Received: 4 January 2024 | Revised: 4 March 2024 | Accepted: 6 March 2024
DOI: 10.1002/asi.24888

RESEARCH ARTICLE

JASIST WILEY

The drivers, features, and influence of first scientific collaboration among core scholars from Chinese library and information field

Xianzhe Peng | **Jin Shi**

School of Information Management, Nanjing University, Nanjing, China

Correspondence
Jin Shi, School of Information Management, Nanjing University, Nanjing 210023, China.
Email: shijin@nju.edu.cn

Funding information
National Social Science Fund of China, Grant/Award Number: 21BTQ012

Abstract

Collaboration among scholars in scientific research is increasingly common, making it important to address how to recommend suitable collaborators, especially for their first cooperation. To address this issue, this study focuses on 1487 core scholars in the field of library and information science in China, and then analyzes the impact of academic differences between these scholars in their first collaboration by using the propensity matching score method. It uncovers potential driving factors for scholars to reach first collaborations, including similar research productivity, contrasting academic influence, aligned research directions, and distinct research focuses. Then, the distribution of features of three types of first cooperation demonstrates that if one or both partners publish the first paper in this collaboration, the collaborative relationship tends to be more enduring and stable. In addition, the subsequent collaboration of scholars and the change in their academic differences are related to the initial academic differences between the two parties in the first collaboration. These patterns can be used to improve the accuracy and effectiveness of the scholar recommendation mechanism, hence promoting research collaboration and knowledge exchange.

1 | INTRODUCTION

Scientific collaboration is defined as "interaction taking place within a social context among two or more scientists that facilitates the sharing of meaning and completion of tasks with respect to a mutually shared, superordinate goal" (Sonnenwald, 2007). It can be thought of as a social communication network made up of researchers who collaborate in groups to produce knowledge and joint publications (Hilário & Grácio, 2017). Collaboration between scholars is becoming more complex and frequent, assisted by the advancement of information technology (Zhai & Yan, 2022). This has manifested a surge in research collaborations across nearly all subject areas, making the collaborative research an essential component of scientific operations and knowledge development (Matveeva & Ferligoj, 2020; Zhao et al., 2021). The ease with which scholars may communicate and interact with one another through modern technologies is a major cause of this increase in cooperation (Ribeiro et al., 2018; Zhai et al., 2014). Furthermore, due to the breadth and complexity of modern research, scientific collaboration is increasingly necessary (Cugmas et al., 2020; Milojević, 2010). Consequently, collaboration has become the mainstream approach to modern scientific development, with coauthorship becoming more prevalent than solitary authorship.

Numerous studies have shown that scientific collaborations have a favorable impact on scholar production and impact (Abramo et al., 2017; Adams, 2013; Beaver &

Rosen, 1978; He et al., 2009; Lee & Bozeman, 2005; Sonnenwald, 2007; Zuo & Zhao, 2018), which can influence their academic career advancement. Collaboration, in detail, can promote research efficiency (Bu et al., 2018; Coccia & Wang, 2016; Petersen et al., 2014; Zhang et al., 2018), the quantity and quality of research outcomes (Guan et al., 2017; Lee & Bozeman, 2005; Scarazzati & Wang, 2019), publication opportunities (Laband & Tollison, 2000; Lopaciuk-Gonczaryk, 2016), cost-effectiveness (Laband & Tollison, 2000), academic reputation (Floyd et al., 1994), knowledge dissemination and creation (Beaver, 2001; Katz & Martin, 1997; Wuchty et al., 2007), resource sharing, and the development of young researchers (He et al., 2021).

Due to the convenience, popularity and importance of research collaboration, the study of the driving factors, characteristic distribution, and subsequent influence of coauthorship is beneficial to promote knowledge production, scholar development and growth, scientific progress, and innovation. It is important to recommend the most potential collaborators who have never worked with the target researcher (Liu, Jaiswal, et al., 2022) and the most valuable collaborators among researchers who have previously collaborated with the target researcher (Liu et al., 2023; Liu, Bu, et al., 2022; Pradhan & Pal, 2020). In particular, the recommendation that leads to the first collaboration between two scholars who have never worked together promotes the collection of a broader scope of resources outside their preexisting relationships and conflicts, so it can increase team freshness (Liu, Jaiswal, et al., 2022), provide scientists with wide reach (Gao et al., 2021), help access more complementary academic resources, and enhance scientific originality and novelty (Porac et al., 2004; Skilton & Dooley, 2010; Yong et al., 2014; Zeng et al., 2021). For accurate recommendation for scientific collaboration, this study utilizes the database from Chinese Social Science Citation Index (CSSCI) involving Library and Information Science (LIS) field, to examine the driving factors of the first collaboration by using the propensity matching score method, explore the distribution of characteristics under various collaboration patterns, and analyze the subsequent influence of this collaboration via correlation analysis.

2 | RELATED WORK

As growing influence and popularization of scientific collaboration (Tang et al., 2008; Xia et al., 2017), many academics and organizations are interested in describing the patterns of scientific collaboration (Cohen & Ebel, 2013; Tang et al., 2008; Tsai & Lin, 2016; Wu et al., 2015). Scientific collaboration has gotten a lot of attention, with studies spanning from contribution (Corrêa Jr et al., 2017; Lu et al., 2020), population (Li et al., 2015), discipline (Moody, 2004; Wagner & Leydesdorff, 2005; Xie et al., 2018), country (Katz, 1994; Perc, 2010), gender (Kwiek & Roszka, 2021), and multination (Glänzel et al., 1999; Gómez et al., 1999; Leclerc & Gagné, 1994; Russell, 1995) to the relationship with citations (Shen et al., 2021). Especially, the analysis and modeling of scientific collaboration dynamics can enhance our comprehension of successful collaborations as well as increase their efficiency and productivity (Barabási et al., 2002; Sinatra et al., 2016; Wang, Cui, et al., 2017). Moreover, the studies about repeat collaboration demonstrate first collaboration can add freshness and novelty to the research team (Liu, Bu, et al., 2022; Liu, Jaiswal, et al., 2022; Skilton & Dooley, 2010). However, previous studies mostly focus on static collaborative habits (Newman, 2001, 2004; Wang, Yu, et al., 2017; Zhang et al., 2018), and few studies have been conducted to follow the entire collaborative process.

From the standpoint of long-term collaborations, Petersen discovered that long-term collaborations, particularly with "life-partner" collaborators, have a significant positive impact on productivity and citation (Petersen, 2015). Then, Coccia and Wang investigate the long-term patterns of international research collaboration across scientific domains, as well as how they evolve structurally over time (Coccia & Wang, 2016). Moreover, Wang et al. developed a model to predict partnership sustainability based on structural similarity indices, authorship attributes, and research objectives (Wang, Cui, et al., 2017), furtherly proposed sustainable collaborator recommendation (Wang, Liu, et al., 2019), and furtherly leverage the early-stage reciprocity to reveal sustainable scientific collaborations (Wang, Ren, et al., 2020). Recently, Hückstädt suggests important factors influencing successful research collaboration include the characteristics of the participating principal investigators and spokespersons, the collaborative development of shared goals and questions, interconnecting the subprojects' research, the synthesis of their research results, and the team climate (Hückstädt, 2023).

While scientific collaboration is not a new phenomenon in scientific practice, there remains a high demand to identify the prerequisites for successful collaboration (Hückstädt, 2023). Surprisingly, there is a lack of systematic research on the conditions that facilitate successful scientific collaboration, with only a few scattered preliminary studies focusing on the differences among collaborators in respect to their age (Liang et al., 2001), gender (Fell & König, 2016), cultural (Sørensen, 2003), academic social network (Lopes et al., 2010), academic rank (Gaughan & Bozeman, 2016), research interests (Kong et al., 2017), and academic ability (Abramo et al., 2017). Moreover, most of previous studies investigates the significance and predictability of collaboration sustainability, and there is a lack of quantitative and microscopic understanding of whole collaboration process including its driving factors, sustainable features, and influence. To reveal the entire collaboration process from the new perspective of focusing the first collaboration, we track core scholars in Chinese LIS from first to ongoing collaboration, and finally get drivers, persistence characteristics and subsequent impact of the first collaboration.

