

Research on Frontier Discovery of Technological Innovation Based on Knowledge Flow

Dechao Wang^a, Yongjie Li^b, Jian Zhu^c and Xiaoli Tang^d

*Institute of Medical Information & Library, Chinese Academy of Medical Sciences and Peking Union Medical College,
Beijing, P. R. China*

Keywords: Technological Innovation Frontier, Knowledge Flow, Science-Technology Linkage, Knowledge Meme.

Abstract: Amidst intensifying technological competition, technological innovation fundamentally arises from the flow of scientific knowledge into technical knowledge. Precisely characterizing the features and connotation of this knowledge flow is therefore crucial for identifying the frontiers of technological innovation. Adopting a semantic flow perspective, this study developed a simulation framework to model semantic flow between documents, progressing from key knowledge elements to the full-text level. Leveraging deep learning models, it then re-identified knowledge flow relationships between documents. The concept of knowledge meme was introduced to quantify the propagation dynamics (intensity and scope) of knowledge units across scientific and technical knowledge systems. Subsequently, a knowledge flow network connecting patents and academic papers in the lung cancer domain was constructed. Building upon this network, the substantive content of the knowledge flows was measured. This research achieved the identification and reconstruction of knowledge flow relationships between scientific and technical documents. Furthermore, by analyzing the content and communication patterns of computable knowledge units, it elucidated the frontiers of technological innovation. This approach holds significant implications for understanding science-technology linkages and identifying emerging technological innovation frontiers.

1 INTRODUCTION

Amidst an increasingly complex and volatile global landscape, rapid scientific advancement and accelerated technological iteration have elevated technological innovation to a critical determinant of national scientific prowess and international competitiveness. The frontiers of technological innovation, characterized by novelty, interdisciplinarity, and high visibility, represent clusters of research achievements that spearhead progress within technological domains (Shibata et al., 2008). These frontiers play a pivotal role in guiding future scientific and technological development. Crucially, science and technology serve as key enablers for discovering innovation opportunities, with most patent innovations drawing upon scientific foundations. The flow of knowledge from scientific

publications into technical domains, exemplified by patents, is instrumental in generating technical knowledge (Roh et al., 2023). Consequently, delving into science-technology linkages offers an effective pathway to uncover technological opportunities and pinpoint innovation frontiers (Robinson et al., 2013).

This study adopts the theoretical lens of knowledge flow to articulate the content and direction of knowledge absorption, growth, and dissemination among distinct entities (Hai, et al., 2006). Within this framework, we define the frontier of technological innovation as a cluster of knowledge content flowing from science to technology.

Leveraging knowledge flow relationships to model science's contribution to technology, this research devised a knowledge flow recognition algorithm capable of identifying semantic inclusion relations. Building upon this algorithm, we

^a  <https://orcid.org/0000-0002-1838-8168>

^b  <https://orcid.org/0009-0004-6306-8288>

^c  <https://orcid.org/0009-0008-8222-0280>

^d  <https://orcid.org/0000-0001-6946-3482>

constructed a knowledge flow network connecting academic papers and patents. Subsequently, utilizing this network, we investigated the propagation pathways of knowledge meme during knowledge dissemination. This analysis aims to elucidate the current frontiers of technological innovation within the target technological field, thereby providing foundational insights and references for related scientific and technological research endeavors.

2 RELATED WORK

2.1 Identification of the Frontiers of Technological Innovation

Research on identifying the frontiers of technological innovation encompasses bibliometric or cluster analyses conducted solely within the technological domain (Roche et al., 2010; Kim et al., 2017; Xu et al., 2022; Li et al., 2015). However, a greater number of studies opt to explore these frontiers by analyzing science-technology linkages. This approach involves examining the relationships among knowledge units across scientific and technological systems, measuring the direction, intensity, and structure of their dissemination and transfer to reveal interactions between science and technology (Ba et al., 2021). It also facilitates research into discovering technological opportunities through these linkages, tracking transformation patterns of scientific and technological knowledge, or identifying innovation frontiers (Han et al., 2022; Du et al., 2019; Tian et al., 2024; Xu et al., 2020; Du et al., 2019). Scholarly papers and patents are widely recognized as representative outputs of scientific and technological research, respectively (Ahmadpoor et al., 2017). Primary methodologies for analyzing science-technology linkages include: citation correlation analysis between papers and patents (Nguyen et al., 2019; Kenan-Flagler et al., 2011; Li et al., 2024), author-inventor analysis (Boyack et al., 2008; Breschi et al., 2010; Ning et al., 2020), and knowledge structure analysis (Xu et al., 2022; Du et al., 2024; Zhang et al., 2022; Ran et al., 2024).

