termed “sleeping beauties,” a concept that has stimulated extensive research [3]. These studies have proposed methods for identifying sleeping beauty research papers (SBRPs) whose scientific value remains underrecognized [4] and have documented such papers across disciplines ranging from physics to strategic management [5,6].

The sleeping beauty phenomenon is also observed in patents. Similarly to research articles, patents embody technological ideas that diffuse through citations; the forward citations a patent receives serve as both indicators of its value and bases for anticipating technological trajectories within sectors [7,8]. However, forward citations are typically concentrated in a small subset of patents [9], whereas most patents attract limited external attention and are thus regarded as having low value [10]. Notably, some long-neglected patents subsequently receive substantial forward citations. For example, Canon filed a patent for a liquid-jet recording head in 1982, but it did not receive widespread citations

until 1990, exemplifying a sleeping beauty patent [11]. A central question, therefore, is why certain sleeping beauty patents (SBPs) suddenly attract renewed attention [12,13]. Because even slightly outdated technologies are often deemed unpromising [14], identifying innovative yet initially under-cited patents that still retain reemergence potential has clear implications for corporate R&D management [15].

This study investigates SBP awakenings through the lens of collective dynamics. In social and economic systems—whether involving ideas, products, individuals, or firms—success is not an isolated event but an outcome of collective dynamics, defined as the interactions among actors, ideas, and objects that generate system-level behaviors [16]. Beyond general aging effects [17], SBP awakenings can also be explained within this framework [16,18]. For example, when collective attention from diverse actors such as inventors and organizations converges on a specific technological field, opportunities may arise to reconsider the technological and economic value of patents in the field, potentially leading to SBP awakenings. Indeed, Kojaku et al. (2025) argue that the proximity of knowledge communities to an SBP within technological space may drive its awakening, highlighting the role of collective dynamics [18]. Taken together, these insights imply that field-level collective shifts can concentrate attention on SBPs, thereby increasing the likelihood of their awakening. Nevertheless, empirical evidence remains limited regarding whether collective shifts within technological fields are associated with SBP awakenings [11,18].

Building on the perspective of collective dynamics [16,18], this study examines the relationship between SBP awakenings and shifts in technological topics. Research on technological change shows that when the interests of relevant actors shift within technological fields, diverse ideas are tested for their potential to shape markets [19,20]. Such shifts can lead to the delayed recognition or reemergence of previously overlooked technologies [21]. The topics represented in newly filed patents are likely to reflect both technological transitions and evolving actor interests. In other words, collective dynamics may leave imprints on topic structures; these imprints are likely to be associated with SBP awakenings. Thus, we ask whether SBPs within a technological field are more likely to awaken when the distribution of patent topics within that field shifts.

Our empirical analysis focuses on U.S. patents. We defined technological fields as International Patent Classification (IPC) subclasses assigned to patents [22,23]. To ensure sufficient coverage of SBPs, we restricted the analysis to subclasses with more than 1000 patents granted between 1980 and 1989, yielding 191 fields and 572,309 patents. We then derived annual topic changes from abstracts of U.S. patents granted between 1989 and 2023 by using BERTopic [24,25] and examined their association with SBP awakenings at the subclass level. We found that Jensen–Shannon Divergence and topic entropy [26,27], which capture temporal change and dispersion in patent topics, are associated with the likelihood that fields experience SBP awakenings.

This study makes two contributions. First, identifying mechanisms underlying SBP awakenings is as important as detecting SBPs themselves [12,13]. We contribute to the SBP literature by empirically examining how field-level manifestations of collective dynamics relate to SBP awakenings, a mechanism suggested in prior work [16,18]. Second, while recent research highlights the reemergence of abandoned or neglected products and services [21], patent studies have largely focused on natural aging effects [17]. This study advances the literature by identifying factors beyond aging that influence forward citation patterns [18].

The remainder of this paper is structured as follows. The next section reviews prior research on SBRPs and SBPs and outlines how changes in patent topics are linked to the

awakening of SBPs. Section 3 details the methodology. Section 4 presents the results, and Section 5 discusses the implications and provides concluding remarks.

2. Literature Review

2.1. *Sleeping Beauty Patents*

Since van Raan (2004) introduced the term “sleeping beauty”—also referred to as “delayed recognition,” “resisted discovery,” or “premature discovery” [4]—research on SBRPs has advanced along two main streams. The first stream develops methods for identifying SBRPs [6,28], a subgroup of “citation classics” whose impact emerges gradually rather than through immediate citation surges [29]. The second stream explores factors underlying SBP awakenings. For example, Dey et al. (2017) identified factors that trigger the awakening of SBRPs in computer science, including the number of keywords, the diversity of referenced research fields, and the type of publication venue [30]. Hartley and Ho (2017) analyzed the characteristics of the “Prince,” a publication that triggers the awakening of an SBP [31].

Similarly to research papers, patents include citations to prior art, either to other patents or to non-patent references. However, most patents receive few citations [9,10], whereas a small subset attracts substantial attention after prolonged dormancy, which indicates SBP awakenings [11]. Accordingly, recent studies have investigated the antecedents of SBP awakenings. For example, analyzing Chinese graphene-related patents, Hou and Yang (2019) proposed that an SBP is considered awakened when its transfer, licensing, or citation frequency exceeds a specified threshold [13]. Tur et al. (2022) found that science linkages of nanotechnology SBPs are not significant predictors of their awakenings [32]. More recently, studies have analyzed SBP families that generated subsequent inventions in polymerase chain reaction, offering a more comprehensive view of technological diffusion [3,15].

Despite these advances, little is known about how shifts at the field level relate to SBP awakenings. In the SBP literature, scholars have examined awakenings triggered by technological change [6]. Li et al. (2014) noted that dormancy duration appears to have changed with advances in science and technology [33]. Haghani and Varamini (2021) further showed that pre-pandemic coronavirus research (e.g., SARS, MERS) became highly cited during COVID-19 [34]. These findings illustrate how exogenous changes can rapidly redirect research attention to SBP [11], implying that technological shifts may also prompt the reevaluation of previously overlooked patents, a gap that remains largely unexplored in existing research.