3 | DATA AND METHODOLOGY

3.1 | Dataset

CSSCI, established by Institute for Chinese Social Sciences Research and Assessment at Nanjing University in 1998s, employs a combination of bibliometric indicators and peer review to index more than 600 journals with high academic and editorial standards from more than 2700 Chinese journals of humanities and social sciences. As CSSCI is widely considered as an authoritative and comprehensive database for bibliometric studies of Chinese social sciences (Gong & Cheng, 2022; Jie et al., 2008; Song et al., 2015; Wang et al., 2016; Yan et al., 2010), we use CSSCI dataset in the field of LIS to evaluate the motivation, perseverance, and impact of first research collaboration. The dataset includes 49,848 authors, 90,174 articles between 1998s and 2022s.

Since not every scholar constantly writes publications, we identify a core group of scholars in the field via analyzing the numeric distribution of scholars and their publications based on Bradford's law. Specifically, in accordance with the steps of determining core journals through Bradford's Law, scholars in a specific field are first ranked in descending order based on the number of their publications. Then, they are divided into three zones with equal total numbers of papers. The number of scholars in these three zones follows a geometric progression of $1{:}n{:}n^2$, where n is the Bradford constant, and the scholars in the first zone are considered the core scholars in this field. In this study, the distribution of scholars in the three zones is approximately in a ratio of 1487:8192:40169, representing a geometric progression of approximately $1{:}5{:}5^2$. The results show that the Bradford constant is around 5, and newly obtained core group of scholars has 1487 persons, 55,640 publications, and 5597 first collaboration pairs (Figure 1). The first collaboration partnerships formed by core scholars with continuous research productivity are more representative, and the results drawn from such collaborative relationships are more reasonable and correct.

Laudel categorizes research collaboration into six types: (1) collaboration involving division of labor, (2) service collaboration, (3) knowledge transmission, (4) provision of research equipment, (5) reliable evaluation, and (6) mutual motivation (Laudel, 2002). The collaboration in the latter five categories is often difficult to be tracked in real-time using publicly available records, and is usually collected through interviews, surveys or questionnaires (Boardman & Corley, 2008; Boardman & Ponomariov, 2009; Ponomariov & Boardman, 2008). This makes it inconvenient to extensively investigate the research collaboration that occurs among different types of scholars. On the other hand, collaboration in the first category often takes the form of coauthored publications. Although coauthorship does not completely equate to research collaboration (Katz & Martin, 1997), it is still considered as a partially effective operationalization of collaboration (de Oliveira et al., 2021; Dehdarirad & Nasini, 2017; Li et al., 2020; Lundberg et al., 2006). Furthermore, there is already a large body of research that uses coauthorship as a representation of research collaboration between scholars (Adams et al., 2005; Ponomariov & Boardman, 2010, 2016; Yoshikane & Kageura, 2004). In addition, non-core scholars in the LIS

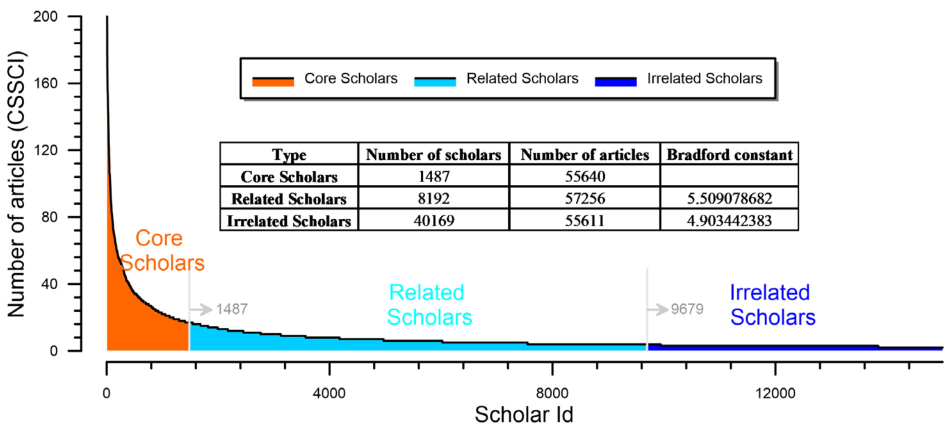

FIGURE 1 Core, related and unrelated scholars according to Bradford's Law by using scholar productivity.

field typically have shorter academic careers (Wang, Yu, et al., 2017), and the coauthorship relationships generated from their published papers are often not persistent and exhibit a certain degree of randomness (Cugmas et al., 2016). To avoid potential interference from this, we specifically focus on the 5597 pairs of coauthorship relationships among core scholars formed based on the bibliography data collected above (see detail in Table S1, Supporting Information). It aims to analyze the specific characteristics of research collaborations among core scholars and identify any patterns that may exist.

3.2 | Propensity score matching

PSM (propensity score matching) is a nonparametric method proposed by Rosenbaum and Rubin that is one of the most successful methods for assessing treatment effects when randomized trials are not possible (Rosenbaum & Rubin, 1984). This randomized trial is an appropriate method for estimating treatment effects since it randomly allocates individuals to the treated and untreated groups, resulting in no variations in their distributions of observable features, unobservable features, and treatment effects within them. In most real-world situations, however, we can only have observational samples instead of running random studies. As a result, researchers are paying close attention to the matching approach, PSM, which is more intuitive and effective at estimating the treatment effect from this type of data by minimizing selection bias.

PSM, in further detail, uses a mapping function to convert the multidimensional variable X into a one-dimensional propensity score (Xi), and then performs matching based on the propensity score. In this procedure, samples with identical or very similar propensity scores will have similar observable confounding covariate values (Austin, 2011), making the treated and untreated groups comparable (Morgan, 2018). PSM has recently been widely used in numerous domains of social sciences, including management, economics, and education, for estimating treatment effects (Caliendo & Kopeinig, 2008; Kim & Park, 2021; Luellen et al., 2005; Zong et al., 2020). As a result, PSM is appropriate for identifying prospective drivers of first research collaboration in this study.

3.3 | Topic model

LDA (Latent Dirichlet Allocation) topic model is a statistical method that employs Bayesian estimation to represent the subjects of each text in a text set as a probability distribution (Blei et al., 2003). The advantage of this model as an unsupervised learning technique is that there is no requirement to label the training set and only the optimal number of topic clusters must be given. In this study, keywords of publications are primarily employed to produce correlations of perplexity, coherence, and number of topics by restricting the range of variation in the number of subject clusters. Based on this correlation and "elbow-folding method" (as in Figure 2), it is determined that optimal number of topic clusters should be 9.

