# Research Frontier Topic Detection and Feature Analysis Based on Global Microscopic Models - Postprint

**Authors:** Cui Yuhong, Wang Sa, Gao Xiaowei, Yang Hui, Cao Xuewei

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

## Abstract

[Purpose/Significance] Accurate identification of research fronts constitutes a strategic imperative at the national macro-level. Bibliometrics, as a quantitative research methodology, finds extensive application in scientific topic detection and research front identification.

[Method/Process] This study systematically reviews the developmental trajectory and methodological models of research front topic detection, introduces the concept of the all-domain micro-model, and provides a detailed exposition of the topic creation methodology employed by the SciVal module, encompassing direct citation-based document clustering, keyword-driven topic nomenclature, and the topic prominence algorithm for research front selection. Furthermore, an empirical analysis is conducted on the characteristics of 96,000 topics generated by SciVal and the top 1% of selected research front topics.

[Results/Conclusion] The all-domain micro-model enables simultaneous identification of all topics across the entire scientific domain in a single operation; however, different disciplines manifest distinct patterns in research front representation, and topic prominence should not be simplistically equated with importance. A moderate correlation exists between topic paper volume and topic ranking. Automatically extracted keyword terminology names and describes topics from both disciplinary domain perspectives and uniqueness dimensions. Evolutionary trend analysis of graphene-related front topics can be utilized to identify critical nodes and emerging themes.

## Full Text

# Detecting and Characterizing Research Fronts Topics Based on Global-Micro Model

**Cui Yuhong**[1], **Wang Sa**[1], **Gao Xiaowei**[2], **Yang Hui**[3], **Cao Xuewei**[2]
[1]Beijing Institute of Technology Library, Beijing 100081 [2]National Academy of Innovation Strategy, Beijing 100012 [3]Relx Group Shanghai District, Shanghai 200040

### Abstract

[**Purpose/Significance**] Accurate identification of research fronts represents a strategic demand at the national macro level. Bibliometrics, as a quantitative research method, has been widely applied in scientific topic detection and research front identification. [**Method/Process**] This study reviews the developmental trajectory and methodological models of research front topic detection, introduces the concept of the global-micro model, and elaborates on the topic creation methodology adopted by the SciVal module, including direct citation-based document clustering, keyword-based topic naming, and the topic prominence algorithm for research front selection. An empirical analysis is conducted on the characteristics of 96,000 topics created by SciVal and the top 1% selected research front topics. [**Result/Conclusion**] The global-micro model can simultaneously identify all topics across the entire scientific domain in a single process, though different disciplines exhibit varying patterns in research fronts. Topic prominence should not be simplistically equated with importance. A moderate correlation exists between the number of papers in a topic and its ranking. Automatically extracted keyword terms name and describe topics at both the disciplinary level and in terms of uniqueness. Evolutionary trend analysis of graphene-related front topics can be used to identify key nodes and emerging topics.

**Keywords:** topic detection; research fronts; global-micro model; SciVal; topic prominence

**Classification Number:** G301

---

## 1. Review of Research Front Topic Detection

The research front of a discipline represents the most significant critical issues that capture its development trends and constrain its current progress. From a macro-level strategic perspective, accurate judgment of research fronts influences policy orientation for national scientific, technological, and innovative development. Since 2006, Japan, the EU, the United States, and Canada have prioritized research fronts as primary research topics to maintain their positions as global scientific and technological leaders, establishing research institutions and special funds oriented toward innovative frontiers to support cutting-edge research [1]. To accelerate implementation of the national innovation-driven development strategy, China' s State Council issued the "National Innovation-Driven Development Strategy Outline" in May 2016, which further emphasized tasks such as "strengthening basic frontier and high-tech research oriented toward national strategic needs"and "enhancing original innovation facing scientific frontiers," highlighting the importance and urgency of scientific frontier research from a national strategic perspective.

