Review of Research Frontier Identification Methods: Postprint

Authors: Luo Rui, Xu Haiyun, Dong Kun

Date: 2023-08-27T00:00:00+00:00

Abstract

[Purpose/Significance] Through comparative analysis of different domain frontier identification methods, this study summarizes the advantages and disadvantages of various methods in identifying and predicting frontiers, and proposes improvement suggestions for the identification and prediction of future research frontiers. [Method/Process] By reviewing domestic and foreign literature related to research frontiers, this paper clarifies the relevant concepts of research frontiers. It analyzes the current main research frontier identification methods, and compared with traditional identification methods, focuses on summarizing the identification methods for transformative research frontiers. It also summarizes the main problems existing in current domain frontier identification methods and proposes improvement suggestions. [Results/Conclusions] In terms of concepts, the conceptual connotations of "research hotspots", "emerging research", and "research frontiers" are distinguished along two dimensions: time and degree of innovation; based on different degrees of innovation, research frontiers can be divided into "conventional research frontiers" and "transformative research frontiers". In terms of identification methods, different methods have their applicable scenarios. Future research needs to deeply explore "research frontiers", particularly focusing on the semantic associations between topics and the fusion of multi-relational data from multiple sources, as well as feature mining of early signs of "transformative research frontiers", and construct corresponding identification and prediction methods.

Full Text

Abstract

[Purpose/Significance] This paper summarizes the advantages and disadvantages of various research frontier identification methods through comparative analysis, and proposes improvements for future identification and prediction of research frontiers. [**Method/Process**] By reviewing domestic and international literature related to research frontiers, this study clarifies relevant concepts and analyzes current mainstream identification methods. Compared with traditional approaches, it focuses particularly on summarizing methods for identifying transformative research frontiers. The paper then outlines major existing problems in current domain frontier identification methods and puts forward suggestions for improvement. [**Result/Conclusion**] Conceptually, this paper distinguishes among "research hotspots," "emerging research," and "research frontiers" along two dimensions: time and degree of innovation. Based on innovation level, research frontiers can be categorized into "conventional research frontiers" and "transformative research frontiers." Methodologically, different approaches have their applicable scenarios. Future research should further excavate the characteristics of "research frontiers," particularly by attending to semantic associations between topics and fusing multiple relationships from multi-source data, as well as mining early signs of "transformative research frontiers" to construct corresponding identification and prediction methods.

2. Conceptual Analysis and Characteristics

2.1 Concept Proposal and Development

D. J. Price first proposed the concept of "research front" in 1965, arguing that a research frontier in a field could be identified from frequently cited papers, as this small body of actively and rapidly developing literature connects previously published works with newly published ones, promoting scientific development. He viewed this as a kind of "growing tip" or "epidermal layer" that distinguishes science from non-science [1]. In English terminology, both "research front" and "research frontier" are translated as "research frontier," but the former uses bibliometric methods for predictive evaluation (a priori assessment representing expected results), while the latter represents actual observed results determined by peer experts in the field (a posteriori assessment representing true domain frontier research) [2]. Minimizing the deviation between these two approaches is the ultimate goal of intelligence analysis.

H. Small and B. C. Griffith conceptualized research frontiers as the result of highly interactive literature clustering, as such paper clusters demonstrate "high-level activity" within scientific fields. They employed co-citation clustering to extract "research frontiers" [3]. O. Persson identified research frontiers through bibliographic coupling, treating citing literature as the research frontier and cited literature as the knowledge base [4]. S. A. Morris et al. shared a similar understanding of research frontiers as O. Persson, identifying them through bibliographic coupling and defining research frontiers as collections of papers citing a fixed set of temporally invariant foundational literature, with cited literature representing the knowledge base. They also argued that research frontiers are discontinuous, likely emerging or disappearing as scientists begin new problems or solve existing ones [5]. Wang Lixue et al. categorized foreign definitions of research frontiers into three types corresponding to three bibliometric identification methods: co-citation analysis (treating highly cited papers as frontiers), bibliographic coupling (treating citing papers as frontiers), and co-word analysis (treating emergent or hot topics as frontiers) [6]. These classic perspectives all start from scientific citations, focusing on active or highly cited papers, thus limiting the definition of research frontiers by data sources.

Subsequent scholars expanded this understanding. Chen Chaomei et al. viewed research frontiers as sets of emergent dynamic concepts and potential research questions, emphasizing their characteristics of new trends and mutations [7]. S. P. Upham and H. Small considered research frontiers as the most dynamically changing research topics that attract the most scientists' attention in scientific and technological fields, representing a fusion of scientific discovery and social concern [8]. Zheng Yanning et al. noted that research frontiers are specific to particular domains and time periods, emphasizing their rapid development and active academic exchange [9]. Feng Jia argued that research frontiers are the latest research topics with high academic attention, characterized by "high attention" and "novelty" [10].

2.2 Differences in Temporal Dimension

2.2.1 "Research Hotspot" versus "Research Frontier" Research hotspots are topics with high popularity that currently attract many researchers' attention. Many scholars view hotspot topics as those with high frequency and can capture hotspot events through social media [11-12]. Zhong Zhen, when exploring the relationship between research hotspots and frontiers, suggested that frontier topics have a high probability of becoming the next stage's hotspot topics [2]. Thus, research hotspot topics partially originate from research frontiers. As research frontiers develop with the continuous joining of the disciplinary community, they may become research hotspots in the next period. Research hotspots focus more on receiving "attention" and having higher "popularity," showing some temporal lag compared to research frontiers. From the perspective of value innovation, "attention" cannot serve as a criterion for "value." High attention does not necessarily mean the research will ultimately promote scientific and social development. The main reason for high attention may be that the research value of a domain topic has gradually been established, attracting numerous researchers, though it could also result from researchers following trends in emerging fields.