Next, we apply Author-Topic Model (ATM), which represents both topics and authors by adding author elements, for mining author-topic relationships from document sets (Steyvers et al., 2004). This model incorporates the analyzed author information of the document into the LDA model in order to establish semantic topic

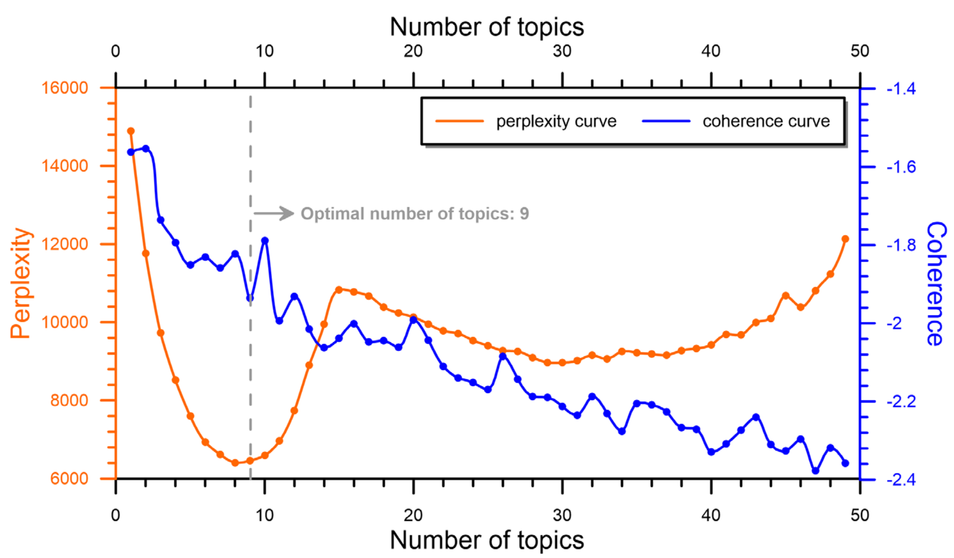

FIGURE 2 Correlations of perplexity, coherence, and number of topics based on LDA topic model.

connection between words, themes, authors, and documents. It uses a multinomial distribution matrix on the word layer to map each author to an implicit topic, next acquires a mixed distribution of authors on every topic. Finally, this ATM can be used to calculate the probability distribution of each scholar on each of nine topics identified above.

3.4 | Louvain algorithm

Louvain algorithm is a community discovery algorithm based on graph data proposed by Blondel et al. (Blondel et al., 2008). As one of the community discovery algorithms based on modularity, it performs better in terms of efficiency and effectiveness (Sattari & Zamanifar, 2018; Wang, Zou, et al., 2019) and aims to discover hierarchical community structures. As a result, this approach is applicable to determine the research team to which the scholar belongs based on collaboration links between scholars.

Specifically, Louvain's algorithm identifies scientific teams in two steps, repeatedly executing those steps until the modularity is maximized. In the first step, a single author (node) is considered as a research team with team size of 1, and the modularity Q is calculated for this division, and then each author is merged with its neighboring authors into a research team in turn, and the modularity gain ΔQ is calculated: if ΔQ is greater than 0, the author is assigned to the research team to which the neighboring author belongs, and if ΔQ is less than or equal to 0, the author is rejected to be assigned to the research team to which the neighboring author belongs. Repeat this step until all authors' research teams no longer change, then step 1 is aborted and we move to step 2. In the second step, all authors of each research team are compressed into one node, and the sum of the edge weights between the newly generated "research team nodes" and the edge weights between all nodes within the "research team nodes" is calculated for the next round of step 1.

Based on the CSSCI database, we apply the above steps to construct the coauthorship network of core scholars over different time periods in the field of Chinese LIS. This allowed us to track the evolution of the academic cliques of these core scholars from 1998s to 2022s and determine which academic clique each core scholar belonged to at different years (Figure 3). Each panel in Figure 3 uses the Fruchterman-Reingold layout, and represents the coauthorship network relationships among core scholars formed by all the papers up until the year indicated by that panel. For example, the panel marked as 2010s represents the coauthorship relationships among core scholars in the Chinese LIS field for all papers up until 2010s. The nodes of different colors represent different research teams, and the division of research teams is achieved by applying the Louvain community detection algorithm to the coauthorship network of scholars up until 2010s. By examining whether the nodes in the coauthorship network of scholars in a specific year, as shown in Figure 3, have the same color, we can determine whether they belonged to the same research team by that year. This enables us to investigate whether core scholars belonged to the same research team when they initially collaborated or continued their collaboration, thereby understanding the impact of belonging to the same research team on their research collaboration behavior.

3.5 | Differential expression analysis

Analysis of differential expression is a method used to compare differences or similarities between two or more different groups, and the method can be applied in a variety of fields, including biology, medicine, social sciences, and economics, to help identify meaningful variables and determine how they relate to specific outcomes. Limma (Linear models for microarray data), as a differential expression analysis method based on a generalized linear model (Law et al., 2014), can compute moderated *t*-statistics, moderated *F*-statistic, and log-odds of differential expression by empirical Bayes moderation of the standard errors towards a global value to obtain differential genes between different comparison groups and controls (Phipson et al., 2016; Ritchie et al., 2015; Smyth, 2004). Based on Limma package from R languages (Smyth, 2005), this method is also adoptable to analyze obvious differences in interpersonal differential indicators between cooperating and non-cooperating scholars from matched samples after PSM in this study.

3.6 | Variables

3.6.1 | Intrapersonal difference

Intrapersonal differences affect the design and validity of a research cooperation that can be connected to the personality traits and features of the collaborating parties (Misra et al., 2011). Therefore, based on previous research, this study focuses on differences between their age (Liang et al., 2001), academic social networks (Lopes et al., 2010), research interests (Kong et al., 2017), and academic abilities (Abramo et al., 2017), while excluding gender and cultural differences that have a minor impact

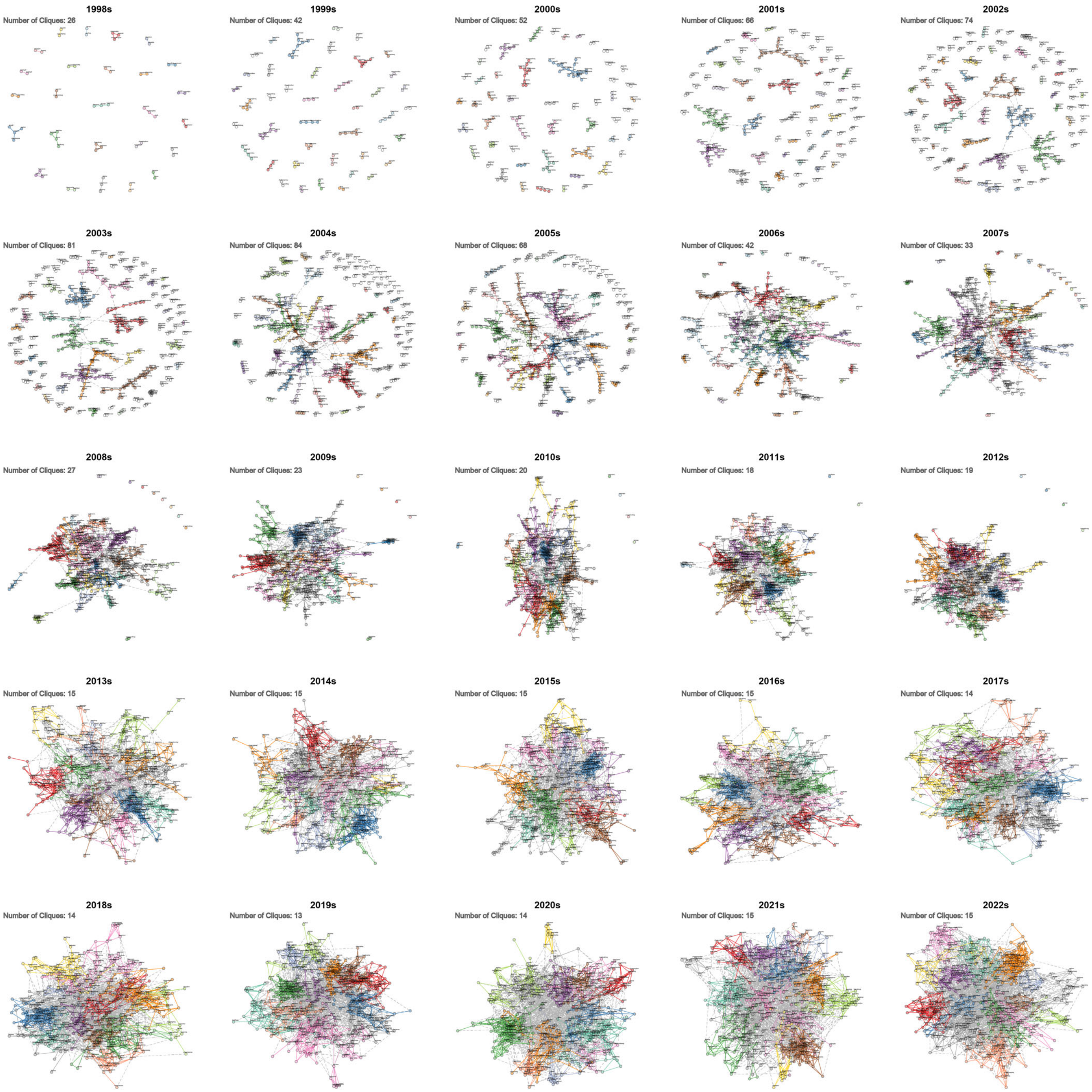

FIGURE 3 Evolution of academic community in Chinese Library and Information Science from 1998s to 2022s based on the CSSCI database.

