Breakthrough Prediction for Emerging Technology Topics Based on Citation Curve Fitting: A Case Study of the Stem Cell Field (Postprint)

Authors: Cao Yiwen, Xu Haiyun, Wu Huawei, Luo Rui

Date: 2023-04-01T16:15:50+00:00

Abstract

[Purpose/Significance] Through fitting analysis of citation curves for emerging technology topics, this study extracts and summarizes the main types and characteristics of citation curves, aiming to provide references for research on prediction methods for breakthrough innovation topics at the micro level. [Method/Process] First, we propose the hypothesis of using citation curves to track and perceive the formation of breakthrough innovations in emerging technology topics, and introduce the concept and measurement method of a transition index. From the two dimensions of knowledge transition and sustained growth, we summarize the criteria for identifying breakthrough innovations and construct a methodological model for breakthrough identification. Second, we utilize temporal citation data from publications on emerging technology topics to construct citation curves and conduct fitting analysis, enabling type classification of citation curves contained within different topics, and extracting the main characteristics of citation curves for breakthrough prediction. [Results/Conclusion] The characteristics of citation curves for emerging technology topics include: recent sustained growth; recent sustained decline; short lifecycle; and consistent change trends in adjacent years. According to the identification criteria for breakthrough innovations, if a citation curve exhibits numerous transitions with large amplitudes, the peak citation occurs relatively late, and maintains high citations with sustained stability or rapid growth in the recent period, then the emerging technology topic possesses the potential to become a breakthrough innovation. By combining expert evaluation of prediction results with research progress on different emerging technology topics, we verify that citation curves can effectively predict the breakthrough potential of emerging technology topics.

Full Text

Preamble

Vol. 64 No. 5 March 2020
ChinaXiv Cooperative Journal

Predicting Breakthroughs in Emerging Technology Topics Based on Citation Curve Fitting: A Case Study of the Stem Cell Field

Cao Yiwen[1],[2], **Xu Haiyun**[1],[3], **Wu Huawei**[1],[2], **Luo Rui**[1],[2]

[1] Chengdu Documentation and Information Center, Chinese Academy of Sciences, Chengdu 610041
[2] Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100190
[3] Institute of Scientific and Technical Information of China, Beijing 100038

Abstract:
[Purpose/Significance] Through fitting analysis of citation curves for emerging technology topics, this study extracts and summarizes the main types and characteristics of citation curves to provide a reference for research on predictive methods for breakthrough innovation topics at the micro level. [Method/Process] First, we propose the hypothesis that citation curves can be used to track and perceive the formation of breakthrough innovations in emerging technology topics, introduce the concept and measurement method of the transition index, and construct an identification model for breakthrough innovations based on two dimensions: knowledge transition and continuous growth. Second, we construct citation curves using temporal citation data from emerging technology topic publications, perform curve fitting analysis, classify the types of citation curves contained in different topics, and extract the main characteristics of citation curves for breakthrough prediction. [Result/Conclusion] The characteristics of citation curves for emerging technology topics include: recent continuous growth, recent continuous decline, short lifecycle, and consistent trends in similar years. According to the identification criteria for breakthrough innovation, if a citation curve has multiple transitions with large amplitudes, a late-occurring peak citation, and maintains high citations with stable or rapid growth in the near term, the emerging technology topic has the potential to become a breakthrough innovation. Combined with expert evaluation of prediction results and research progress across different emerging technology topics, we verify that citation curves can effectively predict breakthroughs in emerging technology topics.

Keywords: breakthrough innovation; technology forecasting; citation curve; emerging technology topic
Classification Number: G250
DOI: 10.13266/j.issn.0252-3116.2020.05.011

With the development of knowledge globalization and intensifying international competition, technological innovation worldwide has entered an unprecedentedly active period. Identifying and predicting disruptive and breakthrough-potential technologies from continuously emerging technology topics is crucial for national and corporate technology strategic planning. Breakthrough innovations exhibit high innovation intensity and often form the foundation of core industrial technologies [1]. Early identification of breakthrough innovations is significant for nations and enterprises to shorten innovation cycles, solve industry technical barriers, build emerging technology industrial clusters, and promote the emergence of breakthrough technologies to enhance national scientific competitiveness. However, because breakthrough innovations adopt new technologies that deviate from existing technological trajectories, they face greater future uncertainty, making their patterns difficult to grasp and identification more challenging than incremental innovations. Emerging technology refers to technology with high novelty and rapid relative growth that has coherence and may significantly impact society and the economy, though it initially carries high uncertainty and ambiguity [2]. Emerging technologies may produce sustained impacts on market structures and industry development, potentially evolving into breakthrough innovations that change existing technological paradigms and trajectories, or they may be fleeting. While emerging technology topics have the potential to develop into breakthrough innovations, this does not guarantee their future research value and prospects [3]. Therefore, identifying potential breakthrough innovation topics from emerging technology topics is of significant importance.

Several methods for identifying breakthrough innovations have been developed, but existing research mostly conducts qualitative or quantitative analysis of breakthrough innovations that have already occurred. Qualitative analysis remains an important means for discovering technology development trends, yet multidisciplinary integration creates limitations in expert wisdom and knowledge structures [4], making the method highly subjective and complex to implement, consuming substantial human, material, and financial resources, and requiring considerable time. Quantitative prediction methods primarily focus on technology evolution and patent analysis. Technology evolution methods require pre-defined identification indicators, but because breakthrough innovation development patterns differ across technology fields, it is difficult to extract universally applicable evaluation metrics. Patent analysis is limited by inventors' citation motivations, as inventors may avoid citing similar patents to ensure novelty.