2.2.2 "Emerging Research" versus "Research Frontier" Emerging research topics are newly appearing topics showing growth trends. In the Wharton School's 2000 publication *Wharton on Managing Emerging Technologies* (Chinese edition 2002), emerging technologies are defined as innovations built on science that may create a new industry or transform an old one, including both discontinuous technologies from radical innovation and more innovative technologies formed by integrating multiple independent past research achievements [13]. Guo Hanning argued that emerging research topics are scientifically or technologically recognized, young, and rapidly growing research topics within the scientific community, while research frontiers can change over time within specific domains [14]. If we differentiate these concepts by observation periods, we lose sight of their essential nature, as both emerging research fields and research frontiers have temporal requirements—they are "newly emerging" and show "growth trends" or represent "frontier domains" relative to the observation period. The essential difference is that emerging research fields, despite their "youth" and "rapid growth," do not necessarily represent valuable and promising research frontiers. Rotolo et al. classified technologies with novelty, rapid growth, coherence, significant impact, and uncertainty as emerging technologies [15]. Therefore, emerging research topics are "newly emerging" and show "growth trends" or represent "frontier domains" relative to the observation period.

2.3 Differences in Innovation Degree

T. Kuhn divided scientific research into two categories based on innovation degree: normal scientific research and innovative research, corresponding to incremental innovation and transformative innovation [16]. Incremental innovation, also called normal scientific research, innovates based on existing scientific development, while transformative innovation causes paradigm shifts in scientific or technological fields. Related concepts include radical technology, disruptive technology, creative destruction, technological breakthroughs, and discontinuous innovation [17].

C. M. Christensen first proposed the concept of "disruptive technology," viewing it as technology that replaces existing mainstream technology, brings new product and service functions, and has destructive, disruptive impacts on industries or market structures [18]. Zhang Jinzhu defined breakthrough innovation technology as technology where methods, products, equipment, materials, and other technical themes undergo discontinuous changes, triggering performance leaps or functional changes that ultimately lead to discontinuous changes in markets, products, services, and business models [17]. Disruptive technology emphasizes the actual effects of science and technology, with characteristics similar to emerging technology—both emphasize "innovation" and impact on "practice." Breakthrough innovation (radical innovation) has connotations very close to disruptive innovation, but differs in that disruptive innovation does not explicitly require "technological innovation" degree. Its core perspective is market segmentation and value systems, not necessarily accompanied by technological breakthroughs [19], whereas breakthrough innovation emphasizes the enormous changes in scientific development, focusing on its own "innovativeness," with uncertain immediate market effects.

2.4 Related Concept Analysis

The above concepts differ from yet connect with "research frontier." This paper distinguishes the connotations of "domain frontier," "research hotspot," and "emerging research" through time and innovation dimensions. Figure 2 [Figure 2: see original paper] is a schematic diagram showing these relationships in a two-dimensional coordinate system with time on the horizontal axis and innovation on the vertical axis, using O as the origin. On the time axis, t_0 represents the observation point, dividing the timeline into past, present, and future.

Overall, from the temporal dimension, "research frontier" is closer to the future and represents higher innovation with greater development value. Due to first-mover advantages in technological innovation, pioneers exploring unknown fields can obtain maximum value. From the innovation degree perspective, compared with popular "research hotspots," both "research frontier" and "emerging research" themes show higher innovation. Additionally, research frontiers can be divided into "incremental innovation" with lower innovation degree and "transformative innovation" with higher innovation degree.

Specifically, each development stage in a research field has corresponding "research hotspots," representing incomplete predictions of "research frontiers." Some past "research hotspots" that have been validated through development further become "emerging research" and "research fronts." Research hotspots accumulate obvious advantages with prominent topic characteristics, making them easy to identify. Compared with research hotspots that appear at each stage, "emerging research" themes refer to newly appearing topics that are current and have future development potential, including explorations of "research frontiers," thus intersecting with them. Due to their temporal novelty, emerging research topics usually lack the prominent features of hotspot topics and require specific intelligence analysis methods for selection and identification. Similarly, as "research frontier themes" continue to develop in the future, they may attract numerous researchers and eventually become "research hotspots," though they may also terminate due to unfavorable development. Meanwhile, "research frontier" themes focus more on detecting future high-value research topics compared to "research hotspots" and "emerging research themes," thus requiring more targeted intelligence analysis and prediction methods with greater identification and prediction difficulty.

From the above conceptual analysis of "research frontier" and classification of scientific discoveries, we can see that "research frontier" has rich connotations with multiple characteristics such as "innovativeness," "transformative nature," "value," and "development potential." Research frontier emphasizes the scientific value after innovation, mainly referring to scientific research or technological development with promising prospects, focusing on themes or technologies that can bring enormous changes to human life. Both "incremental innovation" and "transformative innovation" contain tremendous development potential and innovative value, which is the most important feature of "research frontier."

Since there is no unified definition of "research frontier," researchers have different understandings of its features, leading to various identification methods with different emphases. Traditional scientometric methods have good identification effects for "incremental innovation frontiers," but "transformative innovation frontiers" are full of randomness and mutations. Compared with the traceable patterns of incremental innovation, transformative innovation is more difficult to identify, requiring distinct methods. The following sections compare and analyze current mainstream research frontier identification methods.

3. Analysis of Research Frontier Identification Methods

Existing research frontier identification methods can be broadly categorized into two types: (1) expert judgment methods that utilize expert knowledge for subjective assessment and comprehensive judgment; and (2) quantitative analysis methods that identify disciplinary research frontiers through quantitative analysis of scientific and technological information.