Bibliometrics, as a quantitative research method, has been extensively applied in scientific topic detection and research front identification. However, topic detection and front identification face numerous challenges: How many topics exist in science? Should all topics be identified simultaneously across the entire scientific domain, or can they be identified only in smaller domains according to specific needs? Which method provides the most accurate topic description? What characteristics do research fronts possess? How can research fronts be effectively interpreted and visualized? In recent years, the global-micro model based on direct citation relationships across the entire scientific domain has been introduced and applied to topic creation and frontier identification. In October 2017, SciVal adopted this model and launched its topic prominence module, offering a novel solution to these questions.

Scientific topic detection and research front identification have long attracted scientists' interest. As early as 1955, Garfield noted in the renowned *Science Citation Index* that citation linkage analysis of scientific literature could track emerging ideas and identify emerging scientific fields [2]. In 1965, Price utilized extensive citation data to define what he described as "research fronts" —pioneering research conducted by distinguished scientists at the cutting edge—and measured research fronts through publication density and activity across different periods [3]. In 1970, sociologist Kuhn explicitly suggested using Garfield' s citation data to identify research communities and depict paradigms of scientific revolutions [4].

Building upon these theoretical foundations, bibliometricians across different eras have employed three types of citation relationships—direct citation (DC), co-citation (CC), and bibliographic coupling (BC)—to conduct exploratory and empirical research based on different data sources, literature scales, and clustering algorithms, reflecting the evolution of scientific mapping theories, data visualization techniques, and computer information processing capabilities (see Table 1 ).

As the founder of ISI, although Garfield proposed in 1964 that direct citation analysis could be used to construct historical maps for discovering scientific breakthroughs [5], it was not widely adopted due to the substantial computational demands, which exceeded early processing capabilities. In 1965, Kessler analyzed the bibliographic coupling relationships of 334 papers in *Physical Review*, representing undoubtedly the largest-scale literature clustering study at that time [6].

The first widely adopted research front detection model was proposed in 1974 by ISI' s chief scientist Small and Griffith. Based on co-citation analysis of 1,832 highly cited papers from ISI and using single-link clustering algorithms, they revealed the complete structural map of science [7], which was applied to tracking and predicting emerging scientific fields [8] and continues to be used today. From 1974 to 2010, from Small to Klavans and Boyack, despite the growth in literature scale from thousands to millions [9] and the evolution of clustering algorithms from single-link to VxOrd and DrL/OpenOrd [10], the combination of co-citation analysis and the ISI database remained largely unchanged. Most researchers created literature clusters on local datasets by limiting disciplines, journals, or terms to achieve specific research front topic detection and identification.

In 2001, ESI adopted co-citation analysis to cluster highly cited papers from Web of Science, identifying and dynamically generating nearly 10,000 research fronts across 22 disciplinary fields simultaneously. Domestic scholars have conducted empirical analyses based on ESI front topics, including mapping of nanoscale frontiers [11], temporal evolution of biological science fronts [12], and key research pathways in quantum discord [13].

The challenge in recent years has been how to precisely construct a more granular topic identification model framework across the entire scientific domain. Following the widespread application of co-citation analysis, Klavans and Boyack attempted combinations of bibliographic coupling and co-citation analysis with different clustering algorithms on million-scale literature from ISI [14] and Scopus [15] in 2009 and 2010, respectively. In 2011, they proposed the concept of the global-micro model and continuously refined it [16]. In 2012, Waltman and Van Eck from Leiden University introduced the first methodology for constructing a publication-level classification system for the entire scientific domain based on direct citation models [17]. They demonstrated that direct citation combined with the Smart Local Moving clustering algorithm could precisely classify tens of millions of ISI papers into distinct topics, with a simple, transparent method requiring modest computational resources. In 2014, Boyack and Klavans applied Waltman et al.' s method to over 20 million Scopus documents [18], subsequently comparing and evaluating the effectiveness of three citation types in constructing scientific maps of research topics. They found that direct citation could more accurately map knowledge classification systems at the micro-level research problem layer, better identify emerging interdisciplinary fields, and understand development trends and evolutionary dynamics across the entire scientific domain compared to bibliographic coupling or co-citation analysis [19].

