

# Research on the topic discovery and hotspot evolution of the LIS field from the perspective of papers supported by the National Foundation

Yue Wang\*

School of Economics & Management, Nanjing University of Science & Technology, Nanjing, China

\* Corresponding Author Email: wy\_unicornx@163.com

**Abstract.** Exploring the fine-grained characteristics of the results of papers funded by national funds is helpful to help researchers more accurately grasp the research hotspots and frontier trends of the discipline to identify the frontier topics of scientific research. Starting from the projects approved by the National Natural Science Foundation of China and the Social Science Foundation of China from 2010 to 2019, this study took the state-funded papers published in the core journals in the field of Library and Information Science from 2010 to 2020 as the research object, constructed a domain dictionary according to the funds and keywords of the papers, and conducted topic discovery for the abstracts of the funded papers based on the LDA model. And explore the evolution trend of common keywords in highly cited papers. The results show that in the field of LIS, the social science fund has more diversified project themes than the natural science fund. Information resource construction and library management research are hot topics in the LIS discipline. With the development of The Times, topics such as information service, network public opinion, and big data gradually become the focus of research.

**Keywords:** National foundation papers; LDA model; Topic discovery.

## 1. Introduction

Science fund is one of the most effective and widely used basic research funding methods in scientific research and technology development activities [1], such as the National Natural Science Foundation of China and the National Social Science Foundation of China. Library and Information Science (LIS), as a comprehensive interdisciplinary subject, has made great achievements in applying and approving projects funded by the natural science Foundation and social science foundation. Fund projects are the forward-looking layout of scientific research management institutions, and papers are the embodiment of the phased achievements of scientific research [2]. Therefore, this paper starts from the projects approved by the LIS discipline in the National Natural Science Foundation of China and the National Social Science Foundation of China from 2010 to 2019 and takes the papers published in the core journals in the LIS field of the corresponding fund projects from 2010 to 2020 as the research object and makes an econometric analysis of the paper results. Based on the LDA model, the topic mining of the abstracts of the funded papers is carried out, aiming to help researchers with related disciplinary backgrounds intuitively understand the actual situation of the project proposal and paper results of the discipline of information technology, understand the research themes and frontier trends of the discipline in the past decade and have a basic and holistic understanding of the development direction of the discipline.

## 2. Data and Method

Firstly, 1404 social science fund projects and 232 natural science fund projects in LIS were obtained from the National Social Science Fund Project and Science Network Fund databases, respectively. Secondly, a total of 23 Chinese journals in the LIS field included in CSSCI from 2010 to 2020 were selected. The publication time was set as January 1, 2010, to December 31, 2020. The supporting funds were set as "National Natural Science Foundation of China" and "National Social Science Foundation of China" respectively. The irrelevant literature and English literature were

excluded, and 4457 articles from the natural science Foundation were obtained. There were 10005 papers supported by the Social science Foundation. The project number of the fund was used to match the literature of the self-science foundation and Social science foundation respectively. Finally, 1620 papers were funded by the natural science Foundation, and 6057 papers were funded by the Social science Foundation. The data analysis of the paper results mainly includes the following three steps: the first step is to conduct descriptive statistics on the external characteristics of the paper. The second step is to construct a domain dictionary based on the keywords in the fund project and thesis catalog data, and the generated dictionary contains a total of 17470 domain-related words. The third step is to identify and name the topic of the abstract text based on the LDA model and draw the evolution trend of high-frequency keywords.

### 3. Results

#### 3.1 Analysis of external characteristics of fund papers

##### 3.1.1 Analysis of the annual published quantity

From 2010 to 2020, 23 core journals in the field of mapping published 6057 papers funded by social sciences, and 19 core journals published 1620 papers funded by science. Analysis of Figure 1 shows that the number of papers produced by the social science fund is on the rise on the whole, and the number of papers produced by the self-science fund is less than that of the social science fund but shows a steady upward trend as a whole.