This study conducts prospective analysis of emerging technologies without definitive conclusions. We propose to construct a methodological model for identifying breakthroughs in emerging technology topics from two aspects: knowledge transition and continuous growth as represented by citation curves. Using the stem cell field as an example, we perform time-sliced citation curve fitting analysis to identify and predict breakthrough technologies that may become major key technologies in the field from the characteristics of citation curves themselves, aiming to improve and enrich the theoretical and methodological system for quantitative identification of breakthrough innovations.

2 Research Status

2.1 Connotation of Breakthrough Innovation

Breakthrough innovation (radical innovation) originated from Schumpeter's concept of "creative destruction" [5]. The true rise of breakthrough innovation research occurred after G. Dosi's classic paper "Technological Paradigms and Technological Trajectories," which unified breakthrough and incremental technological innovation within a single theoretical framework [6]. Most existing research on breakthrough innovation focuses on new products, new processes, new technologies, entirely new scientific knowledge bases, product performance improvement, cost reduction, and changes to market competition structures. Breakthrough innovation concerns both the breakthrough characteristics of technology itself and the substantive impact of innovation.

From a technological breakthrough perspective, breakthrough innovation represents an entirely new technology built upon a completely new foundation of engineering and scientific knowledge [7], disrupting existing technology systems and understanding [8], requiring entirely new scientific knowledge and resources. It is a highly discontinuous, revolutionary innovation and a competence-destroying innovation [9]. From a substantive impact perspective, breakthrough innovation can enhance product performance, create new products, and change market structures or competitive situations. On one hand, breakthrough innovation can endow existing products, processes, or services with unprecedented performance or similar features with significantly improved performance and reduced costs [10]. On the other hand, breakthrough innovation can create new markets, products, processes, industries, and business models [7, 11], eliminate existing technologies and products [9], change industrial competitive structures, improve performance, reduce costs [11], change market consumption patterns [12], better satisfy customer needs [13], and have decisive impacts on market rules, competitive situations, and industrial landscapes, even causing industry reshuffling [14].

This paper focuses on breakthrough innovation emphasizing the dynamic, discontinuous, and novel characteristics of technological innovation, concerning the entirely new scientific knowledge foundation upon which innovation relies and its potential to produce major scientific breakthroughs, rather than incremental improvements to existing technologies.

2.2 Existing Quantitative Identification Methods for Breakthrough Research

Scholars domestically and internationally have conducted a series of studies on breakthrough innovation identification and prediction from different perspectives, including patent analysis, technology evolution, topic mutation, network structure, and sleeping beauty literature perspectives.

2.2.1 Patent Analysis Perspective Using patents to identify breakthrough innovation mainly includes patent citation analysis and patent semantic difference analysis. Regarding patent citations, W. Schoenmakers and G. Duysters used semiconductor industry data to verify that highly cited patents are more likely to generate breakthrough innovations [15]. K. B. Dahlin and D. M. Behrens identified breakthrough innovations through differences in patent citation structures, verifying with tennis racket-related patents that larger citation structure differences are more likely to produce breakthrough innovations [16]. Zhang Jinzhu identified potential breakthrough innovation timing, research topics, and disciplinary classification combinations through mutations in scientific knowledge cited by patents, from the perspective of basic research impact on technological innovation [17]. H. Small argued that methods for identifying scientific discoveries are influenced by multiple factors, and co-citation analysis is an effective method for identifying multiple discoveries, with “multiples” being an important phenomenon reflecting competition and cooperation within scientific communities in scientific development [18]. Regarding patent semantic differences, researchers primarily identify, warn, and predict breakthrough innovations from their origins through patent classification clustering and mutation, patent topic clustering and mutation, patent holder cooperation, and cross-boundary cooperation changes. For example, J. Yoon and K. Kim calculated patent text similarity based on SAO semantic vectors to find outlier patents representing novel breakthrough technological innovations [19].

2.2.2 Technology Evolution Perspective The technology evolution perspective seeks to judge the transition period of a new technology trajectory before it takes shape. A. Sood and G. J. Tellis found that some industry technology trajectories experienced random jumps after long-term stable development [20], while P. Anderson and M. L. Tushman found that the evolution of single technology performance in industries also exhibited random jumps [21]. Breakthrough innovation typically emerges during the transition between old and new technology trajectories, changing the original trajectory, whereas incremental innovation represents performance improvements on the original trajectory [20]. However, technology trajectories require pre-defined performance indicators to characterize breakthroughs, and a single performance indicator is insufficient to fully describe technology evolution, with its drastic changes not equivalent to breakthrough innovation occurrence [17]. G. J. Tellis further noted that technology performance improvements do not simply follow S-curves, and the emergence of breakthrough technologies in technology evolution is full of randomness and contingency. Therefore, technology trajectory S-curves are not good predictive tools and should not serve as the basis for strategy formulation [22].

2.2.3 Topic Mutation Analysis Kuhn argued that the trajectory of scientific progress is discontinuous [23]. Topic mutation analysis focuses on the degree of significant change in topics at certain moments, unaffected by external factors such as topic attention thresholds or domain research 热度. Over time, mutation topics may evolve into research hotspots or disappear/transition into ordinary topics [24]. Current research on topic mutation primarily analyzes burst words and outlier data. J. Kleinberg proposed a burst detection algorithm that identifies burst words in each time period based on word frequency change rates and discovers new research hotspots by analyzing changes in burst word status [25]. Zhang Jinzhu conducted breakthrough innovation identification research in scientific and technological fields by measuring the degree of mutation in cited scientific knowledge [26]. Wang Liya argued that pattern anomalies and data with obvious outlier characteristics often lead to new knowledge discovery and introduced the outlier data concept into topic evolution analysis [27].

