publications

MDPI

Article

How Scholars Collaborate on Data Assets Research: A Systematic Comparative Analysis of Chinese and International Publications

Yaqin Li [1,*], Jinyuan Shi [1] and Yuequan Yang [2]

[1] School of Business, Yangzhou University, Yangzhou 225127, China; mz120242209@stu.yzu.edu.cn
[2] School of Information Engineering, Yangzhou University, Yangzhou 225127, China; yang@yzu.edu.cn
* Correspondence: yqli@yzu.edu.cn

Abstract

In the era of data elements, it is extremely necessary and practically important to analyze network characteristics and evolutionary trends in academic research collaboration in the field of data assets research, which can provide valuable insights for promoting deep cooperation of scholars and enhancing their collaborative efficiency. However, existing studies on data assets research rarely delve into key differentiating characteristics and core thematic priorities between Chinese and international samples of collaboration networks. Based on bibliometric methods and social network analysis, a systematic comparative analysis between Chinese collaboration networks and international collaboration networks is conducted via CiteSpace software by using core literature from the China National Knowledge Infrastructure (CNKI) and Web of Science Core Collection, developed by Clarivate Analytics (WoS). Through observation, we find that the number of publications in this field has reached a preliminary scale with distinct differences in research focus and collaborative features between cooperation networks in China (CNCs) and international cooperation networks (ICNs). In recent years, Chinese samples have primarily focused upon research themes related to data value realization, such as data rights confirmation, data assets accounting, and data trusts. The overall connectivity of CNCs seems relatively weak, and a stable core author group has not formed, while collaborations in CNCs are predominantly localized and short-term. In contrast, international samples in recent years have mainly addressed the contextual application of data assets, exhibiting a collaboration network characterized by multi-center, interdisciplinary, and cross-institutional synergy, while core authors in ICNs are closely interconnected and their connectivity and structure are generally stronger than those of CNCs.

Keywords: data assets; bibliometric analysis; collaboration network

check for updates

Academic Editor: Andrew Kirby

Received: 10 July 2025
Revised: 12 August 2025
Accepted: 13 August 2025
Published: 16 August 2025

Citation: Li, Y., Shi, J., & Yang, Y. (2025). How Scholars Collaborate on Data Assets Research: A Systematic Comparative Analysis of Chinese and International Publications. *Publications*, *13*(3), 38. https://doi.org/10.3390/publications13030038

Correction Statement: This article has been republished with a minor change. The change does not affect the scientific content of the article and further details are available within the backmatter of the website version of this article.

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

As the fifth major production factor, data has become increasingly and prominently valuable. The Fourth Plenary Session of the 19th Central Committee of the Communist Party of China explicitly proposed to strengthen the integration and utilization of data resources (see https://www.gov.cn/xinwen/2019-10/31/content_5447245.htm (accessed on 10 July 2025)). The State Council of China has successively issued policies, such as its opinions on establishing a data foundation system to better leverage the role of data elements (see https://www.gov.cn/zhengce/2022-12/19/content_5732695.htm (accessed on 10 July 2025)), and the 14th Five-Year Plan for Digital Economy Development (see https://www.gov.cn/zhengce/content/2022-01/12/content_5667817.htm (accessed on

10 July 2025)), promoting the systematic construction of mechanisms for data resource rights confirmation, circulation, trading, and value realization. Driven by both policy impetuses and practical demands, data have become a crucial type of strategic resource, while data assets have emerged as a hot topic for both in academic communities and in practical fields as a key form of realizing the value of data elements.

Research on data asset-related topics covers core contents, such as data rights confirmation, pricing, governance, risk management and value assessment, and has a highly interdisciplinary nature, involving multiple fields, such as information management, accounting, economics, law and computer science (Tang et al., 2025). Research collaboration refers to the formation of complementary teams by scholars to address specific scientific challenges, achieving efficient scientific synergy through knowledge sharing and resource integration (X. Chen & Zhang, 2019). According to resource-based theory, sustained competitive advantage depends on unique resources (Barney, 1991). Through academic collaboration, scholars build unique and heterogeneous resources, which in turn contribute to a sustainable competitive advantage. Research collaboration helps to promote the dissemination of research findings and is essential in breeding innovation. Collaboration network analysis is a useful tool to study researchers' and institutions' collaborations (Zhang et al., 2022).

With the deepening of related research, collaboration and communication among authors and institutions have been continuously strengthened, and cooperation has gradually become an important driving force for promoting research and practice in this field. Therefore, systematically sorting out the research characteristics of the data assets domain from the perspective of cooperation networks, identifying core authors, high-frequency keywords, key institutions and their cooperation patterns, holds great significance for understanding the current development status and evolution trend of this field.

In recent years, bibliometric analysis and social network analysis methods have been widely applied in academic collaboration research, with the use of visualization tools such as CiteSpace enhancing the ability to reveal research hotspots, collaborative relationships, and knowledge structures (Zhou & Li, 2024). Broad collaboration networks facilitate faster dissemination of research findings and lead to higher citation frequencies (Borgatti et al., 2009). In addition, scholar collaboration network characteristics can predict research performance (Abbasi et al., 2011). However, systematic studies on collaboration networks in the field of data assets remain relatively scarce, particularly comparative research on the structural characteristics, collaboration patterns, and influence of Sino-foreign academic collaboration networks.

Bibliometric analysis, as a quantitative research method, has become indispensable for tracking knowledge development in fast-growing academic disciplines (Klarin & Suseno, 2021). It enables systematic analysis of scientific development at macro and meso levels, revealing historical trajectories, current research trends, and future directions within specific fields (Klarin & Suseno, 2021). Scholars have established that bibliometric analysis fundamentally examines articles, authors, citations, co-citations, co-authorships, affiliations, countries, and journals, along with their interrelationships (Ellegaard & Wallin, 2015). This methodology enables the construction of robust frameworks to map and analyze large-scale collaboration networks across authors, institutions, and nations, providing a comprehensive overview of scientific interaction patterns (Glänzel, 2001; Duan, 2024). It provides quantitative evaluation of scientific productivity and emerging research trends (Nan et al., 2025). This study selects the China National Knowledge Infrastructure (CNKI), which is full-text database for Chinese journals, and the Web of Science Core Collection (WoS) as data sources, and uses the CiteSpace bibliometric tool to conduct, respectively, the cooperation network analysis of Chinese and international literature in the data assets research field. It systematically identifies core authors, high-frequency keywords, core research institutions and cooperative structures, and reveals the similarities and differences in collaboration

patterns, academic influence, and network characteristics derived from Chinese and international scholars in the data assets research field, to provide references for promoting international cooperation and interdisciplinary exchanges in data assets research.

In pursuit of these research goals, this study addresses the following research questions:

RQ1. What distributional disparities exist in data assets research publications between Chinese and international samples?

RQ2. How do core thematic priorities in data assets research diverge between Chinese and international samples?

RQ3. What significant distinctions characterize scholarly collaboration networks between Chinese and international samples in data assets research?

RQ4. What are the key differentiating characteristics between CNCs and ICNs based on Chinese and international samples in data assets research?

The remainder of this study is structured as follows: Section 2 introduces the research methodology, including data sources and research methods. Section 3 analyzes the current research landscape of data assets studies between Chinese and international samples. Section 4 presents a comparative analysis of research collaboration networks in data assets studies between Chinese and international samples, including scholar collaboration networks and institutional collaboration networks. Finally, Section 5 provides the conclusion and discussion.