2.2. *Technological Shifts and Awakening*

From the perspective of collective dynamics, the reemergence of long-neglected patents is not an isolated event but the cumulative outcome of cognitive processes and choices made by interconnected actors across multiple levels [16]. Research on technological change has examined how the collective perceptions and choices of relevant actors (e.g., firms, users, and investors) influence the success of technologies [20,21,35]. These studies argue that technological development often unfolds through an era of ferment in which competing technologies emerge and relevant actors hold divergent interpretations of the value and potential of these technologies [19,20]. During this phase, experiments are conducted on the value and viability of both new and existing technologies, ultimately giving rise to a dominant technology that shapes the industry [19,20]. In such periods of change, previously neglected or abandoned technologies may reenter the focus of actors [21].

Research on technological change implies that linking the manifestations of technological shifts, shaped by actors’ perceptions and choices, to SBP awakenings can yield insights into the relationship between collective dynamics and SBP awakenings. In this regard,

bibliometric research has sought to capture such shifts by identifying technological topics actively pursued by actors [36]. Changes in dominant topics or in topic diversity within an industry may indicate shifts in actors' perceptions of the novelty of existing technologies and their search for alternatives [25,37].

The preceding discussion suggests that changes in technological topics may be associated with SBP awakenings. While Kojaku et al. (2025) argue that collective shifts in knowledge communities can redirect actors' attention to SBPs [18], prior work has not empirically tested whether field-level topic shifts are associated with SBP awakenings. The present study estimates changes in field-level topics over time and examines whether indicators of such changes are related to SBP awakenings.

3. Methods

3.1. Data and Sample

This study utilized granted U.S. patents as the basis of the empirical analysis for three reasons. First, the United States has accumulated long-term, well-maintained patent records that ensure the reliability of citation, abstract, and classification information central to our analysis. Second, the United States has developed cutting-edge technologies and has attracted valuable innovations from around the world; therefore, U.S. patents are well-suited to capturing global technological topics embodied in patent documents. Finally, granted patents embody technological ideas officially determined to be patentable (i.e., usefulness, novelty, and non-obviousness), rendering them analogous to peer-reviewed papers in SBRP studies. Thus, we analyzed granted patents rather than patent applications, which document ideas without formal examination.

Two considerations are essential for defining SBPs and identifying their awakening. First, a clear specification of the time frame is required. Prior research on SBRPs has been criticized for relying on short time windows to approximate long-term citations [6]. A sufficiently long period is required to allow for technological shifts and for the accumulation of enough citations necessary for the awakening of SBPs. Without this consideration, a sudden increase in forward citations just a few years after a patent's grant could be mistakenly identified as the awakening of an SBP. On the contrary, it would be unreasonable to interpret a citation from a patent granted in 2024 to one granted in 1900 as evidence of the latter's awakening. In light of these considerations, we selected U.S. patents granted between 1980 and 1989 as the sample. We measured changes in technological topics by using U.S. patents granted between 1989 and 2023. Allowing a one-year lag between topic change and awakening, we examined the awakenings of sample patents during 1991–2024 as outcomes associated with topic changes computed from patents granted in 1989–2023.

The second consideration concerns the definition of technological fields. Each patent is assigned IPC codes, which categorize patents into a hierarchical technological structure that remains relatively stable over time [38,39]. An IPC code is structured as shown in Figure 1. Prior studies suggest that the technological fields of patents can be represented by IPC subclasses [22,23]. A single patent can be assigned multiple IPC codes. Examiners typically designate the code that best reflects a patent's core technological content as the main IPC, while additional codes represent supplementary technological features. In this study, a patent's technological field is defined by the subclass of its main IPC code.

We collected forward citations and IPC information for U.S. utility patents granted between 1980 and 1989 from PatentsView, which is supported by the Office of the Chief Economist at the U.S. Patent and Trademark Office. Missing information was supplemented from Espacenet, a database managed by the European Patent Office. A few patents for which information was unavailable were excluded from the analysis. To capture SBPs effectively, we limited the sample to subclasses with more than 1000 patents granted

between 1980 and 1989, yielding 191 subclasses and 572,309 patents for SBP identification. To trace topic shifts within the same subclasses, we collected U.S. utility patents granted during 1989–2023; because some records were unavailable, the final dataset for topic measurement comprised 5,745,577 patents. Table 1 presents the basic statistics of the SBP dataset and the dataset used for measuring topic shifts at the IPC subclass level.

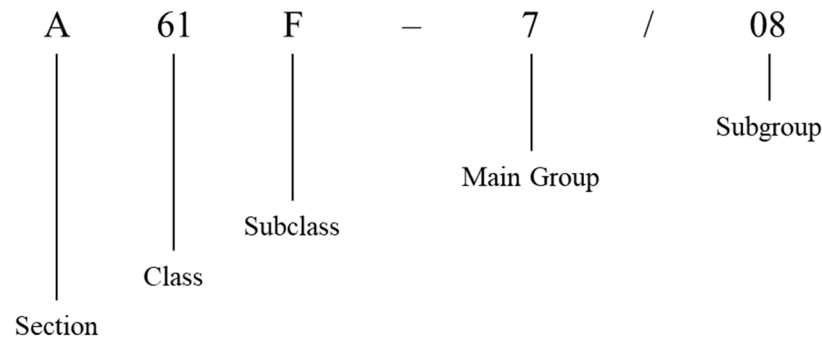


Figure 1. Patent IPC structure.

Table 1. Basic statistics for patents across subclasses.