2.2.4 Network Structure or Turning Point Identification New technologies emerge in technological environments composed of other technologies [28-29]. The impact of a new technology depends not only on its own influence but also on how it adapts to existing trajectories. R. J. Funk used network analysis methods to explore the impact of new technologies on previous technologies and entire technology trajectories [30]. In fields of technological change, using quantitative methods to distinguish whether innovations depart from or enhance established technology trajectories would enable empirical research to better match basic theory and may promote new concepts. From a network dynamic development perspective, the entry of new information affects structural stability [31]; therefore, measuring the degree of network structure change can identify the impact of knowledge innovation on original knowledge, where information causing huge network structure changes may represent potential breakthrough innovation. C. Chen et al. detected potentially important new publications through centrality change rate Δcentrality and modularity change rate Δmodularity, finding that high Δmodularity node paper scores depend on whether connections are added between network clusters—that is, whether the paper creates a bridge or boundary fusion connection between originally unrelated knowledge modules—and argued that potential major scientific discoveries often appear at the intersection of different scientific knowledge fields [32].

2.2.5 Sleeping Beauty Literature Identification Forward-looking research often exceeds existing cognitive domains. Due to disrupting existing research paradigms, the scientific community is unaware of its existence or fails to recognize its potential knowledge value, maintaining significant psychological distance and underestimating its knowledge value [33-34]. Therefore, compared with incremental innovation, breakthrough innovation is more easily overlooked. History of science shows that some major scientific discoveries and innovative achievements were not timely accepted by scientific community members and were neglected or resisted, only to be rediscovered years later. Such literature is called “sleeping beauties” in science [35]. Both science and technology have delayed recognition phenomena, manifested as sleeping beauty literature in science and sleeping patents in technology [36]. Du Jian proposed a new parameter-free indicator—Bcp index—to identify sleeping beauty literature and shorten the recognition lag for important scientific discoveries [37].

The above methods mostly analyze and observe the discontinuous changes in technology evolution processes for breakthrough innovations that have already occurred. Patent analysis methods also mostly use breakthrough innovations that have occurred to validate methods, and patent citations themselves avoid similar patents or those threatening their innovativeness, thus having limitations in identifying breakthrough innovation through patent citations. Sleeping beauty literature identification aims to shorten the recognition lag for breakthrough innovation achievements, but sleeping beauty literature itself represents a delayed recognition phenomenon that cannot provide timely warning. Topic mutation identification and network structure identification for breakthrough innovation are based on research content of scientific papers themselves, analyzing title, abstract, keywords, authors, and other feature information.

Most existing research evaluates and predicts breakthrough innovation at the meso research field level. However, with increasingly fine disciplinary divisions, breakthroughs in macro or meso research fields often require coordination from multiple micro subfields. Since each micro subfield has different development processes and breakthrough degrees, it is necessary to conduct evaluation and prediction at the micro topic level. Compared with existing research, this paper aims to predict breakthrough innovation at the micro technology topic level. Meanwhile, because breakthrough innovation represents a major innovation type among numerous technological innovations with relatively small numbers, this study starts from emerging technology topics to increase identification effectiveness. Additionally, from the perspective of knowledge creation and diffusion driving technology mutation, we track knowledge diffusion characteristics of micro topics and identify and predict topics with breakthrough innovation potential by observing and extracting types and characteristics of citation curves constructed by different emerging technology topics, aiming to supplement and enrich breakthrough innovation identification methods.

3 Method for Predicting Breakthroughs in Emerging Technology Topics Based on Citation Curve Fitting Analysis

3.1 Main Types of Citation Curves

A citation curve refers to the curve of citation counts over time after a literature’s publication [37]. A. Avramescu divided citation curves into five types: the first three are monotonically increasing then decreasing, with different recognition levels based on peak heights; the fourth is continuously growing; the fifth is curves that are initially recognized but suddenly negated and no longer cited [38]. Li Jiang et al. proposed five citation curves through curve fitting: the first two are regular citation curves including classic and exponential growth curves; the last three are irregular including sleeping beauty, bimodal, and wave-shaped curves [39], where regular curves correspond to Avramescu's first four types.

Building on Li Jiang et al.'s framework, this paper proposes four possible citation curve types: classic, wave-shaped, bimodal, and exponential growth curves (see Figure 1 [Figure 1: see original paper]), and classifies and studies citation curves of emerging technology topics in the stem cell field through curve fitting.

3.2 Predicting Breakthroughs in Emerging Technology Topics Based on Different Citation Curves

This paper predicts breakthroughs in emerging technology topics by observing the growth trends of emerging technology topics and the representation capability of citation curves for these trends. We first propose three hypotheses for using citation curves to track and perceive breakthrough innovation formation in emerging technology topics.

3.2.1 Citation Curves Can Represent the Dynamic Process of Knowledge Innovation Emerging technology topics with high innovation intensity generally receive significant attention from the scientific community, with both publication and citation volumes increasing noticeably over time. Scientific papers are important achievements of technological innovation, citations represent knowledge flow between disciplines, and citation curves represent the dynamic value of scientific papers over time, reflecting the dynamic process of knowledge creation and dissemination in a field. The term "mutation" emphasizes discontinuity or sudden transformation in change processes [17], and breakthrough innovation processes are similarly dynamic. Forward-looking scientific knowledge and research achievements can drive scientific and technological development and serve as sources and drivers for major scientific discoveries and breakthrough innovations. Therefore, identifying breakthrough innovation by observing citation frequency patterns of scientific papers published in emerging technology fields is theoretically feasible.

