Postprint: Micro-level Knowledge Flow Analysis Framework for Science and Technology Based on Multilayer Networks

Authors: Wu Ning, Yang Yanping

Date: 2024-04-03T00:00:00+00:00

Abstract

Purpose/Significance How knowledge flows between science and technology and drives innovation represents a prominent issue in the field of information science. This study reveals the structural characteristics of knowledge flow networks between science and technology from a micro-level perspective, offering novel insights for discovering new knowledge flow propagation pathways, understanding the connections between science and technology, and identifying innovation breakthrough points. Method/Process Based on patent and paper data and employing multi-layer network methods, this study constructs a micro-level network analysis framework for two-dimensional science-technology knowledge flow, and conducts an empirical study using transgenic corn as a case to verify the feasibility of fine-grained revelation of structural characteristics in science-technology knowledge flow. Results/Conclusion The analysis framework proposed in this study can reveal knowledge flow characteristics from a micro perspective, and is capable of identifying two types of knowledge flow structures in the empirical case of transgenic corn, thereby expanding and enriching knowledge flow related research. In the future, the science-technology two-dimensional knowledge flow research framework proposed in this study can be further extended to achieve broader applications by altering knowledge subject types, enriching knowledge attribute factors, and extending knowledge flow paths.

Full Text

Preamble

Science and technology constitute two fundamental forces that determine the direction of technological innovation, and their interaction and integration have profoundly impacted social progress and human development. Both science and technology possess the essential attribute of knowledge: science serves as the primary source of knowledge creation and dissemination, while technological breakthroughs and innovations directly propel scientific advancement. The interaction between science and technology is fundamentally a process of knowledge flow and innovation, wherein their cross-penetration is accompanied by interactive knowledge collisions that facilitate new knowledge accumulation and the formation of knowledge advantages, thereby becoming a wellspring of major innovations. From an informetrics perspective, examining the patterns of knowledge propagation between science and technology has thus emerged as a critical research area.

Knowledge flow between science and technology is essentially a process of knowledge creation and utilization. The characteristics of knowledge linkages across domains influence the identification of important scientific fields, the direction of technological innovation, and the optimization of science and technology policies to enhance innovation efficiency. Science provides theoretical foundations for technology and indicates development directions and evolutionary pathways, while technology's utilization of scientific knowledge also affects knowledge flows between technological domains. Technology domains do not exist in isolation; in an era of increasing disciplinary differentiation and more specialized theoretical systems, technological knowledge increasingly originates from interdisciplinary and convergent fields. The knowledge association characteristics between domains affect the knowledge sending and receiving processes, making the study of science-technology linkages crucial for understanding innovation mechanisms.

1.1 Science-Technology Knowledge Flow Linkages

From a techno-economics perspective, relevant research has focused on regional innovation systems, revealing the correlation between science and technology, as well as the relationships between enterprises and basic research institutions in the innovation chain. Studies have shown that the geographic proximity between basic research institutions and technology development organizations correlates with the strength of their knowledge associations and flows. Jaffe et al. [?] analyzed the geographic agglomeration phenomenon of innovation entities in the United States, revealing knowledge diffusion trajectories from specific scientific knowledge to different technology fields, thereby reflecting science's role in technological research and development. AUTANT-BERNARD [?] examined whether externalities from public and private sector research activities are geographically constrained and influenced by inter-sectoral proximity, demonstrating that the degree of science-technology knowledge linkages within the U.S. research community doubles every six years. Liu et al. [?] focused on collaborative relationships between industry, academia, and research institutions, revealing the knowledge relationships among collaborative innovation entities and scientific governance models.

From a scientometrics and bibliometrics perspective, research primarily uses papers and patents as carriers to represent knowledge between science and technology. Scholars argue that scientific literature and other media provide bridges for knowledge transmission. Innovation entities can cite literature from other regions or countries to achieve cross-regional knowledge dissemination. Papers and patents can represent scientific and technological knowledge, serving as the basis for predicting trends and planning development strategies, tracing innovation propagation paths, and assessing innovation value. Wen [?] noted that knowledge is transferred from cited to citing documents during the citation process, and innovations are generated based on reference to and learning from cited literature. Narin et al. [?] pointed out that almost all technological achievements require previous research as a foundation, and patent citation relationships constitute an important manifestation of knowledge flow. Lai [?] considered patent citations a behavior with greater significance for technological diffusion, representing researchers' absorption of scientific literature and forming a bridge between scientific and technological knowledge.