2. Data Sources and Methods

2.1. Data Sources and Data Processing

Publicly published papers are generally considered as the most direct manifestation of research achievements. By analyzing the publication of co-authored papers in core journals and the collaboration networks among core authors and their institutions, the characteristics and trends of current research collaboration can be revealed. To explore the features of research collaboration in the field of data assets research, this study selects CNKI and WoS as the primary data sources. We employ these two complementary authoritative databases to ensure comprehensive coverage of both international and Chinese research samples. The Web of Science (WoS) Core Collection, recognized as the premier global citation database with rigorous journal selection standards (Garfield, 2006), provides essential multidisciplinary coverage for the field of data assets research, while the China National Knowledge Infrastructure (CNKI), the dominant Chinese scholarly platform, systematically archives peer-reviewed journals, including the Peking University Chinese Core (PUCC) journals database, the Chinese Social Sciences Citation Index (CSSCI) database, and the Chinese Science Citation Database (CSCD), and gives access to China's domestic scholarship. This dual-database approach enables robust analysis of both international trends and region-specific research developments (F. Liu et al., 2025; J. Wang et al., 2025).

According to Bradford's Law, the core literature in many fields tends to be concentrated in a few core journals. Therefore, we limit research samples from Chinese journal sources to PUCC, CSSCI, and CSCD journals in the CNKI database, and employ the retrieval strategy TS = ('data asset*' OR 'data resource*') with a publication year range of 2000–2024. Through manual screening, this study excludes non-scholarly publications such as conference abstracts and news reports, and removes duplicate records, ultimately obtaining 5832 eligible publications as Chinese samples for subsequent analysis. All retrieved records are exported as RefWorks files and then systematically converted into the required data structure for network analysis in CiteSpace via our designed translation program.

In contrast, after data cleaning, this study identifies 4788 valid international publication samples from the WoS database using the same search strategy. Note that non-research documents, including conference proceedings and news reports, are systematically ex-

cluded from our samples. All retrieved records are exported with complete bibliographic metadata in plain text format for subsequent analysis.

2.2. Research Methods

The study employs CiteSpace (6.4 R1), which is one of the most well-known tools for bibliometric analysis (C. Chen, 2004, 2006) to conduct bibliometric and social network analysis, with a focus on constructing and examining collaboration network characteristics in the field of data assets research. CiteSpace is a widely used tool for bibliometric analysis, featuring robust collaboration network analysis capabilities. It can reveal cooperative relationships and network characteristics among scholars and research institutions, effectively identifying core authors, key collaborators, and collaboration patterns, which thereby can present dynamic evolution and hotspots in specific research fields. The cross-platform compatibility of Citespace (supporting WoS, Scopus, CNKI, and other major databases) enables robust multidisciplinary comparisons. Its temporal evolution graphs and burst detection charts offer powerful visual analytics for tracking disciplinary development trends (F. Liu et al., 2025). The study aims to uncover cooperative relationships among scholars and institutions, identify core authors, pivotal collaborators, and central research organizations, and to analyze the features and trends, and their similarities and differences, between Chinese collaboration networks and international collaboration networks. Through observation and analysis, the findings are expected to facilitate efficient academic collaboration in the data assets research domain, as well as enhancing collaboration quality between Chinese and international scholars.

3. Research Status of Data Assets

3.1. Publication Analysis

In the field of data assets research, scholars from both China and international samples have achieved numerous results. Due to differences in economic development stages, policy environments, and technological applications among various countries, there are significant variations in research contents. The growth trends in publication numbers show no significant differences, as illustrated in Figure 1. Since 2000, research on data assets in China has been increasing year by year. Especially since 2019, the number of published papers in Chinese core journals has experienced a sharp increase. This is closely related to the implementation of China's "Digital China" strategy and the practical issues arising from the advancement of data asset-related fields, which have consequently attracted widespread attention in Chinese academia.

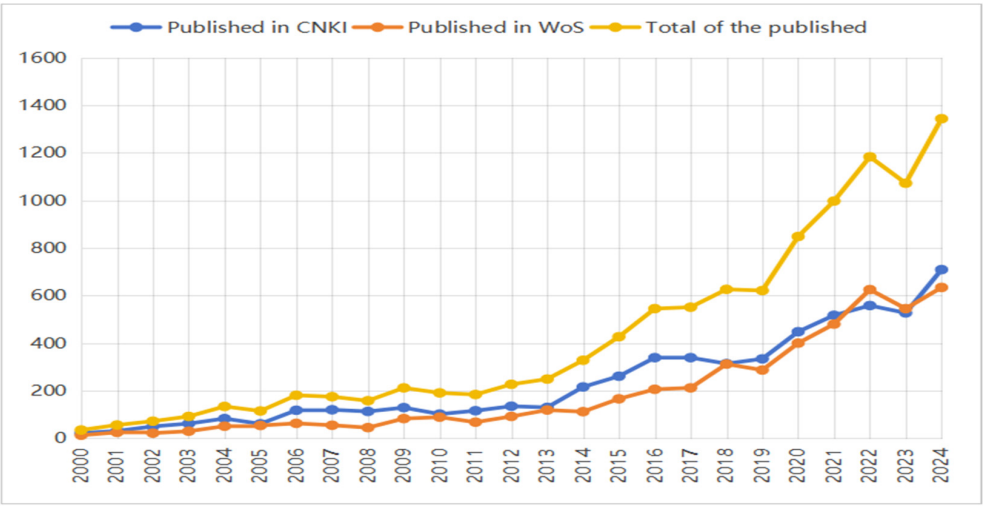


Figure 1. The distribution of published papers on data assets research.

In the WoS sample, a total of 131 countries/regions have published research papers related to data assets. Among them, 47 countries/regions contributed at least 15 publications, forming three distinct clusters. Table 1 presents the top 10 most productive countries/regions in terms of publication output. International research collaboration in this field is relatively strong, while Chinese scholars have shown a consistent increase in publication output, as illustrated in Table 1.

Table 1. Top 10 countries/regions in terms of publication output in WoS sample.

NO.	Country	Publications	Citations	Total Link Strength
1	USA	1816	93,316	1551
2	PEOPLES R CHINA	1360	31,621	561
3	ENGLAND	651	46,875	1192
4	GERMANY	313	19,733	708
5	CANADA	250	16,383	462
6	AUSTRALIA	241	7294	417
7	NETHERLANDS	160	10,491	453
8	FRANCE	155	6767	442
9	ITALY	143	4866	513
10	SPAIN	140	4057	421

3.2. Research Themes

To identify the core research foci in data assets studies, this study conducts keyword co-occurrence analysis. The WoS sample of 4788 publications contained 11,969 author keywords, with 79 keywords appearing ≥15 times, while the Chinese sample (5832 publications) yields 11,620 keywords, of which 141 meet the ≥15 frequency threshold, indicating greater keyword concentration in Chinese literature. Table 2 presents the top 20 keywords from each dataset.

Table 2. Top20 keywords.

	Chinese Sample			International Sample		
NO.	**Keyword**	**Frequency**	**Total Links**	**Keyword**	**Frequency**	**Total Links**
1	big data	688	425	genpred	217	522
2	data assets	269	231	shared data resources	144	348
3	metadata	240	133	big data	137	103
4	data elements	238	291	genomic selection	128	326
5	data governance	223	270	machine learning	125	134
6	digital economy	221	269	genomic prediction	119	327
7	data resources	194	165	deep learning	90	86
8	data sharing	188	182	database	79	51
9	library	107	105	shared data resource	78	177
10	digital transformation	97	89	epidemiology	69	47
11	scientific data	97	85	artificial intelligence	53	62
12	open data	93	112	COVID-19	50	32
13	government data	90	134	bioinformatics	47	27
14	data management	84	93	data sharing	44	31
15	cloud computing	80	67	blockchain	40	27
16	public data	78	74	databases	39	23
17	data	78	51	ontology	36	23
18	data security	78	110	data integration	32	30
19	data mining	76	43	cancer	31	19
20	artificial intelligence	71	61	d3r	31	38

In the international samples, the co-occurrence network of 79 high-frequency keywords formed five clusters, but primarily converged into two core research directions: (1) emerging technologies related to data assets, including keywords such as AI, big data,

cloud computing, data governance, blockchain, and machine learning; (2) biomedical and healthcare data research, featuring terms like Alzheimer's disease, bioinformatics, gene expression, transcriptomics, cancer, and Drug Design Data Resource (D3R), while in the Chinese sample there are 72 keywords with a frequency of at least 25. The keyword co-occurrence analysis formed five clusters, primarily converging into four core research directions: (1) technology-related research on data assets management, including keywords such as cloud computing, metadata, linked data, big data, data centers, data integration, and data sharing; (2) applied research on data assets, featuring terms like digital humanities, digital transformation, smart libraries, and digital government; (3) research on data assets circulation and value realization, covering digital economy, data trading, data ownership confirmation, data elements, new quality productive forces, and high-quality development; (4) research on data security, including open sharing, open data, data security, data quality, and related keywords.