3.1 Expert Judgment Method

Currently, expert judgment is an important means for identifying and predicting scientific and technological development trends. Many industries and institutions conduct frontier predictions to grasp domain development directions. Representative products include China's "Top Ten Scientific Advances," the Chinese Academy of Sciences' "Science Development Report," *Science* magazine's "Top Ten Scientific Breakthroughs," and MIT Technology Review's "Top Ten Breakthrough Technologies," all of which primarily rely on expert wisdom supplemented by data analysis to determine current status and future directions of science and technology development [17].

Expert judgment can fully utilize experts' wisdom and experience, but experts have independent mental models and subconsciously tend to pay more attention to their own areas of interest, making results highly subjective. M. Bengisu and R. Nekhili found that expert subjective judgment could produce significant biases when using patent and bibliometric methods to validate breakthrough innovation technologies [20]. Especially when research objects are highly dynamic and feature items are difficult to extract, expert wisdom may not perform optimally, failing to monitor research frontiers quickly and efficiently with limited advantages in speed, efficiency, and accuracy. With the emergence of data-intensive knowledge innovation paradigms and their gradual deepening, the role of data analysis has become increasingly prominent as an important auxiliary foundation for expert judgment.

3.2 Citation-Based Analysis Methods

Citation-based analysis methods primarily build citation networks based on citation relationships between documents and identify research frontiers through visual analysis of co-citation knowledge networks. Few studies use this method alone; most compare it with other citation methods [21]. Co-citation and bibliographic coupling analyses are more common citation methods for domain frontier identification. Specific interpretations are shown in Table 1 .

Co-citation analysis has temporal lag—only when a paper reaches a certain citation count does it gain attention, often overlooking potential research frontiers. M. H. Huang and C. P. Chang compared bibliographic coupling and co-citation analysis for detecting OLED research frontiers and found that bibliographic coupling was superior in detection quantity and speed [28]. Compared with co-citation analysis, coupling analysis improves the time lag issue to some extent.

Bibliographic coupling analysis focuses on citing documents, with reference literature remaining fixed. Coupling relationships between citing documents can be obtained immediately after publication and are less dynamically variable than co-citation analysis, making results more static and stable [29]. However, when using external feature items like authors as coupling objects, coupling relationships change dynamically over time, allowing more comprehensive frontier analysis from a developmental perspective [30].

Beyond these two common citation network analyses, D. Fajardo-Ortiz et al. identified paradigm research frontiers and analyzed their dynamics by constructing sub-networks of highly inter-cited papers combined with network clustering and text mining [31]. However, this method is rarely applied in research frontier identification.

In summary, citation relationship networks are classic and important methods for identifying and predicting domain research frontiers. Direct citation, co-citation, bibliographic coupling, and mutual citation each reflect different understandings of research frontiers, with each citation network having its own advantages and disadvantages. In practical applications, the choice of which citation network method to use for predictive analysis depends on the overall prediction method system and the specific functions of different citation analysis networks.

3.3 Knowledge Unit-Based Domain Frontier Identification Methods

Knowledge unit-based methods approach research frontier identification from the perspective of subject terms, primarily from two angles: (1) examining term frequency change rates to identify emergent terms, especially those transitioning from low to high frequency; and (2) constructing keyword co-occurrence networks and mining research frontiers through co-word knowledge network mapping. Specific interpretations are shown in Table 2 .

Term frequency analysis is the most convenient quantitative method, identifying and monitoring frontier topics by statistically analyzing keyword frequencies or change rates. Yang Xuanhui et al., when analyzing emerging trends in "deep reading," categorized mutation terms into four types: mature, declining, emergent, and developing, arguing that emergent and developing mutation terms can reveal emerging trends in a field [32]. Most research frontier identification studies focus on emergent terms. However, using term frequency analysis or emergent term monitoring alone breaks semantic connections between documents and subject terms, lacks coherence, and makes it difficult to analyze the knowledge structure of research frontiers, especially when revealing domain knowledge structures.

Co-word analysis compensates for the isolation of knowledge units in term frequency analysis by effectively connecting terms and revealing structural changes in knowledge to some extent. Moreover, compared with citation analysis, co-word network structure not only reveals entity relationship characteristics within scientific knowledge systems at the micro level but also reflects growth patterns of scientific concepts and propositions through its evolution [38]. Compared with term frequency analysis, co-word analysis constructs relationships between terms, better mining structural relationships among knowledge units and handling larger data volumes. The key challenge lies in identifying important nodes and extracting valuable knowledge structures and communities in co-word networks. Compared with citation networks, co-word networks belong to cognitive-level knowledge networks that interpret knowledge structures from the perspective of micro knowledge units, with more refined and accurate knowledge units.

3.4 Time Series-Based Domain Frontier Identification Methods

3.4.1 Path-Based Research Frontier Identification Technological innovation does not emerge from nothing but develops through cross-fusion based on existing research. Therefore, analyzing technological innovation paths can help better predict scientific and technological development directions. A technological innovation path connects innovations chronologically, outlining the evolution of a technological innovation from its first appearance to continuous development [39]. Path evolution trend analysis has long been a research focus in intelligence analysis, dating back to 1965 when "father of scientometrics" D. J. Price identified "innovation knowledge points" in scientific development paths when defining research fronts [1], demonstrating the importance of citation link-based technological innovation path analysis for frontier identification and prediction.

