Study on Innovation Path Identification Based on Topics Association of Science and Technology

Haiyun Xu, Chunjiang Liu, Rui Luo, Ziqiang Liu and Yan Qi

Chengdu Library and Information Center, Chinese Academy of Sciences
University of Chinese Academy of Sciences

2018.9.11.Leiden.Netherlands

OUTLINE

- 1. Background
- 2. Methodology
- 3. Empirical Study
- 4. Discussion & Conclusion

1. BACKGROUND

- **Our research work**

Our study focuses on both science and technology (S-T) through bibliometric analysis, exploring the identification method of innovation paths based on the topic association of science and technology. Our study process is as follows:

- Firstly, we define the connotations, concepts and summarize the research status of scientific innovation path.
- Secondly, we discuss the relationship between science and technology. In order to identify the innovation evolution path in an interactive perspective, we reveal the interaction of science and technology by analyzing the relevance among science and technology innovation topics in a micro level.
- Thirdly, the genetic engineered vaccine (GEV) is selected as our empirical field.
- Finally, we summarize the contributions and limitations of this study.

1. BACKGROUND

- **Innovation Path Identification**
  - The evolution path of science and technology (S-T) refers to the development and evolution of innovation topics, reflecting the emergence, diffusion and evolution of technological innovation.
  - There are two innovative elements in the innovation process, one spreads from science to technology, and the other from technology to science. Two innovative elements jointly determine direction of scientific development（Figure1）.

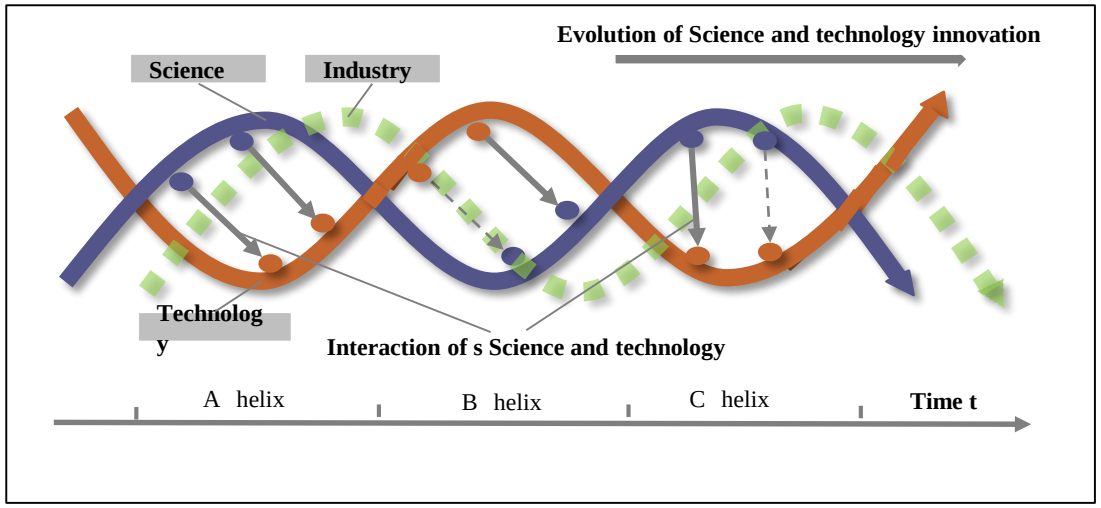

Figure 1 Interaction of Science and Technology

It is important to study the identification method of the innovation path based on science-technology topic association.

2. RESEARCH STATUS OF SCIENTIFIC INNOVATION PATH

2.1 Research on Innovation Evolution Path Identification Method

（1）Innovative Path Identification Based on Citation Analysis

- The Citation Main Path
- Citation Cluster

（2） Identification Based on Text Topic Word Analysis

- Co-word analysis
- Semantic-enhanced Co-word Analysis
- Dynamic Evolution of Keyword Analysis

（3）Identification Based on Multivariate-relationship Fusion

In summary, the existing bibliometric methods mostly regarded research papers as representing the scientific research and patents as representing of the technological innovation.

1. BACKGROUND

- **Status of scientific innovation path**

- There are mainly two problems in the identification of scientific innovation paths research.

  - First, previous identification methods mainly focus on single innovation element of science or technology, ignoring the intrinsic relevance them.

  - Second, the quantitative researches on the relevance of science and technology are limited to the numeral feature of literatures, which can't reveal the inherent relevance of science and technology.

- The two problems makes it difficult to fully grasp the features of scientific innovation, and then affect the accuracy of the innovation evolution path identification.

2. METHODOLOGY

➢ What's the correlation of science and technology

- **Herbert Simon thinks scientists care about--how thing are, while engineer focus on how things ought to be.**

- Research papers and patent data are now widely used in hot and cutting-edge identification in interdisciplinary and technology integration forecasting. With the rapid accumulation of scientific literatures, the scientific measurement method has become the main method for quantitatively analyzing the science-technology association.

- In many scientific and technological documents, scientific papers are the fruits of scientific research, carrying the basic scientific knowledge. While patent documents are the fruits of technological.