To identify the research priorities and the evolution paths of themes in the field of data assets, the study conducts keywords co-occurrence analysis on the sample documents in the CNKI and WoS datasets, respectively. The results are shown in Figures 2 and 3. As shown in Figure 2, the keywords in the Chinese literature exhibit a progressive evolution along the timeline, shifting from data governance to assetization mechanisms, such as the role of data assetization (L. Chen, 2024). Early research predominantly featured keywords such as "information sharing", "e-government" and "data security", which are related to data governance, indicating that the studies on data assets in the early 21st century primarily focused on data management and governance. After 2015, high-frequency keywords such as "big data" "blockchain" and "data mining" emerged, reflecting the integration of technological foundations and data value discovery in data assets research. Post-2020, keywords such as "personal data", "data rights confirmation", "data assets accounting", "data trusts", "digital economy" and" data elements" have gained prominence, which signaled a research trend toward data ownership clarification, data assets valuation and data assets capitalization. In the Chinese sample, the keywords "data governance" generally serve as a central axis connecting multiple time periods, threading through the research themes of institutional design and technological support. "Personal data" focuses on ownership definition, data security, and individual data management, representing the most significant growth area in recent research. Meanwhile, "blockchain" functions as a technological implementation pathway, intersecting with multiple themes and highlighting the strong correlation between data assets and technological applications.

As shown in Figure 3, the WoS sample literatures demonstrate a typical technology-driven evolutionary path. Early research predominantly focused on core keywords related to data modeling, such as "ontology" and "semantic web". After 2015, keywords like "machine learning" and "data integration" gradually became dominant, reflecting a shift in research focus from data organizational structures to processing capabilities and modeling methods. Post-2020, scenario-specific application keywords such as "bioinformatics" "prediction" and "drug discovery" have emerged continuously, which significantly indicate that research in the field of data assets has become deeply embedded in vertical domain applications, like life sciences. This demonstrates that international research exhibits a trend toward the valorization and contextualization of data assets.

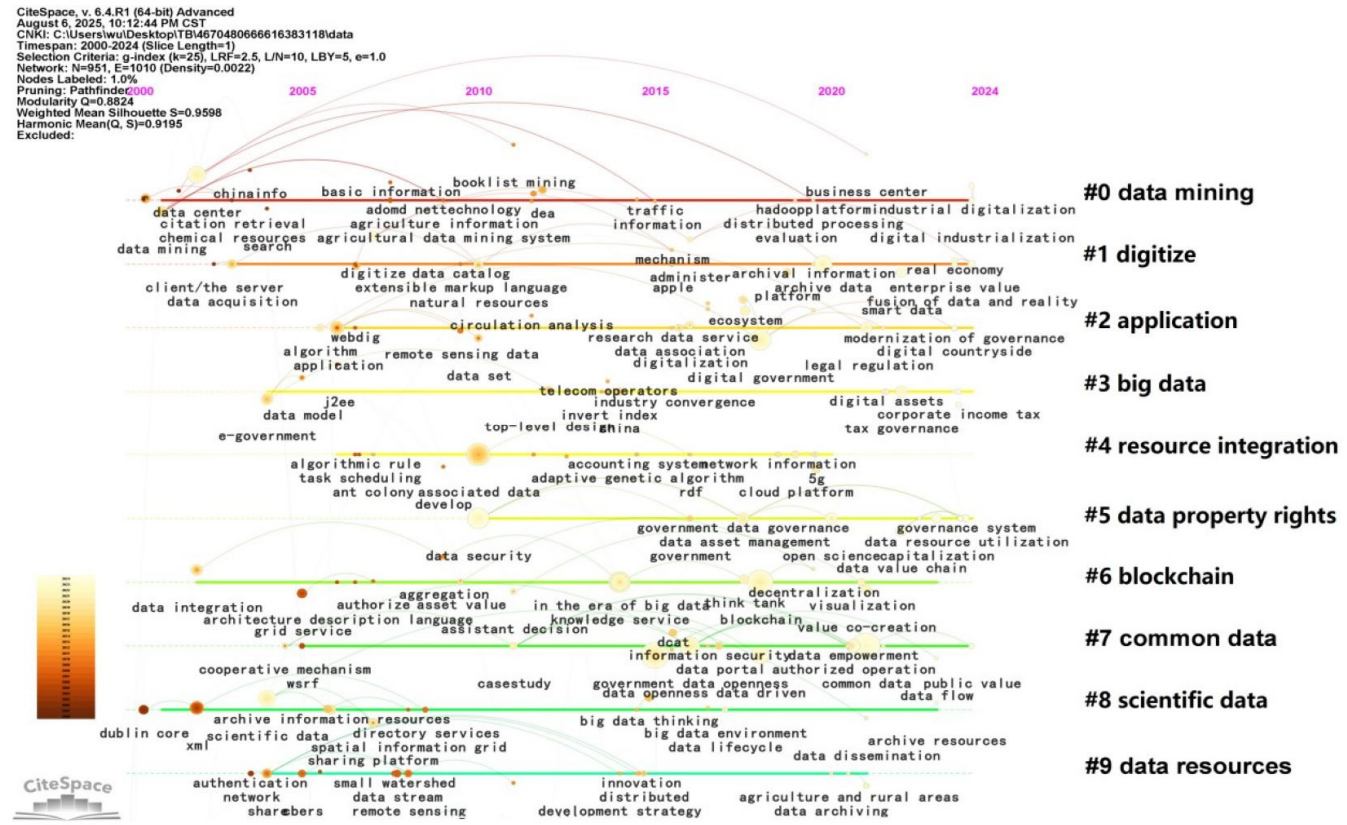


Figure 2. The evolution process regarding Chinese keywords in the CNKI research sample.