Combining citation links with visualization maps to construct technological innovation paths has become a common method for identifying and predicting scientific frontiers. E. Garfield generated a historical evolution map of a knowledge domain using direct citation networks and predicted its development trends [40]. N. P. Hummon et al. first proposed a method focusing on connections between nodes in citation networks—the main path analysis method—based on their summary of citation analysis methods in 1989 [41]. Main path analysis is significant for identifying key documents and extracting mainstream research clues [42]. With rapid text mining development, analyzing technological innovation paths based on subject terms has become another major method for predicting scientific frontiers. A. Kontostathis et al. proposed a text mining-based automatic detection method called ETD (Emerging Trend Detection), which first represents topics using time-characteristic-associated features, then uses text mining for topic extraction, associates topics over time using evaluation criteria, constructs topic evolution paths, and judges trends to predict emerging trends (scientific frontiers) [43]. This knowledge evolution path approach is also called "data stream analysis." Zhu Na identified scientific and technological innovation topics based on the LDA model, constructed knowledge evolution paths by refining stages according to research life cycles, and demonstrated through empirical research in the 3D printing field that time series-based knowledge evolution paths can effectively detect scientific frontier topics [39]. Liang Li et al. used data stream analysis to interpret disciplinary topic evolution processes [44].

3.4.2 Timeline-Based Research Frontier Identification Timeline-based methods can intuitively display the dynamic changes of research frontiers. Timeline technology uses the X-axis for time and Y-axis for other features, establishing a research topic function with time as a variable and reflecting it in maps to clarify the changing 脉络 of scientific research topics [45]. Timeline analysis involves dividing the entire analysis object's timeline to explore the evolution of research frontiers throughout their life cycle. When examining how research objects change over time, this method is applied. For example, when Gartner predicts breakthrough innovation technologies, it divides new technology life cycles into five stages: embryonic, introduction, growth, maturity, and decline [17].

Currently, this method is mostly used as a supplement after research frontier identification—that is, after identifying topics using common methods, timeline maps present the evolution of research fields over time [46]. Some studies directly combine timeline map evolution analysis with time-zone view analysis to speculate on future research trends [47-48]. Zhang Ting improved upon traditional time series research frontier analysis by applying timeline and topographic visualization methods, extending from two-dimensional timeline maps to three-dimensional dynamic topographic maps for better visualization effects [49]. With information technology development, researchers now comprehensively use visualization technology to conduct in-depth textual content analysis of topic evolution, tracking complex evolutionary processes such as splitting and merging of frontier topics over time, which can compensate for limitations of citation and content analysis methods [50].

3.5 Multi-Source Data-Based Domain Frontier Identification Methods

In quantitative analysis methods for identifying research frontiers, data source selection is fundamental. The reliability and sufficiency of data sources significantly affect identification effectiveness. Single data sources usually have limited information, making multi-source data integration important in domain frontier identification. In terms of data types, Wang Xianwen et al. divided scientometric research data systems into four major objects: publication data, citation data, usage data, and altmetric data [51]. Liu Ziqiang categorized research frontier identification data sources into: scientific planning data, funding project data, patent data, and paper data [50]. Most current research concentrates on analyzing publication and citation data from papers, though some studies combine these data types using various methods for frontier identification. For example, I. Park et al. identified energy field research frontiers using both patent and academic literature as core documents [26]. D. Zhao and A. Strotmann compared author co-citation analysis results from XML domain journal papers and web papers, finding that web papers could reveal domain evolution trends earlier [27].

Some researchers build domain frontier prediction methods from knowledge association relationships between different data sources. Bai Rujiang used scientific planning texts and project data to identify frontiers through a scientific research frontier identification model based on semantic computation [52]. Sun Zhen integrated multiple data types including publication citation data, download usage data, and altmetric data, combining citation analysis, term frequency analysis, and co-word analysis to construct an integrated research frontier identification model [53]. Xu Xiaoyang et al. combined papers and patents to establish an integrated basic research and applied research frontier identification method [54]. Zhang Jinzhu integrated patent and scientific paper data to study breakthrough innovation identification from the perspective of basic research influencing technological innovation, using patent citations of scientific papers as the link [17].

However, these data source integrations remain at the shallow level of data source fusion. Multi-source data fusion can be divided into three types according to fusion stage: early-stage data source and data type fusion, mid-stage data relationship fusion, and late-stage clustering fusion [55]. Current multi-source data fusion mostly remains at early-stage data type fusion and mid-stage data relationship fusion, with late-stage clustering fusion not yet widely applied.

3.6 Multi-Dimensional Indicator-Based Domain Frontier Identification Methods

Due to the one-sidedness of single indicators in domain frontier identification, researchers have attempted to use multi-dimensional indicators. These can be divided into two types: (1) multiple indicators due to diverse analysis data sources, and (2) multi-dimensional measurement indicators built from research frontier characteristics. The latter may be based on single or multiple data sources. Table 3 lists representative multi-dimensional indicator methods for domain frontier identification.

Currently, this method is mostly applied as a supplement after research frontier identification—that is, after topic identification using common methods, multi-dimensional indicators are used for secondary evaluation and selection. For example, Zheng Yanning et al. constructed a comprehensive evaluation index system for research frontier measurement from four aspects: topic age, timeliness index, inheritance index, and breakthrough index [9]. Zhang Lihua constructed a research frontier identification index system from three dimensions: interdisciplinary index, topic age, and topic attention [57]. Fan Shaoping built a risk index, timeliness index, and applicability index for research frontiers in the context of scientific fund application evaluation [56]. Liu Ziqiang constructed a comprehensive frontier index from four sub-indices: topic attention, topic evolution rate, topic evolution intensity rate, and frontier feature evolution index [50]. Zhang Jinzhu constructed a CD index and CDt index from the perspective of basic research's impact on technological innovation, capturing the influence of inventions on existing technology [17]. Feng Jia and Zhang Yunqiu integrated indicators based on the LDA model and ontology, calculating the ratio of a topic's total weight in all scientific literature to total literature volume as topic novelty [59].