##### 3.1.2 Analysis of high fund paper output institutions

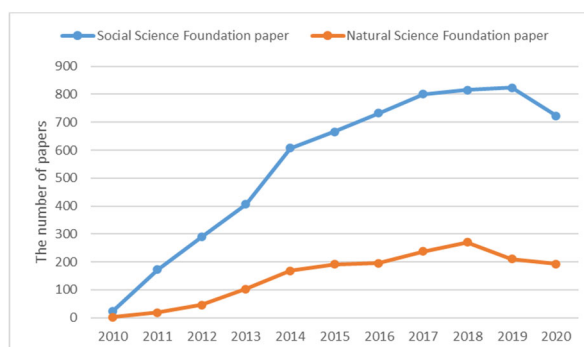


Fig. 1 Annual distribution of LIS papers funded by national funds

To explore the institutions with strong scientific research strength and the discipline construction of each institution in the field of LIS, the top 20 institutions with the highest number of published papers funded by social science were analyzed. Due to the relatively small number of papers funded by natural science, the top 10 institutions with the highest number of published papers funded by natural science were analyzed.

Among the institutions with a high yield of papers funded by social sciences, except for four research institutions (Institute of Scientific and Technical Information of China, Information Resources Research Center of Wuhan University, Institute of Scientific and Technological Information of Shandong University of Technology, and Information Resources Management Research Center of Heilongjiang University), the other 16 agencies are from secondary school in colleges and universities. Among them, School of Information Management of Nanjing University, School of Information Management of Wuhan University, and School of Information Management of Central China Normal University are leaders in the LIS field with strong discipline development levels and scientific research strengths. Among the institutions with a high yield of natural science foundation papers, except for the Information Resources Research Center of Wuhan University, the remaining 9 institutions are all secondary colleges. Among them, the top 3 institutions with the number of published papers are from Wuhan University and Nanjing University, which are the same as the institutions with the high production of social science-funded papers, and the top 4 institutions

with the high production of natural science-funded papers all have more than 100 papers. These four institutions have high scientific research output capacity and a level of the discipline development.

### 3.2 Literature topic mining based on the LDA model

This research uses the python scikit - learn Latent Dirichlet Allocation method to construct the LDA model in the library, in 2010-2019 national fund project corresponding areas of core journals in 2010-2020 for topic discovery. Finally, the topic clustering results of papers funded by the social science Fund and papers funded by the natural science fund are shown in Table 1 and Table 2, respectively.

**Table 1.** Topic clustering of Social science Foundation papers based on the LDA model

Topic ID	Topic	Keywords
#Topic0	Data mining and algorithm design	data, user, model, application, information, feature, experiment, foundation, resource, effective, concept, linked data, algorithm, knowledge, domain
#Topic1	Theory building and policy making	information, theory, technology, foundation, service, intelligence, big data, policy, application, government, enterprise, construction, perspective, environment, knowledge
#Topic2	Information Services and User Research	user, influence, library, service, university library, factor, behavior, influence factor, information, model, reader, use, data, salient, reading
#Topic3	Scientometrics and evaluation	field, literature, discipline, paper, index, topic, Journal, data, author, evaluation, domestic, scholar, domestic and foreign, important, identification
#Topic4	Organization and management of book archives	library, construction, service, public library, archives, practice, work, important, resources, content, culture, management, society, project, reference

**Table 2.** Topic clustering of Natural science papers based on the LDA model

Topic ID	Topic	Keywords
#Topic0	Text mining and algorithm design	model, experiment, recognition, user, feature, data, algorithm, effect, domain, text, topic, label, semantic, computational, effective
#Topic1	Empirical research on user behavior	user, influence, behavior, salience, search, information, model, factor, influence factor, theory, use, network, data, positive, cognition
#Topic2	Scientometrics and evaluation	fields, papers, disciplines, literature, networks, indicators, relationships, knowledge, topics, patents, theories, data, features, journals, Important
#Topic3	Information resources construction and services	service, application, technology, foundation, data, digital library, system, knowledge, demand, library, theory, management, emergency, organization, construction

In the past 11 years, there have been five major topics, including data mining and algorithm design, theory building and policy making, information services and user research, scientometrics and evaluation of science and technology, and organization and management of book archives. Natural Science Foundation papers mainly include four categories: text mining and algorithm design, empirical research on user behavior, scientometrics and evaluation of science and technology, and information resource construction and services.