Research front detection comprises two stages: topic creation and front selection. In October 2017, Elsevier' s SciVal adopted the global-micro model in its topic creation process, clustering 70 million papers and references across all scientific fields in Scopus from 1996 to 2016, identifying nearly 96,000 research topics. Research shows that the most common characteristics of research fronts are high attention and novelty [20]. For instance, ESI research fronts use high citations as the basis for calculating high attention, while the *2017 Research Fronts* report identifies the "youngest" research fronts by sorting core papers by publication year for in-depth analysis [21]. In contrast, SciVal calculates each topic' s prominence percentile based on citations, views, and journal quality indicators from the most recent two years of papers. The prominence percentile thus incorporates both high attention and novelty features. Therefore, utilizing SciVal' s topic prominence data to select the largest and hottest research fronts across all disciplines and verifying its effectiveness constitutes the primary objective of this study.

## 2. Model and Methods

This section introduces the global-micro conceptual model, direct citation topic clustering model, keyword-based topic naming method, and topic prominence calculation for selecting research fronts from all created topics.

**2.1 Global-Micro Conceptual Model** "Global" refers to constructing datasets from all literature in the entire database rather than subsets. In contrast, local models are completed on literature subsets, as researchers cannot access all data in the database and can only download partial data locally —for example, all literature under a specified physics classification, several journals in information science, or highly cited papers retrieved with the term "graphene." Research indicates that global models outperform local models in precision and recall, being more suitable for describing and discovering unexpected emergent topics [22]. Relevant scientific and technological fields may include discovery pathways, such as new physics findings originating from chemistry, computer science, or instrumental technology development. Therefore, it is necessary to generate a model covering as much scientific and technological literature as possible.

In hierarchical scientific classification systems, previous research primarily aggregated literature at the domain, discipline, or specialty level. Domains sit at the top of the tree-like classification, with quantities ranging from several to dozens, such as ESI's 22 disciplinary fields or Scopus's 27 disciplinary fields, each comprising hundreds of thousands of publications annually. Disciplines reside at the next level down, with annual literature quantities ranging from hundreds to thousands, typically based on journal clustering and often equivalent to the subject directories of Web of Science or Scopus. Specialty-level analyses are frequently based on literature samples generated from term search results.

Topics aggregate literature at the research problem level or micro-level, representing a group of articles sharing the same research foundation and sitting at the bottom of the scientific classification hierarchy. Research problems represent the detailed issues researchers actually work on. For example, while "dye-sensitized solar cells" is considered a specialty, "counter electrode materials for dye-sensitized solar cells" constitutes a research problem. Topics vary in size, with annual literature quantities ranging from a few to several hundred papers.

SciVal' s research topics created based on direct citation relationships persist permanently once generated. A small number of new topics emerge each year, with subsequent literature added to topics based on citation relationships. More recent literature indicates newer topics, while older topics may become dormant but do not disappear.

**2.2 Topic Clustering Model** Topic creation involves two steps: forming document clusters through direct citation and clustering these clusters into distinct topics. Direct citation cluster creation is conceptually and practically straightforward. Figure 1 [Figure 1: see original paper] provides a simplified schematic. When processing large-scale literature collections, to reduce computational resources, the number of related document pairs should be minimized. Therefore, undirected direct citation is adopted—that is, if document i cites j or j cites i, then Cij = 1; otherwise, Cij = 0.

After creating document clusters through direct citation links, the VOS method developed by Waltman and Van Eck is applied for clustering. The VOS algorithm uses modularity clustering variables to maximize the accuracy of inter-cluster similarity, weighting more similar clusters higher and less similar clusters lower [24]. To calculate paper relevance, each link is normalized by the number of references in the citing document, and all matrices' K50 (adjusted cosine) values are calculated. To reduce computational scale, the maximum number of links per article is set to 15 (highest K50 value). The paper set and filtered links are then input into the VOS clustering algorithm. The VOS algorithm implementation can be freely accessed.