1.2 Measurement of Science-Technology Knowledge Flow

Existing research primarily employs citation relationships between papers and patents, using methods such as topic mining to abstractly summarize patterns or conduct statistical numerical analyses of their associations. However, these approaches lack concrete structural identification of knowledge flow from a finer-grained network perspective. This study treats science and technology as an integrated whole and constructs a multi-dimensional knowledge flow micro-structure analysis framework using multi-layer network theory to mine the relationships in knowledge flow between science and technology and discover potential knowledge flows, thereby better promoting their integration and transformation.

1.2.1 Patent-to-Paper Citations Patent-to-paper citations reflect the flow of theoretical knowledge from scientific research to technological development, demonstrating science's influence and promotion of technology. Narin et al. [?] analyzed citation relationships between patents and papers, revealing the mutual influence and promotion between technology and science. Meyer et al. [?] found that patents citing papers reflect the utilization and diffusion of scientific knowledge to technological research and development. Clarivate Analytics incorporates the influence of papers cited by patents into its evaluation system to measure the impact of basic research on commercial development in academic contexts [?]. Schmoch [?] analyzed representative patent citations from multiple perspectives, proving the existence of multi-faceted interactions with varying intensities.

1.2.2 Paper-to-Patent Citations Paper-to-patent citations reflect scientific research's absorption of technological knowledge and demonstrate technology's 推动作用 on scientific research. The frequency of patents cited by papers indicates the scope of knowledge dissemination; higher citation frequency suggests greater technological originality and value [?, ?]. Meyer et al. [?] discovered that papers citing patents have higher academic impact and are more likely to be cited by other papers. Compared to patent-to-paper citations, paper-to-patent citations are more time-sensitive. Glanzel et al. [?] analyzed SCI papers and USPTO patent data, finding that chemical patents are the most frequently cited type, followed by pharmaceutical and medical patents.

1.2.3 Mixed Citations Unidirectional citation analysis struggles to reflect the fusion, penetration, and interaction between science and technology. Mixed citation analysis based on bidirectional citation relationships between patents and papers has been widely adopted to measure science-technology linkages. Gao et al. [?, ?, ?] proposed a paper-patent hybrid co-citation analysis method using spectral clustering algorithms to reveal knowledge structures in technology diffusion, summarizing the functions and dominant roles of science and technology in different periods. Huang et al. [?] analyzed cross-citation patterns of papers and patents in the fuel cell field, finding stronger science-technology linkage convergence. Yu et al. [?] combined path analysis and machine learning methods to construct linkages between science and technology from a knowledge flow perspective, achieving integration of citation and semantic analysis.

2 Research Design and Methodology

This study defines knowledge flow as the diffusion of knowledge from knowledge creators to knowledge recipients, using network structures to characterize knowledge flow patterns. The analysis framework involves four aspects: data collection and relationship extraction, model construction, and knowledge flow analysis, as illustrated in Figure 1 [Figure 1: see original paper].

2.1 Data Collection

Data sources include the Incopat patent database, Lens database, and Web of Science Core Collection. The data processing workflow is shown in Figure 2 [Figure 2: see original paper]. In the Incopat database, target domain patents are retrieved to extract IPC classification numbers and corresponding forward citation patents, with patent publication numbers recorded. Using these publication numbers, the Lens database is queried to obtain correspondence between patents and referenced papers (DOI). In Web of Science Core Collection, referenced papers are retrieved by DOI to extract their WOS category information. Based on the correspondence between target patents and forward citations, three types of relationships are extracted and deduplicated: target patent IPC-forward citation IPC, target patent IPC-paper WOS category, and paper WOS category-paper WOS category.

2.2 Network Mapping

Multi-layer networks consist of multiple single-layer networks, with each layer representing a network stratum and inter-layer connections representing multi-type relationships [?, ?]. This approach overcomes the homogeneity constraints of nodes and edges in single-layer networks, offering advantages in describing and analyzing complex systems with multi-attribute associations. This study constructs single-mode network layers for patent citation networks and scientific knowledge intersection networks, forming interdependent networks.