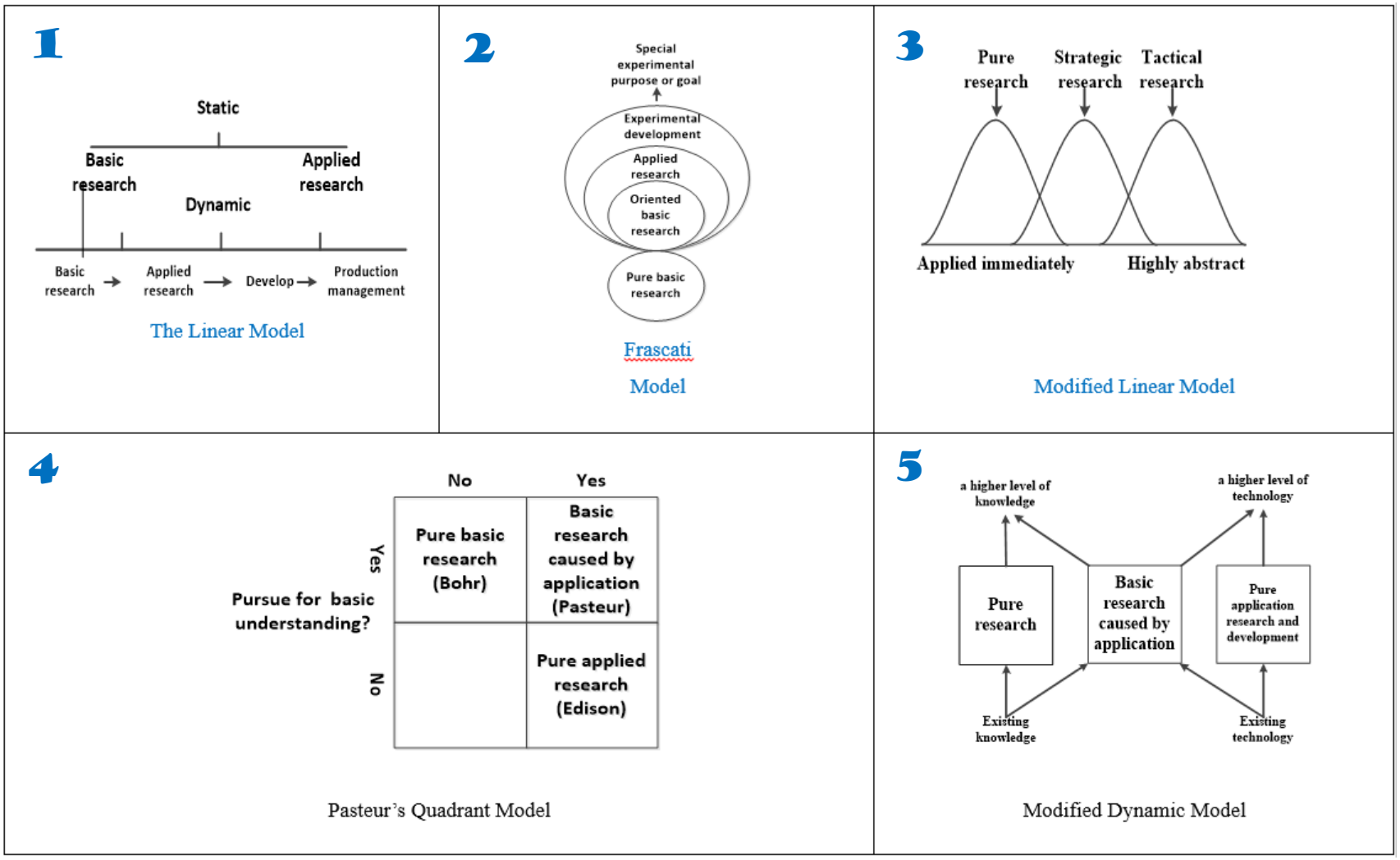

Figure 2 Relationship Model Between Science Research and technology Research

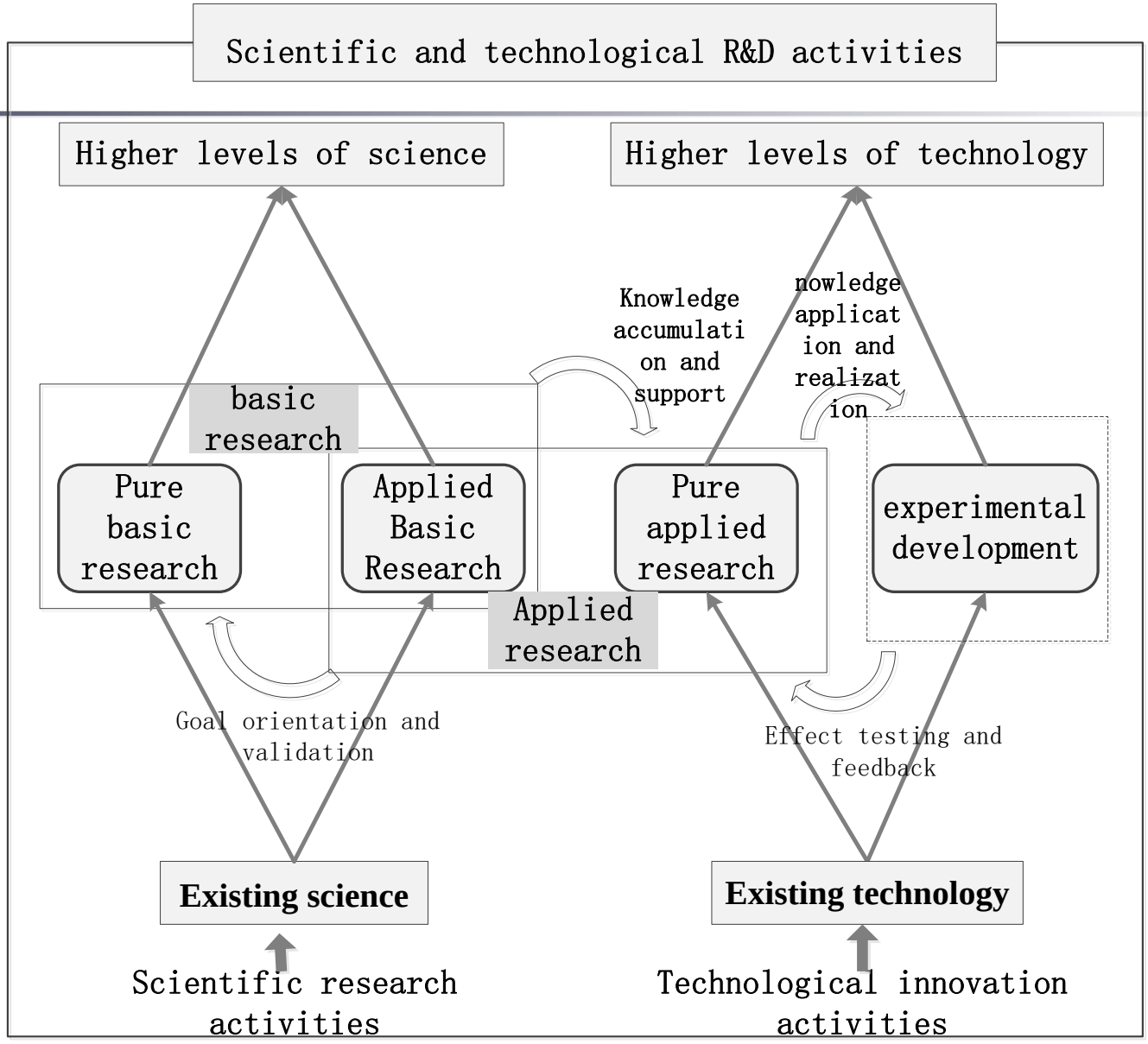

Figure 3 Schematic Diagram of the Relationship Between Science and Technology

3 The Research on the Relationship Between Science and Technology

- **The Linkage Between the Dynamic Mechanism of Science and Technology**

  - Firstly, scientific advancement and technological evolution follow the same logic, which means they share the same self-organizing dynamics mechanism with the features of the ordering and logic of knowledge structure.
  - Secondly, both the theoretical systems of science and technology are purposeful, so they follow the same logic. Both science and technology are based on the existing research and present a chain.
  - Thirdly, this consistency does not mean that science and technology are totally the same. Actually, the development of science and technology is very different in the development mechanism, because scientific development rely on more reasoning, argumentation and logical support, while technological development depends on the combination and superposition of existing technologies.