**2.3 Topic Naming Method** Topics created through the above steps typically consist of dozens to thousands of documents and require computer-automated extraction of keywords or phrases for naming. The topic naming method integrates Elsevier' s Fingerprinting Technology (EFT) and special phrases through a three-step process: (1) Applying natural language processing techniques to mine title and abstract information from papers in the topic. (2) Matching a set of words against thesauri from all major disciplines to obtain conceptual terms. Elsevier integrates several general and specialized thesauri, such as the Medical Subject Headings (MeSH) and the Unified Astronomy Thesaurus (UAT). (3) Selecting unique keywords for each document based on inverse document frequency to reduce the weight of high-frequency words in the document set and increase the importance of rarely occurring words. Each keyword is assigned a relevance value between 0 and 1 based on its ratio to the highest-frequency term' s occurrence, with a relevance value of 1 indicating the most frequently occurring keyword.

In practice, the system automatically provides three terms to name each topic. The first two are generated using EFT, generally selecting high-frequency words to provide high-level descriptions of the topic at the research field or specialty level. The third selects special phrases about the topic to provide more specific descriptions at the research problem level. For example, a topic named "Graphene; Energy storage; Graphene fibers" involves the research direction of "graphene and energy storage," with specific research content being "graphene fibers," thus describable as "graphene fibers in energy storage."

**2.4 Topic Prominence Calculation** Topic prominence is an indicator measuring topic visibility and momentum [25], comprehensively considering three parameters: recent citation count, recent view count, and journal CiteScore. The prominence Pj for each topic j in year n is calculated as follows:

$$P_j = 0.495\frac{C_j - \text{mean}(C_j)}{\text{stdev}(C_j)} + 0.391\frac{V_j - \text{mean}(V_j)}{\text{stdev}(V_j)} + 0.1149\frac{CS_j - \text{mean}(CS_j)}{\text{stdev}(CS_j)}$$

where Cj represents the citation count of papers published in topic j in years n and n-1, Vj represents the Scopus view count of papers published in topic j in years n and n-1, and CSj represents the average CiteScore of papers published in topic j in year n. The raw data undergoes logarithmic transformation:

$$C_j = \ln(C_j + 1), \quad V_j = \ln(V_j + 1), \quad CS_j = \ln(CS_j + 1)$$

Prominence calculation uses standardized scores to eliminate dimensional differences among the three indicators, then weights and sums the deviation of each topic' s recent two-year paper citation count, view count, and journal evaluation index from their means. Therefore, higher prominence values indicate that more researchers are paying attention to the topic and that its growth momentum is stronger. In practice, SciVal calculates each topic' s percentile indicator based on prominence values.

## 3. Results Analysis

**3.1 Selection of Research Fronts** In the new version of SciVal launched by Elsevier in October 2017, competitiveness analysis was replaced by topic prominence. Based on approximately 70 million documents and 1 billion citation links from the Scopus database, the aforementioned model and methods were applied for topic clustering, yielding nearly 96,000 topics across the entire scientific domain, with prominence percentiles assigned to each topic. Table 2 lists the top 10 research front topics by prominence percentile.

Percentile indicators, as relative metrics, have been widely applied in excellence performance evaluation [26]. In practice, appropriate percentile thresholds are selected based on needs–for example, ESI highly cited papers and research fronts use the top 1% threshold by citation count within each discipline, the Ministry of Education' s discipline evaluation extends ESI highly cited papers to the top 3%, and the *2017 Research Fronts* report further extracts the top 10% most citation-influential research fronts from ESI [21].

This study aims to select the newest and hottest research questions across all disciplines from the 96,000 research topics, then categorize them into corresponding disciplinary fields for interpretation and analysis to showcase the latest advances in current scientific frontiers. Considering publication requirements, interpretation workload, and the need to cover different disciplines as comprehensively as possible, experiments determined that a topic prominence percentile threshold of 99% (i.e., the top 1% of topics as research fronts) was appropriate, yielding 963 research front topics covering 23 disciplinary fields except Arts and Humanities, Veterinary, Health Professions, and Multidisciplinary.