3.2.2 Citation Curves Can Represent Knowledge Transition Emerging technology topics are novel and rapidly growing research topics concerning scientific or technological problems [42]. Novelty is not only a main characteristic of emerging technology topics but also an important feature of breakthrough innovation [4]. The concept of "transition" originally comes from quantum transition theory in physics, referring to atoms transitioning from low-energy initial states to high-energy final states after absorbing light quanta, and vice versa [40]. Later, "transition" was metaphorically used for major breakthroughs in science and technology—when science and technology experience major breakthroughs, their technology trajectories often show leapfrog development, forming new technological paradigms and trajectories [41]. Breakthrough innovation differs from incremental innovation in having discontinuous technology trajectories [14]. Technology breakthroughs require accumulation processes; when large amounts of scientific research activities cluster around a research topic, new scientific theories and methods may emerge, and when innovation accumulation reaches a certain level, technological innovation breaks through original trajectories to achieve transition, eventually evolving into breakthrough innovation.

Li Baizhou et al. analogized the knowledge creation process to quantum transition—through organizational learning, knowledge energy levels increase, and the knowledge system enters an excited state; when new knowledge is generated at high energy levels, knowledge transition is completed [40]. The scientific community creates new knowledge to solve new scientific problems, and when new knowledge is sufficient to fundamentally change existing knowledge system paradigms, knowledge transition is achieved. Meanwhile, scientific knowledge evolution is a continuous process of accumulation, diffusion, and transition, with knowledge diffusion represented by citation relationships between documents. The difference between highest citation peaks in citation curves can represent the degree of knowledge transition to some extent.

New discoveries of scientific principles are not only important conditions for knowledge transition but also key to new technologies replacing old ones and breaking through technology bottlenecks to achieve upgrades. Therefore, emerging technology topics with clearly layered citation curves achieving multiple large transitions are more likely to produce breakthrough innovation achievements.

The larger the difference between adjacent years' highest citation peaks, the more obvious the fault between citation curves, indicating higher knowledge innovation and more likely qualitative change and discontinuous development. This study uses the transition index—the difference between adjacent years' highest citation peaks—to represent the layering degree of citation curves, which is also the standard for determining knowledge transition.

The transition index calculation formula is:

$$X_{ik} = \frac{x_{ik} - x_{i(k-1)}}{x_{ik}^{\max} - x_{ik}^{\min}} \quad (1)$$

where X_{ik} is the transition index, representing the difference between the highest citation peaks of topic i ($i = 1, 2, 3, \ldots, n$) in year k and year $k-1$ ($k = 1, 2, 3, \ldots, m$); x_{ik} represents the highest citation peak of topic i in year k; and $x_{ik}^{\max}$ and $x_{ik}^{\min}$ are the maximum and minimum values of all highest citation peaks for topic i, respectively.

3.2.3 Citation Curves Can Represent Development Potential of Emerging Technology Topics "Transition" refers to the process where emerging technology topics, if they can maintain high citations in the near term, indicate that the latest scientific knowledge created and produced by the topic can quickly gain attention and recognition. The higher the novelty of scientific knowledge, the greater the probability of breakthrough innovation. Therefore, for emerging technology topics, citation curves with continuous climbing characteristics suggest greater potential to become breakthrough innovations.

The study measures the novelty of scientific knowledge through the timing of the highest citation peak across all citation curves and the recent change trends of citation curves. The later the highest citation peak appears and the more recent the growth, the higher the attention from the scientific community, the more accumulation of related research resources, and the more likely new theories and methods will be created to 催生 breakthrough innovation.

Based on the three hypotheses above, this paper identifies breakthroughs in emerging technology topics themselves—new and growing topics whose sustainable development is constrained by multiple factors. We analyze citation curves of emerging technology topics from knowledge transition and continuous growth perspectives, and extract main criteria for identifying breakthrough innovation: citation curves show obvious layering with many transitions and large amplitudes; citation curves maintain high citations with stable or rapid growth in the near term, with late-occurring highest citation peaks. Figure 2 [Figure 2: see original paper] shows the methodological model and analysis process for identifying breakthroughs in emerging technology topics using citation curve characteristics.

3.3 Curve Fitting Analysis Method

Curve fitting uses appropriate mathematical functions or curve types to fit temporal citation data and is a common method for studying citation curves. This paper performs lognormal function fitting on time-sliced citation curves of emerging technology topics to determine whether topics contain classic citation curve types, with other types identified through line graph observation.

Using the stem cell field as an example, we conduct breakthrough prediction for emerging technology topics in this field. First, we identify emerging technology topics. By retrieving the total number of documents in the field and applying identification indicators—topic novelty and growth rate—we ultimately screen 26 emerging technology topics as analysis objects for breakthrough prediction. Second, we collect citation data for each topic to construct citation curves, perform curve fitting, and conduct breakthrough prediction. By collecting temporal citation data for each of the 26 topics from 2000 to 2017, we construct citation curves for each topic, perform lognormal function fitting, and classify citation curve types for each topic combined with line graphs. Finally, we extract main characteristics of citation curves for different emerging technology topics to ultimately identify and predict topics with potential to develop into breakthrough innovations.

Community detection is a common method for understanding large complex network structures. V. A. Traag et al. found that the Louvain algorithm may produce arbitrarily poorly connected communities and even cause community disconnection, and developed the Leiden algorithm [43], which guarantees connected communities, runs faster, and discovers better partitions. Under the Leiden algorithm, all papers in the WOS database are divided into three levels: 22 clusters at the macro level, 868 at the meso level, and 4,047 at the micro level. This study adopts micro-level clustering topics based on the Leiden algorithm for further analysis.

Stem cells are a class of cells with high self-renewal and multi-directional differentiation potential, providing new ideas for regenerative medicine and disease treatment, with important research value and development prospects, making them a key research area in biomedicine.