on collaboration (Birnholtz, 2007; Gaughan & Bozeman, 2016), as well as differences in academic ranking represented by differences in available academic capabilities (Abramo et al., 2011). Specifically, scholars' academic age can be indicated by the cumulative duration of their involvement in scientific research (i.e., career age) (Xu et al., 2022). Their academic social networks are represented by their previous collaborator and scientific cliques they belong to. Furthermore, their research interests are reflected by the keyword lists of their published papers as well as their probability distribution on each of nine topics calculated from ATM.

In general, the productivity of a scholar is often measured by the volume of their publications, and the resulting citation count to some extent represents their academic influence. The scholar's h-index is defined as the number of his or her papers with citation number $\geq$ h, giving particular emphasis to the high-impact papers (Hirsch, 2005). This metric serves to quantify the influence and number of a scholar's high-quality academic publications and is widely utilized to gauge scholarly impact (Saad, 2010). To provide a more comprehensive assessment of a scholar's academic productivity and impact, the p-index has been suggested (Prathap, 2010). By considering the total output of a scholar's papers and the number of citations, it takes into account all of the scholar's publications, thereby achieving a balance between quantity and quality of these publications and offering a more holistic evaluation of the scholar's comprehensive academic ability (Prathap, 2010, 2011). Therefore, we use the scholar's number of publications, total citations, h-index, and p-index to reflect their academic abilities from multiple aspects.

The variance in the time span of journal articles published by both academics at the time of scholars' first collaboration explains the variety in the age of their research at this time (simplified as "age_dif"), which also represents their difference in academic seniority. The difference of the collaborative network between scholars can be measured by Jaccard similarity of their previous collaborators (simplified as "coauthor_jc"), and whether they belong to the same clique (simplified as "community_sim"). Similarly, the difference of research directions between scholars can be indicated by the Jaccard similarity of keywords (simplified as "key_jc") from their publications, the Jensen-Shannon divergence (simplified as "topic_js") and cosine similarity (simplified as "topic_cos") of the probability distribution of scholars' topics. Simply, the difference of their academic ability is measured as the difference between the scholar's number of publications (simplified as "total_dif"), total citations (simplified as "cited_dif"), h-index (simplified as "h_dif"), and p-index (simplified as "p_dif"). The meaning of these indicators and others is illustrated in Table 1.

3.6.2 | Collaborative persistence

Usually, scientific collaboration is not a one-time event in which two researchers work together only once. To initially quantifying the continuity of research collaborations, we count times and duration of cooperation. Further, we propose the notion of cooperation half-life to measure the persistence of the partnership using the calculation method of the half-life of the literature (Burton & Kebler, 1960). The following is the precise formula for calculating cooperation half-life:

TABLE 1 Meaning of intrapersonal difference.

Indicators	Definition	Usage
age_dif	Differences in the duration of scholar *i* and *j* have been involved in research at the time of their first collaboration	Measurement for intrapersonal difference at the time of the scholars' first collaboration
coauthor_jc	Jaccard similarity of the set of collaborators owned by both scholar *i* and *j* at the time of their first collaboration	
community_sim	Whether scholar *i* and *j* belong to the same scientific clique, with 1 indicating yes and 0 indicating no	
key_jc	Jaccard similarity of the set of keywords in articles published by scholar *i* and *j* before they first collaboration	
topic_js	Jensen-Shannon divergence of the probability of topic distribution when scholar *i* and *j* first collaborate	
topic_cos	Cosine similarity of the probability of topic distribution when scholar *i* and *j* first collaborate	
total_dif	Difference in the number of papers published by scholar *i* and *j* before they first collaborated	
cited_dif	Difference in the total number of citations of papers published by scholar *i* and *j* when they first collaborated	
h_dif	Difference in h-index between scholar *i* and *j* before they first collaborate	
p_dif	Difference in p-index between scholar *i* and *j* before they first collaborate	
final_coauthor_jc	Jaccard similarity of the set of collaborators owned by both scholar *i* and *j* until 2022s	Measurement for intrapersonal difference until 2022s
final_key_jc	Jaccard similarity of the set of keywords in articles published by scholar *i* and *j* until 2022s	
final_topic_js	Jensen-Shannon divergence of the probability of topic distribution of scholar *i* and *j* until 2022s	
final_topic_cos	Cosine similarity of the probability of topic distribution of scholar *i* and *j* until 2022s	

$$\text{Halflife} = Y + \frac{50\% - C}{D - C},$$

where C is the cumulative percentage of cooperation for the year with the closest cumulative frequency of cooperation and less than 50%, D is the cumulative percentage of cooperation for the year with the closest cumulative frequency of cooperation and greater than 50%, and Y is the time interval from the year with the cumulative cooperation percentage of C to the year with the most recent cooperation. It is worth noting that the above method cannot calculate situations where all research collaborations are concentrated in the same year or when over 50% of research collaborations are concentrated in the last year. These types of research collaborations do not exhibit a decaying pattern, so their half-life is set to 0. In summary, we use times, duration and half-life of cooperation to measure collaborative persistence.

3.7 | Overview of the research framework

To explore the drivers, features and subsequent influence of the first collaboration, we design a research framework (Figure 4). In this framework, the steps of data processing include identifying core scholars, constructing and calculating scholar attribute indicators, measuring the