**3.2 Subject Distribution of Topics** According to the principle of assigning each topic to the discipline with the most papers, topics were assigned to Scopus' s 27 subject categories. Table 3 statistics show the number of all topics and research front topics across different disciplines, the percentage of research front topics among all topics, and the relative strength of research front topics. Relative strength is the ratio between the number of research fronts in a discipline and the maximum number in any discipline—for example, Medicine has the most research front topics (188), giving it a relative strength of 1; Chemistry ranks second with 171, giving it a relative strength of 0.91.

Significant variations exist among disciplines in terms of research topics and fronts. Figure 2 [Figure 2: see original paper] presents a disciplinary analysis matrix constructed from two indicators: relative strength and percentage of research front topics. The 27 disciplines can be divided into four clusters: (1) Medicine (marked O) has the highest number of both total topics and research front topics, with a research front percentage of 0.85%, slightly below the expected 1% value. (2) Chemistry and Materials Science (marked ×) have research front topic numbers second only to Medicine, with relative strengths of 0.91 and 0.67, respectively, but their research front percentages significantly exceed the expected 1% at 3.26% and 4.14%, respectively. (3) Seven disciplines including Energy, Chemical Engineering, Immunology and Microbiology, Environmental Science, Neuroscience, Biochemistry, Genetics and Molecular Biology, and Physics and Astronomy (marked +) have relatively few research front topics (relative strength below 0.4) but research front percentages exceeding the expected 1%. (4) Seventeen disciplines including Engineering and Computer Science (marked  ) have relatively low numbers and percentages of research front topics, with Arts and Humanities, Veterinary, Health Professions, and Multidisciplinary having zero research front topics.

**3.3 Relationship between Topic Size and Ranking** Statistics for 95,769 topics show paper counts (within a 5-year window from 2012-2016) ranging from a maximum of 4,574 to a minimum of 1, with a median of 56. In contrast, the

963 research front topics have a minimum of 122 papers and a median of 1,119. Figure 3 [Figure 3: see original paper] illustrates the relationship between paper count and topic ranking (with logarithmic axes), showing a moderate positive correlation ($R^2 = 0.692$, $p < 0.0001$)—topics with more papers tend to have higher prominence indicators and rank higher.

Special attention should be paid to outliers with high rankings but few papers. For example, topic T67927 has only 35 papers but ranks 984th (near the top 1% of research fronts), with keywords "diagnosis," "blood," and "computer microscopy." Citations primarily come from one highly cited article—the American Cancer Society' s 2016 Cancer Statistics report [27], which has been cited 4,242 times (as of October 30, 2017). Such papers are generally highly cited but cannot be considered core papers of the topic. In another example, topic T67378 has only 122 papers but ranks 440th (within the top 1% of research fronts), with keywords "disease," "health services," and "diabetes," featuring multiple highly cited papers that could be considered core papers of the topic. These examples indicate that highly cited topics require further detailed investigation.

**3.4 Keyword-Based Topic Naming** According to topic naming rules, the first two terms are keywords extracted from titles and abstracts to describe the research field or direction level, while the third term uses special phrases to describe the topic' s specificity. Statistical analysis of the first two keywords for the top 1% of research front topics shows frequent terms including tumor, photocatalysis, graphene, lithium-ion, apoptosis, carbon nanotubes, DNA, hydrates, proteins, solar cells, biofuels, catalysts, energy management, hydrogels, electrocatalysis, metagenomics, nanoparticles, and phosphates, each appearing more than 10 times. These terms primarily involve Medicine, Chemistry, Materials Science, Life Sciences, Energy, Environmental Science, Physics, and Engineering, consistent with the aforementioned disciplinary distribution of research fronts.

To validate the effectiveness of system-automatically extracted keyword terms in describing and naming topics, all topic keyword sets were queried using "Graphene" as the search term across all scientific fields, yielding 19 graphene-related research fronts within the top 1%, belonging to Chemistry, Materials Science, Physics, and Engineering. Based on these three keywords, topics were preliminarily named. After consulting graphene field experts, most keyword terms were found to describe topics effectively from both macro research directions and micro-level content uniqueness, helping professionals quickly understand and preliminarily judge a topic' s content. However, some keywords lack precision—for example, topic T31540 has three keywords: "electrolytic capacitors," "graphene," and "area specific capacitance." Experts noted that "area specific capacitance" is overly narrow and rarely used, requiring further interpretation of core papers to provide a more accurate description: "graphene supercapacitors."