These methods collectively aim to construct specific relationships between patents and papers, enabling the identification and analysis of knowledge clusters flowing between science and technology domains, thereby uncovering innovation frontiers. Nevertheless, these approaches exhibit significant limitations. Citation correlation analysis suffers from inherent constraints of citation relationships

themselves – including the questionable semantic correlation between citing and cited documents (Li et al., 2014; Meyer, 2000), the scarcity of citations between patents and papers (Xu et al., 2022; Callaert et al., 2006), the difficulty in constructing potential citation links, and the potential intentional concealment of patent citations (Wu et al., 2017) – resulting in compromised data quality. Author-inventor analysis establishes connections via researcher identities but neglects the underlying semantic association between science and technology, failing to accurately capture deep data linkages. Among knowledge structure analysis methods, vocabulary- or topic-based approaches are more effective at revealing semantics; however, they face challenges in systematically constructing knowledge structure networks (Ba et al., 2021) and delineating propagation pathways.

2.2 Knowledge Flow Relationships

Knowledge flow describes the process by which knowledge disseminates and transfers among different entities, domains, or systems. Knowledge flow relationships are conventionally measured using citation data (Criscuolo et al., 2008; Lyu et al., 2022; Zhao et al., 2022) and can also be inferred through indirect citation chains (Feng et al., 2023). However, the implicit, ambiguous, and complex nature of technological interactions complicates the revelation of intrinsic relationships within scientific and technological knowledge (Chen et al., 2023). Crucially, citation relationships cannot fully represent knowledge flow (Meyer, 2000); even when incorporating indirect citations, they inadequately capture the knowledge contribution and academic influence across disciplines (Roh et al., 2023).

From a knowledge flow perspective, its core elements encompass the subject, content, and direction. To address the shortcomings of the aforementioned methods, researchers require methodologies capable of constructing accurate knowledge flow relationships between scientific and technological entities and analyzing the substantive content of these flows through effective semantic techniques (Kang et al., 2022; Zhang et al., 2024).

2.3 Knowledge Meme

The term "meme" originated in Dawkins' seminal work *The Selfish Gene*, conceptualized as the functional unit of knowledge inheritance and variation, reflecting the process of knowledge flow and dissemination (Yang et al., 2021). Within the

scientific domain, extracting meme facilitates the discovery of semantic information embedded in academic papers (Zeng et al., 2023), and their propagation serves to quantify the diffusion patterns of knowledge flow (Mao et al., 2024; Kamada et al., 2021). Fundamentally, innovation constitutes a process of meme-based search, combination, experimentation, and adjustment. Kuhn et al. (2014) defined scientific meme as short text units within scientific publications whose semantics are replicated when the publications are cited. Similarly, Sun et al. (2018) defined technical meme as short text units within patents whose semantics are replicated upon citation. However, research on scientific or technological meme predominantly relies on citation relations (Araújo et al., 2018). Consequently, tracking meme propagation remains largely confined to citation analysis, failing to transcend its inherent limitations. Nevertheless, this research provides a novel conceptual framework for reflecting knowledge flow—specifically, considering the direction and volume of knowledge movement at the level of semantic content.

Building upon the concepts of scientific and technological meme, this study defines knowledge meme as "short text units within scientific or technological publications whose semantics are replicated when knowledge flow occurs among publications."

3 DATA AND METHODS

3.1 Data

This study selected the field of lung cancer as the empirical research domain. The primary dataset comprised lung cancer patents (Dataset A) retrieved from the Dimensions database, covering patent grants issued between January 2019 and December 2023. Following the consolidation of patent families, Dataset A contained 6,671 unique patents. Building upon this foundation, the referenced patents (source: incoPat) and referenced scientific publications (source: PubMed) cited by these lung cancer patents were collected by matching patent numbers with PMID identifiers. These referenced documents were amalgamated into Dataset B, which contained 19,453 referenced patents and 12,394 referenced publications.