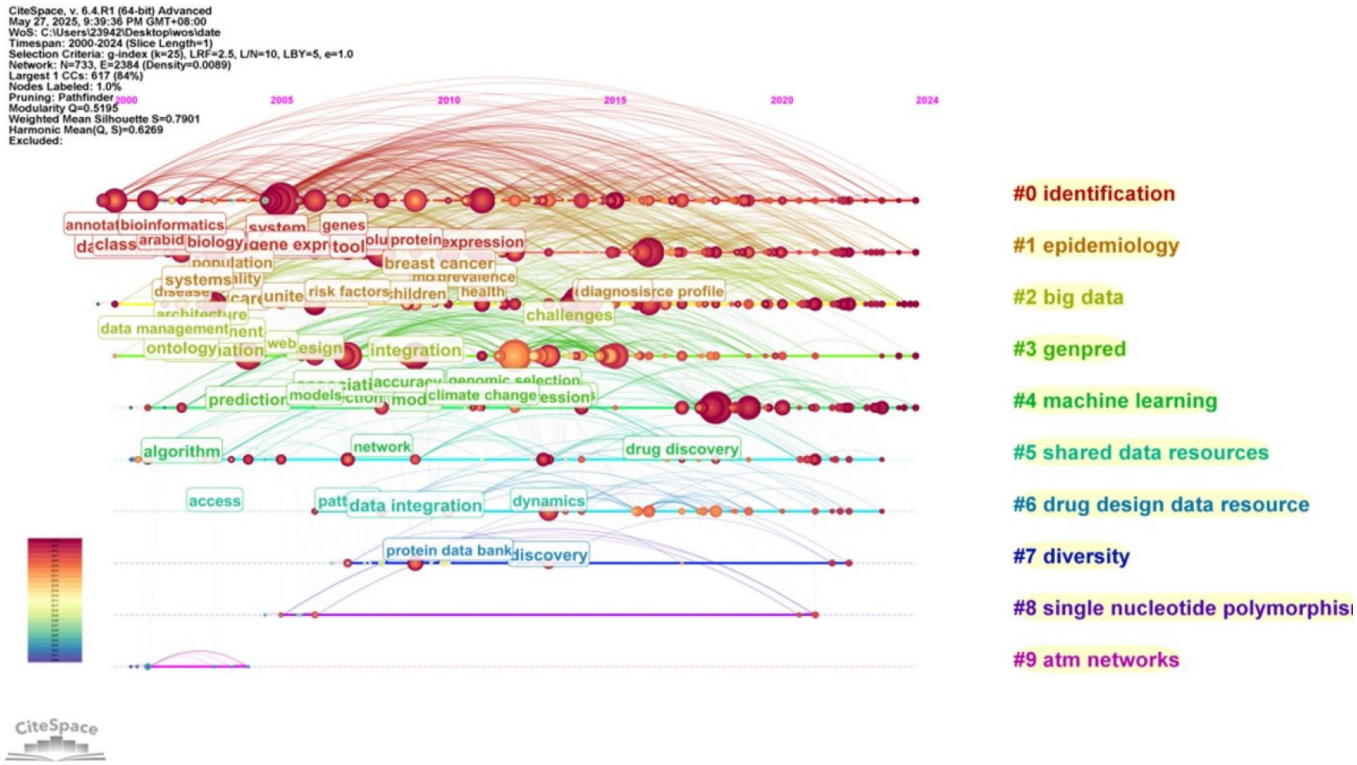


Figure 3. The evolution process regarding English keywords in the WoS sample.

4. Comparative Analysis of Research Collaboration

The situation of scientific research cooperation can usually be measured by indicators such as the degree of author collaboration and the collaboration rate (Li & Wei, 2023). From the indicator of collaboration degree, i.e., average number of authors per paper, there are significant differences between Chinese and international samples. Among the 5832 papers in the Chinese sample, the number of solo-authored papers is 1464, and the total number of authors is 11,230, with an average of 1.93 co-authors per paper. In contrast, in the 4788 papers included in WoS, the number of solo-authored papers is 1247, and the total number of authors is 25,658, with an average of 5.36 co-authors per paper. In the field of data assets research, there is a significant difference in the number of collaborators per research paper between Chinese and international studies. The collaboration degree of 1.93 in Chinese samples is significantly lower than that of 5.36 in international samples, while the number of co-authors in international published papers is relatively higher with a larger team size. This may to some extent reflect the differences in scientific culture, academic evaluation systems, and magazine-related requirements between Chinese and international collaborative communities. Considering the large annual span of the samples, to facilitate the analysis of the trend in paper collaboration, combined with the growth of publications as shown in Figure 1, the study divides the research period from 2000 to 2024 into three phases. The first phase of 2000–2010 is a period of slow growth, with the second phase of 2011–2019 showing steady growth and the third phase of 2020–2024 showing rapid growth, as shown in Table 3. From the indicator of collaborative rate, i.e., the proportion of collaborative papers to total papers, the collaborative rate of Chinese scholars has slightly decreased. In contrast, the collaborative rate of international papers generally shows an increasing trend. However, the collaborative rate of Chinese and international scholars fluctuates but is fairly close, which indicates a general trend of global academic collaboration. Moreover, due to the multidisciplinary and comprehensive nature of data assets research, this cooperative tendency will continue strengthening.

Table 3. Collaboration between Chinese and international scholars in the field of data assets.

Time	Chinese Sample				International Sample			
	Publications	Non-Collaborations	Collaborations	Collaborative Rate (%)	Publications	Non-Collaborations	Collaborations	Collaborative Rate (%)
2000–2010	891	204	687	77.10	530	145	385	72.64
2011–2019	2632	675	1957	74.35	1974	557	1417	71.78
2020–2024	2309	585	1724	74.66	2284	545	1739	76.13
Total	5832	1464	4368	75.24	4788	1247	3541	73.96

4.1. Analysis of Authors

4.1.1. Core Authors and Academic Influence

Price's Law is an important theory in bibliometrics, which is widely used to measure the distribution pattern of scientific research achievements among authors and identify core authors in related fields (D. Wang et al., 2024). Research outputs in academic fields typically exhibit a highly concentrated distribution, with a small number of core authors contributing the majority of research results. The minimum publication threshold for core authors is $M_P = 0.749\sqrt{N_{PMAX}}$, where N_{PMAX} represents the publication count of the most prolific author, which can effectively identify core authors who have made outstanding contributions and have had an extremely significant influence in a certain field. When the total publications of core authors account for 50% of the total output, it indicates the formation of a stable core author group in that research area. By analyzing the core author group, the backbone of academic research and development in a specific field

can be effectively identified, thereby providing young scholars with guidance for selecting potential collaborators.

Table 4 presents the collaboration between and influence of the top 10 core authors within the Chinese samples. As shown in Table 4, Baoping Yan has become the most prolific author, having published 22 papers. According to Price's Law, authors with at least four publications are considered core authors in the data assets research field. Within the Chinese samples, there are 376 core authors in total, collectively contributing 2239 papers and accounting for 38.39% of the total sample literature. The index does not meet the threshold set by Price's Law for a stable core author group, indicating that Chinese scholars in the data assets research field have not yet formed a stable core authors community. The composition of core authors remains dynamic, further suggesting that this research area is still in its developmental stage with ample room for exploration.

Table 4. The collaboration between and influence of the top 10 core authors in Chinese publications.

No.	Author	Publications	Sole-Authored	Collaborative Rate (%)	H-Index
1	Baoping Yan	22	0	100.00	26
2	Xiaomi An	21	1	95.24	39
3	Jiandong Wang	20	2	90.00	22
4	Wenhong Zhou	19	7	63.16	20
5	Yang Li	18	0	100.00	17
6	Ruhua Huang	18	2	88.89	45
7	Li Si	16	0	100.00	33
8	Yunqiang Zhu	15	0	100.00	30
9	Haiqun Ma	15	1	93.33	43
10	Yikun Xia	14	8	42.86	28

Table 5 displays the collaboration between and influence of the top 10 core authors within international samples. The most prolific author, Crossa Jose, has published 40 papers. Based on Price's Law, authors with 5 or more publications qualify as core authors in this field. The international sample includes 236 core authors, collectively contributing 1654 papers and accounting for 34.54% of the total sample literatures. This mirrors the characteristics observed in the Chinese samples, demonstrating that neither Chinese nor international research samples in the data assets field have established stable core author groups. The vast potential for research development in this domain suggests that interested scholars should further explore this area.

Table 5. The collaboration between and influence of the top 10 core authors in international publications.