3.7 Transformative Research Frontier Identification Methods

3.7.1 Identifying Research Frontiers by Measuring Scientific Structure Changes C. Chen et al. proposed a research framework for scientific knowledge uncertainty, using scientific mapping technology to study evolutionary processes in scientific development and identify milestones, key paths, turning points, and boundary ranges in scientific knowledge [60]. When identifying knowledge breakthrough points based on Kuhn's paradigm theory, they argued that scientific revolutions are important components of science and used progressive visualization methods to improve three aspects: clarity of individual networks, highlighting transitions in adjacent networks, and identifying potential nodes. For potential node identification, they constructed landmark nodes, hub nodes, and pivot nodes [61].

K. W. Boyack and R. Klavans compared four citation methods for biomedical research frontier identification accuracy: co-citation analysis, bibliographic coupling, direct citation, and hybrid coupling-text methods. Results showed: (1) computational costs of the first three methods were similar, while the hybrid method was more expensive; (2) co-citation analysis had the largest clustering coverage, followed by hybrid and coupling methods; (3) among pure citation methods, bibliographic coupling had the best coherence and clustering, making it the most accurate. However, the hybrid method improved upon bibliographic coupling in both aspects, making it the best method [21]. N. Shibata et al. compared co-citation networks, bibliographic coupling, and direct citation for monitoring research frontiers, finding direct citation performed best while co-citation performed worst [71], differing from Boyack and Klavans' results. Bai Rujiang et al. compared citation analysis and keyword analysis, noting three problems: citation analysis has time lag; keyword analysis lacks semantic support; and data sources cannot be effectively integrated [72].

3.7.2 Identifying Research Frontiers through Weak Relationship Analysis Weak relationships (or "weak ties") are early characteristics of interdisciplinary cross-fertilization and important entry points for predicting breakthrough innovations. The term originated in sociology, contrasting with strong relationships. In network structures, weak ties refer to relationships whose strength is below a threshold. M. Granovetter argued that strong ties maintain intra-organizational relationships, while weak ties serve as information transmission bonds between groups and organizations, facilitating information flow between different groups, disseminating information people would otherwise be unlikely to see, and making larger network structures more cohesive [62]. Weak ties exist not only between human groups but also between topic groups. Research frontiers are more forward-looking than other topics, making them likely to be overlooked in knowledge development 脉络. From this perspective, weak ties play important roles in knowledge network structure formation and are easily overlooked, representing content requiring attention in research frontier identification.

Zhang Yingjie detected weak tie data evolution in networks from both whole-network and individual-network perspectives to identify potential development frontiers [63]. L. Wei et al. effectively identified intelligence science interdisciplinary frontiers by combining weak tie analysis [64]. Currently, weak tie analysis receives limited attention in transformative research frontier identification but should be strengthened for effective information mining of knowledge weak ties.

3.7.3 Identifying Research Frontiers through Sleeping Beauty Literature Transformative research, by overturning existing research paradigms, creates large psychological distance within the scientific community, leading to underestimation of its knowledge value and thus resistance [65-66]. Both science and technology exhibit delayed recognition phenomena—"sleeping beauty literature" in science and "sleeping patents" in technology [67]. The essence of sleeping beauty literature and delayed recognition phenomena is the 超前性 or transformative nature of scientific research. Transformative research frontiers belong to non-linear development frontiers, similar to the pre-explosion germination state of sleeping beauty literature with citation curves showing no early citations followed by later high citations [68], indicating sleeping beauty literature is a major source for transformative research frontier identification. W. Glänzel et al. defined sleeping beauty literature as papers where over 80% are first cited within 3 years and over 90% within 5 years [69]. J. Li et al. applied the Gini coefficient from economics to develop the G5 index for sleeping beauty identification [70].

3.9 Main Problems in Domain Frontier Identification Methods

3.9.1 Lack of Deep Association Analysis of Frontier Topics in Innovation Paths The difficulty in research frontier identification lies in measuring knowledge novelty, involving knowledge states before and after scientific discoveries and measuring their novelty and unpredictability. To discover emerging disciplinary trends, we must first understand the evolution processes, patterns, and trends of research topics in disciplinary fields. Mainstream research frontier identification methods mostly remain at the topic identification stage, with insufficient research on topic evolution processes and dynamic change patterns. Although some studies have used time series and community evolution algorithms for topic evolution analysis, they mostly remain at descriptive pattern summarization without further interpretation of topic emergence and disappearance reasons.

3.9.2 Data Singleness and Shallow Multi-Source Data Fusion Most frontier identification research analyzes citations or keywords from scientific papers and patents, with limited data sources. Patent data, funding data, experimental data, research papers, and analysis reports all reflect current scientific development status and contain future trends, but remain underutilized. Common data sources also suffer from time lag and insufficient dynamism, with each having its own advantages and disadvantages. While some studies attempt multi-source data fusion, most remain at early-stage fusion. With rich entity types and increasingly diversified data relationships under multi-source data, current research frontier identification rarely considers multi-relationship fusion and has not deeply processed and integrated textual content entity relationships, making it difficult to fully explain entity associations.

3.9.3 Insufficient Attention to Transformative Research Frontiers Current domain frontier identification methods focus more on incremental innovation frontiers, with limited methods for transformative innovation frontier identification. Incremental innovation research usually follows stable development paths, while transformative innovation development patterns are harder to obtain, full of uncertainty and non-linear characteristics. Many scientific breakthroughs and creative discoveries show no early signs for detection. The question of early signs and how to measure transformative potential at project initiation stages is extremely challenging. Transformative research frontier identification requires different methods from incremental frontier identification, meaning unconventional methods are needed to identify these uncertain scientific topics.