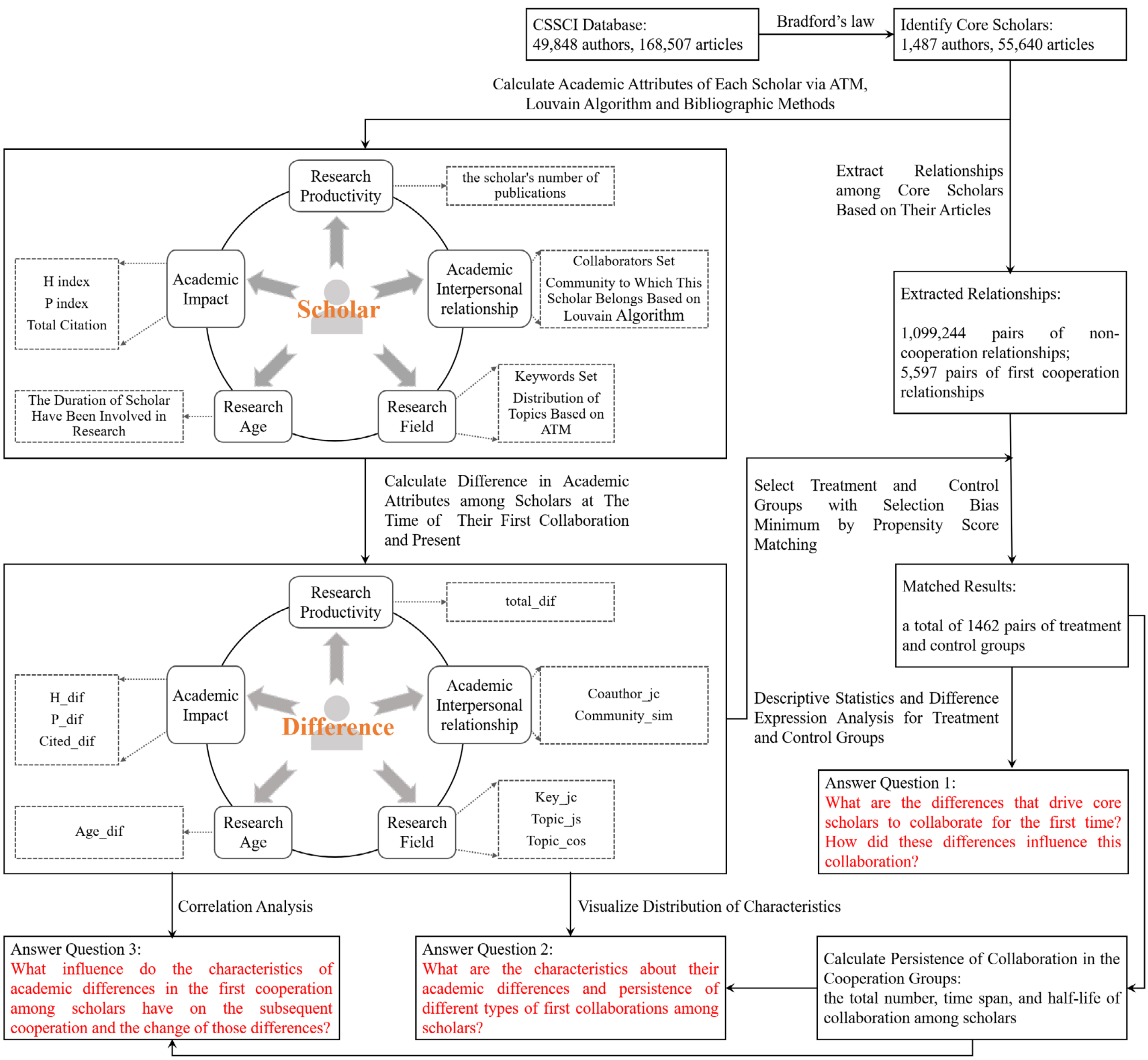

FIGURE 4 Research framework of the present studies.

difference degree of scholars' academic attributes, matching the treatment and control group, and establishing indicators for measuring collaborative persistence. Based on difference expression analysis and correlation analysis for above processed data, we will answer the following three questions:

1. What are the differences that drive core scholars to collaborate for the first time? How did these differences influence this collaboration?
2. What are the characteristics about their academic differences and persistence of different types of first collaborations among scholars?
3. What influence do the characteristics of academic differences in the first cooperation among scholars have on the subsequent cooperation and the change of those differences?

4 | RESULTS AND ANALYSIS

4.1 | Difference analysis based on PSM

This study first uses logit model to estimate the fitted values of the conditional likelihood of scholars reaching first collaboration, and adopts the nearest neighbor matching method to choose treatment and control groups, with whether or not to reach first cooperation as the dependent variable and 10 indicators of intrapersonal difference as independent variables. Based on 1,099,244 pairs of non-cooperation and 5597 pairs of first cooperation relation extracted from 55,640 publications from core scholar, we analyze the differences in the distributions of the 10 variables before and after matching, as well as the changes in covariates to validate the use of PSM. Consistency of the distribution of dependent variables between the treatment and control groups after matching shows that the influence of confounding factors is successfully reduced by PSM, enhancing the reliability and correctness of inference for the driver of first cooperation (Figure 5).

Using PSM method, a total of 1462 pairs of treatment and control groups were matched from 1,099,244 non-cooperative relationships and 5597 first cooperative relationships. According to the findings of the statistical analysis of the corresponding factors in the matched sample (as in Table 2), seven independent variables were found to be substantially different, whereas three independent variables were not. In the cooperative group as opposed to the non-cooperative group, coauthor_jc, topic_js, topic_cos, h_dif, p_dif, total_dif, and key_jc are

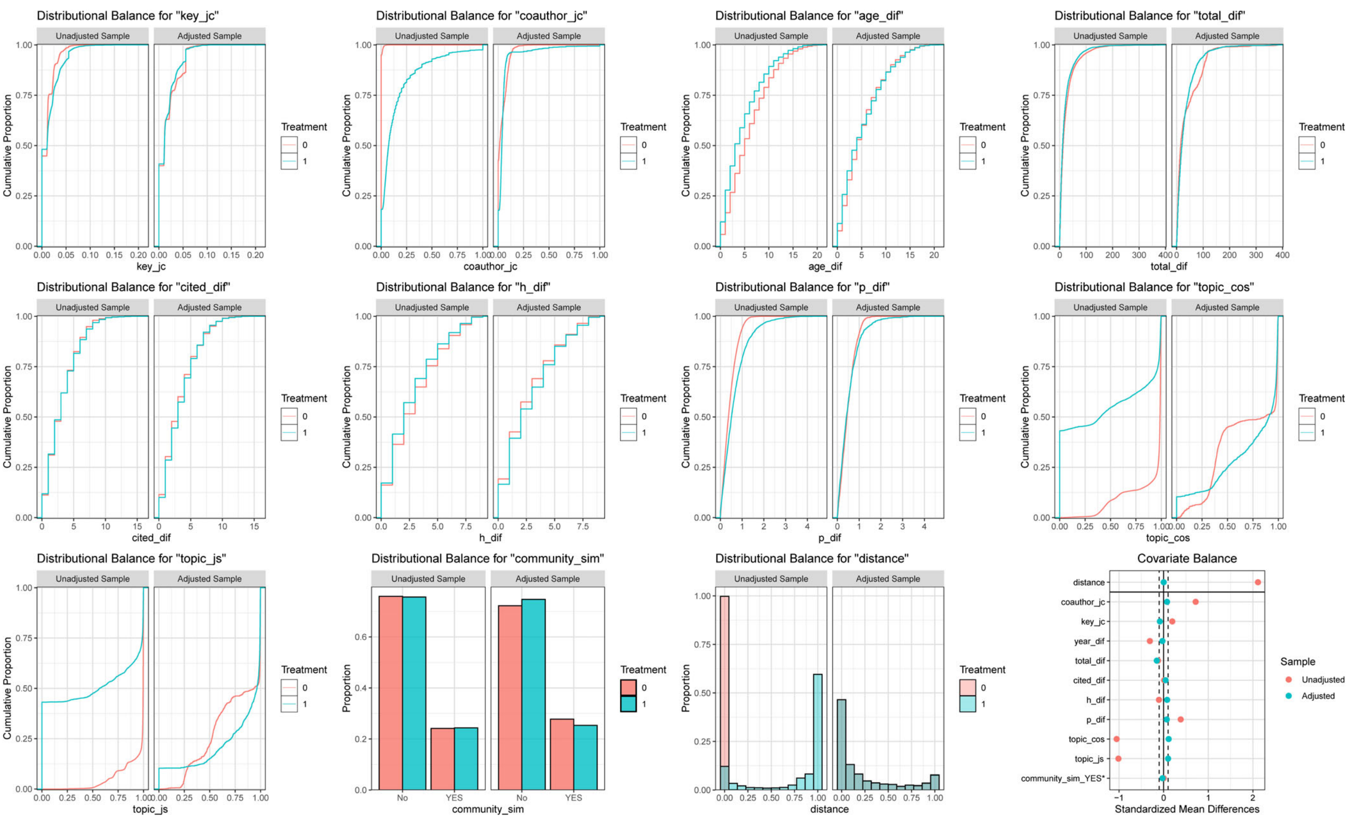

FIGURE 5 Differences in the distributions of the 10 variables before and after matching, as well as the changes in covariates and distance.