3.2 Research Design

This research is predicated on three fundamental assumptions:

- 1) Knowledge flow occurs within citation relationships between patents and publications.
- 2) This knowledge flow is effectively captured by the propagation of knowledge meme semantics.
- 3) Knowledge flow extends beyond the existing citation network, encompassing relationships between patents and publications where semantic transfer of knowledge meme occurs independently of direct citation links.

Aligned with the characteristics of knowledge meme dissemination, this study conceptualizes citation-based knowledge flow as the semantic containment of key knowledge elements (represented by meme) from the knowledge-outflow entity (publication) within the text of the knowledge-inflow entity (patent). Consequently, a knowledge flow identification model was designed. This model learns the characteristics of knowledge flow within the known citation network (Dataset B) to predict knowledge flow relationships existing outside this network.

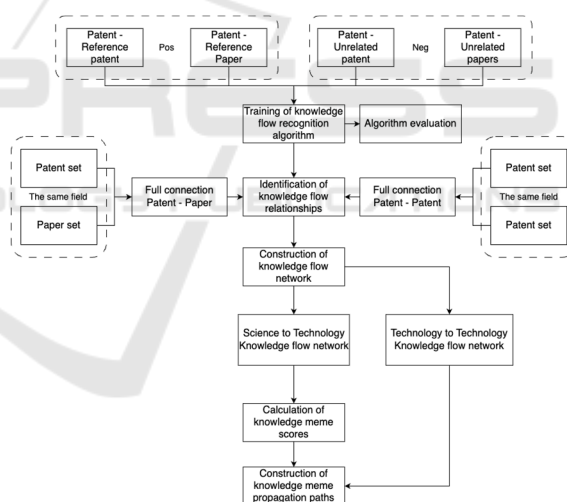


Figure 1: Frontier discovery of technological innovation based on knowledge flow.

This study designed and trained a deep learning algorithm, based on PubmedBERT+Bi-LSTM, to identify whether the semantics of knowledge meme derived from a publication are contained within the text semantics of a patent. This identification process enables the construction of knowledge flow relationships between publications and patents. The model was trained using the lung cancer patent data (Dataset A) and their corresponding references (Dataset B). Subsequently, the trained model was

applied to identify potential knowledge flow relationships between all patents and publications within the lung cancer domain. This facilitated the construction of a comprehensive knowledge flow network for lung cancer. Finally, leveraging this network, the propagation intensity and propagation scope of knowledge meme were calculated. This analysis aimed to assess the impact of science on technological development and identify the frontiers of technological innovation within the lung cancer field. The overall research workflow is depicted in Figure 1.

3.3 Knowledge Meme Analysis

Traditionally, meme extraction employs rule-based methods, often utilizing high-frequency words from the corpus after removing stop words, functional terms, and meaningless tokens to form a candidate meme set. However, such rule-based approaches are semantically constrained and struggle to identify and merge candidate meme sharing identical semantics.

Within natural language processing (NLP), the keyword extraction task automatically selects phrases from a document to summarize its content. Keyword extraction algorithms leveraging pre-trained language models capture semantic information within documents, enabling effective identification and output of key information as keywords. This capability aligns well with the defining characteristics of meme and offers the potential for enhanced meme recognition. Therefore, this study employed the PromptRank model – a current state-of-the-art (SOTA) algorithm for keyword extraction (Kong et al., 2023) – to extract keywords from all patent and publication documents. The resulting keywords were treated as the knowledge meme for subsequent knowledge flow analysis and computation.

Driven by the features of meme propagation mechanisms, this study adopted the method proposed by Kuhn et al. to calculate a comprehensive score for candidate meme. To address variations in term expression, including synonyms and lexical variants, we incorporated an additional word sense disambiguation step. Specifically, a word embedding generation model was utilized to generate embeddings for candidate meme words. Knowledge meme exhibiting a cosine similarity score exceeding 0.9 were subsequently merged.