In the science dimension, different disciplinary identifiers in papers imply knowledge cross-fertilization between disciplines. This study uses WOS categories to represent scientific research and captures knowledge flow relationships between scientific fields through disciplinary co-occurrence relationships, constructing a discipline-discipline co-occurrence network (single-mode network A). In the technology dimension, this study uses patent citation relationships to represent technological knowledge flow, constructing a patent-patent citation network (single-mode network B). Combining both dimensions, this study constructs a patent-paper citation network (bipartite network X) based on patent-to-paper citations, connecting the two single-mode networks to form a multi-layer network system for analyzing knowledge flow between technology fields, as shown in Figure 3 [Figure 3: see original paper].

2.3 Model Construction

To quantitatively analyze multi-layer networks, this study employs Multi-level Exponential Random Graph Models (MERGMs). MERGMs are among the few methods that can directly assess inter-layer dependencies in multi-layer networks [?]. The model extends general ERGMs to multi-layer contexts, where the existence of one network layer may influence another. The probability formula for two-layer networks is:

$$Pr(A=a,X=x,B=b)=\frac{1}{K(\theta)}\exp\{\theta^T_{ZQ}Z_Q(a)+\theta^T_{ZQ}Z_Q(b)+\theta^T_{ZQ}Z_Q(x)+\theta^T_{ZQ}Z_Q(a,x)+\theta^T_{ZQ}Z_Q(b,x)+\theta^T_{ZQ}Z_Q($$

where $Z_Q(a)$ and $Z_Q(b)$ are ERGM statistics for single-mode networks (including structural and exogenous attribute effects), $Z_Q(x)$ is the ERGM configuration for bipartite networks, and $Z_Q(a,x)$, $Z_Q(b,x)$, and $Z_Q(a,x,b)$ are two-layer ERGM configurations following Wang et al. [?]. θ_{ZQ} represents parameter vectors corresponding to effects Z_Q, and $K(\theta)$ ensures probability distribution normalization. The model uses Markov Chain Monte Carlo Maximum Likelihood Estimation for parameter estimation and goodness-of-fit testing through simulation-based modeling, implemented in MPNet software.

2.4 Knowledge Flow Pattern Analysis

When MERGM indicator calculations yield positive results, the represented graph structure has a high probability of occurrence; negative results indicate the opposite. Therefore, the sign of graph structure indicators can assess the likelihood of knowledge flow between nodes. Based on extracted relationship data, specific technology and disciplinary domains are selected for case analysis.

3 Empirical Study: Genetically Modified Corn

Genetically modified corn represents a major technological breakthrough in corn breeding and production, playing a key role in maintaining stable and increased yields. Its development involves cross-integration among life sciences, bioinformatics, and other fields, accompanied by knowledge flows between different domains. The technological innovation network for genetically modified corn exhibits trends of virtualization, internationalization, and intellectualization.

3.1 Data Sources and Processing

Data were collected from Incopat, Web of Science Core Collection, and Lens databases. In Incopat, 5,430 genetically modified corn patents and 7,583 forward citation patents were retrieved. In Lens, 37,062 patent-to-paper correspondence records were obtained. In Web of Science, 10,277 referenced papers were retrieved. Results are summarized in Table 2 .

3.2 Multi-Layer Network Construction

Based on extracted relationships, three single-layer networks were constructed: (1) Patent-patent citation network (single-mode network B) with patent IPC classifications as nodes and citation relationships as edges, yielding 78 nodes and 423 edges; (2) Discipline-discipline co-occurrence network (single-mode network A) with WOS categories as nodes and co-occurrence relationships as edges, yielding 132 nodes and 844 edges; and (3) Patent-paper citation network (bipartite network X) with patent IPC classifications and WOS categories as nodes and citation relationships as edges, yielding 1,005 edges. These networks were combined into a multi-layer model (A&X&B), as shown in Figure 4 [Figure 4: see original paper].

3.3 MERGM Model Specification

Six parameters were selected based on network node connection directionality and parameter graph structure similarity, as detailed in Table 3 . The model identifies two knowledge flow structures: Structure 1 (knowledge flow tendency between technology fields generated from cross-disciplinary knowledge) and Structure 2 (knowledge flow tendency between technology fields generated from two cross-disciplinary subjects). Parameters L3AXBin and L3AXBpath test whether technology nodes citing cross-disciplinary knowledge more easily absorb or diffuse technological knowledge. Parameters AinASXAinBS and AinASXAoutBS examine knowledge absorption and diffusion when cross-disciplinary scope is extensive. Parameters C4AXBexchangeAreciprocity and C4AXBexchangeBreciprocity test unidirectional and interactive knowledge flow tendencies between technology nodes citing corresponding cross-disciplinary knowledge.