TABLE 2 Statistical analysis of the corresponding factors in the matched and unmatched sample.

Process	Before matching					After matching				
Indicator type	**Non-cooperation ($N = 1,099,244$)**	**Cooperation ($N = 5597$)**	***t*-value**	**SD**	***p*-value**	**Non-cooperation ($N = 1462$)**	**Cooperation ($N = 1462$)**	***t*-value**	**SD**	***p*-value**
coauthor_jc										
Mean (SD)	0.000896 (0.00487)	0.166 (0.314)	<0.001	0.0042	<0.001***	0.0371 (0.0468)	0.0535 (0.109)	<0.001	0.00311	<0.001***
Median [min, max]	0 [0, 0.333]	0.0741 [0, 7.00]				0.0138 [0, 0.333]	0.0357 [0, 1.40]			
key_jc										
Mean (SD)	0.0110 (0.0132)	0.0148 (0.0195)	<0.001	<0.001	<0.001***	0.0175 (0.0202)	0.0159 (0.0187)	2.27	<0.001	0.0236*
Median [min, max]	0.0109 [0, 0.147]	0.0109 [0, 0.211]				0.0115 [0, 0.0986]	0.0114 [0, 0.104]			
age_dif										
Mean (SD)	5.91 (4.38)	4.64 (4.13)	23	0.0554	<0.001***	5.24 (4.21)	5.13 (4.43)	0.698	0.16	0.485
Median [min, max]	5.00 [0, 21.0]	3.00 [0, 20.0]				4.00 [0, 21.0]	4.00 [0, 20.0]			
total_dif										
Mean (SD)	24.7 (34.1)	21.0 (29.4)	9.38	0.395	<0.001***	37.2 (47.6)	32.7 (39.1)	2.8	1.61	0.00517**
Median [min, max]	13.0 [0, 384]	11.0 [0, 358]				15.0 [0, 384]	20.0 [0, 358]			
cited_dif										
Mean (SD)	3.13 (2.41)	3.16 (2.50)	<0.001	0.0335	0.364	3.31 (2.65)	3.42 (2.62)	<0.001	0.0974	0.261
Median [min, max]	3.00 [0, 16.0]	3.00 [0, 15.0]				3.00 [0, 14.0]	3.00 [0, 13.0]			
h_dif										
Mean (SD)	2.86 (2.31)	2.62 (2.26)	7.82	0.0302	<0.001***	2.61 (2.30)	2.78 (2.33)	<0.001	0.0857	0.0388*
Median [min, max]	2.00 [0, 9.00]	2.00 [0, 9.00]				2.00 [0, 9.00]	2.00 [0, 9.00]			
p_dif										
Mean (SD)	0.429 (0.317)	0.654 (0.586)	<0.001	0.00783	<0.001***	0.490 (0.328)	0.531 (0.478)	<0.001	0.0152	0.00697**
Median [min, max]	0.369 [0, 2.27]	0.510 [0, 4.68]				0.442 [0, 1.75]	0.413 [0, 3.48]			
topic_cos										
Mean (SD)	0.909 (0.181)	0.449 (0.438)	78.7	0.00585	<0.001***	0.675 (0.327)	0.724 (0.334)	<0.001	0.0122	<0.001***
Median [min, max]	0.987 [0.0122, 1.00]	0.402 [0, 1.00]				0.863 [0.0122, 0.999]	0.907 [0, 1.00]			
topic_js										
Mean (SD)	0.951 (0.112)	0.492 (0.453)	75.8	0.00606	<0.001***	0.745 (0.281)	0.790 (0.315)	<0.001	0.011	<0.001***
Median [min, max]	0.995 [0.0419, 1.00]	0.568 [−0.0000267, 1.00]				0.935 [0.0431, 0.999]	0.964 [0, 1.00]			
community_sim										
No	834023 (75.9%)	4233 (75.6%)			0.684	1056 (72.2%)	1092 (74.7%)			0.143
Yes	265221 (24.1%)	1364 (24.4%)				406 (27.8%)	370 (25.3%)			

Note: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

higher and total_dif, key_jc are lower, whereas age_dif, cited_dif, and community_sim are not substantially different.

Next, the differential analysis was used to generate the volcano plot for the matched samples. In this plot (Figure 6), the *p*-value is primarily used to assess whether there is a significant difference between the two sets of data, while the Fold Change value is used to quantify the degree of the difference. The results show that the significant differences in the independent variables are consistent with the descriptive analysis statistics, and the independent variables are arranged in descending order according to the degree of significance, in the order of coauthor_jc, topic_js, topic_cos, total_dif, p_dif, key_jc, and h_dif, and in descending order according to the degree of the difference between the independent variables in the two groups, in the order of topic_cos, topic_js, h_dif, key_jc, total_dif, p_dif, and coauthor_jc. Moreover, the level of significance and degree of differences are obvious by directly comparing nine normalized independent variables of the control and treatment groups in the heatmap (Figure 7).

4.2 | Distribution of characteristics and effects of cooperation

First collaborations are classified into three types based on whether the collaborative paper published by both or single scholars of this collaboration is the first paper of their research careers (as in Table 3). Afterwards, we analyze the persistence characteristics and subsequent effects of these three types of first collaborations via statistics analysis.

From this analysis result (Figure 8), the current coauthor_jc, key_jc, topic_js, and topic_cos between the

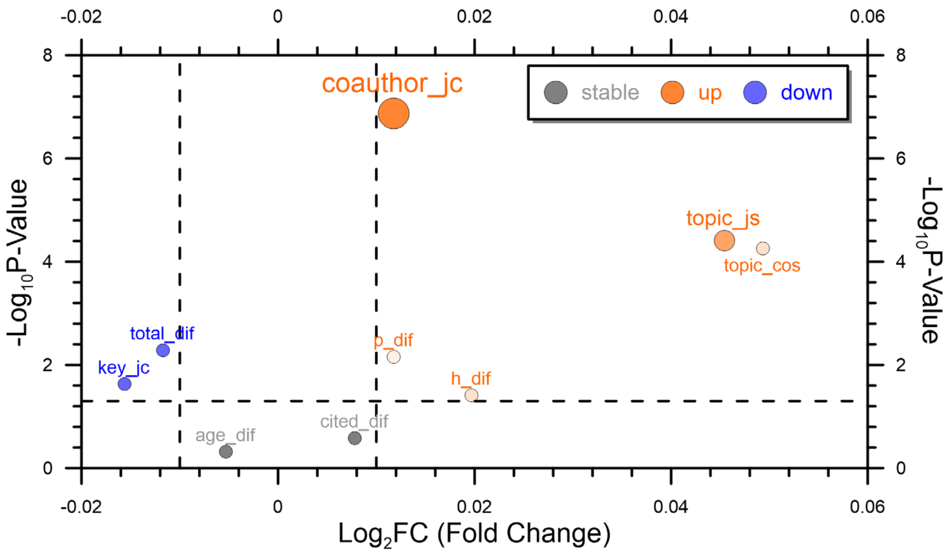

FIGURE 6 Volcano plot based on *p*-value and Fold Change calculated by the empirical Bayesian adjusted *t*-test method.