The keywords extracted from all patents within the lung cancer dataset (Dataset A) served as the initial candidate knowledge meme. Following this, the comprehensive score for each candidate knowledge meme was calculated sequentially based

on the constructed knowledge flow network, utilizing Formulas (1) and (2).

$$P_m = \frac{d_{m \rightarrow m}}{d_{\rightarrow m} + \delta} / \frac{d_{m \rightarrow \bar{m}} + \delta}{d_{\rightarrow \bar{m}} + \delta} \quad (1)$$

$$M_m = f_m \cdot P_m \quad (2)$$

Where, P_m is the propagation score of knowledge meme; f_m is the document frequency of knowledge meme; M_m is the composite knowledge meme score. $d_{m \rightarrow m}$ indicates the number of knowledge-inflow documents containing m and at least one of the associated knowledge-outflow documents containing m ; $d_{\rightarrow m}$ indicates the number of knowledge-inflow documents with at least one associated knowledge-outflow document containing knowledge meme m ; $d_{m \rightarrow \bar{m}}$ indicates that the knowledge meme m are contained, but the associated knowledge outflow documents do not contain the number of knowledge inflow documents of the knowledge meme; $d_{\rightarrow \bar{m}}$ indicates that all associated knowledge-outflow documents do not contain the number of knowledge-inflow documents of the knowledge meme. To prevent the denominator from being zero, set δ as the smoothing factor.

3.4 Knowledge Flow Identification Algorithm

3.4.1 Algorithm Design

As illustrated in Figure 2, the knowledge flow identification algorithm developed in this study comprises three primary stages:

Embedding Generation: Producing sentence-level and word-level embeddings.

Sequence Data Processing: Handling the sequential nature of input features.

Classification Prediction: Determining the presence of a knowledge flow relationship.

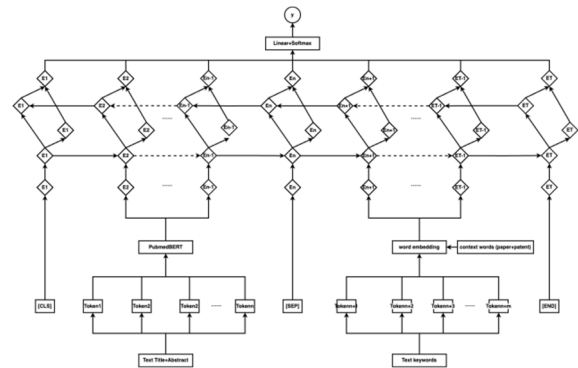


Figure 2: The structure of knowledge flow identification algorithm.

Stage 1: Embedding Generation

To enhance the model's capacity for capturing domain-specific semantics and terminology within the biomedical field, the PubMedBERT pre-trained language model was utilized.

The text of the knowledge-inflow document (typically a patent) was directly processed by PubMedBERT to generate sentence-level embeddings.

The keywords (representing knowledge meme) extracted from the knowledge-outflow document (typically a publication) were initially embedded using PubMedBERT to generate initial word embeddings. Subsequently, these initial embeddings were further optimized by training a dedicated word embedding vector model. This optimization incorporated the full-text contextual information from both patents and publications to refine the semantic embedding representation of the keywords.

Stage 2: Sequence Data Processing

Given the sequential dependence and contextual relevance inherent in the input features for knowledge flow identification, the generated word embeddings (representing outflow meme) and sentence embeddings (representing inflow document context) were concatenated and fed into a Bidirectional Long Short-Term Memory (Bi-LSTM) layer. The Bi-LSTM network effectively processes this sequential input to capture complex dependencies.

Stage 3: Classification Prediction

The output vectors from the Bi-LSTM layer were sequentially passed through a linear layer and a Softmax activation layer. This process yielded a probability score indicating the likelihood of a semantic containment relationship (i.e., a positive knowledge flow relationship) existing between the input pair. All predicted results were ranked by their probability scores. The top n results, based on the highest probability scores, were selected as positive predictions.

3.4.2 Construction of Algorithm Dataset

The algorithm training and evaluation datasets were constructed from the lung cancer patent dataset (Dataset A) and the reference dataset (Dataset B, containing cited publications and patents).

Positive Samples: Pairs were formed based on explicit citation relationships (e.g., a lung cancer patent citing a paper or another patent).

Negative Samples: For a given document (Feature A), documents within the reference dataset (Dataset B) that had no citation relationship with it were randomly selected to form negative pairs (Feature B).

Patents within the lung cancer dataset (Dataset A) that possessed citations were partitioned into training, validation, and test sets using a 6:2:2 ratio. Reflecting the typical imbalance in such tasks, the number of randomly generated negative samples in both the training and validation sets was set to four times the number of positive samples. Within the test set, negative samples were generated at ratios of either five or ten times the number of positive samples.