4 Empirical Analysis

4.1 Data Acquisition and Analysis

The empirical research object is the stem cell field. On October 20, 2018, we retrieved 422,101 papers in the stem cell field. Using 2004-2018 as the research period with 11 stages, the number of micro-topic classifications for 422,101 documents showed a growth trend over time slices. To obtain topics belonging to the stem cell field, we further calculated the proportion of stem cell field papers in each micro-topic's total WOS database papers, deleting topics with low publication proportions to screen candidate emerging technology topics.

Combining emerging technology topic identification indicators—topic novelty and growth rate—we further screened candidate topics. Novelty and growth rate include four indicators: average growth rates of paper count, journal count, funding count, and author count for each topic. Paper count represents topic research 热度, journal count represents attracted submissions, funding count represents funding status and national emphasis/investment, and author count represents scholar attention.

We calculated and analyzed topics ranking in the top 50 for paper count in each of the 11 stages. The screening process includes three steps:

Step 1: Investigate growth trends of each topic. First, calculate the growth rate of paper counts for each topic across stages and the average growth rate across all stages ($i = 10$) to grasp topic emergence timing and overall development trends. Focus on topics appearing recently (after 2010) and exclude recently disappeared topics considered insufficiently novel, initially screening 54 emerging technology topics.

Step 2: Calculate average growth rates ($i = 10$) for journal count, funding count, and author count related to each topic. Exclude topics with empty average growth rates, indicating the topic did not rank in the top 50 for journals, funding, or authors, suggesting insufficient research 热度 or attention. For example, Topic 1297 had high paper count growth rate ranking but empty values for journals, funding, and authors, indicating low research 热度, thus deleted. This eliminated 10 topics, leaving 44 emerging technology topics.

Step 3: Analyze recent growth trends of each topic. Select the last 5 stages as the analysis window ($i = 4$) and calculate average growth rates of paper counts for remaining topics. Exclude topics with negative average growth rates in the last 5 stages, indicating declining paper counts. This eliminated 18 topics. Finally, 26 emerging technology topics were screened, with partial topic labels shown in Table 1 .

After screening 26 emerging technology topics in the stem cell field, we collected from the WOS database each topic's annual publication data and temporal citation data from 2000-2017, with citation counts measured from publication year to 2018. Using Topic 814 as an example, partial temporal citation data is shown in Table 2 .

4.2 Types and Characteristics of Citation Curves

4.2.1 Citation Curve Types We used Python for lognormal function fitting and performed Kolmogorov-Smirnov (K-S) goodness-of-fit tests to determine whether citation curves of the 26 emerging technology topics belong to classic types, with significance level at 0.05. Remaining curve types were judged by observing line graph shapes and trends.

Among 467 fitted citation curves, 454 were classic curves (97.2%), 9 were wave-shaped (1.9%), 1 was bimodal, and 3 were other types. Sixteen topics contained only classic citation curves, which dominate in quantity (see Table 3). Forty-six curves had good fitting effects with P-values above 0.9, with Topic 461 containing the most good-fitting classic curves (6). Besides classic curves, wave-shaped curves ranked second in quantity, found in Topics 60, 142, 254, 581, 648, 921, and 1046. Bimodal curves appeared only in Topic 1142, with no exponential growth curves in any topic.

Simple citation curve classification is insufficient for identifying breakthrough potential. We further observe citation curve characteristics from knowledge transition and continuous growth perspectives for breakthrough prediction.

4.2.2 Citation Curve Characteristics and Breakthrough Innovation Prediction **Knowledge transition measurement** includes transition index and transition frequency. By calculating transition indices for 26 topics, we found all citation curve transition indices range between -0.71 and 0.85. Combining overall change amplitudes, we set transition index > 0.14 as the criterion for citation curve layering. All topics showed layering with 2-6 transitions, with Topic 261 having the highest transition index of 0.85 in 2012. Transition index distribution is shown in Figure 3 [Figure 3: see original paper].

Continuous growth measurement includes: late occurrence of highest citation peak, ability to maintain high citations recently, and recent growth trends.

Recent growth refers to citation curves maintaining upward trends within three years after publication during 2014-2017. Due to citation delays, incomplete 2018 citation data does not necessarily indicate declining trends.

Among 26 topics, 11 have potential to develop into breakthrough innovations, while 15 have lower possibilities. Due to space limitations, we analyze representative topics in detail.

(1) Recent Continuous Growth. Many topics show recent continuous growth. Topic 1460 (Figure 4 [Figure 4: see original paper]) experienced four transitions: 2000-2005 with low citations (<150) and stable trends; 2006-2008 with first transition and slow decline after peak; 2009 with second transition exceeding 300 citations; 2010-2014 with sharp growth exceeding 700 citations (transition index 0.51); and 2014 with another transition. The highest citation peak occurred late (2014 paper cited in 2018), with recent high citations and sustained growth, indicating many transitions with large amplitudes and growing scientific community attention, suggesting breakthrough potential. Topic 1460 studies decellularized stem cells in tissue regenerative medicine. In September 2010, Professors J. Fisher and E. Ingham from the University of Leeds successfully created naked scaffolds from deceased donors or animal parts to make new body parts using patient stem cells. In October 2010, Professor S. Soker's team at Wake Forest University Medical School created miniature human livers using human stem cells. This technology retains tissue/cell-specific ECM protein components and microstructures while removing rejection-triggering materials, gaining wide application [44]. In 2016, J. P. Guyette et al. used human induced pluripotent stem cells seeded onto decellularized heart scaffolds from organ donors to construct tissue-engineered hearts [45].