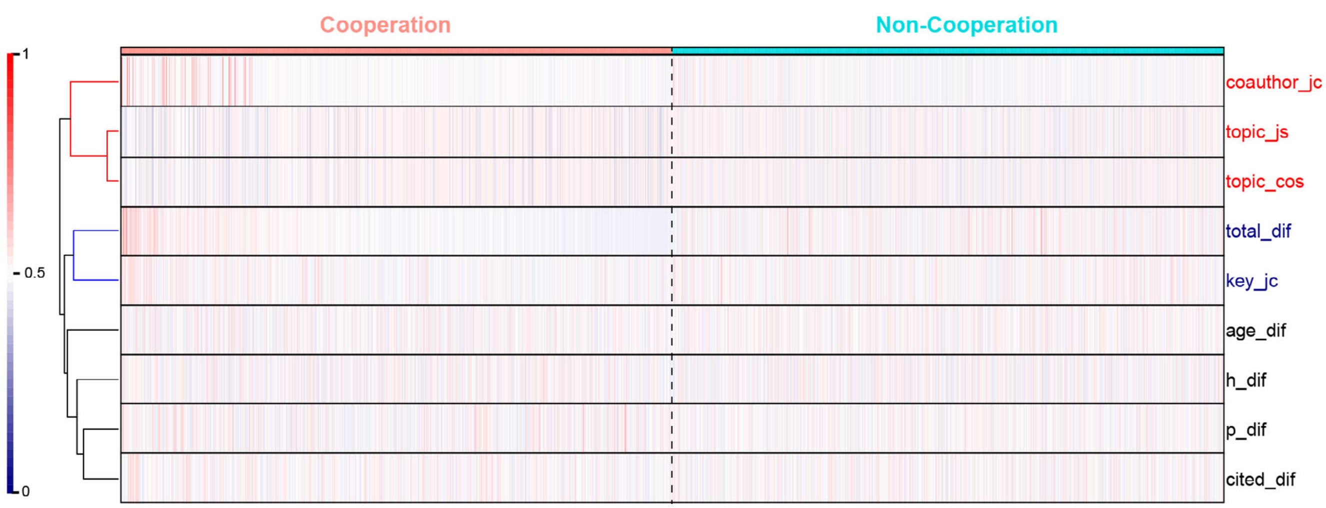

FIGURE 7 Heatmap of normalized independent variable distribution in control and treatment groups.

TABLE 3 Type of first collaboration.

Type of first collaboration	Description about cooperative characteristic
Type_0	The first collaborative paper is the first paper in the research careers of both collaborating parties
Type_1	The first collaborative paper is only the first paper in the research career of one of the collaborating parties
Type_2	The first collaborative paper is not the first paper in the research career of any of the collaborating parties

scholars of the 3 first cooperation types, Type_0, Type_1, and Type_2, are decreasing in this order. Moreover, the times, duration, and half-life of collaboration among scholars of the three first collaboration types, Type_0, Type_1, and Type_2, decreases accordingly.

4.3 | Correlation analysis of cooperative characteristics

Since most of the first collaboration types are Type_2, and there are nulls in the scholar variability indicators of the 2 first collaboration types, Type_0 and Type_1, we

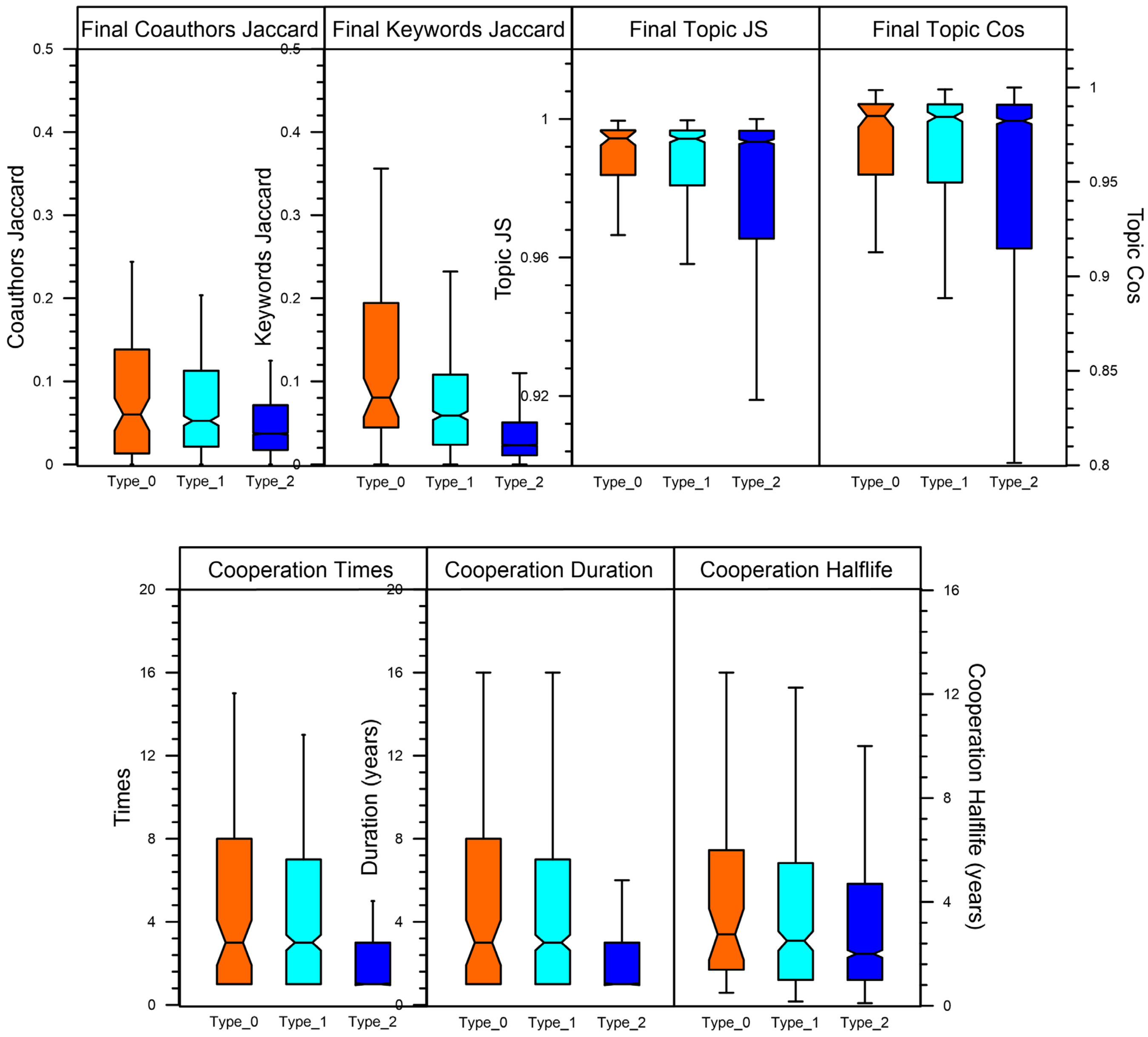

FIGURE 8 Distribution of subsequent effects and persistence characteristics of three types of first cooperation.

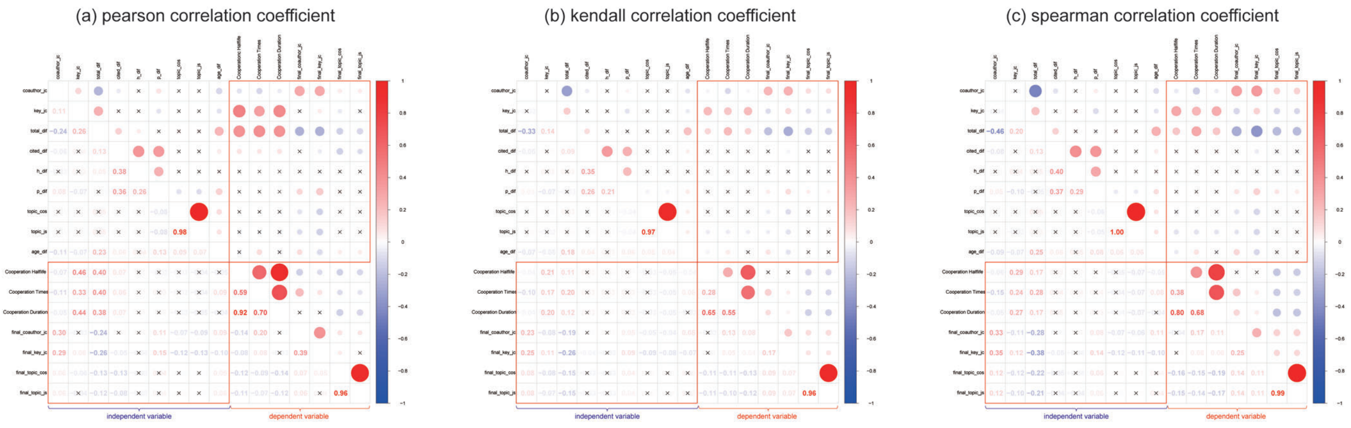

FIGURE 9 The correlation among initial interpersonal differences of both sides of the first collaboration, the characteristics of collaboration persistence, and their current differences.