The neural network model was trained using input pairs consisting of the two sets of embedding features along with their corresponding classification labels (indicating positive or negative knowledge flow relationship). The model's objective was to identify the semantic containment relationship between the embeddings of Feature A and Feature B.

Post-processing for Chronological Consistency: To ensure temporal validity of the predicted knowledge flow relationships, a post-processing step was applied. Any positive prediction where the publication date of the knowledge-outflow document (Feature B) occurred later than the publication date of the knowledge-inflow document (Feature A) was reclassified as a negative result, as knowledge cannot logically flow from a future document to a past one.

4 RESULTS

4.1 Algorithm Evaluation

The model was evaluated on two distinct test sets, characterized by class imbalance ratios (positive to negative samples) of 1:5 and 1:10, respectively. This design reflects the inherent scarcity of knowledge flow relationships compared to non-flow pairs within the data. Given that the model outputs a probability score for the existence of a knowledge flow relationship between any document pair, the top n ranked predictions by probability were selected as positive identifications.

As the knowledge flow identification algorithm functions as a ranking task, standard information retrieval metrics were employed for evaluation: Mean Reciprocal Rank (MRR), Mean Average Precision (MAP), Recall, Precision, and F1-score. The performance metrics of the algorithm on both test sets are presented in Table 1.

Table 1: The algorithm test results.

Positive/ negative sample ratio	MRR	MAP	Recall	Precision	F1
1:5	0.8309	0.6601	0.8038	0.7692	0.7861
1:10	0.7579	0.5713	0.8029	0.6967	0.7460

Despite potential data quality limitations inherent in the training samples (notably, the non-correlative nature of some citation relationships), the knowledge flow identification algorithm demonstrated robust performance on the citation-based dataset. This validates its applicability for constructing interactive networks of scientific and technological knowledge.

4.2 Memetic Comprehensive Score Calculation

An empirical dataset was constructed comprising patents granted between 2022 and 2023 and publications issued between 2019 and 2023 within the lung cancer domain. The trained model was then applied to identify potential knowledge flow relationships between these publications and patents. For each patent, the top 5 publications (published prior to the patent application date) with the highest predicted probability scores were identified as potential knowledge-outflow sources.

Applying a smoothing factor of 0.5, the comprehensive scores for all candidate knowledge meme were computed and ranked. Table 2 lists the top 10 ranked candidate knowledge meme alongside their comprehensive scores.

Table 2: Ranking of Candidate Knowledge meme by Comprehensive Score (Top 10).

Keywords	Comprehensive score
anti-pd-l1 antibody	2.013221
peptides	1.423944
anti-tumor immune responses	1.312993
bronchial asthma	1.214736
anti-pd-1 antibody	1.182252
compounds	0.990032
tumor immunotherapy	0.912560
nucleic acid aptamer	0.899915
air pollution	0.895928
cell peptide epitopes	0.883943

Subsequent to calculating the comprehensive meme scores, and leveraging the knowledge flow relationships identified among patents within the lung cancer dataset by the trained model, a "patent-patent knowledge flow network" was constructed. This network enables the exploration of meme propagation dynamics within the technological knowledge system after their initial flow from scientific publications into patents.

5 DISCUSSION

This study addresses critical challenges in identifying the frontiers of technological innovation—namely, the inadequacy of capturing science-technology linkages and the limited semantic recognition capabilities prevalent in current methodologies. The ability of traditional approaches to accurately reveal deep-level semantic flow relationships is fundamentally constrained by inherent data limitations and methodological shortcomings.

To overcome these limitations, this research proposed a semantic-driven approach centered on knowledge flow. A pivotal contribution is the introduction of knowledge meme as computable units of knowledge transfer. By enabling the extraction and quantitative measurement of these knowledge meme, this study provides novel semantic-level perspectives and methods for analyzing the frontiers of technological innovation.

The knowledge flow identification model was trained based on citation relationships. However, the representativeness of citations as proxies for genuine knowledge flow relationships is inherently limited (Meyer, 2000; Chen et al., 2023), potentially introducing bias into the model's predictions.

Future research will focus on several key directions:

Enhancing Meme Interpretation: Delving deeper into the semantic information embedded within knowledge meme, employing techniques such as classification, combinatorial analysis, and logical deduction to enhance the interpretability of the results.