3.4 Micro-Structure Characteristics of Science-Technology Knowledge Flow

MPNet software calculations yielded five significant parameters (three positive, two negative), revealing two micro knowledge flow structures.

3.4.1 Structure 1: Cross-Disciplinary Technology Field Knowledge Flow

This structure exhibits two patterns based on cross-disciplinary quantity differences:

Pattern 1: Single Cross-Disciplinary Knowledge Utilization and Gene Manipulation Technology Absorption

When citing a single cross-disciplinary field, technology nodes show poor knowledge absorption and diffusion. As shown in Table 4 , parameters L3AXBin (-0.0225*) and L3AXBpath (-0.0285*) are significantly negative, indicating that genetically modified corn technology fields citing cross-disciplinary knowledge do not easily absorb or diffuse technological knowledge. Case analysis reveals that patent A1 identified an optimal natural genomic locus, developing a new gene sequencing tool method that absorbed zinc finger protein positioning and gene targeting methods from other technology fields, referencing interdisciplinary papers in mathematical and computational biology, biotechnology, and applied microbiology (Table 6).

Pattern 2: Multi-Disciplinary Knowledge Utilization and Gene Screening Technology Diffusion

When citing multiple cross-disciplinary fields, technology nodes exhibit strong diffusion. Patent B1 identified a herbicide-related gene and designed recombinant vectors, referencing interdisciplinary papers in pharmacology, biotechnology, and immunology regarding streptomycin resistance gene cloning and characterization (Table 7). This technology showed strong diffusivity, with its resistance gene screening method cited by other patents for improvements in corn genomic loci and herbicide tolerance.

3.4.2 Structure 2: Two Cross-Disciplinary Technology Field Knowledge Flow

This structure shows that technology fields based on extensive cross-disciplinary knowledge have strong absorption and diffusion capabilities. As shown in Table 5 , parameters AinASXAinBS (0.0825*) and AinASXAoutBS (0.0874*) are significantly positive, indicating that when disciplinary cross-scope is large, genetically modified corn technology fields demonstrate strong knowledge absorption and diffusion. Parameter C4AXBexchangeAreciprocity (0.0161*) is significantly positive, showing that when corresponding scientific fields intersect, technology nodes more easily absorb domain knowledge and diffuse to multiple genetically modified corn technologies.

Case Analysis: Multi-Disciplinary Knowledge Utilization and Genetic Trait Transmission

Patent C1 discovered a Class II EPSPS gene for enhanced glyphosate herbicide tolerance in bacteria and plants, referencing interdisciplinary papers in biochemistry, molecular biology, and genetics regarding aromatic amino acid biosynthesis and broad host-range plasmid nucleotide sequences (Table 9). This technology not only absorbed research methods from paper C1 but also inherited nucleotide sequence analysis methods from patent C2, ultimately forming the commercially valuable genetically modified corn variety MON87403.

4 Outlook

This study constructs a two-layer science-technology knowledge flow network and random graph model, extending traditional single-layer bibliometric networks to multi-layer structures and providing new perspectives for bibliometric research. The framework can visualize abstract knowledge diffusion and absorption processes and identify micro knowledge flow structures. Based on the identified structures, innovation entities can focus on cross-disciplinary scientific nodes and associated technologies to discover potential technological breakthroughs.

However, limitations exist. The network construction relies on patent-paper citation relationships, requiring high citation density. The method is more suitable for fields like information science and bioscience with frequent patent-paper citations, but less effective for mathematics or social sciences with sparse citation relationships. The current two-dimensional framework could be extended by incorporating additional knowledge subject types, enriching knowledge attribute factors, and expanding knowledge flow paths. Future research could explore relationships between knowledge flow structures and domain characteristics, providing theoretical support for innovation policy and technology development strategies.

References

[1] SHIBATA N, KAJIKAWA Y, SAKATA I. Extracting the commercialization gap between science and technology - Case study of a solar cell [J]. Technological forecasting and social change, 2010, 77(7): 1148-1155.

[2] LING G, DUAN Y Z, LI S. Creating mechanism of strategic emerging technologies based on knowledge flows[J]. Pioneering with science & technology monthly, 2021, 34(5): 43-46.

[3] JAFFE A B. Real effects of academic research [J]. The American economic review, 1989, 79(5): 957-970.

[4] AUTANT-BERNARD C. Science and knowledge flows: Evidence from the French case[J]. Research policy, 2001, 30(7): 1069-1078.