Category	Mean	SD	Min	Max
SBP dataset	2996.4	2732.3	1006	17,166
Topic measurement dataset	30,081.6	62,809.2	1528	695,918

3.2. Sleeping Beauty Patents

Rule-based identification of SBRPs can be arbitrary [6]. To address this, Ke et al. (2015) proposed the sleeping beauty coefficient, a parameter-free metric designed to quantify the extent to which a research paper can be considered an SBRP [6]. This metric has also been applied in recent analyses of SBPs [18]. The present study employed the beauty coefficient to identify the awakening of SBPs.

For a given patent, c_t denotes the number of citations it received in year t . We rescale time so that the grant year of the given patent is $t = 0$, with $(0, c_0)$ as the first point on the time–citation plane. The grant year is chosen as t_0 because the citation-age profile is better described when measured from the grant date [17]. Let t_m be the year in which the patent receives its maximum annual citations c_{t_m} . The linear reference line l_t connects $(0, c_0)$ and (t_m, c_{t_m}) . The sleeping beauty coefficient (B) is defined as:

$$B = \sum_{t=0}^{t_m} \frac{l_t - c_t}{\max(1, c_t)},$$

which is interpretable as the area between the reference line and the citation history [6,18]. A larger B indicates a more abrupt peak; $B \approx 0$ corresponds to approximately linear citation growth; $B < 0$ indicates early surges followed by decline. Because citation-age profiles vary by field [40], we determined field-specific B thresholds. As a theoretical cut-off is absent, patents with B in the top 0.1% within each IPC subclass were classified as exhibiting abrupt awakenings.

The awakening time t_a is

$$t_a = \arg \left\{ \max_{t \leq t_m} d_t \right\},$$

where d_t is given by

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

This definition works well for cases in which no citations occur until a spike, and it captures the qualitative notion of awakening time when strong SB-like behavior is present [6].

Finally, at the subclass level, we made a binary variable $Awakening_{it}$, which takes the value of 1 if at least one patent in the subclass i experienced a sudden awakening in year t , and 0 otherwise.

3.3. BERTopic

Topic modeling is widely used to extract thematic structures from documents [41]. Among various algorithms, Latent Dirichlet Allocation (LDA) has been extensively applied to studies on technological topics [42]. By simultaneously estimating document–topic and topic–word distributions with Dirichlet priors [43], LDA uncovers topic structures in large document collections [44,45]. However, LDA treats documents from different time periods independently [46] and, as a bag-of-words model, is not able to capture contextual semantics that should not be ignored when modeling topic dynamics [47,48]. Dynamic Topic Model (DTM) adaptively updates its estimation of underlying topics as new documents are added, thereby capturing evolving trends and patterns within a corpus [49]. However, DTM often represents topics solely as multinomial distributions over words, thereby failing to adequately capture the semantic regularities of documents [50].

Word embedding models have enabled the computation of semantic relationships between words, under the assumption that words appearing in similar contexts tend to share similar meanings [51,52]. BERTopic generates document embeddings with pretrained transformer-based language models, clusters these embeddings, and represents topics using a class-based TF–IDF (Term Frequency–Inverse Document Frequency) procedure [24,25]. In addition, BERTopic employs HDBSCAN, which infers the number of clusters from data density and labels outliers as noise. Hence, recent research has utilized BERTopic to track technological topic evolution within fields such as 6G technologies and wave-and-tidal energy [25,53].

Leveraging these advantages, the present study applied BERTopic to patent abstracts to uncover topics over time at the IPC subclass level. The procedure was as follows. First, the abstracts were preprocessed (e.g., removal of URLs). Second, each abstract was embedded into a document vector by a specified pretrained model (all-MiniLM-L6-v2). Third, dimensionality reduction was performed with UMAP (Uniform Manifold Approximation and Projection) to project the high-dimensional embeddings into a lower-dimensional space while preserving local and global structure. Fourth, HDBSCAN was applied to identify persistent high-density clusters as topics and to classify low-density documents as noise. Finally, patents within each subclass were grouped by grant year to construct a year–topic matrix, and the counts were normalized to probabilities within each year.

3.4. Yearly Change in Topic Distribution

To quantitatively assess year-to-year changes in topic distribution, we computed the Jensen–Shannon Divergence (JSD) and *topic entropy*. *Entropy* quantifies within-year topic diversity, and JSD represents year-to-year change in topic distributions; both have been used to capture trends in topic shifts [26,54]. These two indicators play complementary roles in the analysis.

JSD measures the difference between two probability distributions and can be regarded as a symmetric version of the Kullback–Leibler Divergence [26,27]. JSD is calculated by comparing two distributions, P and Q , to their average distribution M as follows:

$$JSD(P \parallel Q) = \frac{1}{2}KL(P \parallel M) + \frac{1}{2}KL(Q \parallel M), \quad KL(P \parallel M) = \sum_i p_i \log \frac{p_i}{m_i}.$$

A JSD value of zero indicates identical distributions, while larger values indicate greater dissimilarity. In this study, if the *JSD* of a given subclass equals 1 at year t , it implies that the distribution of technological topics in that subclass is entirely different from that at year $t - 1$.

Topic entropy is an information-theoretic measure of the diversity of a probability distribution; it increases when the distribution is spread evenly across topics within a subclass and decreases when probability mass concentrates in a few topics [26]. The *entropy* (H) of a topic distribution is generally defined as follows:

$$H(P) = -\sum_{i=1}^K p_i \log p_i,$$

where p_i denotes the probability of topic i and K is the total number of topics.

4. Results

First, we examined the time required for SBPs to awaken across different IPC subclasses. Across all awakened patents, the mean time-to-awakening was 29.8 years (SD = 4.92), with substantial variation across IPC subclasses. In particular, F02C (gas-turbine plants; air intakes for jet-propulsion plants; controlling fuel supply in air-breathing jet-propulsion plants) exhibited the slowest awakening (longest recognition delay), averaging 40.5 years. By contrast, B63B (ships or other waterborne vessels; equipment for shipping) exhibited the fastest awakening, averaging 20 years. These findings suggest field heterogeneity in SBP awakenings, supporting our choice to use the subclass-specific B as the threshold for identifying SBP awakenings.