No.	Author	Publications	Sole-Authored	Collaborative Rate (%)	H-Index	WoS Cited Number
1	Crossa Jose	40	0	100.00	80	21,301
2	Burley Stephen K	20	2	90.00	96	40,191
3	Montesinos-Lopez Abelardo	20	0	100.00	19	1352
4	Birney Ewan	17	0	100.00	127	140,274
5	Montesinos-Lopez Osval A	17	0	100.00	24	3095
6	Rocca Walter A	17	0	100.00	105	36,730
7	Berman Helen M	15	1	93.33	84	58,652
8	Lawson Catherine L	15	1	93.33	35	6913
9	Perez-Rodriguez Paulino	15	0	100.00	37	6649
10	Sahebi Ali	15	0	100.00	13	764

As shown in Tables 4 and 5, this not only identifies the core authors in the field of data assets research in China and international samples during the same period based on the number of published papers, but also evaluates the influence of high-productivity

authors from aspects such as collaborative rate and H-index. In Table 4, the collaborative rate of the authors is generally high. Baoping Yan, Yang Li, Li Si and Yunqiang Zhu have a collaborative rate of 100%, and all their research papers are collaborative achievements. In contrast, in Table 5, the collaborative rate of scholars is generally higher, all above 90%, and there are fewer solo-authored works. The H-index is an important indicator for measuring a scholar's academic influence. It takes into account the number of published papers and the number of citations, and can objectively reflect the citation situation and academic influence for a scholar (Masic & Begic, 2016). From Tables 4 and 5, it can be seen that the academic research influence of international core authors is generally higher than that of Chinese core authors. The findings suggest that the influence of core Chinese researchers needs to be further expanded. Effective international collaboration could be leveraged to broaden research perspectives and enhance the impact of academic achievements.

4.1.2. Author Collaboration Network Analysis

The influence of the data assets research field relies heavily on senior authors' sustained contributions and collaborative efforts within the domain. By analyzing co-authorship networks among publishing authors, we can clearly map collaborative relationships and identify research teams in this field (Li & Wu, 2023). The k-core filtering threshold is set based on observed samples, ensuring balance between network clarity and information retention. Considering the need to highlight the core author group, we make this appropriate choice through the multiple tests. In this work, we employ k-core filtering and retain only nodes (authors) with collaborations involving at least five others. This approach yields distinct Chinese and international core author collaboration networks, illustrated in Figures 4 and 5.

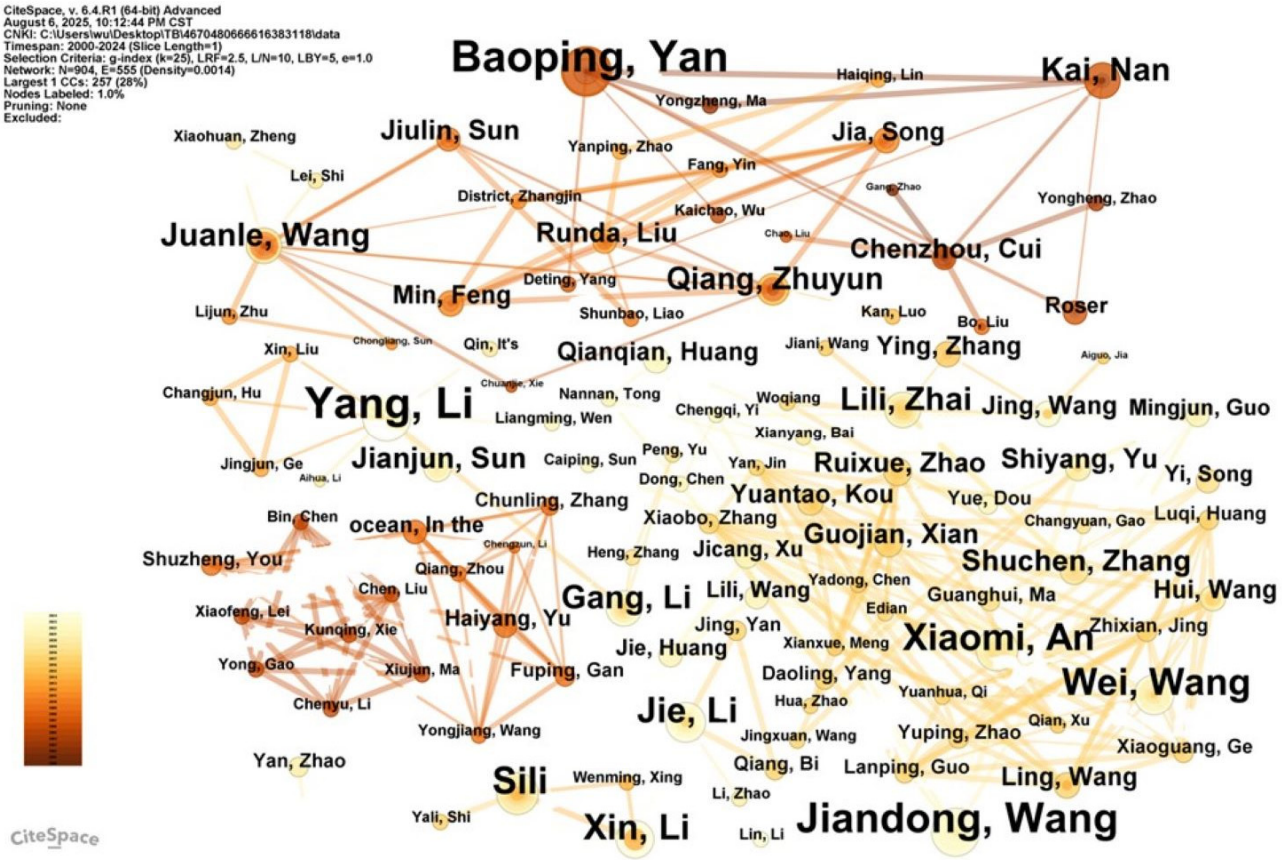


Figure 4. Co-occurrence graph of core Chinese authors in Chinese sample.

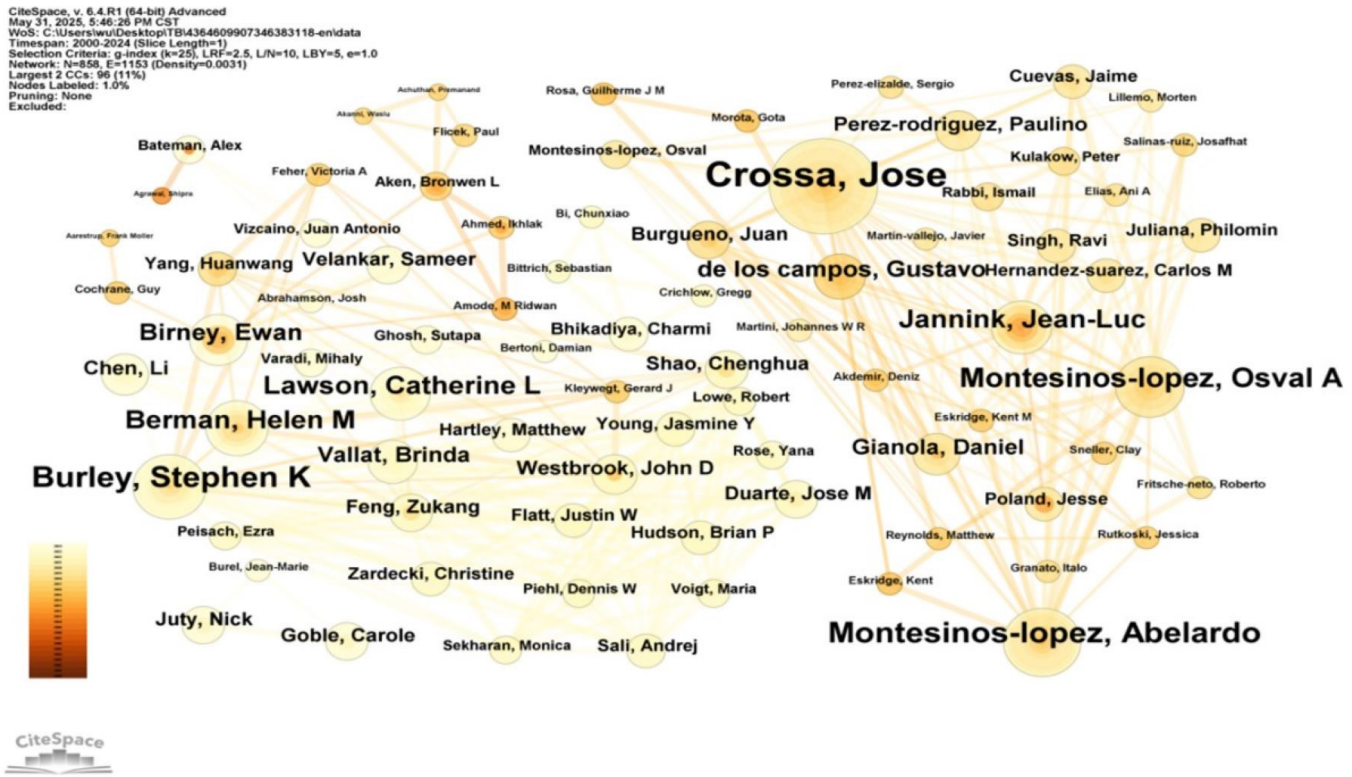


Figure 5. Co-occurrence graph of core authors in international sample.