4. Future Development Trends of Domain Frontier Identification Methods

4.1 Enhancing Research on Frontier Topic Associations in Topic Paths

A feasible approach for frontier topic identification is using paths to represent the process of technological innovation, placing topics within their generation and development paths to display emergence, development, evolution, fusion, and extinction of innovative themes. Throughout topic paths, we can summarize interaction patterns and development models between topics to identify frontier themes. This process involves the key issue of interdisciplinary topic identification, requiring discussion of interdisciplinary dynamics mechanisms and analysis of growth patterns of interdisciplinary topics to extract topic features and set corresponding measurement indicators.

Furthermore, semantic association-based interpretation is needed in topic paths. In topic evolution paths, one topic's development changes can cause changes in other related topics. We cannot view a topic's development from a single static perspective but should dynamically observe evolution patterns through topic associations. The deep-level association is semantic relationship. Identifying domain research frontiers through semantic association-based innovation evolution path analysis can further identify frontiers and judge innovation development trends. This requires strengthening interpretation of associations between topics.

Currently, entity semantic relationship identification methods based on machine learning have developed significantly, and introducing them into semantic association identification under technological path development is meaningful. When judging relationships between research topics, we must combine specific disciplinary backgrounds, deeply analyzing multi-dimensional information including disciplinary positioning, academic goals, development history, research methods, and researchers to construct comprehensive semantic association networks for accurate frontier prediction.

4.2 Enhancing Multi-Source Data and Multi-Relationship Fusion Analysis

Research frontier identification requires comprehensive judgment by integrating various domain information. Traditional research used papers and patents due to information carrier and technology limitations. In contemporary society, rapid information technology development has spawned multiple information carriers that affect domain development and reflect its prospects. However, multi-source data collection must distinguish useful from useless information.

Future frontier identification methods need enhanced mid- and late-stage multi-source data fusion to mine more topic association information. Information relationship structures map scientific knowledge structures, and only by clarifying scientific knowledge structures can we make reasonable frontier judgments. Citation relationships, coupling relationships, co-occurrence relationships, and other data relationships all reflect knowledge development 脉络. The breakthrough point lies in mining knowledge development trends from these multiple data relationships. The difficulty lies in integrating different data sources, different data relationships, and clustering after multi-relationship fusion. With continuous improvement of big data analysis methods, efficient and in-depth analysis of multi-source dynamic data is expected to advance frontier identification timeliness and accuracy to new levels.

4.3 Capturing Uncertainty in Transformative Research

Most existing research frontier detection methods detect frontiers during the progress or maturity stages when many researchers have already entered the field, making detection results less forward-looking [57]. Transformative innovation requires long accumulation, and earlier identification stages are better for technology implementation and industrial layout. However, earlier stages have less information and greater difficulty. Many scientific breakthroughs and creative discoveries show no early signs for utilization. Deep mining of transformative innovation's early characteristics is prerequisite for accurate identification. The innovation process is accompanied by uncertainty, and transformative research, as a non-linear innovation, is particularly affected. Grasping this far-reaching uncertainty is crucial.

We need to deeply mine early characteristics of transformative innovation and design corresponding screening indicators for early signs. Future research frontier identification from the perspective of scientific uncertainty represents a more micro-level understanding of transformative research features. When capturing uncertainty information, weak relationships deserve attention. Frontier topics often exist as weak interdisciplinary relationships before or during germination, making weak ties an entry point for exploring transformative frontier early characteristics. They represent data objects different from high-frequency terms and strong relationships, and ignoring them may cause important information loss. Future research could attempt to capture uncertainty information like weak relationships during transformative research processes to help identify research frontiers.

References

[1] PRICE D J. Networks of scientific papers[J]. Science, 1965, 149(3683): 510-515. [2] ZHONG Zhen. Distinguishing Research Front from Research Frontier from citation structure differences between highly cited and zero cited papers[J]. Library and Information Service, 2015, 59(8): 87-96. [3] SMALL H, GRIFFITH B C. The structure of scientific literatures I: identifying and graphing specialties[J]. Science studies, 1974, 4(1): 17-40. [4] PERSSON O. The intellectual base and research fronts of JASIS 1986-1990[J]. Journal of the American Society for Information Science, 1994, 45(1): 31-38. [5] MORRIS S A, YEN G, WU Z, et al. Timeline visualization of research fronts[J]. Journal of the Association for Information Science & Technology, 2003, 54(5): 413-422. [6] WANG Lixue, LENG Fuhai. Discussion on research frontier and its bibliometric identification methods[J]. Information Studies: Theory & Application, 2010, 33(3): 54-58. [7] CHEN Chaomei, CHEN Yue, HOU Jianhua, et al. CiteSpace II: detecting and visualizing emerging trends and transient patterns in scientific literature[J]. Journal of the China Society for Scientific and Technical Information, 2009, 28(3): 401-421. [8] UPHAM S P, SMALL H. Emerging research fronts in science and technology: patterns of new knowledge development[J]. Scientometrics, 2010, 83(1): 15-38. [9] ZHENG Yanning, XU Xiaoyang, LIU Zhihui. Research frontier identification method based on keyword co-occurrence[J]. Information Science, 2018, 36(6): 142-148. [10] FENG Jia. Research on research frontier identification and analysis methods[D]. Changchun: Jilin University, 2017. [11] ISHIKAWA S, ARAKAWA Y, TAGASHIRA S, et