To illustrate how SBP awakenings relate to topic shifts, Figure 2 displays complementary sparklines for the top 20 IPC subclasses, selected on the basis of both the number of awakenings and the variability of the topic-shift metrics. Figure 2a shows temporal changes in *JSD*, and Figure 2b shows temporal changes in *entropy*; awakening events are overlaid on each panel. Notably, awakenings tend to cluster in years with decreases in *JSD* and elevated *entropy*—that is, when attention expands to a broader range of topics or the topic composition changes little relative to earlier periods. This pattern implies a negative association between *JSD* and awakenings and a positive association for *entropy*.

To statistically test whether topic shifts affect SBP awakening, we estimated a panel logistic model with year and subclass fixed effects:

$$\Pr(\text{Awakening}_{it} = 1) = \frac{1}{1 + e^{-(\alpha + \beta_1 \text{JSD}_{it} + \beta_2 \text{H}_{it} + \delta_t + \mu_i)}},$$

where δ_t and μ_i denote year and subclass fixed effects, respectively. Table 2 shows the results of the logistic regression. Model 2 reports the relationship between *JSD* and the awakening of SBPs. Consistent with the patterns observed in Figure 2a, the result shows that an increase in *JSD* reduces the likelihood of SBP awakening ($\beta_1 = -4.381$; $p < 0.05$). In other words, when technological topic structures undergo substantial variation over time, the likelihood of SBP awakening diminishes. Model 3 reports the relationship between *entropy* and the awakening of SBPs. The result indicates that *entropy* is positively associated with SBP awakening ($\beta_2 = 0.414$; $p < 0.1$). This suggests that the awakening of SBPs is more likely to occur in technological fields where exploration and experimentation span a broad range of topics.

When *JSD* and *entropy* are included simultaneously in the same model (Model 4), the coefficients retain the same signs as in the separate models, although their statistical significance diminishes. The reduced significance in the full model reflects conceptual

complementarity rather than problematic multicollinearity between the two variables, as variance inflation factors (VIFs) for both variables were close to 1, which is below the conventional threshold of 10 used to assess multicollinearity [55]. This indicates that *JSD* and *entropy* capture related but distinct aspects of topic shifts—stability versus diversity. Specifically, *JSD* quantifies year-to-year structural shifts in topic composition, whereas *entropy* reflects within-year breadth of technological exploration. Their joint inclusion tests whether each variable has an independent effect when controlling for the other, while the separate models highlight their total association with SBP awakenings. In this sense, estimating both individual and combined models provides a fuller picture: the separate models illustrate the overall relevance of each variable, whereas the combined specification clarifies that their explanatory power is partly shared.

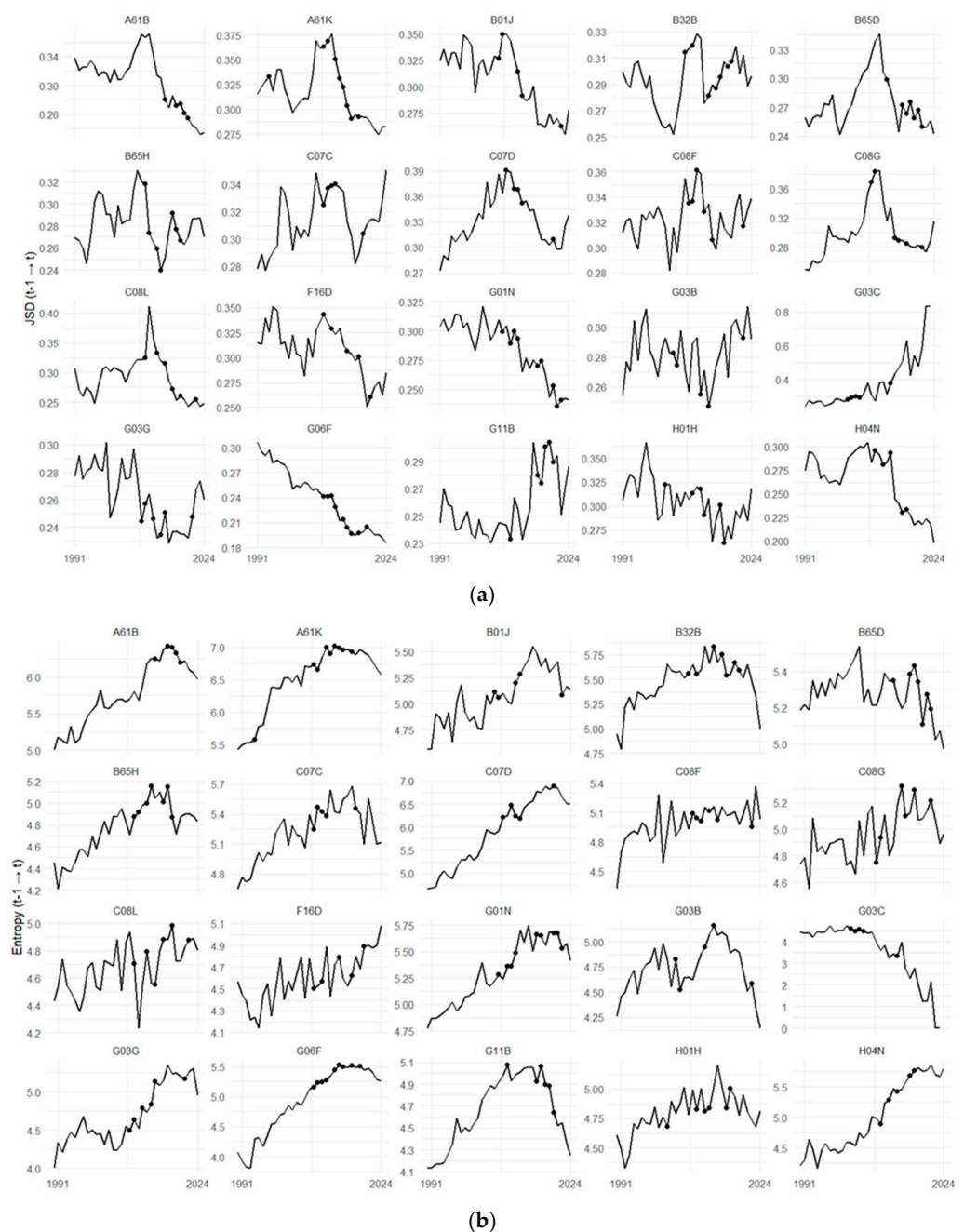


Figure 2. (a) JSD trajectories across selected IPC subclasses with awakening events. (b) Entropy trajectories across selected IPC subclasses with awakening events.