In Figure 4, there are 904 nodes and the connection number is 555. The overall network density is 0.0014, indicating that the cooperative relationships among the core authors in the data assets research field are relatively weak, and the degree of cooperation among scholars is low. There are not any dense academic cooperation groups. Among them, the largest connected subgraph contains 28% of the nodes, i.e., 257 nodes, which is lower than the common 30% to 60% proportion in mature research fields. The analysis reveals that, while some Chinese core authors in the data assets research field have formed relatively stable collaborative groups, a significant number of scholars remain outside the core network, indicating relatively weak collaboration density among researchers.

As shown in Figure 4, from the perspective of core scholars' collaboration patterns, Baoping Yan's cooperative network exhibits dense internal connections with a high degree of collaboration among co-authors, indicating that his research relies on long-term stable teamwork with substantial knowledge sharing within the group. The collaboration networks of both Jiandong Wang and Xiaomi An are more widely distributed, spanning multiple research clusters. The overall structure remains more open, demonstrating potential for cross-institutional and interdisciplinary cooperation. In contrast, scholars like Feng Liu show relatively limited collaborative ties without stable academic partnerships. Comparatively, Kai Nan and Ze Luo have integrated into Baoping Yan's core collaborative group with higher network integration.

Furthermore, the Collaboration network in Figure 4 exhibits a distinct clustered structure, and indicates that academic cooperation primarily revolves around a small number of highly productive scholars. The betweenness centrality of all authors is less than 0.01, suggesting the absence of key cross-team connectors in the network. The collaboration pattern tends toward internally stable, closed small groups with limited connections between different research clusters. This structure may arise because the relevant research topics are confined to specific academic circles, with restricted cross-team exchanges. Additionally, the weights of collaboration edges between authors are relatively low, dominated by infrequent

and sporadic cooperation, while long-term, stable, and frequent academic collaborations are scarce. Consequently, the entire network appears structurally fragmented and loosely connected, with only a certain degree of internal collaboration within individual teams and extremely limited exchanges or cooperation between different teams. This low-density, low-cohesion network structure somewhat hinders the dissemination of research outcomes and the effective diffusion of knowledge. The overall Chinese co-authorship network suggests that academic collaboration patterns in the data assets research field remain at a nascent stage, with both the depth and breadth of cross-institutional cooperation requiring further enhancement.

Figure 5 presents the co-occurrence map of international samples of core author collaboration network, which exhibits a highly centralized structure centered around scholar Crossa Jose, demonstrating strong overall connectivity. Although multiple collaborative groups exist within the network, a certain degree of cross-team cooperation can still be observed, with some bridging authors effectively serving as links between teams. This indicates that academic collaboration in international samples in the data assets research field exhibits structural characteristics dominated by core authors with localized diffusion. The collaboration network comprises 858 nodes with 1153 connections, a network density of 0.0031, and a largest connected component accounting for 11%, lower than that of the Chinese samples of core author collaboration networks, which suggests the existence of a certain scale of core collaborative groups. However, the overall network integration remains relatively weak, with research collaborations primarily concentrated among a few groups, presenting a "local clustering, global dispersion" structural feature. Although several cross-group authors exist in Figure 5, most nodes exhibit low betweenness centrality, which apparently reflects the limited number of "bridge-type" scholars in the network, and there is much room for strengthening cross-team collaborative linkages. From Figures 4 and 5, it can be observed that there are significant differences in the structural characteristics of the author collaboration patterns between Chinese samples and international samples. In the international samples of collaboration network, core authors such as Crossa Jose and Montesinos-Lopez Osval A. demonstrate a typical radial structure in their cooperative relationships. There are a large number of collaborators, but the connections between them are loose. Although the overall network density is slightly higher than that of the Chinese samples of collaboration network, there are still multiple relatively isolated sub-groups in the international samples, and the number of bridge nodes across teams is limited. The degree of cooperation integration needs to be improved.

Comparative analysis reveals that prominent Chinese sample research teams, such as those led by Baoping Yan and Jiandong Wang, demonstrate stronger internal cohesion, with densely interconnected collaborators. However, the Chinese sample of co-authorship network exhibits more severe fragmentation overall. The overall network density is only 0.0014, indicating sparse collaborative connections beyond core teams. The network lacks substantial cross-institutional or inter-group collaboration, resulting in limited connectivity and poor scalability considering the network as a whole.

4.2. Institution Network Analysis

4.2.1. Core Institution Analysis

Identifying the core research institutions in a specific field facilitates the rapid discovery of potential collaborative partners within that domain (Li & Wu, 2023). Table 6 lists the top 10 institutions by the number of publications in both Chinese and international samples, respectively. Among the top 10 institutions for the Chinese sample, apart from the Institute of Scientific and Technical Information of China and the Agricultural Information Institute of the Chinese Academy of Agricultural Sciences, the rest are all higher education

institutions. The top three are the School of Information Management at Wuhan University, the School of Information Resource Management at Renmin University of China, and the School of Information Management at Nanjing University. Overall, the research on data assets in China is primarily concentrated in schools of information management at major universities, with particular emphasis on data assets management, resource optimization, data governance and, especially, the effective management and utilization of information resources.

Table 6. The distribution of institutions with the top 10 publications between Chinese and international samples.

No.	Chinese Sample		International Sample	
	Institution	**Publications**	**Institution**	**Publications**
1	School of Information Management, Wuhan University	123	University of California	263
2	School of Information Resource Management, Renmin University of China	82	University of London	158
3	School of Information Management, Nanjing University	70	Harvard University	156
4	Center for Information Resources Research, Wuhan University	57	Chinese Academy of Sciences	149
5	University of Chinese Academy of Sciences	49	European Molecular Biology Laboratory (EMBL)	128
6	Computer Network Information Center, Chinese Academy of Sciences	45	European Bioinformatics Institute	124
7	Institute of Scientific and Technical Information of China	39	National Institutes of Health (USA)	114
8	School of Business and Management, Jilin University	30	Harvard Medical School	92
9	Agricultural Information Institute, Chinese Academy of Agricultural Sciences	28	Mayo Clinic	89
10	School of Public Administration, Sichuan University	28	University of Texas	88

Among the top 10 institutions for international samples, as shown in Table 6, apart from the Chinese Academy of Sciences, the rest are international higher education institutions, predominantly located in the United States, which clearly indicates America's absolute leading position in data assets research. At the same time, Chinese scholars have been increasingly active in this field, with the Chinese Academy of Sciences ranking fourth with 149 publications, demonstrating the gradual rise of China's academic influence in data assets research.

4.2.2. Analysis of Collaborative Institutions

Through the academic collaboration and information exchange among the authors' institutions and organizations, the characteristics of research collaboration among different institutions can be effectively identified. Figures 6 and 7, respectively, illustrate the institutional collaboration co-occurrence in the Chinese and international samples.