al. Hot topic detection in local areas using Twitter and Wikipedia[C]//ARCS workshops. Muenchen: IEEE, 2012: 1-5. [12] KIM H G, LEE S, KYEONG S. Discovering hot topics using Twitter streaming data: social topic detection and geographic clustering[C]//IEEE/ACM international conference on advances in social networks analysis and mining. Niagara Falls: IEEE, 2014: 1215-1220. [13] SCHOEMAKER. Wharton on managing emerging technologies[M]. SHI Ying, et al., trans. Beijing: Huaxia Publishing House, 2002. [14] GUO Hanning. Exploration of emerging research field identification from multiple scientific indicator perspectives[D]. Dalian: Dalian University of Technology, 2013. [15] ROTOLO D, HICKS D, MARTIN B R. What is an emerging technology?[J]. Research policy, 2015, 44(10): 1827-1843. [16] KUHN T. The structure of scientific revolution[M]. Chicago: University of Chicago Press, 1999. [17] ZHANG Jinzhu. Identifying breakthrough innovation using cited scientific knowledge mutation[M]. Beijing: Science Press, 2017. [18] CHRISTENSEN C M. The innovator's dilemma[M]. Boston: Harvard Business School Press, 1997. [19] SUN Qigui, DENG Xin, XU Fei. Concept definition and model construction of disruptive innovation[J]. Science and Technology Management Research, 2006, 26(8): 175-178. [20] BENGISU M, NEKHIILI R. Forecasting emerging technologies with the aid of science and technology databases[J]. Technological forecasting & social change, 2006, 73(7): 835-844. [21] BOYACK K W, KLAVANS R. Co-citation analysis, bibliographic coupling, and direct citation: which citation approach represents the research front most accurately?[J]. Journal of the American Society for Information Science & Technology, 2010, 61(12): 2389-2404. [22] XU Zhenliang, GUO Xiaochuan. Scientometric analysis of international technological innovation research frontiers[J]. Library and Information Service, 2011, 55(8): 49-53. [23] DENG Mingjun, LUO Wenbing. Visualization of international carbon footprint research frontier evolution[J]. Journal of Hunan University of Science & Technology (Social Science Edition), 2013, 16(3): 94-98. [24] PAN Li, HOU Jianhua. Hot topics and research frontiers in international higher education research: analysis based on co-citation networks of 8 SSCI higher education journals from 2000-2011[J]. Educational Research, 2012(6): 136-143. [25] HOU Jianhua, LI Lianji, YANG Xiucai. Identification and visualization of big data research new dynamics based on citation network structure transformation[J]. Journal of the China Society for Scientific and Technical Information, 2009, 28(3): 401-409. [26] PARK I, LEE K, YOON B. Exploring promising research frontiers based on knowledge maps in the solar cell technology field[J]. Sustainability, 2015, 7(10): 13660. [27] ZHAO D, STROTMANN A. Can citation analysis of web publications better detect research fronts?[J]. Journal of the American Society for Information Science & Technology, 2007, 58(9): 1285-1302. [28] HUANG M H, CHANG C P. A comparative study on detecting research fronts in the organic light-emitting diode (OLED) field using bibliographic coupling and co-citation[J]. Scientometrics, 2015, 102(3): 2041-2057. [29] SUN Taotao, YANG Liying. Research frontier monitoring method based on literature coupling and its application[C]//Proceedings of 2010 academic annual conference of Medical Information Institute/Library, Chinese Academy of Medical Sciences/Peking Union Medical College. Beijing: Chinese Academy of Medical Sciences, 2011: 564-568. [30] CHEN Yuan, WANG Feifei. Author bibliographic coupling analysis in domestic informatics based on CSSCI[J]. Information and Documentation Services, 2011(5): 6-12. [31] FAJARDO-ORTIZ D, LOPEZ-CERVANTES M, DURAN L, et al. The emergence and evolution of the research fronts in HIV/AIDS research[J]. Plos One, 2017, 12(5): e0178293. [32] YANG Xuanhui, DU Xinyu, CAI Zhiqiang. Emerging trend analysis of deep reading based on mutation detection and co-word analysis[J]. Library Development, 2018(5): 48-53. [33] LIU G Y, JIANG R, JIN Y. Sciatic nerve injury repair: a visualized analysis of research fronts and development trends[J]. Neural regeneration research, 2014, 9(18): 1716-1722. [34] CHEN L. Information visualization analysis of hot research topics and research fronts in the field of public information resources management (PIRM) based on Web of Science[C]//2009 international conference on public administration. Chengdu: UESTC Press, 2009: 944-948. [35] HOU Jianhua, LIU Zeyuan. Visual analysis of nanotechnology research frontiers and their evolution[J]. Science of Science and Management of S.& T., 2009, 30(5): 23-30. [36] JIANG Bingquan, XU Zhenliang. International ecological civilization research frontiers based on knowledge mapping: co-word analysis perspective[J]. Science & Technology and Economy, 2009, 22(5): 55-58. [37] XU Zhenliang, ZHAO Yuna, ZHOU Wenxia. Knowledge mapping analysis of international ecological city research frontiers: based on co-word network analysis[J]. Urban Development Studies, 2013, 20(2): 115-124. [38] WANG Xiaoguang. Formation and evolution of scientific knowledge networks (I): proposal of co-word network method[J]. Journal of the China Society for Scientific and Technical Information, 2009, 28(4): 599-605. [39] ZHU Na. Research on semantic-enhanced scientific and technological innovation path identification[D]. Zibo: Shandong University of Technology, 2015. [40] GARFIELD E. Historiographic mapping of knowledge domain literatures[J]. Journal of information science, 2004, 30(2): 119-145. [41] HUMMON N P, DEREIAN P. Connectivity in a citation network: the development of