1. BACKGROUND

- **Relationship of Science and Technology**

In summary, science and technology are both unique and interrelated. So just as figure 1, the interaction of science and technical knowledge forms a complex system of knowledge evolution in the process of innovation.

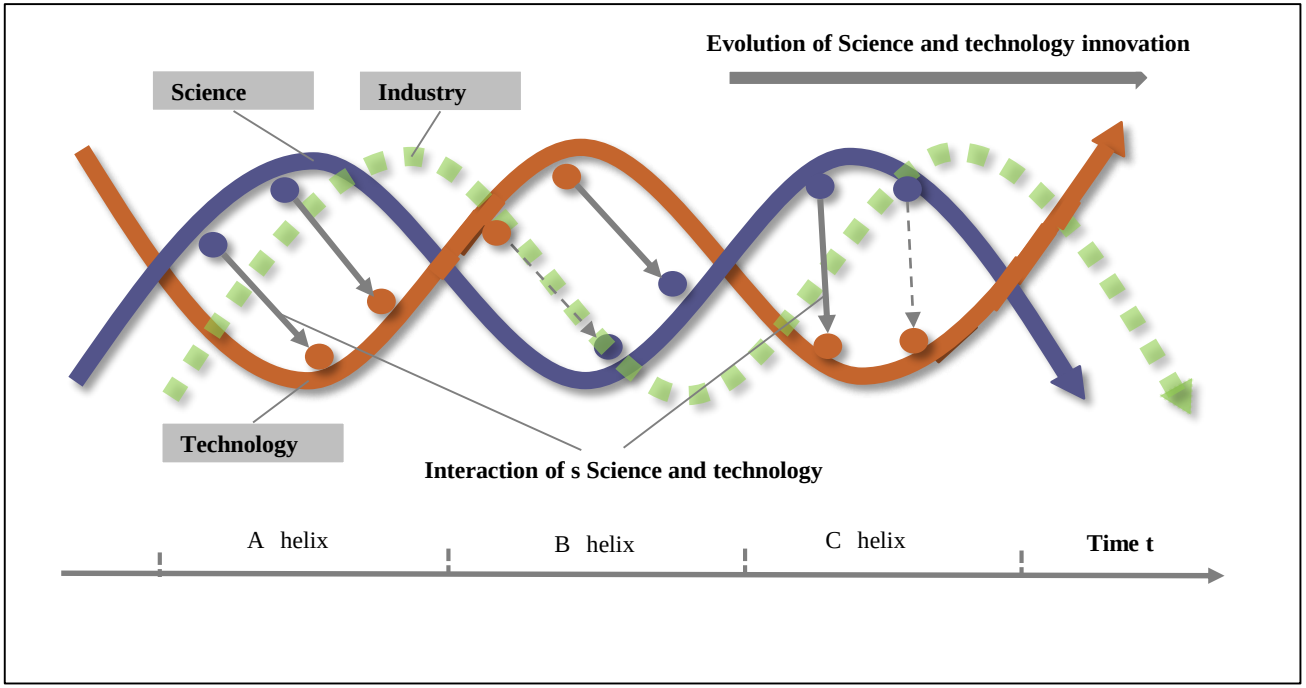

Fig4 Interaction of Science and Technology

4.1 Path Identification Model based on of Topic Relevance of S-T

- **How to evaluate the correlation of science and technology**
- Study on the science-technology relationship only shows the current topic and the internal activities, cannot present the dynamic relation of science and technology in a specific field. So there is Time Variable Added.
- The schematic diagram of identifying the science-technology interaction pattern (Fig5). S_1... S_n are scientific topics, T_1... T_n are technical topics. All the scientific topics and technical topics are distributed on the timeline t, the line thickness indicates the different intensity of topic relevance.

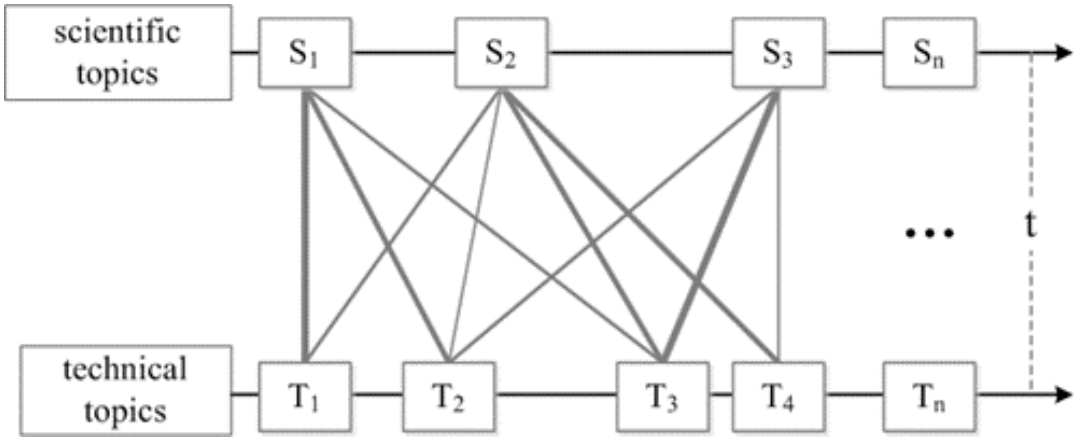

Figure 5 The schematic diagram of identifying the science-technology interaction pattern

4.2.1 Multi-relationships in the Topic Relevance of Science and Technology

- **Text topic identification and similarity calculations are the key points in the study of association between science and technology innovation topics.**