Table 2. Logistic regression results for the awakening of SBPs.

Variables	Model 1	Model 2	Model 3	Model 4
JSD		−4.381 (1.710)		−3.758 (1.827)
Entropy			0.414 (0.225)	0.238 (0.243)
Subclass dummies	Included	Included	Included	Included
Year dummies	Included	Included	Included	Included
Intercept	−19.960 (757.378)	−18.515 (754.567)	−21.380 (757.462)	−19.543 (754.779)
AIC	3165.564	3158.764	3164.103	3159.796
BIC	4697.279	4683.852	4702.524	4691.663
log-likelihood	−1357.782	−1354.382	−1356.051	−1353.898
Pseudo R ² (McFadden)	0.254	0.249	0.255	0.249

Note: The values in parentheses indicate standard errors. The estimation results for year and subclass dummy variables are omitted to conserve space. Full coefficient tables are available from the first author upon request.

Table 3 presents the results of additional analyses conducted as robustness checks. First, because citation-age profiles vary substantially across IPC subclasses, we adopted a relative *B* cut-off (top 0.1% within each subclass) instead of an absolute threshold. This approach ensures comparability across technological fields with different citation profiles. Nevertheless, using the top 0.1% *B* within each IPC subclass may be arbitrary. Our intention is to capture extreme cases of sudden awakenings; for this reason, we deliberately chose a conservative threshold of the top 0.1% *B*. To address concerns of arbitrariness, we performed supplementary analyses by using a top 1% cut-off in Models 5 and 6. The coefficients of *JSD* and *entropy* retained the same signs as the findings in Models 2 and 3. However, while *JSD* remained statistically significant, *entropy* lost statistical significance once the sample was broadened. This suggests that including less extreme cases introduces additional noise, weakening the entropy effect. Although the conservative 0.1% threshold isolates the most pronounced awakenings and provides more stable results, some caution is warranted in interpreting the entropy effect, as its significance appears sensitive to threshold specification. Second, we re-estimated the models by replacing year fixed effects with a continuous time specification (linear and polynomial year trends). While year fixed effects non-parametrically capture any temporal shocks common to all subclasses, continuous year trends impose a more parsimonious structure on temporal dynamics. In Models 7 and 8, our main findings do not differ substantially from those in Models 2 and 3, indicating that the results are not driven by ways temporal variation is modeled.

Table 3. Additional analysis results for robustness checks.

Variables	Model 5	Model 6	Model 7	Model 8
JSD	−3.025 (1.037)		−4.293 (1.590)	
Entropy		0.136 (0.141)		0.467 (0.219)
Subclass dummies	Included	Included	Included	Included
Year dummies	Included	Included		
Year trend (linear)			1.101 (0.076)	1.045 (0.077)
Year trend (quadratic)			−0.023 (0.002)	−0.022 (0.002)

Table 3. Cont.

Variables	Model 5	Model 6	Model 7	Model 8
Intercept	−18.079 (258.666)	−20.546 (429.076)	−13.315 (1.241)	−16.032 (1.330)
AIC	5301.018	5310.543	3205.705	3208.581
BIC	6826.107	6848.964	4520.670	4529.172
log-likelihood	−2425.509	−2429.272	−1408.852	−1410.290
Pseudo R ² (McFadden)	0.415	0.425	0.219	0.225

Note: The values in parentheses indicate standard errors. The estimation results for year and subclass dummy variables are omitted to conserve space. Full coefficient tables are available from the first author upon request.

5. Discussion and Conclusions

Although most patents receive few citations [9,10], some dormant patents can attract renewed attention through collective dynamics [16]. Given that forward citations serve as proxies for patents' value and help trace technological trajectories [7,8], it is both theoretically and practically important to link SBP awakenings to the outcomes of collective dynamics. By extracting technological topics with BERTopic, quantifying yearly topic shifts by using JSD and entropy, and estimating their associations with SBP awakenings, this study found that JSD is negatively associated with SBP awakenings, whereas entropy is positively associated with them. These findings suggest that SBP awakenings are not merely the outcome of random citation shocks or the attributes of individual patents but an emergent phenomenon rooted in collective shifts in technological fields [18].

Prior work has emphasized SBP-specific characteristics in explaining their awakening [3,13,15,32] but has often overlooked the relationship between the manifestations of complex interactions among actors, objects, and ideas and SBP awakenings within technological fields [18]. This study empirically links field-level shifts in technological topics—reflecting the research interests and experimental agendas of actors engaged in patenting—to SBP awakenings. The findings contribute to the SBP literature by providing empirical evidence that complements Mariani et al.'s (2024) claim that the awakening of SBPs can arise through collective dynamics [16] and Kojaku et al.'s (2025) proposition that shifts in knowledge communities may redirect attention to SBPs [18].

The current analysis could be extended to examine sector-specific effects, such as whether topic entropy and stability influence SBP awakenings in domains such as biotechnology and ICT. Collective dynamics at the sector level are likely to be broader in scale than those observed at the field level. Alternatively, topic changes may occur less frequently at the sector level than at the field level. This raises the question of whether topic changes at the sector level are also associated with SBP awakenings. If future research adapts the framework developed in this study to sectors and finds results consistent with those reported here, it could enhance the framework's applicability for domain-level forecasting.