**3.5 Comparison of Related Topic Development Trends: Graphene as an Example** Research fronts were formed by clustering literature from 1996-2016, with updated recent papers assigned to existing topics based on direct citation relationships. Therefore, analyzing the temporal distribution of papers across multiple related topics can reveal not only the developmental trajectory of a research problem but also key nodes and emerging trends within a research field or direction.

Figure 5 [Figure 5: see original paper] shows the evolution of paper publication numbers from 1996-2016 for the top six graphene research fronts. In 2004, researchers at the University of Manchester isolated single-layer graphene through a simple method, leading to a significant increase in graphene research papers from 2006 onward. The 2010 Nobel Prize in Physics greatly accelerated graphene research, making it one of the hottest current research fields. In contrast, the two research front topics of graphene oxide and graphene plasmonics have maintained around 600 papers annually in recent years, but graphene plasmonics shows stronger growth momentum, surpassing graphene synthesis in 2013 and graphene oxide in 2015 to become the most prominent topic in the graphene field. Graphene drug delivery papers have maintained high growth, reaching approximately 400 papers in 2016. Meanwhile, graphene fibers in energy storage and graphene aerogels represent interdisciplinary applications between materials science and energy/environmental fields, achieving high prominence rankings despite relatively low paper counts.

## 4. Discussion and Outlook

This study demonstrates from theoretical and empirical perspectives the methods and processes of the global-micro model in creating, identifying, selecting, and describing research fronts across the entire scientific domain. Many topics themselves possess interdisciplinary characteristics. For instance, mapping papers from graphene-related research topics to Scopus' s subject classification system reveals coverage of multiple disciplines including Materials Science, Chemistry, Physics, Engineering, Chemical Engineering, Energy, Biochemistry, and Environmental Science, reflecting the current state of graphene research. As interdisciplinary research becomes increasingly common, emerging frontier issues or major scientific breakthroughs are believed to occur at disciplinary intersections and margins. SciVal' s prominence topics, unrestricted by predefined searches, facilitate identification of interdisciplinary topics, providing researchers and decision-makers with an effective tool for detecting emerging research fronts, setting priority development directions, and allocating funding.

The correlation between topic prominence ranking and topic paper count indicates that topics ranking higher among research fronts tend to have more papers. Compared to ESI research fronts clustered from highly cited papers, SciVal research front topics average over 1,000 papers, raising issues of how to accurately and efficiently identify core papers and interpret topics. Meanwhile, analyzing multiple related topics to identify emerging themes and discover potential paths for technology transfer are questions worthy of deeper exploration.

Future research will conduct more empirical analyses based on published research topics from three aspects: (1) Introducing Altmetrics indicators to compare differences and impacts between media attention and academic influence in topic detection; (2) Applying prominence topics to competitiveness analysis and evaluation of institutions and disciplines in research fronts; (3) Conducting knowledge evolution mapping and technology transfer prediction research in specific research fields or directions.