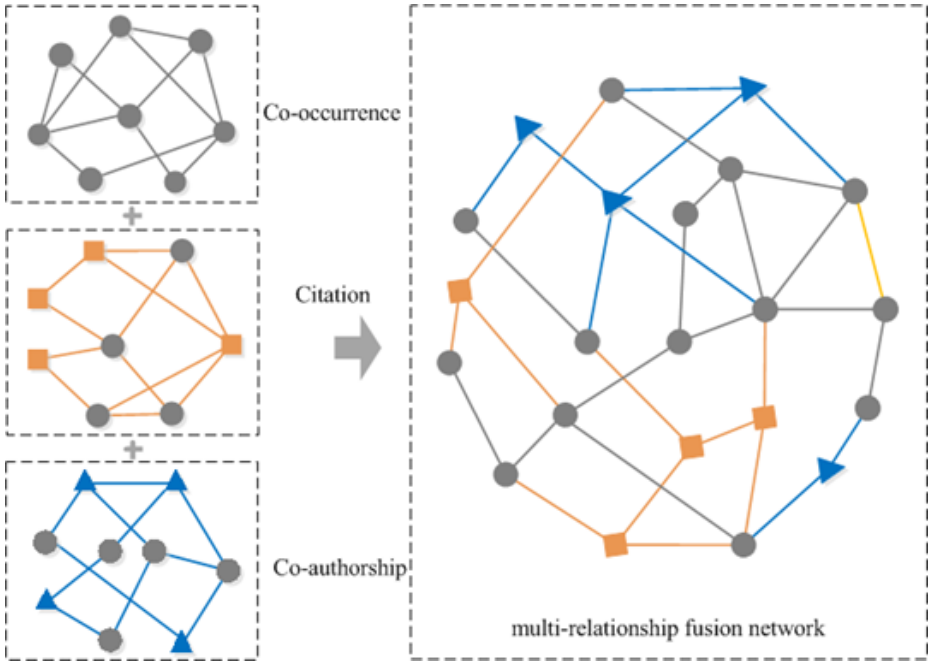

Figure 6 Entity relationship fusion

- The semantic relations linked by the research topic is among the topic terms, the author and the citation of the scientific literature, the subject term co-occurrence relationships, the citation relationship, and the co-authorship all reveal thematic relationships from different perspectives.
- Topic relevance to be fused includes three types:(fig 6).

4.2.2 Topic Relevance Calculation of S-T Based on the Multi-relation Fusion

In figure 7, the topic S contains N_S topic terms (blue dots), and the topic T contains N_T topic terms (yellow dots), and the correlation between the topic S and the topic T is a comprehensive calculation based on the topic correlation degree of the co-word relevance, the author relevance and the citation relevance.

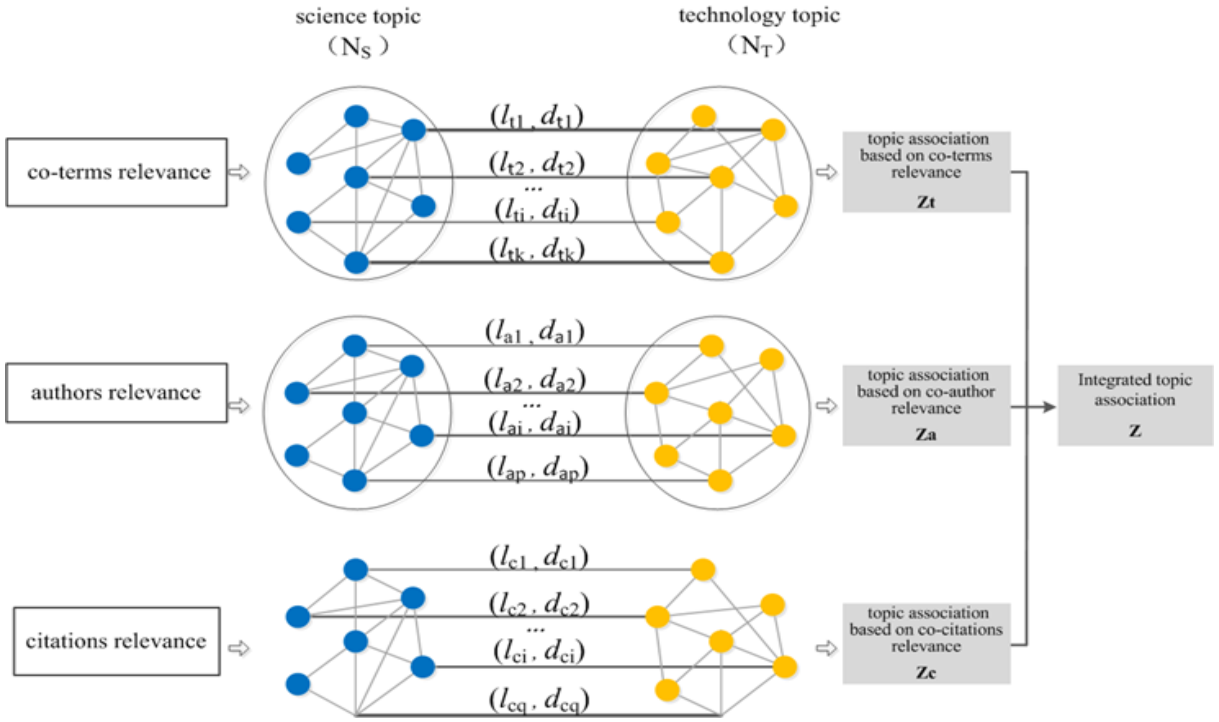

Figure 7 Calculation method of topics relevance

4.2.3 The Calculation of Topic Distribution time

- This study regards the number of the scientific documents contained in the topic as the impact of the topic.
- There is a breakthrough in the key issues of the topic.
- People's attention continuously will go to the study of the topic in the following time, with the number of the published topic-related documents continuing to increase.
- If the finding is expressed by a growth function curve, this point is the point at which the positive growth rate changes, which means the **inflection point of the growth function** and it has the highest impact (Figure 8).
- The topic near the time of the inflection point is in a period of rapid growth, and the published scientific documents have accumulated to a certain quantity.

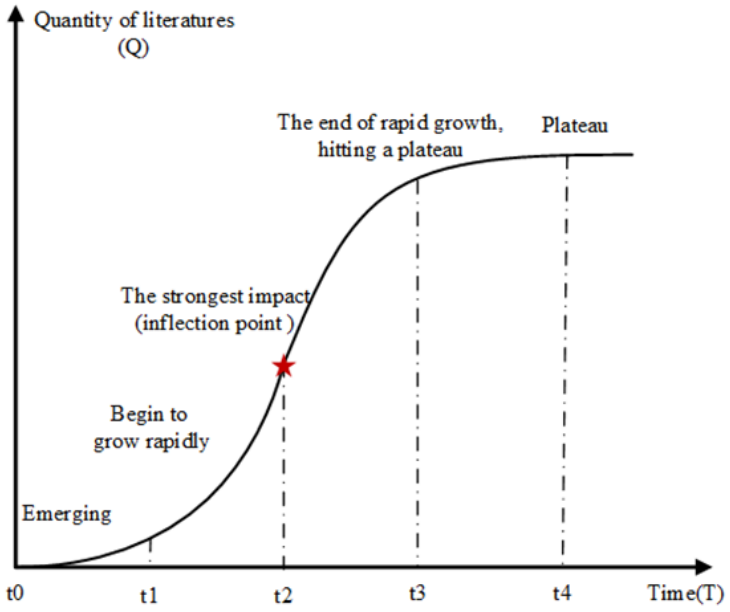

Figure 8 The Schematic Diagram of Discovering Topic Distribution Time

3. EMPIRICAL STUDY

- **Data sources and analysis tools**
  - In this study, Papers and patent documents are collected in the GEV field for a period as analysis data sets.
  - Web of Science database of is selected to search the scientific papers, and the Derwent Innovations Index database is used to search the patent documents.
  - Retrieval strategy

    TI=((Genetic* adj engineer* or DNA adjengineer*) and (vaccine* or antigen*)) or TS=((Genetic* adj engineer* or DNA adj engineer*)and (vaccine* or antigen*)) not TI=(test or immunoassay adj detect* or detect*) or TI=((nucleic* adj acid* or RNA) same (vaccine* or antigen*)) or TI=((plasmid* adj DNA) and (vaccine* or antigen*))

➢The search date is January 6, 2018, and the publication year is up to 2017, and the paper record is 4146 and patent is 4050.