Through the analysis of the institutional cooperation networks of Chinese and international samples, significant differences are observed. As shown in Figure 6, the institutional collaboration network of the Chinese samples consists of 815 nodes with a connection number of 333, while the proportion of the largest connected subgraph is just 18% and the network density of 0.001. It is easy to see that the cooperation is mainly concentrated among a few core institutions, such as the Information Management School of Wuhan University, the Information Management School of Nanjing University, and the Information Resources Management School of Renmin University of China, etc. While these institutions demonstrate relatively strong internal collaboration ties, the overall network exhibits weak connectivity, limited cross-institutional cooperation, and constrained academic outreach.



Figure 6. Collaboration co-occurrence graph of institutions from the Chinese sample.

In comparison, the international sample of institution collaboration network exhibits a larger scale, shown in Figure 7, comprising 594 nodes and 3322 connections, with the largest connected component accounting for 80% of the network. The institutional collaboration network displays a polycentric structure, spanning multiple countries and disciplines, with institutions such as the University of California and Stanford University. The network density of 0.0189 in the international sample of institutions is significantly higher than that of 0.001 for the Chinese sample.

It is noteworthy that there are more intensive collaborations among the international sample of institutions and more frequent transnational cooperation, and this reflects the more extensive nature of global academic collaboration.

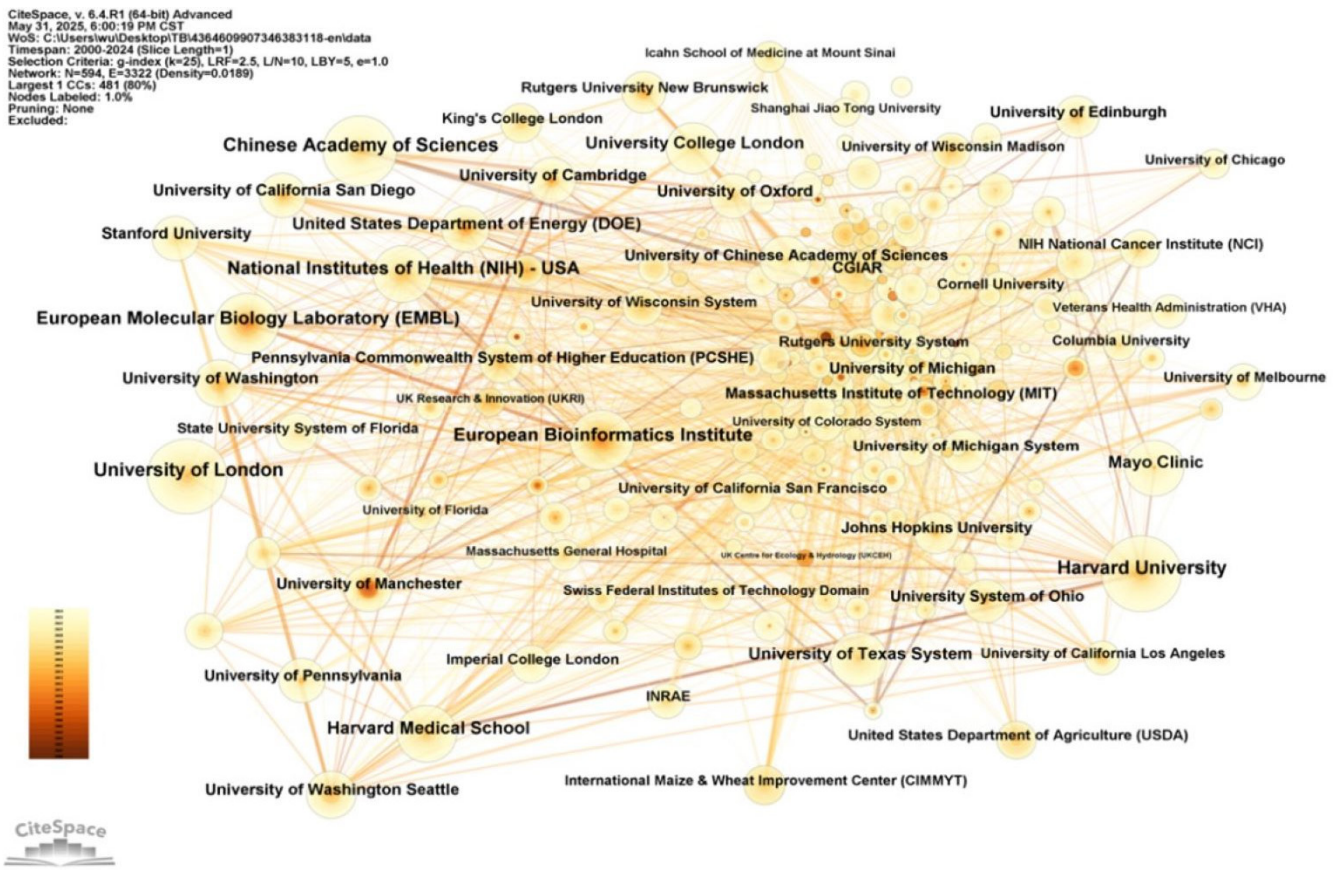


Figure 7. Collaboration co-occurrence graph of institutions from the international sample.

Overall, the Chinese sample of collaboration network is relatively centralized, exhibiting cohesive partnerships but lacking extensive cross-institutional cooperation. In contrast, the international sample demonstrates more polycentric and international characteristics, with tighter interdisciplinary connections that facilitate global academic exchange and knowledge dissemination. Furthermore, the status and influence of core institutions within the collaboration network can be reflected by their centrality measures (H. Liu & Li, 2023). In the research collaboration network of the data assets field, an institution's centrality indicates its academic influence and pivotal role within the network. In Table 7, institutions such as Erasmus MC with its centrality of 0.28 and Helmholtz Association with its centrality of 0.20 occupy central positions, highlighting their key roles in the international sample of collaboration network. Although these institutions produce relatively fewer publications, their significant contributions to interdisciplinary collaboration grant them substantial influence in the global academic network. By comparison, the centrality of core institutions in the Chinese sample is generally lower. Although China's publication output continues to grow, the current collaboration network among research institutions demonstrates limited connectivity. Future efforts should prioritize enhancing extensive inter-institutional cooperation, as shown in Table 7.

To facilitate comparison, the density of the collaboration network is presented in Table 8.

Table 7. The distribution of institutions with the top 10 centralities.

No.	Chinese Samples			International Samples		
	Institution	**Centrality**	**Publications**	**Institution**	**Centrality**	**Publications**
1	School of Information Resource Management, Renmin University of China	0.03	82	Erasmus MC	0.28	14
2	School of Information Management, Wuhan University	0.02	123	Helmholtz Association	0.20	42
3	Institute of Scientific and Technical Information of China	0.02	39	Centre National de la Recherche Scientifique (CNRS)	0.20	40
4	School of Information Management, Central China Normal University	0.02	27	University of Leeds	0.18	7
5	Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences	0.02	22	Swansea University	0.17	14
6	School of Management, Hebei University	0.02	9	Massachusetts Institute of Technology (MIT)	0.16	55
7	School of Information Management, Nanjing University	0.01	70	Curtin University	0.16	7
8	University of Chinese Academy of Sciences	0.01	49	University of Leicester	0.15	12
9	Department of Information Management, Peking University	0.01	24	University of Western Australia	0.15	7
10	Big Data Development Department, National Information Center	0.01	23	Harvard Medical School	0.14	92

Table 8. Collaboration network density.

Collaboration Network	CNC	ICN
Co-authorship Network	0.0014	0.0031
Institutional Collaboration Network	0.0010	0.0189

5. Conclusions and Outlook

5.1. Conclusions

In the field of data assets research, more and more researchers from different fields participate, and the number of publications continues to increase. Based on bibliometric and social network analysis, this study explores the characteristics of the collaboration networks among high-frequency keywords, core authors and core research institutions between Chinese and international samples.