Multidimensional Validation: Employing bibliometric methods to analyze the constructed knowledge flow networks and cross-validating the findings derived from knowledge meme analysis with these network-based insights.

REFERENCES

- Ahmadpoor, M.A., & Jones, B.F. (2017). The dual frontier: Patented inventions and prior scientific advance. *Science*, 357, 583 - 587.
- Araújo, T., & Fontainha, E. (2018). Are scientific meme inherited differently from gendered authorship? *Scientometrics*, 117, 953–972.
- Ba, Z., & Liang, Z. (2021). A novel approach to measuring science-technology linkage: From the perspective of knowledge network coupling. *Journal of Informetrics*, 15(3), 101167.
- Boyack, K. W., & Klavans, R. (2008). Measuring science-technology interaction using rare inventor-author names. *Journal of Informetrics*, 2, 173–182.
- Breschi, S., & Catalini, C. (2010). Tracing the links between science and technology: An exploratory analysis of scientists' and inventors' networks. *Research Policy*, 39(1), 14–26.
- Callaert, J., Van Looy, B., Verbeek, A., et al. (2006). Traces of Prior Art: An analysis of non-patent references found in patent documents. *Scientometrics*, 69, 3–20.
- Chen, X., Ye, P., Huang, L., et al. (2023). Exploring science-technology linkages: A deep learning-empowered solution. *Information Processing & Management*, 60(2), 102.
- Criscuolo, P., & Verspagen, B. (2008). Does it matter where patent citations come from? Inventor vs. examiner citations in European patents. *Research Policy*, 37, 1892–1908.
- Du, C., Yao, K., Zhu, H., et al. (2024). Mining technology trends in scientific publications: A graph propagated neural topic modeling approach. *Knowledge and Information Systems*. Advance online publication.
- Du, J., Li, P., Guo, Q., et al. (2019). Measuring the knowledge translation and convergence in pharmaceutical innovation by funding-science-technology-innovation linkages analysis. *Journal of Informetrics*, 13, 132–148.
- Du, J., Sun, Y., Li, Y., et al. (2019). Identifying innovation frontier at the interface of science and technology: A bibliometric framework and empirical study [In Chinese]. *Information Studies: Theory & Application*, 42(1), 94–99.
- Feng, S., Li, H., & Qi, Y. (2023). How to detect the sleeping beauty papers and princes in technology considering indirect citations? *Journal of Informetrics*, 17, 101431.
- Hai, Z. (2006). Discovery of knowledge flow in science. *Communications of the ACM*, 49(5), 101–107.
- Han, X., Zhu, D., & Wang, X. (2022). Research on the method of technology opportunity discovery promoted by science [In Chinese]. *Library and Information Service*, 66(10), 19–32.
- Kamada, M., Asatani, K., Isonuma, M., et al. (2021). Discovering interdisciplinarily spread knowledge in the academic literature. *IEEE Access*, 9, 124142–124151.
- Kang, X., Jia, X., Deng, L., et al. (2022). Research on the characteristics of high-impact patent knowledge diffusion based on all generation citation network [In Chinese]. *Library and Information Service*, 66(22), 83–94.
- Kenney, M. R. (2011). Lens or prism? A comparative assessment of patent citations as a measure of knowledge flows from public research. *Management Science*, 59(2), 504–525.
- Kim, G., & Bae, J. (2017). A novel approach to forecast promising technology through patent analysis. *Technological Forecasting and Social Change*, 117, 228–237.
- Kong, A., Zhao, S., Chen, H., et al. (2023). PromptRank: Unsupervised keyphrase extraction using prompt. *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics* (Vol. 1, pp. 9788–9801).
- Kuhn, T., Perc, M., & Helbing, D. (2014). Inheritance patterns in citation networks reveal scientific meme. *Physical Review X*, 4(4), 041002.
- Li, B., & Chen, X. (2015). Identification of emerging technologies in nanotechnology based on citing coupling clustering of patents [In Chinese]. *Journal of Intelligence*, 34(5), 35–40.
- Li, B., Ding, K., Sun, X., et al. (2024). Research on the diffusion speed and diffusion effects of scientific papers into the technological domain [In Chinese]. *Information Studies: Theory & Application*, 47(7), 35–47.
- Li, R., Chambers, T., Ding, Y., et al. (2014). Patent citation analysis: Calculating science linkage based on citing motivation. *Journal of the Association for Information Science and Technology*, 65.
- Lyu, H., Bu, Y., Zhao, Z., et al. (2022). Citation bias in measuring knowledge flow: Evidence from the web of science at the discipline level. *Journal of Informetrics*, 16(4), 101338.
- Mao, J., Liang, Z., Cao, Y., et al. (2024). Quantifying cross-disciplinary knowledge flow from the perspective of content: Introducing an interdisciplinary distance indicator. *Journal of Informetrics*, 17(2), 101092.
- Meyer, M. S. (2000). Does science push technology? Patents citing scientific literature. *Research Policy*, 29, 409–434.
- Nguyen, A. L., Liu, W., Khor, K. A., et al. (2019). The golden eras of graphene science and technology: Bibliographic evidences from journal and patent publications. *Journal of Informetrics*, 14, 101067.
- Ning, Z., & Wei, L. (2020). Research on the relationship between patent documents and academic papers based on patent subjects: A case study of data mining [In Chinese]. *Library and Information Service*, 64(12), 106–117.
- Robinson, D., Huang, L., Guo, Y., et al. (2013). Forecasting innovation pathways (FIP) for new and emerging science and technologies. *Technological Forecasting and Social Change*, 80(2), 267–285.
- Roche, I., Besagni, D., François, C., et al. (2010). Identification and characterisation of technological topics in the field of molecular biology. *Scientometrics*, 82, 663–676.
- Roh, T., & Yoon, B. (2023). Discovering technology and science innovation opportunity based on sentence