### 3.3 Analysis of high-frequency keyword evolution

Keywords are the concentration and refinement of the core content of an article, high-frequency keyword collection can represent the research hotspots in a discipline [3], and citation frequency is one of the important indicators to evaluate the academic influence of literature [4]. Based on this, analyzing the keywords of highly cited papers in national funded papers can reflect the research hotspots of highly concerning literature in the discipline to a certain extent. Figure 2 and Figure 3 show the evolution trend of keywords in highly cited papers of the Social Science Foundation and Natural science Foundation. It can be seen in Figure 2 that the research hotspot of the social science fund in the LIS field has gradually transited from the early focus on library construction to the research on digital resource services and users. As can be seen from Figure 3, the popularity of high-frequency keywords in natural science foundation papers shows the characteristics of fluctuation and change, focusing on information resource construction and service, network public opinion, and domain ontology construction.

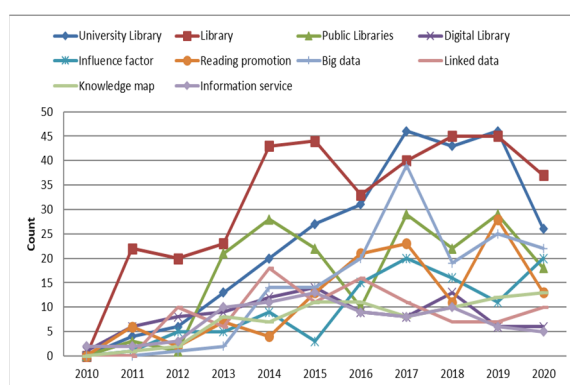


Fig. 2 Evolution trend of high-frequency keywords in social science foundation papers

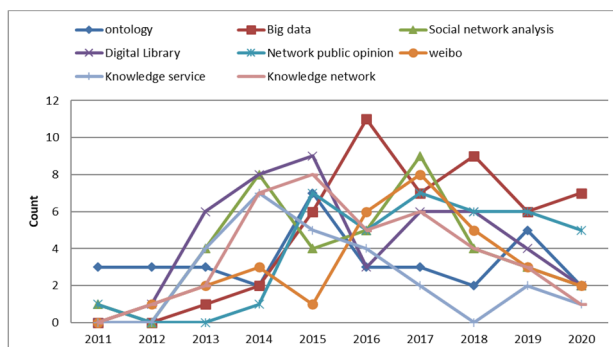


Fig. 3 Evolution trend of high-frequency keywords in NSFC papers

## 4. Summary

Based on the number of fund papers, the level and strength of scientific research in LIS fields are continuously strengthening in recent years. From the perspective of the thesis theme, the project theme of the social science fund is more diversified than that of the natural science fund. The natural science fund projects tend to the direction of technology realization and technology application, while the social science fund has both technology and humanities research. From the perspective of keyword evolution, information resource construction and library management have always been hot topics of research. In recent years, topics such as information services, online public opinion, and big data have become the focus of attention. Research hotspots are highly compatible with the background of The Times and policy guidance. The development of science depends on the complementarity of theory and practice. Therefore, when applying for projects, researchers should closely combine the new trends in the discipline field, closely adapt to the development of The Times, and find emerging themes.

## References

- [1] Chen Jian-xin, MA Qiang. The Role and Function System in National Innovation System of China. *China Soft Science*,2001(08):44-49.
- [2] Liu Bowen, Bai Rujiang, Zhou Yanting et.al, Identifying Frontier Topics from Funding and Paper--Case Study of Carbon Nanotube. *Data Analysis and Knowledge Discovery*, 2019, 3(08): 114-122.
- [3] Gao Haiyan. Research Status of Library and Information Science from the Perspective of National Social Science Fund Project Achievement. *Library Work and Study*, 2020(01): 35-43.
- [4] Qiu Junping, Miao Wenting. Application of h-index in Evaluating Individual's Performances. *Science Focus*, 2007(03): 17-22.