Through keyword co-occurrence analysis, this study reveals that Chinese studies in the data assets research field primarily focus on two key themes: data assets management technologies and promoting data assets circulation/value realization. In contrast, international studies not only examine data management technologies but also place particular emphasis on the management and application of data in biological, medical, and disease-related contexts.

The study also found that the core authors who dominated in the collaboration networks of the Chinese sample are from distinguished universities, with a relatively concentrated overall structure, while the largest connected subgraph accounts for a relatively high proportion, reaching 28%, and no stable core author groups have been constructed. Moreover, the cooperative relationships are mostly low-frequency and sporadic, lacking key connectors across teams, and the international influence still needs to be improved. In contrast, we observe that the international sample scholar collaboration network presents multi-center, interdisciplinary, and cross-institutional characteristics, where core authors cooperate closely

and some scholars play a bridging role between teams. The connectivity and structural stability of the cooperation network as a whole are better than in the Chinese sample.

In the analysis of cooperative institutions, the cooperation network of the Chinese sample of institutions is mainly concentrated in information management-related departments and schools, with a relatively concentrated cooperation scope and overall low connectivity. Furthermore, the breadth and depth of institutional cooperation are still insufficient. In contrast, the international sample institutional cooperation network presents a multi-center and high-density structural feature with a relatively high proportion of the largest connected subgraph, and the collaboration relationships between institutions are close, where cross-national and interdisciplinary cooperation is frequent. Research institutions such as the University of California and Harvard University, represented by these institutions, are at the core position in the global cooperation network and have promoted international cooperation and research progress in the field of data assets research.

The findings of this study provide valuable insights for scholars and teams newly engaged in data assets research. Firstly, the findings offer a quick understanding of the main research topics and directions in this field, both domestically and internationally, as well as key themes for future research. Secondly, the research conclusions help in clearly locating leading scholars, research teams, and major institutions in the field of data assets worldwide, enabling young researchers to efficiently locate potential collaborators. This, in turn, enhances the efficiency and effectiveness of their scientific collaboration.

5.2. Limitations and Outlook

This study provides young scholars with a systematic understanding of current research themes in the data asset field, facilitating their rapid comprehension of key research priorities, and helping young scholars to find in a timely fashion core authors with high academic influence, collaborative relationships, and key research institutions, and can provide references for seeking potential collaborators and partner organizations. However, this study has several limitations that should be noted: (1) This study only used the publicly available core literature included in the databases, such as CNKI and WoS, as samples, without considering the cooperative relationships of non-paper form research results, such as projects, research topics, books and patents. (2) This study was constrained by its sole consideration of English-language literature available in the WoS database; in addition, other databases (e.g., Scopus) were not included in the samples, potentially affecting both the breadth of included evidence and the external validity of the results. (3) The analysis of author and institutional collaboration networks currently lacks fine granularities, such as strength of collaboration (Kurata et al., 2023), intensity of collaboration and manner of collaboration. Future studies could consider incorporating more diversified data to enhance the generalizability and representativeness of the research conclusions.

Author Contributions: Methodology, Y.L.; validation, Y.L.; formal analysis, J.S.; investigation, Y.L.; data curation, J.S.; writing—original draft preparation, Y.L. and J.S.; writing—review and editing, Y.L. & Y.Y.; visualization, J.S. and Y.Y.; supervision, Y.L.; funding acquisition, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the Humanities and Social Sciences Foundation of the Ministry of Education of China, *Online Academic Social Networks Users' Research Sharing Behavior Study*, under Grant No. 20YJAZH059 and the College Student Science and Technology Innovation Project of Yangzhou University under Grant No. XCX20251001 and No. SXYSJCX202544.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

Abbasi, A., Altmann, J., & Hossain, L. (2011). Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures. *Journal of Informetrics*, *5*(4), 594–607. [CrossRef]

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management*, *17*(1), 99–120. [CrossRef]

Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, *323*(5916), 892–895. [CrossRef]

Chen, C. (2004). Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, *101* Suppl. 1, 5303–5310. [CrossRef]

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, *3*, 359–377. [CrossRef]

Chen, L. (2024). Does data assetisation improve corporate liquidity and corporate growth? Evidence from "hidden champion" SMEs in China. *Borsa Istanbul Review*, *6*, 1205–1226. [CrossRef]

Chen, X., & Zhang, Z. (2019). A review of the impact of international collaboration on research performance. *Library and Information Service*, *63*(15), 127–139. [CrossRef]

Duan, C. (2024). Analyses of scientific collaboration networks among authors, institutions, and countries in Fintech studies: A bibliometric review. *FinTech*, *3*, 249–273. [CrossRef]

Ellegaard, O., & Wallin, J. A. (2015). The bibliometric analysis of scholarly production: How great is the impact? *Scientometrics*, *105*, 1809–1831. [CrossRef]

Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA*, *295*(1), 90–93. [CrossRef]

Glänzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics*, *51*, 69–115. [CrossRef]

Klarin, A., & Suseno, Y. (2021). A state-of-the-art review of the sharing economy: Scientometric mapping of the scholarship. *Journal of Business Research*, *126*(1), 250–262. [CrossRef]

Kurata, K., Miyashita, S., Sengoku, S., Kodama, K., & Lim, Y. J. (2023). A comparative analysis of social entrepreneurship and entrepreneurship: An examination of international co-authorship networks. *Sustainability*, *15*(22), 15873. [CrossRef]

Li, Y., & Wei, J. (2023). Visualization analysis of domestic data assets management based on bibliometric methods. *Straits Science*, *11*, 115–119. (In Chinese)

Li, Y., & Wu, Y. (2023). A comparative study on the characteristics of research collaboration networks in the field of accounting between China and foreign countries. *Accounting and Finance*, *6*, 1–7.

Liu, F., Lv, X., Yang, S., Yu, M., Chen, J., & Peng, S. (2025). A comparative analysis of research interest in medical students: An international and Chinese perspective over the past decade using CiteSpace. *Cogent Education*, *12*(1), 2510021. [CrossRef]

Liu, H., & Li, Y. (2023). Analysis of scientific research cooperation trends and research dimension of international academic social networks. *Information Science*, *3*, 119–126+144.

Masic, I., & Begic, E. (2016). Scientometric dilemma: Is H-index adequate for scientific validity of academic's work? *Acta Informatica Medica*, *24*(4), 228–232. [CrossRef]

Nan, D., Zhao, X., Chen, C., Sun, S., Lee, K. R., & Kim, J. H. (2025). Bibliometric analysis on ChatGPT research with CiteSpace. *Information*, *16*(1), 38. [CrossRef]

Tang, X., Zhu, K., & Liu, C. (2025). Data assets: A new frontier in accounting research—Literature review and research trends. *Commercial Accounting*, *1*, 4–11. (In Chinese)

Wang, D., Deng, C., Zhao, X., Lu, D., & Guo, X. (2024). Study on structure evolution and influencing factors of national agriculture science, technology and innovation collaboration network of the Belt and Road. *Science and Technology Management Research*, *5*, 32–39.

Wang, J., Cao, Q., Guo, S., Li, Y., Li, G., Yan, C., Zheng, M., Wang, D., & Xi, X. (2025). A comprehensive bibliometric analysis of brucellosis research: Insights from CNKI and Web of Science databases (2014–2023). *Journal of Multidisciplinary Healthcare*, *18*, 2921–2939. [CrossRef] [PubMed]

Zhang, Y., Pan, R., Wang, H., & Su, H. (2022). Community detection in attributed collaboration network for statisticians. *Stat*, *1*, e507. [CrossRef]

Zhou, D., & Li, R. (2024). Research hotspots and frontier trends in data assets accounting—A CiteSpace knowledge graph analysis. *Friends of Accounting*, *2*, 102–111. (In Chinese)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.