Our findings also contribute to the interpretation of citation-age profiles. On average, patents receive most of their citations within two to four years after grant, followed by a gradual decline [17], which masks the existence and awakening of SBPs. Consistent with this view, Kojaku et al. (2025) argue that modeling forward citations requires going moving aging effects to account for SBPs [18]. We provide a basis for understanding SBP awakenings as exceptional cases that deviate from the general aging patterns of patents.

Finally, the results of this study raise an important question about patent expiration. Patent lifespan partly depends on maintenance-fee decisions; when expected benefits fall below costs, patent owners may choose to terminate their rights early by not paying the fees [56,57]. Because early expiration often signals low value to external observers [10], such patents tend to receive fewer citations. Nevertheless, some prematurely expired

patents later received substantial citations [58], suggesting that valuable patents may have been terminated prematurely (Type II error). Building on our results, future research should examine the late emergence of prematurely expired patents through the lens of collective dynamics.

Our findings have practical implications for patent analytics and R&D management. High entropy within a technological field indicates broad exploration by relevant actors, thereby creating fertile ground for previously overlooked patents to reemerge. Conversely, low divergence reflects stability in the field's thematic structure, facilitating the consolidation of collective attention. For R&D managers and patent analysts, under-cited patents in high-entropy yet low-divergence topic spaces may be prioritized for investment or reactivation, as they are situated in environments conducive to delayed recognition.

This study has several limitations, which suggest directions for future research. First, we did not examine whether the topics of citing patents reveal an additional dimension of collective attention that may help explain delayed recognition. Second, this study did not distinguish between different citation types. Some citations may not reflect actors' interest in a patent but instead result from patent examiners' additions to compensate for citations missed by applicants [59]. Future research should develop topic measures—e.g., distinguishing applicant- versus examiner-added citations—that more accurately capture genuine interest. Third, we did not consider additional forms of collective dynamics. Shifts in inventor collaboration networks, changes in the composition of assignees, and industry-level processes (e.g., standardization) may help explain when dormant patents are rediscovered. Another limitation is that our study does not explicitly address the “Prince” mechanism often discussed in the Sleeping Beauty literature [31]. Rather than analyzing such micro-level triggers, our framework emphasizes field-level topic shifts. In this sense, our approach complements rather than replaces the Prince perspective. Future research could contribute to the Sleeping Beauty literature by integrating both perspectives, examining how individual Princes interact with collective topic shifts to affect delayed recognition. Final limitation concerns the reliance on a single set of BERTopic parameters. Although our indicators (JSD and entropy) are probability-based and normalized within each year—making them inherently less sensitive to clustering choices—the topic structures may still vary under alternative parameterizations. Future research could examine whether the findings hold across different embedding models, dimensionality reduction methods, or clustering thresholds.

This study examined whether field-level changes in technological topics are associated with the awakening of SBPs. By applying BERTopic to U.S. patents and quantifying topic dynamics with JSD and entropy, we found that SBP awakenings are more likely when exploratory attention spans a broader range of topics while year-to-year topic composition remains relatively stable. Theoretically, the results support the perspective of collective dynamics, suggesting that SBP awakenings partly emerge from topic shifts in technological fields. Practically, the findings suggest that firms should monitor field-level topic diversity and stability as early signals of delayed recognition for their SBPs. Future research should address our limitations to better explain when dormant patents are rediscovered.