5.2 Scientific and Technical Classification of Papers

The proportion of papers biased to technical features is getting higher in recent years，and this is mainly due to the early research focusing more on the basic theory, but with the mature of the theory research, the number of the paper about applied research are gradually increasing (Fig.9).

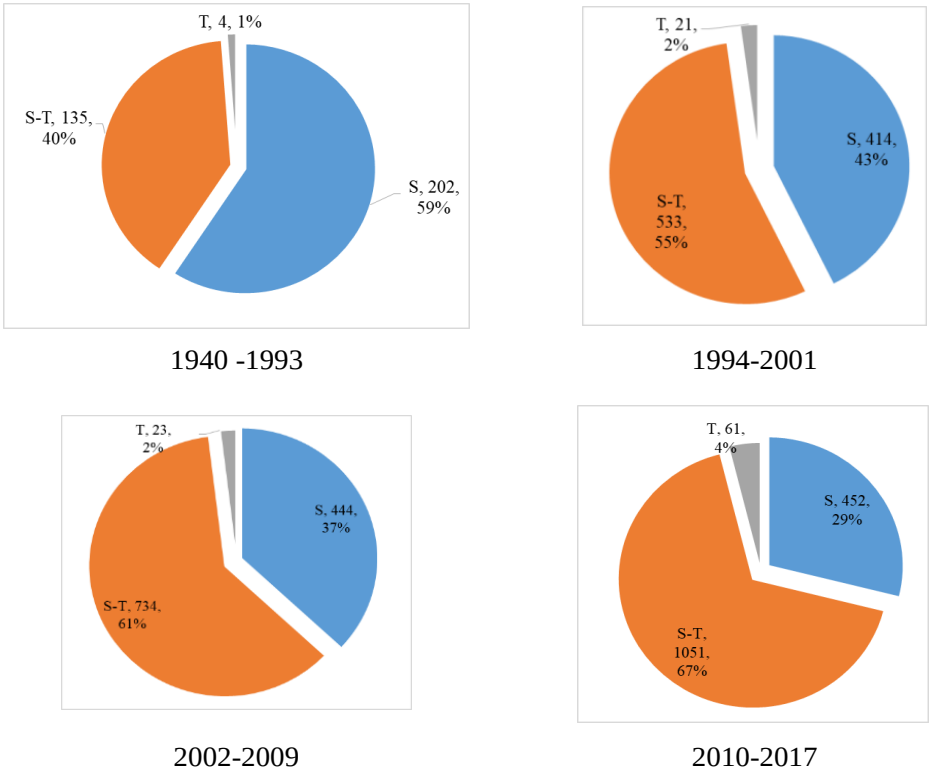

Figure 9 Scientific - technical Classification Results of Papers

5.3 Analysis Result

5.3.1 The Results of Topic Identification

LDA (Latent Dirichlet Allocation) topici model, we finally identify 15 scientific topics and 15 technical topics (Table 1 and Table 2).

Table 1 The Identification of Scientific Topics (15 topics)

wos	Feature Words
0	genetic, vaccines, engineering, development, based, applications, host, clinical, molecular, production, therapeutic, disease, bacillus, bacterial, analysis, artificial, strategies, research, control.
1	virus, vaccine, influenza, engineered, genetically, attenuated, live, vaccines, viruses, type, encephalitis, infection, herpes, salmonella, mice, strain, simplex, system, strains.
2	protein, antigen, rna, cell, binding, nuclear, activity, proteins, region, large, sequence, factor, barr, epstein, mediated, regulation, recognition, dna, intracellular.
3	coli, escherichia, toxin, recombinant, mice, expression, protein, production, subunit, producing, gondii, lactis, lactococcus, fusion, heat, toxoplasma, brucella, protection, expressed.
4	cells, engineered, genetically, human, stem, cell, derived, vitro, bone, alpha, gene, experimental, marrow, vivo, transgenic, mice, transplantation, autoimmune, cultured.
5	hepatitis, virus, rna, antigen, delta, hcv, core, viral, surface, antigenomic, replication, infected, genomic, patients, hbv, negative, antigens, ns, synthesis.

5.3 Analysis Result

5.3.1 The Results of Topic Identification

Table 2 The Identification 15 technical Topics (15 topics)

DII	Feature Words
0	antigen, antibody, detecting, nucleic, sample, acid, comprises, detection, reaction, enzyme, specific, target, involves, antigens, probe, antibodies, reagent, assay, presence.
1	vaccines, nucleic, acid, acids, encoding, proteins, diagnosis, infections, antibodies, antigens, treatment, related, infection, protein, bacterial, polypeptides, diagnostic, polypeptide, derived.
2	cancer, tumor, treating, antigen, prostate, preventing, rna, comprises, molecule, preparing, human, factor, specific, cell, diagnosing, inhibiting, growth, subject, expression.
3	treating, disease, diseases, disorders, autoimmune, cancer, preventing, treatment, human, inflammatory, polypeptides, arthritis, polypeptide, diabetes, encoding, immune, diagnosing, cancers, disorder.
4	vaccine, virus, disease, protein, preparing, engineered, gene, genetically, strain, preventing, comprises, type, treating, infectious, mouth, subunit, foot, porcine, recombinant.
5	comprises, surface, polymer, material, agent, active, comprising, substance, lipid, cationic, delivery, acid, drug, carrier, compound, device, composition, substrate, membrane.

5.3 Analysis Result

5.3.1 The Results of Topic Identification

It can be seen that most of the topics are concentrated in 2000.