Topic 814 (Figure 5 [Figure 5: see original paper]) experienced five transitions through four stages: low citations with stable trends → high citations with fluctuating slow decline → very high citations with fluctuating rapid decline → very high citations with continuous growth. Recent papers maintain high citations with rapid growth, and the highest peak occurred late (2014 paper cited in 2017), indicating recent papers quickly gained peer recognition and contributed importantly to stem cell field development, with strong breakthrough potential. Topic 814 studies stem cell therapy for spinal cord injury. According to ClinicalTrials.gov, over 50 clinical trials on stem cell therapy for spinal cord injury exist internationally, involving bone marrow stem cells, mesenchymal stem cells, and autologous stem cell transplantation [46]. In May 2018, E. Curtis et al. published new results on neural stem cell therapy for spinal cord injury, with 3 of 4 subjects showing significant improvement [47]. In August 2018, K. Hiromi et al. published research in Nature Methods creating spinal cord-derived neural stem cells from human pluripotent stem cells (hPSCs) for potential clinical application [48].

Similar to Topic 814, Topic 353's citation curve has four stages (Figure 6 [Figure 6: see original paper]), but with fluctuating growth in stage two and slower decline in stage three, with stage four showing continuous growth (the monotonic increase portion of classic curves). Overall, Topic 353 shows clear layering with five transitions, late highest peak (2014 paper cited in 2017), long lifecycle, and recent high citations with continuous growth, indicating lasting impact and potential for breakthrough innovation. Topic 353 studies extracellular matrix regulation of stem cell fate. In September 2018, W. J. Sullivan et al. published in Cell that extracellular matrix can control cell movement in vivo by regulating cellular glucose metabolism, important for revealing cancer cell metastasis [49].

Topic 161 (Figure 7 [Figure 7: see original paper]) shows three stages: rapid rise 2-3 years after publication followed by fluctuating slow decline; transition in 2009 with continued fluctuating decline; and transition in 2014 with highest peak that year. With six transitions and recent relatively high citations showing upward trends, this topic maintains high research 热度 and breakthrough potential. Topic 161 studies clinical application of hematopoietic stem cells, with increasing patients receiving allogeneic hematopoietic stem cell transplantation for acute myeloid leukemia.

(2) Recent Continuous Decline. Some topics show declining research 热度 despite multiple transitions. Topic 142 (Figure 8 [Figure 8: see original paper]) experienced five transitions across three stages: 2000-2006 with citations under 500 and stable linear trends; 2007-2010 with three transitions and rapid growth followed by slow decline; and 2011+ with sharp growth for fourth transition, all curves peaking 2-4 years after publication then rapidly declining. The highest peak occurred early (2011) with recent obvious decline and no sustained high citations, indicating that despite early attention and knowledge transitions, recent research 热度 has decreased, possibly encountering technical bottlenecks, with low breakthrough potential. Topic 142 studies DNA methylation in stem cells, an epigenetic regulatory mechanism for lineage classification and quality control of embryonic, induced pluripotent, and mesenchymal stem cells. Domain experts note that as mechanistic research, it is unlikely to evolve into breakthrough innovation.

Topic 2 (Figure 9 [Figure 9: see original paper]) experienced four transitions across three stages: 2001-2004 with low citations (<500) and stable linear trends; 2005-2007 with growth followed by fluctuating slow decline; and 2008-2017 with surge to peak then fluctuating rapid decline. Despite many large-amplitude transitions, recent overall decline with recent growth far below peak levels indicates reduced research 热度 and low breakthrough potential. Topic 2 studies microRNA regulation in stem cells, which as a regulatory factor remains mechanistic research with low breakthrough potential.

(3) Short Lifecycle. Topics with short citation curve lifecycles show rapid citation growth to peak followed by quick decline. Topics 469 and 261 (Figures 10 [Figure 10: see original paper] and 11 [Figure 11: see original paper]) gained quick recognition and high citations after publication, generally peaking 2-3 years post-publication, but citations quickly fell back to low levels. While these topics contributed to the field, they lacked lasting impact. With four and six transitions respectively, and highest peaks in 2015 and 2012, neither maintained recent high citations or growth trends. High-innovation topics generally maintain high attention and research 热度, making these unlikely to become breakthrough innovations. Topic 469 studies stem cell applications in liver regeneration, with basic research on liver regeneration mechanisms still exploratory and unlikely to produce breakthrough innovation [50]. Topic 261 studies stem cell therapy for Hodgkin lymphoma, where high-dose chemotherapy followed by autologous hematopoietic stem cell transplantation has become standard treatment for relapsed/refractory Hodgkin lymphoma [51], making further breakthrough innovation unlikely.

(4) Consistent Trends in Similar Years. Some topics show similar growth and decline rates in similar years. Topic 254 (Figure 12 [Figure 12: see original paper]) shows 2004 and 2005 curves nearly overlapping, 2006-2011 curves with similar shapes, 2011-2013 curves climbing together, 2013-2014 gradual decline, 2014-2015 rebound, then gradual decline. Similar-year curves share mutation patterns and similar change rates, possibly because papers published in similar years face similar research environments and attention levels. Topic 254 studies mesenchymal stem cells and applications. As of March 2019, China had 38 approved stem cell clinical research projects, with 29 using mesenchymal stem cells [53].

Topic 648 (Figure 13 [Figure 13: see original paper]) also shows consistent trends in similar years, with similar curve shapes within stages. Whether this similarity helps predict recent curve development requires future observation. Notably, Topics 254 and 648 maintain rapid citation growth in recent years, potentially surpassing previous peaks, indicating sustained research 热度 and breakthrough potential. Topic 648 studies limbal stem cells and applications, where autologous limbal stem cell expansion and transplantation has been approved in the EU [54], showing broad application prospects.

5 Discussion and Conclusion

This paper proposes three hypotheses for using citation curves to track breakthrough innovation formation in emerging technology topics: citation curves can represent the dynamic process of knowledge innovation, knowledge transition, and development potential of emerging technology topics. From knowledge transition and continuous growth dimensions, we summarize identification criteria, propose the transition index concept and measurement, and construct a methodological model for identifying breakthroughs in emerging technology topics. Using temporal citation data for each topic's publications, we construct citation curves, perform curve fitting, classify curve types, and extract main characteristics for breakthrough prediction. Results align with expert evaluations and research progress in related stem cell topics.