DNA theory[J]. Social networks, 1989, 11(1): 39-63. [42] HAN Yi, JIN Bihui. Formation and evolution of main path analysis method in citation networks[C]//Proceedings of the 6th China Science and Technology Policy and Management Academic Annual Conference. Beijing: China Association for Science of Science and S&T Policy, 2010: 155-164. [43] KONTOSTATHIS A, GALITSKY L M, POTTENGER W M, et al. A survey of emerging trend detection in textual data mining[M]. New York: Springer, 2004. [44] LIANG Li, XIE Fengjie, CHI Lixu, et al. Empirical analysis of hot and frontier topic research methods in specific disciplines[J]. Library Journal, 2018(1): 19-26. [45] ZHANG Ting. Visual analysis of science communication research[D]. Dalian: Dalian University of Technology, 2009. [46] YU Xiaoyan, LU Xiaozhong. A glimpse into the development 脉络 of Pan Maoyuan's educational research: knowledge mapping analysis based on academic papers[J]. Shandong Higher Education, 2015(10): 83-96. [47] WANG Dan, GONG Jingjing. Visual research on domestic construction safety based on knowledge mapping[J]. Journal of Engineering Management, 2016, 30(6): 43-48. [48] LAN Guoshuai. Construction of knowledge mapping for international educational technology research frontiers: visual analysis based on 18 SSCI journals from 1960-2016[J]. Modern Distance Education, 2017(3): 57-76. [49] ZHANG Ting. Timeline and topographic visualization maps: analysis of science communication research frontier evolution trends[J]. Journal of the China Society for Scientific and Technical Information, 2009, 28(6): 923-928. [50] LIU Ziqiang. Research on research frontier topic identification and evolution patterns based on NSF data[D]. Zibo: Shandong University of Technology, 2017. [51] WANG Xianwen, FANG Zhichao, HU Zhigang. Scientificometric analysis of scientific papers: integration framework of data, methods and uses[J]. Library and Information Service, 2015, 59(16): 74-84. [52] BAI Rujiang. Research on scientific research frontier identification based on semantic computation[D]. Beijing: University of Chinese Academy of Sciences, 2015. [53] SUN Zhen. Research on integrated identification model of research frontier based on multi-source data of scientific papers[J]. Journal of Intelligence, 2016, 35(8): 95-100. [54] XU Xiaoyang, ZHENG Yanning, LIU Zhihui. Research on research frontier identification method combining papers and patents[J]. Library and Information Service, 2016, 60(24): 97-106. [55] XU Haiyun, DONG Kun, WEI Ling, et al. Review of multi-source data fusion methods in scientometrics[J]. Journal of the China Society for Scientific and Technical Information, 2018, 37(3): 318-328. [56] FAN Shaoping. Research frontier indicator system based on bibliometrics and its reference significance for NSFC review[J]. China Science Fund, 2016(2): 171-176. [57] ZHANG Lihua. Research frontier detection and its evolution analysis methods and empirical research[D]. Beijing: University of Chinese Academy of Sciences, 2015. [58] FUNK R J, OWEN-SMITH J. A dynamic network measure of technological change[J]. Management science, 2017, 63(3): 791-817. [59] FENG Jia, ZHANG Yunqiu. Scientific frontier identification and analysis method based on LDA and ontology[J]. Information Studies: Theory & Application, 2017, 40(8): 49-54. [60] CHEN C, SONG M. Representing scientific knowledge: the role of uncertainty[M]. Cham: Springer International Publishing AG, 2018. [61] CHEN C. Searching for intellectual turning points: progressive knowledge domain visualization[J]. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101(S1): 5303-5310. [62] GRANOVETTER M. The strength of weak ties: a network theory revisited[J]. Sociological theory, 1983, 1: 201-233. [63] ZHANG Yingjie. Research on frontier measurement detection methods in science and technology fields[D]. Beijing: National Science Library, Chinese Academy of Sciences, 2011. [64] WEI L, XU H, WANG Z, et al. Topic detection based on weak tie analysis: a case study of LIS research[J]. Journal of data and information science, 2016, 1(4): 81-101. [65] CAMPANARIO J M. Rejecting and resisting Nobel class discoveries: accounts by Nobel Laureates[J]. Scientometrics, 2009, 81(2): 549-565. [66] FANG H. An explanation of resisted discoveries based on construal level theory[J]. Science & engineering ethics, 2015, 21(1): 41-50. [67] PALOMERAS N. Sleeping patents: any reason to wake up?[J]. IESE research papers, 2003, 20(35): D506. [68] DU Jian. Research on identification methods and awakening mechanisms of "sleeping beauty" literature[D]. Nanjing: Nanjing University, 2017. [69] GLÄNZEL W, SCHLEMMER B, THIJS B. Better late than never? On the chance to become highly cited only beyond the standard bibliometric time horizon[J]. Scientometrics, 2003, 58(3): 571-586. [70] LI J, SHI D, ZHAO S X, et al. A study of the "heartbeat spectrum" for "sleeping beauties"[J]. Journal of informetrics, 2014, 8(3): 493-502. [71] SHIBATA N, KAJIKAWA Y, TAKEDA Y, et al. Comparative study on methods of detecting research fronts using different types of citation[J]. Journal of the Association for Information Science & Technology, 2009, 60(3): 571-580. [72] BAI Rujiang, LENG Fuhai, LIAO Junhua. Comparison of main methods for scientific research frontier detection and study on development trends[J]. Information Studies: Theory & Application, 2017(5): 33-38.

Author Contributions

XU Haiyun: Supervised paper organization, writing, and revision;
LUO Rui: Responsible for data collection, paper writing, and revision;
DONG Kun: Supervised paper writing.

Note: Figure translations are in progress. See original paper for figures.