Table 3 The Years of Topics' Distribution

Scientific topics	Year	Records	technical topics	Year	Records
0	2015	148	0	2016	295
1	2017	303	1	2001	526
2	2017	206	2	2001	302
3	2015	148	3	2003	246
4	2013	181	4	2013	306
5	2002	218	5	2000	178
6	2015	278	6	2006	356
7	2000	381	7	2012	274
8	1996	135	8	2009	183
9	2007	665	9	2000	402
10	2009	183	10	2008	552
11	2000	810	11	2008	229
12	1982	444	12	2016	198
13	2017	275	13	2003	89
14	1994	141	14	2001	188

5.3.3 topic-related path of science and technology

- Figure 10 shows the evolution of innovation path in the GEV field with the threshold value of 0.5 based on the Wordnet semantic vocabulary.
- In order to obtain a clearer trend of topics evolution, we further increases the threshold displayed by the topic nodes.

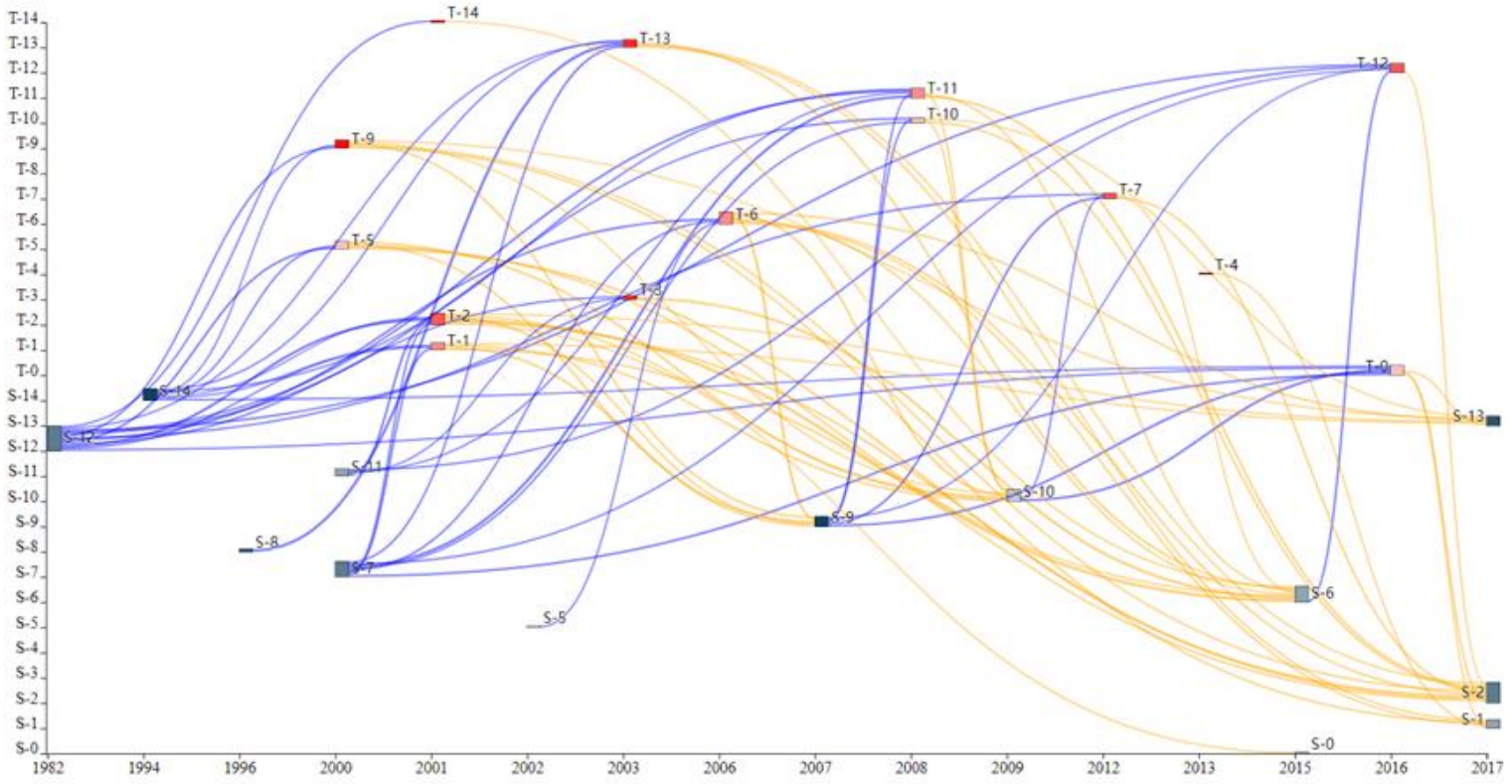

Figure 10 Scientific - technical Classification Results of Papers（threshold value of 0.5）

5.3.3 topic-related path of science and technology

Figure 11 shows evolution of innovation path with a threshold of 0.6 based on the Wordnet semantic vocabulary.