Citation curves for emerging technology topics mainly show four characteristics: recent continuous growth, recent continuous decline, short lifecycle, and consistent trends in similar years. Among these, topics with recent continuous growth, clear layering, many transitions with large amplitudes, late highest peaks, and recent high citations have more lasting impact and higher research 热度, making breakthrough innovation more likely. Topics with recent decline show reduced research 热度 despite knowledge transitions, making breakthroughs unlikely. Short-lifecycle topics rise and fall quickly, lacking lasting impact and breakthrough potential. Some topics show extremely similar curves in similar years, possibly due to similar research environments, with similar mutation patterns and change rates.

Due to breakthrough innovation complexity, unified indicators are difficult. Transition index, frequency, and peak timing serve only as references—many transitions or late peaks don't guarantee breakthroughs. Recent high citations and continuous growth are primary prediction standards; topics with both conditions have greater breakthrough potential. Single standards are insufficient; comprehensive multi-criteria assessment improves prediction accuracy.

Breakthrough innovation identification and prediction remain core tasks of scientific and technological intelligence analysis. Using citation curves to identify breakthroughs in emerging technology topics provides new intelligence analysis ideas, enriching and expanding breakthrough innovation prediction methods to support domain experts and planners. Combined with expert evaluation and stem cell field research progress, citation curves can automatically and efficiently track and perceive breakthroughs in emerging technology topics.

This study has limitations. Predicting breakthroughs from citation curves relies on citation data, which has time lags and is affected by database entry times, creating some errors. Future work will consider setting citation time windows and using average citations per paper to reduce errors. Given breakthrough innovation prediction complexity, future research should also incorporate technology policy layouts for auxiliary analysis to identify breakthrough-potential emerging technology topics earlier.

References

[1] Zhang Hongshi. Research on the Drivers and Organizational Models of Breakthrough Innovation [D]. Hangzhou: Zhejiang University, 2005.
[2] Rotolo D, Hicks D, Martin BR. What is an emerging technology? [J]. Research policy, 2015, 44(10): 1827-1843.
[3] Luo Rui, Xu Haiyun, Dong Kun. Review of Frontier Identification Methods [J]. Library and Information Service, 2018, 62(23): 119-131.
[4] Zhang Jinzhu, Zhang Xiaolin. Review of Breakthrough Innovation Identification Research Based on Patent Science Citations [J]. Journal of the China Society for Scientific and Technical Information, 2016, 35(09): 955-962.
[5] Wan Ning. Analysis of the Relationship Among Disruptive, Destructive, and Breakthrough Innovation [J]. Business, 2015(30): 122-123.
[6] Chen Ao, Liu Xielin. Where Does Breakthrough Technology Come From?—A Literature Review [J]. Studies in Science of Science, 2011, 29(9): 1281-1290.

[7] Dess GG, Beard DW. Dimensions of organizational task environments [J]. Administrative science quarterly, 1984, 29(1): 52-73.
[8] Koshland DE. The cha-cha-cha theory of scientific discovery [J]. Science, 2007, 317(5839): 761-762.
[9] Tushman ML, Anderson P. Technological discontinuities and organizational environments [J]. Administrative science quarterly, 1986, 31(3): 439-465.
[10] Kotelnikov V. Radical innovation versus incremental innovation [M]. Boston: Harvard Business School Press, 2000.
[11] Doyle JF. Radical innovation: how mature companies can outsmart upstarts [J]. Research-technology management, 2000, 43(10): 706-707.
[12] Zhou KZ, Yim CK, Tse DK. The effects of strategic orientations on technology- and market-based breakthrough innovation [J]. Journal of marketing, 2005, 69(2): 42-60.
[13] Chandy RK, Tellis GJ. Organizing for radical product innovation: the overlooked role of willingness to cannibalize [J]. Journal of marketing research, 1998, 35(4): 474-487.
[14] Fu Yuxiu, Zhang Hongshi. Breakthrough Innovation: Concept Definition and Comparison [J]. Journal of Quantitative & Technical Economics, 2004, 21(3): 73-83.
[15] Schoenmakers W, Duysters G. The technological origins of radical inventions [J]. Research policy, 2010, 39(8): 1051-1059.
[16] Dahlin KB, Behrens DM. When is an invention really radical?: Defining and measuring technological radicalness [J]. Research policy, 2005, 34(5): 717-737.
[17] Zhang Jinzhu. Identifying Breakthrough Innovation Using Cited Scientific Knowledge Mutation [M]. Beijing: Science Press, 2017.
[18] Small H. Referencing as cooperation or competition [M]. Berlin: Walter de Gruyter, 2016.
[19] Yoon J, Kim K. Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks [J]. Scientometrics, 2011, 88(1): 213-228.
[20] Sood A, Tellis GJ. Technological evolution and radical innovation [J]. Journal of marketing, 2005, 69(3): 152-168.
[21] Anderson P, Tushman ML. Technological discontinuities and dominant designs: a cyclical model of technological change [J]. Administrative science quarterly, 1990, 35(4): 604-633.
[22] Tellis GJ. Disruptive technology or visionary leadership? [J]. Journal of product innovation management, 2006, 23(1): 34-38.
[23] Kuhn TS. The structure of scientific revolution [M]. Chicago: University of Chicago Press, 1999.
[24] Li Yong, An Xinying, Zhao Yingguang, et al. Research on Burst Topic Detection Combined with Knowledge Organization Systems [J]. Information Studies: Theory & Application, 2013, 36(5): 120-123.
[25] Kleinberg J. Bursty and hierarchical structure in streams [J]. Data mining & knowledge discovery, 2003, 7(4): 373-397.
[26] Zhang Jinzhu, Zhang Xiaolin. Identifying Breakthrough Innovation Using