[5] LIU Y M, ZHENG S P. Bridging science, engineering science and engineering in innovation chain taking national-level collaborative innovation centers as examples [J]. Studies in science of science, 2022, 40(10): 1745-1755.

[6] PAN W W, JIAN L R, LIU T. Knowledge generation and diffusion in science & technology: An empirical study of SiC-MOSFET based on scientific papers and patents[J]. Technology analysis & strategic management, 2022: 1-17.

[7] WEN J Y. A critical thinking on commercialization of scientific and technological knowledge[J]. Science research management, 2019, 40(5): 175-181.

[8] NARIN F, HAMILTON K S, OLIVASTRO D. The increasing linkage between U.S. technology and public science [J]. Research policy, 1997, 26(3): 317-330.

[9] LAI K K, ZHANG S B. Constructing a business method technology diffusion model: Integrating patent citations and Bayesian model[J]. Journal of science and technology management, 2004, 9(1): 1-34.

[10] VERBEEK A, DEBACKERE K, LUWEL M. Science cited in patents: A geographic "flow" analysis of bibliographic citation patterns in patents[J]. Scientometrics, 2003, 58(2): 241-263.

[11] GAO J P, DING K, TENG L, et al. Analysis on knowledge flow in hybrid documents co-citation network[J]. Studies in science of science, 2011, 29(8): 1184-1189, 1146.

[12] ZHAO L M, GAO Y, HAN Y. Application of patent citation analysis to the research of knowledge-transfer mechanism[J]. Studies in science of science, 2002, 20(3): 297-300.

[13] NARIN F, NOMA E. Is technology becoming science?[J]. Scientometrics, 1985, 7(3): 369-381.

[14] SCHMOCH U. Tracing the knowledge transfer from science to technology as reflected in patent indicators[J]. Scientometrics, 1993, 26(1): 193-211.

[15] Clarivate Analytics. Web of Science [EB/OL]. https://clarivate.com/webofsciencegroup/solutions/web-of-science/

[16] STOLPE M. Determinants of knowledge diffusion as evidenced in patent data: The case of liquid crystal display technology[J]. Research policy, 2002, 31(7): 1181-1198.

[17] PARK G, PARK Y. On the measurement of patent stock as knowledge indicators [J]. Technological forecasting and social change, 2006, 73(7): 793-812.

[18] MEYER M, DEBACKERE K, GLÄNZEL W. Can applied science be good science? Exploring the relationship between patent citations and citation impact in nanoscience[J]. Scientometrics, 2010, 85(2): 527-539.

[19] GLANZEL W, MEYER M. Patents cited in the scientific literature: An exploratory study of 'reverse' citation relations[J]. Scientometrics, 2003, 58(2): 415-428.

[20] GAO J P, DING K, TENG L, et al. Implementation and application of co-citation analysis for "publications+patents literature": A case with Derwent innovation index [J]. Journal of the China society for scientific and technical information, 2012, 31(3): 317-324.

[21] LI C C. Dynamic mechanism of science and technology systematization and eight sources of knowledge on technological sciences[J]. Innovation science and technology, 2022, 22(11): 12-21.

[22] GAO J P, DING K, TENG L, et al. Hybrid documents co-citation analysis: Making sense of the interaction between science and technology in technology diffusion[J]. Scientometrics, 2012, 93(2): 459-475.

[23] HUANG M H, YANG H W, CHEN D Z. Increasing science and technology linkage in fuel cells: A cross citation analysis of papers and patents[J]. Journal of informetrics, 2015, 9(2): 237-249.

[24] YU D J, YAN Z P. Combining machine learning and main path analysis to identify research front: From the perspective of science-technology linkage[J]. Scientometrics, 2022, 127(7): 4251-4274.

[25] BATTISTON F, NICOSIA V, LATORA V. Structural measures for multiplex networks[J]. Physical review E, Statistical, nonlinear, and soft matter physics, 2014, 89(3): 032804.

[26] WANG P, ROBINS G, PATTISON P, et al. Exponential random graph models for multilevel networks[J]. Social networks, 2013, 35(1): 96-115.

[27] BERLINGERIO M, COSCIA M, GIANNOTTI F, et al. Foundations of multidimensional network analysis[C]//2011 International Conference on Advances in Social Networks Analysis and Mining. Piscataway, New Jersey: IEEE, 2011: 485-489.

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