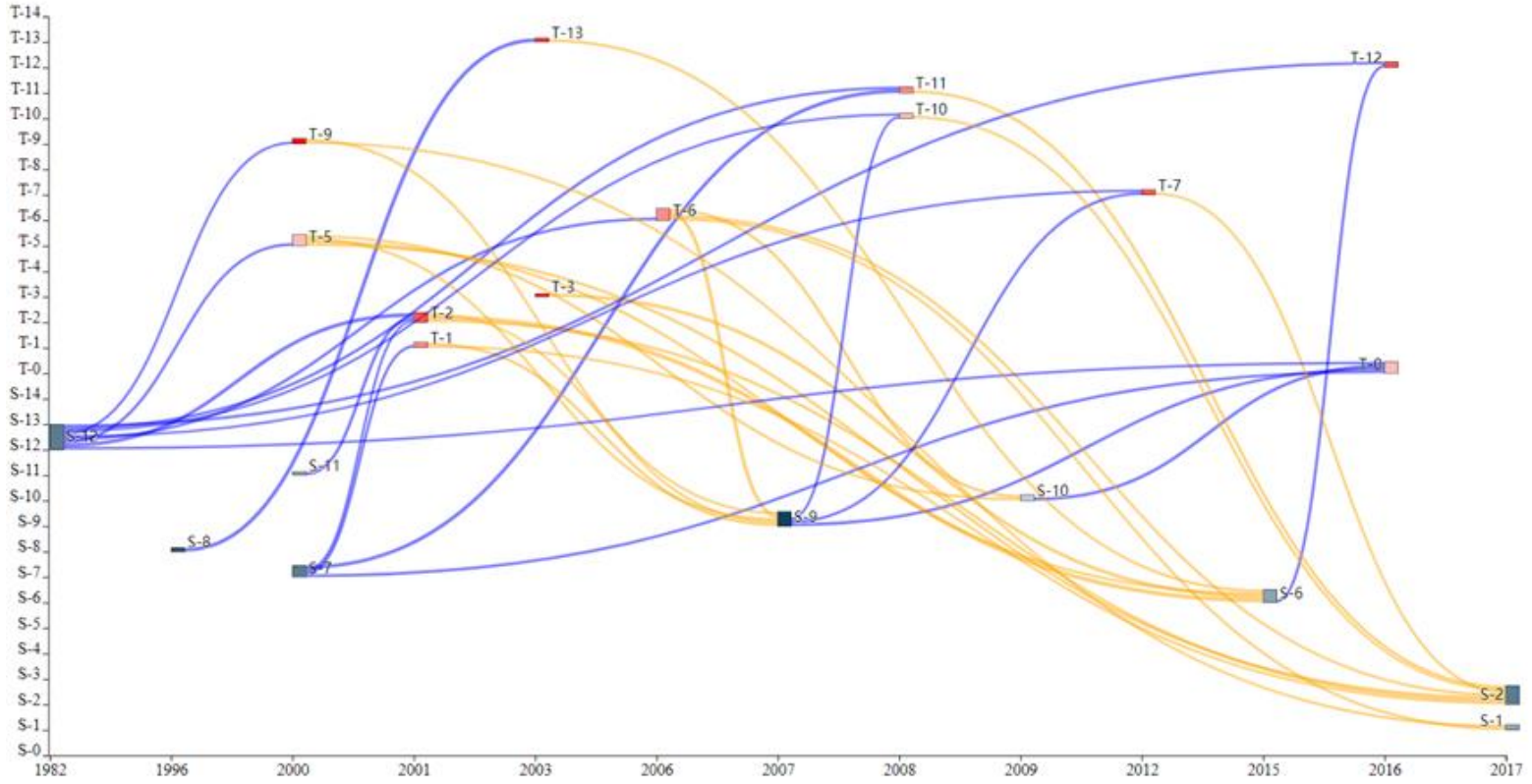

Figure 11 Scientific - technical Classification Results of Papers（threshold value of 0.6）

5.3.3 topic-related path of science and technology

- The Figure 12 shows the evolution of innovation path in GEV with the threshold of 0.5 after the fusion of four fused data type.
- In order to obtain a clearer trend of topics evolution, we also further increases the threshold displayed by the topic nodes.

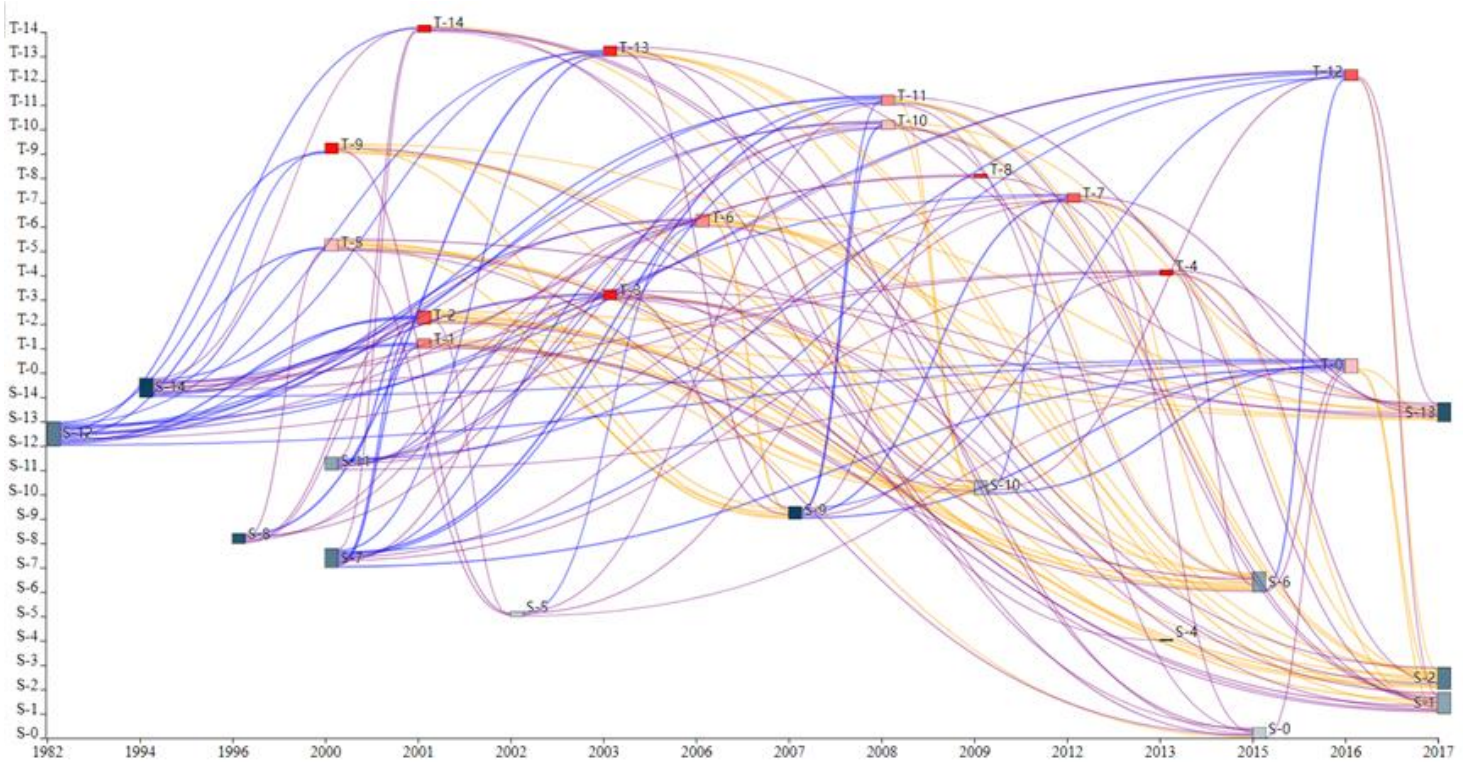

Figure 12 Scientific - technical Classification Results of Papers（threshold value of 0.5）

5.3.3 topic-related path of science and technology

Figure 13 shows the evolution of innovation path in GEV fild with the threshold of 0.6 after the fusion of four fused data.