## References

[1] Wu Feifei, Yang Zi, Huang Lucheng. A research front detection model based on innovation and interdisciplinarity—A case study in smart materials [J]. *Studies in Science of Science*, 2015, 33(1): 11-20. [2] GARFIELD E. Citation indexes for science: A new dimension in documentation through association of ideas [J]. *Science*, 1955, 122(3159): 108-111. [3] PRICE D J. Networks of scientific papers [J]. *Science*, 1965, 149(3683): 510-515. [4] KUHN T S. *The Structure of Scientific Revolutions* [M]. Chicago: University of Chicago Press, 1970. [5] GARFIELD E, SHER I H, TORPIE R J. *The Use of Citation Data in Writing the History of Science* [M]. Philadelphia: Institute for Scientific Information, 1964. [6] KESSLER M M. Comparison of the results of bibliographic coupling and analytic subject indexing [J]. *American Documentation*, 1965, 16(3): 223-233. [7] SMALL H, GRIFFITH B C. The structure of scientific literatures I: Identifying and graphing specialties [J]. *Social Studies of Science*, 1974(4): 17-40. [8] SMALL H. Tracking and predicting growth areas in science [J]. *Scientometrics*, 2006, 68(3): 595-610. [9] SMALL H. Visualizing science by citation mapping [J]. *Journal of the Association for Information Science & Technology*, 1999, 50(9): 799-813. [10] KLAVANS R, BOYACK K W. Quantitative evaluation of large maps of science [J]. *Scientometrics*, 2006, 68(3): 475-499. [11] Wang Xiaomei, Deng Qiping, Li Guopeng, et al. Scientific mapping of ESI research fronts and application in nanoscale fields [J]. *Library and Information Service*, 2017, 61(12): 106-112. [12] Zhou Qun, Zhou Qiuju, Leng Fuhai. Temporal analysis of biological science research front evolution [J]. *Bulletin of Chinese Academy of Sciences*, 2017(4): 405-412. [13] Leng Fuhai, Zhu Qingsong. Optimization and application of critical path analysis—A case study in quantum discord [J]. *Information Science*, 2016(4): 3-6, 12. [14] BOYACK K W. Using detailed maps of science to identify potential collaborations [J]. *Scientometrics*, 2009, 79(1): 27-44. [15] KLAVANS R, BOYACK K W. Toward an objective, reliable and accurate method for measuring research leadership [J]. *Scientometrics*, 2010, 82(3): 539-553. [16] KLAVANS R, BOYACK K W. Using global mapping to create a publication-level classification system of science [J]. *Journal of the Association for Information Science & Technology*, 2012, 63(1): 1-18. [17] WALTMAN L, VAN ECK N J. A new methodology for constructing a publication-level classification system of science [J]. *Journal of the American Society for*

*Information Science & Technology*, 2012, 63(12): 2378-2392. [18] BOYACK K W, KLAVANS R. Creation of a highly detailed, dynamic, global model and map of science [J]. *Journal of the Association for Information Science & Technology*, 2014, 65(4): 670-685. [19] KLAVANS R, BOYACK K W. Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? [J]. *Journal of the Association for Information Science & Technology*, 2017, 68(4): 984-998. [20] SMALL H, BOYACK K W, KLAVANS R. Identifying emerging topics in science and technology [J]. *Research Policy*, 2014, 43(8): 1450-1467. [21] 2017 Research Fronts [EB/OL]. [2018-03-17]. https://clarivate.com.cn/research_{{{fronts}}*{{2017}}}/2017*{{{research}}_{{front}}}.pdf. [22] BOYACK K W, KLAVANS R, SMALL H, et al. Characterizing the emergence of two nanotechnology topics using a contemporaneous global micro-model of science [J]. *Journal of Engineering & Technology Management*, 2014, 32(32): 147-159. [23] Delving deeper into topic prominence in science [EB/OL]. [2018-03-17]. https://www.elsevier.com/___{data}/assets/pdf_{file}/0006/548313/Topic-Prominence-Advanced-Webinar.pdf. [24] ECK N J V, WALTMAN L, DEKKER R, et al. A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS [J]. *Journal of the American Society for Information Science & Technology*, 2010, 61(12): 2405-2416. [25] KLAVANS R, BOYACK K W. Research portfolio analysis and topic prominence [EB/OL]. [2017-11-25]. https://arxiv.org/ftp/arxiv/papers/1709/1709.03453.pdf. [26] Xu Shumei. Research on percentile indicators application for excellence [D]. Beijing: Beijing Institute of Technology, 2015. [27] SIEGEL R L, MILLER K D, JEMAL A. Cancer statistics [J]. *CA: A Cancer Journal for Clinicians*, 2016, 66(1): 7-30.

## Author Contributions

Cui Yuhong: Methodology research and paper writing;
Wang Sa: Data analysis;
Gao Xiaowei, Cao Xuewei: Topic interpretation;
Yang Hui: Data provision.

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