Author Contributions: Conceptualization, H.J.M. and S.L.; Methodology, H.J.M.; Software, H.J.M.; Validation, H.J.M.; Formal Analysis, H.J.M.; Investigation, H.J.M.; Resources, H.J.M.; Data Curation, H.J.M.; Writing—Original Draft Preparation, H.J.M.; Writing—Review and Editing, S.L.; Visualization, H.J.M.; Supervision, S.L.; Project Administration, S.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The data supporting the conclusions of this article will be made available by the first author on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Dotter, C.T.; Judkins, M.P. Transluminal Treatment of Arteriosclerotic Obstruction. *Circulation* **1964**, *30*, 654–670. [\[CrossRef\]](#)
- Gorry, P.; Ragouet, P. “Sleeping Beauty” and Her Restless Sleep: Charles Dotter and the Birth of Interventional Radiology. *Scientometrics* **2016**, *107*, 773–784. [\[CrossRef\]](#)
- Hou, J.; Yang, X.; Song, H.; Yao, H. Will Patent Family Be Dormant? Research on the Identification and Characteristics of Sleeping Beauty’s Patent Family. *Scientometrics* **2023**, *128*, 5361–5387. [\[CrossRef\]](#)
- van Raan, A.F.J. Sleeping Beauties in Science. *Scientometrics* **2004**, *59*, 467–472. [\[CrossRef\]](#)
- Fahimifar, S.; Janavi, E.; Fadaei, F. Awakening the Beauty: A Journey through Dormant Gems in Strategic Management Literature. *Qual. Quant.* **2024**, *58*, 3331–3362. [\[CrossRef\]](#)
- Ke, Q.; Ferrara, E.; Radicchi, F.; Flammini, A. Defining and Identifying Sleeping Beauties in Science. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 7426–7431. [\[CrossRef\]](#)
- Fontana, R.; Nuvolari, A.; Verspagen, B. Mapping Technological Trajectories as Patent Citation Networks. An Application to Data Communication Standards. *Econ. Innov. New Technol.* **2009**, *18*, 311–336. [\[CrossRef\]](#)
- Hall, B.H.; Jaffe, A.; Trajtenberg, M. Market Value and Patent Citations. *RAND J. Econ.* **2005**, *36*, 16–39.
- Singh, J.; Fleming, L. Lone Inventors as Sources of Breakthroughs: Myth or Reality? *Manag. Sci.* **2010**, *56*, 41–56. [\[CrossRef\]](#)
- Schwall, A.; Wagner, J. The Persistence of Worthless Patents? *World Pat. Inf.* **2023**, *72*, 102179. [\[CrossRef\]](#)
- Xu, W.; Zhang, N.; Jiang, H.; Fan, S.; Zhu, B. Uncovering Gold in Ash: Identifying Sleeping Beauties among Massive Unprofitable Patents. *J. Informetr.* **2025**, *19*, 101674. [\[CrossRef\]](#)
- Braun, T.; Glänzel, W.; Schubert, A. On Sleeping Beauties, Princes and Other Tales of Citation Distributions. *Res. Eval.* **2010**, *19*, 195–202. [\[CrossRef\]](#)
- Hou, J.; Yang, X. Patent Sleeping Beauties: Evolutionary Trajectories and Identification Methods. *Scientometrics* **2019**, *120*, 187–215. [\[CrossRef\]](#)
- Rosenberg, N. *Perspectives on Technology*; Cambridge University Press: Cambridge, UK, 1976.
- Song, H.; Hou, J.; Yang, X.; Liu, R. Wake-up of Sleeping Beauty Patent Families: The Global Non-Equilibrium Diffusion of Technological Knowledge. *Technol. Soc.* **2024**, *79*, 102706. [\[CrossRef\]](#)
- Mariani, M.S.; Battiston, F.; Horvát, E.-Á.; Livan, G.; Musciotto, F.; Wang, D. Collective Dynamics behind Success. *Nat. Commun.* **2024**, *15*, 10701. [\[CrossRef\]](#) [\[PubMed\]](#)
- Mehta, A.; Rysman, M.; Simcoe, T. Identifying the Age Profile of Patent Citations: New Estimates of Knowledge Diffusion. *J. Appl. Econom.* **2010**, *25*, 1179–1204. [\[CrossRef\]](#)
- Kojaku, S.; Lee, J.; Ahn, Y.-Y. Community-Centric Modeling of Citation Dynamics Explains Collective Citation Pat-terns in Science, Law, and Patents. *arXiv* **2025**, arXiv:2501.15552v2.
- Anderson, P.; Tushman, M.L. Technological Discontinuities and Dominant Designs: A Cyclical Model of Technological Change. *Adm. Sci. Q.* **1990**, *35*, 604–633. [\[CrossRef\]](#)
- Kaplan, S.; Tripsas, M. Thinking about Technology: Applying a Cognitive Lens to Technical Change. *Res. Policy* **2008**, *37*, 790–805. [\[CrossRef\]](#)
- Raffaelli, R. Technology Reemergence: Creating New Value for Old Technologies in Swiss Mechanical Watchmaking, 1970–2008. *Adm. Sci. Q.* **2019**, *64*, 576–618. [\[CrossRef\]](#)
- Chattopadhyay, S.; Bercovitz, J. When One Door Closes, Another Door Opens . . . for Some: Evidence from the Post-TRIPS Indian Pharmaceutical Industry. *Strateg. Manag. J.* **2020**, *41*, 988–1022. [\[CrossRef\]](#)
- Zhu, K.; Malhotra, S.; Li, Y. Technological Diversity of Patent Applications and Decision Pendency. *Res. Policy* **2022**, *51*, 104364. [\[CrossRef\]](#)
- Grootendorst, M. BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure. *arXiv* **2022**, arXiv:2203.05794v1.
- Jiang, J.; Ying, F.; Dhuny, R. Unveiling Technological Evolution with a Patent-Based Dynamic Topic Modeling Framework: A Case Study of Advanced 6G Technologies. *Appl. Sci.* **2025**, *15*, 3783. [\[CrossRef\]](#)
- Hall, D.; Jurafsky, D.; Manning, C.D. Studying the History of Ideas Using Topic Models. In Proceedings of the Conference on Empirical Methods in Natural Language Processing, Honolulu, HI, USA, 25–27 October 2008; Association for Computational Linguistics: Kerrville, TX, USA, 2008; pp. 363–371.
- Lin, J. Divergence Measures Based on the Shannon Entropy. *IEEE Trans. Inf. Theory* **1991**, *37*, 145–151. [\[CrossRef\]](#)
- Li, J.; Shi, D.; Zhao, S.X.; Ye, F.Y. A Study of the “Heartbeat Spectra” for “Sleeping Beauties”. *J. Informetr.* **2014**, *8*, 493–502. [\[CrossRef\]](#)