- generation algorithm. *Journal of Informetrics*, 17, 101403.
- Ran, C., Tian, W., & Jia, Z. (2024). Modeling of scientific paper-patent technology association relationship based on mixed methods: Taking the biomedical field as an example [In Chinese]. *Information Science*. Advance online publication.
- Shibata, N., Kajikawa, Y., Takeda, Y., et al. (2008). Detecting emerging research fronts based on topological measures in citation networks of scientific publications. *Technovation*, 28, 758–775.
- Sun, X., & Ding, K. (2018). Identifying and tracking scientific and technological knowledge meme from citation networks of publications and patents. *Scientometrics*, 116, 1735–1748.
- Tian, C., Dong, K., Guo, R., et al. (2024). Research on measurement method of transformation efficiency of scientific and technical achievements based on knowledge association analysis [In Chinese]. *Information Studies: Theory & Application*, 47(5), 123–130.
- Wu, H., & Ji, F. (2017). Empirical research on the evaluation effectiveness of patent citation based on patent application and patent censorship [In Chinese]. *Library and Information Service*, 61(19), 89–95.
- Xu, H., Winnink, J. J., Yue, Z., et al. (2020). Topic-linked innovation paths in science and technology. *Journal of Informetrics*, 14, 101014.
- Xu, H., Yue, Z., Pang, H., et al. (2022). Integrative model for discovering linked topics in science and technology. *Journal of Informetrics*, 16, 101265.
- Xu, X., Wu, F., & Wang, B. (2022). Research on identification of key core technology based on international patent classification [In Chinese]. *Journal of Intelligence*, 41(10), 74–81.
- Yang, F., Qiao, Y., Wang, S., et al. (2021). Blockchain and multi-agent system for meme discovery and prediction in social network. *Knowledge-Based Systems*, 229, 107368.
- Zeng, J., Cao, S., Chen, Y., et al. (2023). Measuring the interdisciplinary characteristics of Chinese research in library and information science based on knowledge elements. *Aslib Journal of Information Management*, 75, 589–617.
- Zhang, B., Wu, H., Gao, D., et al. (2022). Research on identification of innovation fronts based on potentially high cited papers and high value patents [In Chinese]. *Library and Information Service*, 66(18), 72–83.
- Zhang, J., Kang, L., & Sun, J. (2024). The influence of recency and time-span in the scientific and technological knowledge convergence [In Chinese]. *Journal of Information Resources Management*, 14(4), 86–102.
- Zhao, H., & Wang, X. (2022). Characteristics and evolutionary trends of knowledge flow in interdisciplinary research under the background of open science: Taking the study of "Five-Metrics" as an example [In Chinese]. *Information Science*, 40(4), 107–117.