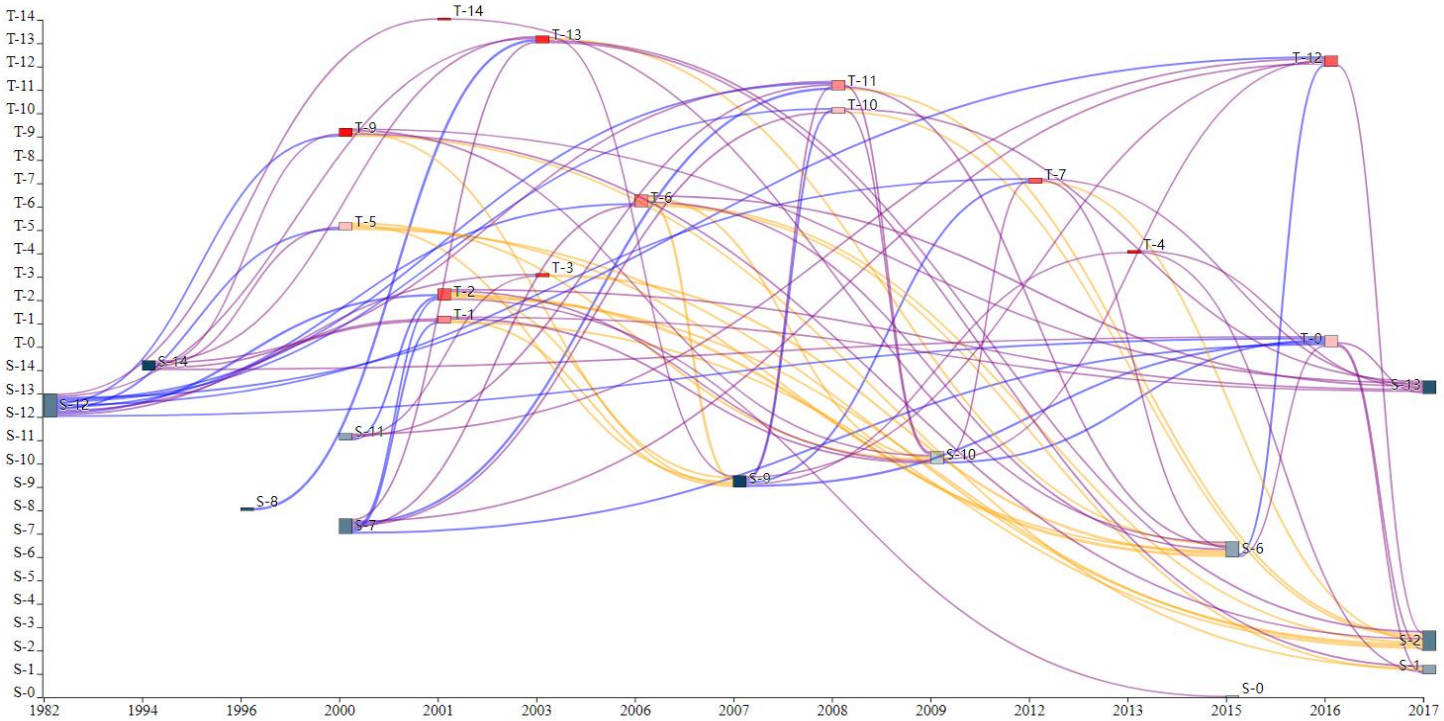

Figure 13 Scientific - technical Classification Results of Papers （threshold value of 0.6）

4. DISCUSSION & CONCLUSION

- Comparative analysis shows that compared to a single path of science or technology innovation, innovation paths associated with S-T topics can quantitatively reveal S-T interactions from a micro level and more fully demonstrate the development laws and evolution characteristics of technological innovation, thus eliminating the current one-sidedness of only one innovative elements in the innovative path.
- **Text topic similarity calculations are the key points in the study of association between science and technology innovation topics.** The method based on the word correlation can reflect the correlation degree between S-T to some extent, but the result of this method is not such reliable due to the inconsistencies in using words and concepts between the patents and the documents.

4. DISCUSSION & CONCLUSION

- In the future, the integration of S-T semantic associations is a good choice to avoid the preceding deficiency. What's more, scientific-technical semantic association integration can be achieved through feasible algorithm design lately.
- With the explosive growth of the number of scientific literature and the continuous enrichment of the types of documents, the relationships types that can be analyzed by scientific measurement are constantly expanding. In addition to citation analysis and text topic analysis, multiple coupling analysis among different entities has been successfully applied, and multi-source heterogeneity has become an important and common form of data existence. Therefore, how to deal with multi-source heterogeneous data through data fusion has become a new problem in scientific measurement.

PART OF REFERENCES

- Zhang Y, Guo Y, Wang X, et al. A hybrid visualisation model for technology roadmapping: bibliometrics, qualitative methodology and empirical study[J]. Technology Analysis & Strategic Management, 2013, 25(6): 707-724.
- Zhang Y, Zhang G, Chen H, et al. Topic analysis and forecasting for science, technology and innovation: Methodology with a case study focusing on big data research[J]. Technological Forecasting & Social Change, 2016, 105: 179-191.
- Dong K, Xu H, Luo R, et al. A Review of the Research on the Relationship between Science and Technology[J]. Journal of The China Society for Scientific and Technical Information, 2018, in press.
- Kostoff R N, Schaller R R. Science and technology roadmaps[J]. Engineering Management IEEE Transactions on, 2001, 48(2): 132-143.
- Hummon N P, Dereian P. Connectivity in a citation network: The development of DNA theory ☆[J]. Social Networks, 1989, 11(1): 39-63.
- Martinelli A. An emerging paradigm or just another trajectory? Understanding the nature of technological changes using engineering heuristics in the telecommunications switching industry[J]. Research Policy, 2012, 41(2): 414-429.
- Lu L Y Y, Liu J S. A survey of intellectual property rights literature from 1971 to 2012: The main path analysis[C]. Portland International Conference on Management of Engineering & Technology, 2014: 1274-1280.
- Pilkington A, Meredith J. The evolution of the intellectual structure of operations management—1980–2006: A citation/co-citation analysis[J]. Journal of Operations Management, 2009, 27(3): 185-202.
- Lai R J, Li M F. Technology Evolution of Lower Extremity Exoskeleton from the Patent Perspective[J]. Key Engineering Materials, 2015, 625: 536-541.
- Coulter N, Monarch I, Konda S. Software engineering as seen through its research literature: a study in co-word analysis[M]. John Wiley & Sons, Inc., 1998: 1206-1223.
- Kim Y G, Suh J H, Sang C P. Visualization of patent analysis for emerging technology[J]. Expert Systems with science?[J]. Scientometrics, 1985, 7(3-6): 369-381.

ACKNOWLEDGEMENT

This work in this paper was supported by Natural Science Foundation of China (Grant No. 71704170), supported by China Postdoctoral Science Foundation funded project）(Grant No. 2016M590124) and supported by Youth Innovation Promotion Association (Grant No. 2016159).

.

Thanks for your attention!

Haiyun Xu

(Chengdu Library and Information Center,

Chinese Academy of Sciences)

Add: P.O. Box 20, No. 16 South Sec. 2 Yihuan Rd., Chengdu 610041, P.R. China

Phone: +86-28-85228846

Fax: +86-28-85220439

Email: xuhy@clas.ac.cn