29. Ye, F.Y.; Bornmann, L. “Smart Girls” versus “Sleeping Beauties” in the Sciences: The Identification of Instant and Delayed Recognition by Using the Citation Angle. *J. Assoc. Inf. Sci. Technol.* **2018**, *69*, 359–367. [\[CrossRef\]](#)
30. Dey, R.; Roy, A.; Chakraborty, T.; Ghosh, S. Sleeping Beauties in Computer Science: Characterization and Early Identification. *Scientometrics* **2017**, *113*, 1645–1663. [\[CrossRef\]](#)
31. Hartley, J.; Ho, Y.-S. Who Woke the Sleeping Beauties in Psychology? *Scientometrics* **2017**, *112*, 1065–1068. [\[CrossRef\]](#)
32. Tur, E.M.; Bourellos, E.; McKelvey, M. The Case of Sleeping Beauties in Nanotechnology: A Study of Potential Breakthrough Inventions in Emerging Technologies. *Ann. Reg. Sci.* **2022**, *69*, 683–708. [\[CrossRef\]](#)
33. Li, S.; Yu, G.; Zhang, X.; Zhang, W. Identifying Princes of Sleeping Beauty—Knowledge Mapping in Discovering Princes. In Proceedings of the 2014 International Conference on Management Science & Engineering 21th Annual Conference Proceedings, Helsinki, Finland, 17–19 August 2014; pp. 912–918.
34. Haghani, M.; Varamini, P. Temporal Evolution, Most Influential Studies and Sleeping Beauties of the Coronavirus Literature. *Scientometrics* **2021**, *126*, 7005–7050. [\[CrossRef\]](#)
35. Garud, R.; Rappa, M.A. A Socio-Cognitive Model of Technology Evolution: The Case of Cochlear Implants. *Organ. Sci.* **1994**, *5*, 344–362. [\[CrossRef\]](#)
36. Xu, H.; Winnink, J.; Yue, Z.; Zhang, H.; Pang, H. Multidimensional Scientometric Indicators for the Detection of Emerging Research Topics. *Technol. Forecast. Soc. Change* **2021**, *163*, 120490. [\[CrossRef\]](#)
37. Wang, B.; Liu, S.; Ding, K.; Liu, Z.; Xu, J. Identifying Technological Topics and Institution-Topic Distribution Probability for Patent Competitive Intelligence Analysis: A Case Study in LTE Technology. *Scientometrics* **2014**, *101*, 685–704. [\[CrossRef\]](#)
38. Schmoch, U. *Concept of a Technology Classification for Country Comparisons*; Fraunhofer Institute for Systems and Innovation Research: Karlsruhe, Germany, 2008.
39. Silvestri, D.; Riccaboni, M.; Della Malva, A. Sailing in All Winds: Technological Search over the Business Cycle. *Res. Policy* **2018**, *47*, 1933–1944. [\[CrossRef\]](#)
40. Jee, S.J.; Kwon, M.; Ha, J.M.; Sohn, S.Y. Exploring the Forward Citation Patterns of Patents Based on the Evolution of Technology Fields. *J. Informetr.* **2019**, *13*, 100985. [\[CrossRef\]](#)
41. Abdelrazek, A.; Eid, Y.; Gawish, E.; Medhat, W.; Hassan, A. Topic Modeling Algorithms and Applications: A Survey. *Inf. Syst.* **2023**, *112*, 102131. [\[CrossRef\]](#)
42. Zhu, Y.; Li, Z.; Li, T.; Jiang, L. Topic Recognition and Refined Evolution Path Analysis of Literature in the Field of Cybersecurity. *PLoS ONE* **2025**, *20*, e0319201. [\[CrossRef\]](#)
43. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
44. Rosen-Zvi, M.; Griffiths, T.; Steyvers, M.; Smyth, P. The Author-Topic Model for Authors and Documents. *arXiv* **2012**, arXiv:1207.4169.
45. Wu, Q.; Zhang, C.; Hong, Q.; Chen, L. Topic Evolution Based on LDA and HMM and Its Application in Stem Cell Research. *J. Inf. Sci.* **2014**, *40*, 611–620. [\[CrossRef\]](#)
46. Li, D.; He, B.; Ding, Y.; Tang, J.; Sugimoto, C.; Qin, Z.; Yan, E.; Li, J.; Dong, T. Community-Based Topic Modeling for Social Tagging. In Proceedings of the 19th ACM International Conference on Information and Knowledge Management, Toronto, ON, Canada, 26–30 October 2010; Association for Computing Machinery: New York, NY, USA, 2010; pp. 1565–1568.
47. Gao, Q.; Huang, X.; Dong, K.; Liang, Z.; Wu, J. Semantic-Enhanced Topic Evolution Analysis: A Combination of the Dynamic Topic Model and Word2vec. *Scientometrics* **2022**, *127*, 1543–1563. [\[CrossRef\]](#)
48. Hu, K.; Qi, K.; Yang, S.; Shen, S.; Cheng, X.; Wu, H.; Zheng, J.; McClure, S.; Yu, T. Identifying the “Ghost City” of Domain Topics in a Keyword Semantic Space Combining Citations. *Scientometrics* **2018**, *114*, 1141–1157. [\[CrossRef\]](#)
49. Egger, R.; Yu, J. A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Front. Sociol.* **2022**, *7*, 886498. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Yang, M.; Qu, Q.; Chen, X.; Tu, W.; Shen, Y.; Zhu, J. Discovering Author Interest Evolution in Order-Sensitive and Semantic-Aware Topic Modeling. *Inf. Sci.* **2019**, *486*, 271–286. [\[CrossRef\]](#)
51. Hamilton, W.L.; Leskovec, J.; Jurafsky, D. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. *arXiv* **2016**, arXiv:1605.09096.
52. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. *arXiv* **2013**, arXiv:1301.3781.
53. Pazhouhan, M.; Karimi Mazraeshahi, A.; Jahanbakht, M.; Rezanejad, K.; Rohban, M.H. Wave and Tidal Energy: A Patent Landscape Study. *J. Mar. Sci. Eng.* **2024**, *12*, 1967. [\[CrossRef\]](#)
54. Gan, J.; Qi, Y. Selection of the Optimal Number of Topics for LDA Topic Model—Taking Patent Policy Analysis as an Example. *Entropy* **2021**, *23*, 1301. [\[CrossRef\]](#)
55. O’Brien, R.M. A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Qual. Quant.* **2007**, *41*, 673–690. [\[CrossRef\]](#)
56. Svensson, R. Commercialization, Renewal, and Quality of Patents. *Econ. Innov. New Technol.* **2012**, *21*, 175–201. [\[CrossRef\]](#)

57. Yun, S.; Song, K.; Kim, C.; Lee, S. From Stones to Jewellery: Investigating Technology Opportunities from Expired Patents. *Technovation* **2021**, *103*, 102235. [[CrossRef](#)]
58. Dong, H.-R.; Chen, D.-Z.; Huang, M.-H. Are Invalid Patents Still Cited? *Proc. Assoc. Inf. Sci. Technol.* **2019**, *56*, 639–641. [[CrossRef](#)]
59. Alcácer, J.; Gittelman, M.; Sampat, B. Applicant and Examiner Citations in U.S. Patents: An Overview and Analysis. *Res. Policy* **2009**, *38*, 415–427. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.