Cited Scientific Knowledge Mutation [J]. Journal of the China Society for Scientific and Technical Information, 2014, 33(3): 259-266.
[27] Wang Liya. Research on Topic Evolution Based on Outlier Data [D]. Beijing: University of Chinese Academy of Sciences, 2012.
[28] Arthur WB. The structure of invention [J]. Research policy, 2007, 36(2): 274-287.
[29] Arthur WB. The nature of technology: what it is and how it evolves [M]. New York: Free Press, 2009.
[30] Funk RJ, Owen-Smith J. A dynamic network measure of technological change [J]. Management Science, 2016, 63(3): 587-605.
[31] Chen C. Turning points: the nature of creativity [J]. Nature of creativity contemporary psychological perspectives, 2011, 18(1): 87-98.
[32] Chen C, Chen Y, Horowitz M, et al. Towards an explanatory and computational theory of scientific discovery [J]. Journal of informetrics, 2009, 3(3): 191-209.
[33] Campanario JM. Rejecting and resisting Nobel class discoveries [J]. Scientometrics, 2009, 81(2): 549-565.
[34] Fang H. An explanation of resisted discoveries based on construal-level theory [J]. Science & engineering ethics, 2015, 21(1): 41-50.
[35] Raan AFJ. Sleeping beauties in science [J]. Scientometrics, 2004, 59(3): 467-472.
[36] Palomeras N. Sleeping patents: any reason to wake up? [J]. Ie working paper, 2003, 20(35): D506.
[37] Du Jian. Research on Identification Methods and Awakening Mechanisms of Sleeping Beauty Literature [D]. Nanjing: Nanjing University, 2017.
[38] Avramescu A. Actuality and obsolescence of scientific literature [J]. Journal of the American Society for Information Science & Technology, 2010, 30(5): 296-303.
[39] Li Jiang, Jiang Mingli, Li Yuting. An Analytical Framework for Citation Curves: Taking Nobel Laureates' Citation Curves as Examples [J]. Journal of Library Science in China, 2014, 40(2): 41-49.
[40] Li Baizhou, Zhao Jianyu, Su Yi. Dynamic Model of Organizational Learning-Knowledge Creation Process Based on Energy Level Transition [J]. Studies in Science of Science, 2013, 31(6): 913-9225.
[41] Zhang Lichao, Liu Yijun. Technology Trajectory Transition and Technological Innovation Evolution [J]. Studies in Science of Science, 2015, 33(1): 137-145.
[42] Guo Hanning. Exploration of Emerging Research Field Identification from Multiple Scientific Indicators Perspective [D]. Dalian: Dalian University of Technology, 2013.
[43] Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities [J]. Scientific reports, 2019, 9(1): 5233.
[44] Bioon. [Review] Stem Cells and New Materials in Regenerative Medicine Research and Application (1) [EB/OL]. [2019-09-16]. https://meeting.bioon.com/2018StemCells/news-detail/b04d4462ec057124.
[45] Guyette JP, Charest JM, Mills RW, et al. Bioengineering human my-

ocardium on native extracellular matrix [J]. Circulation research, 2016, 118(1): 56-72.
[46] ClinicalTrials.gov [EB/OL]. [2019-09-16]. https://clinicaltrials.gov/ct2/results?cond=stem+cell+cord+inju
[47] Curtis E, Martin JR, Gabel B, et al. A first-in-human, phase I study of neural stem cell transplantation for chronic spinal cord injury [J]. Cell stem cell, 2018, 22(6): 941-950.
[48] Hiromi K, Ken K, Adler AF, et al. Generation and post-injury integration of human spinal cord neural stem cells [J]. Nature methods, 2018, 15(9): 723-731.
[49] Sullivan WJ, Mullen PJ, Schmid EW, et al. Extracellular matrix remodeling regulates glucose metabolism through TXNIP destabilization [J]. Cell, 2018, 175(1): 117-132.
[50] International Liver Disease. APASL 2019 Focus | Academician Wang Fusheng: Stem Cell Therapy for Liver Cirrhosis and Liver Failure—Progress and Research Hotspots [EB/OL]. [2019-09-16]. http://www.ihepa.com:8088/default/htmlDocument/2019-02-22/detail_{16360}.html.
[51] Wang H. Bendamustine, Gemcitabine, Vinorelbine as Salvage Chemotherapy Before Autologous Hematopoietic Stem Cell Transplantation for Relapsed/Refractory Hodgkin Lymphoma [EB/OL]. [2019-09-16]. https://www.bjcancer.org/Mobile/Article/Inde
[52] Min Chao, Ding Y, Li Jiang, et al. Citation Diffusion of Single Papers [J]. Journal of the China Society for Scientific and Technical Information, 2018, 37(04): 341-350.
[53] Youzhuwang. Trend | 37 Stem Cell Clinical Projects Approved by National Filing, Treating Extremely Wide Range of Diseases (Full List Attached) [EB/OL]. [2019-09-16]. http://www.sohu.com/a/317270049_{827862}.
[54] Yiou. Stem Cell Therapy Research Progress and Industry Analysis [EB/OL]. [2019-09-16]. https://www.iyiou.com/p/95934.html.

Author Contributions

Cao Yiwen: Participated in discussion of research ideas, responsible for literature review, writing, and revision.
Xu Haiyun: Determined topic selection, proposed research ideas, guided paper revision.
Wu Huawei: Participated in discussion of research ideas, responsible for data processing and figure creation.
Luo Rui: Participated in emerging technology topic identification.