mainly investigate the correlation between the scholar variability characteristics of the first collaboration type, Type2, and the number of subsequent collaborations, collaboration time span, collaboration half-life, and current scholar variability characteristics. The correlation between initial interpersonal differences of both sides of the first collaboration and the characteristics of collaboration persistence reveals that key_jc and total_dif have a significant positive correlation with collaboration persistence, while coauthor_jc and topic_js have a weak negative correlation (Figure 9).

Besides, the correlation results from initial and current interpersonal differences between the two sides of the first collaborating scholars show that final_coauthor_jc is positively correlated with coauthor_jc, p_dif, and age_dif, while it is negatively correlated with total_dif, topic_js, and topic_cos. The results also show that final_key_jc is positively correlated with coauthor_jc, key_jc, and p_dif, and negatively correlated with age_dif, total_dif, topic_js, and topic_cos. In addition, final_topic_js and final_topic_cos are weakly positively correlated with coauthor_jc, and negatively correlated with key_jc and total_dif (Figure 9).

5 | DISCUSSION

5.1 | Driver of first research collaboration

From results of differential analysis based on PSM, the more common collaborators there are among scholars, the more similar their research topics, the greater their difference in academic impact, the closer their research productivity, and their less overlap in research directions, the stronger the willingness of the two to first collaborate. In addition, the difference in whether the scholars belonged to the same clique and the age of the research did not have a significant effect on the two reaching the first collaboration. In detail, the scholars who completed the first collaboration usually have a large intersection of academic networks, and the differences in academic influence between the two tend to be large, and the research themes and research productivity are closer, but the overlap in specific research directions tends to be less. To conclude, the two sides of the first collaboration have a similar academic social network, and although the general research direction and research productivity are generally closer, the greater differences in academic influence and the little overlap of keywords, which indicates that there is no direct competition between them and the collaborative intention to assist each other is clear.

This result shares some similarities with previous studies exploring the driving factors of successful collaborations. Specifically, if two parties who have not collaborated before have more common previous collaborators and similar research fields, they are more likely to achieve initial collaboration. This finding aligns with existing scholar collaboration recommendation mechanisms based on research content (Huang et al., 2021; Kong et al., 2017), scientific social networks (Li et al., 2014; Wang, Liu, et al., 2020; Wang, Ren, et al., 2020), and mixed methods (Chen et al., 2021; Liu et al., 2023). Compared to previous research, this study takes a more diversified perspective and explores more specific factors and unrelated factors influencing initial collaboration success. For example, although achieving initial collaboration between two parties who have not previously collaborated requires having more common existing collaborators, they do not necessarily need to belong to the same scientific social clique. This indicates that the number of common previous collaborators required for initial collaboration is not high, and the initial collaboration between two parties may be facilitated by a few common previous collaborators who introduce them to each other (Zhang et al., 2018). Additionally, while the research fields of the two parties need to be somewhat similar for initial collaboration, it is also important for their specific research topics to have certain differences. This suggests that scholars seek initial collaboration partners who have overall alignment in their research directions but can complement each other in terms of specific research points, making the collaborative cooperation mutually more beneficial.

From the perspective of bibliometric indicators, scholars with similar publication counts but varying academic influence are more likely to achieve initial collaboration. This may be because collaboration with highly cited and renowned scientists is more beneficial for scholars' career development (Li et al., 2019). Furthermore, contrary to previous research suggesting that research collaborations often occur between scholars with similar academic ages (Wang, Yu, et al., 2017), the degree of difference in academic age between the two scholars does not have a significant impact on their possibility of achieving initial collaboration. The likelihood of achieving initial collaboration is influenced more by the differences in their research directions, academic abilities, and scientific social networks.

5.2 | Persistence and impact of three types of first collaboration

Based on the statistical results of the persistence characteristics of collaboration for three types of first collaboration and the current interpersonal differences among scholars, it is found that whether the paper published in first collaboration is the first paper of both parties or single scholars affects times, duration, and stability of subsequent collaborations, as well as the difference of the two in terms of subsequent academic interpersonal relationships, specific research points, and research topics. If scholars publish the first paper of one or more of them in the first collaboration, the more frequent, long-lasting and stable their subsequent collaboration is, and at the same time the more similar the future development of their academic interpersonal relationships, specific research points and research themes. This indicates that there is a certain regularity in the future development trajectory between scholars with three types of first collaboration, and this potential law can be used to optimize and enrich the recommendation mechanism of scientific collaboration in future study.

When scholars publish their first paper during their early academic careers, they typically seek to collaborate with scholars such as their mentors or colleagues from the same laboratory, with whom they have intimate working relationships (Wang, Yu, et al., 2017). As the above results show, the collaborative relationships between them are often more stable, frequent, and enduring. This makes the early collaborators of scholars have a significant influence on their future development, such as having high similarities in their future research directions and social network relationships. By utilizing this regular phenomenon, recommendations can be made for scholars about papers or other scholars related to their early collaborators. This can help enhance the precision of paper reading recommendations and scholar collaboration recommendations.

5.3 | Subsequent effects of initial interpersonal differences among scholars who first collaborated

Correlation analysis of cooperative characteristics indicates that the more the overlap of specific research points and the greater the difference in research productivity between the two scholars at the time of first collaboration, the higher the likelihood and stability of their continued collaboration, but the more the overlap of academic networks and the more similar the research themes, the less favorable the subsequent collaboration. Besides, the results display that the greater the difference in the research age, academic impact, and interpersonal relationship of scholars at the time of first collaboration, the less likely their academic interpersonal networks will overlap in future research, while the distinct difference in research productivity and similarity in research topics are not conductive to promote more overlap of interpersonal network.

The results of the correlation analysis also reveal the factors influencing the subsequent specific research points and overlap of research topics between the two collaborating scholars. The greater their difference in academic networks and impact in the first collaboration, the more likely it is that the specific points of subsequent research will overlap, while the greater the difference in academic age, research productivity, and research themes, the less likely it is that they will overlap. In addition, the similarity of their subsequent research topics shows a weak positive correlation with the similarity of scholars' academic networks at the time of first collaboration, but a weak negative correlation with the degree of overlap of specific research points and difference in research productivity.

The above results reflect the factors influencing the continuity of the scientific collaboration, which differ significantly from the conditions for initial collaboration. If two parties who have not collaborated before have more previous common collaborators, similar research fields, certain differences in specific research points, and similar publication output, the likelihood of them achieving initial collaboration is higher. However, once they have achieved initial collaboration, the likelihood of the collaboration continuing is lower. On the other hand, for scholars who find it difficult to establish initial collaboration, that is, those with significant differences in their academic social networks, research fields, and publication output, but have specific research points that overlap, once they have established initial collaboration, the likelihood of the collaboration continuing is often higher. This suggests that even though there is a lower likelihood of initial collaboration between two scholars with significantly different academic backgrounds, once they establish initial collaboration due to a specific intersection of research, this type of collaboration often has a higher likelihood of continuing. This to some extent ensures the openness and inclusiveness of academic teams for both parties as collaborators from different research backgrounds bring a higher level of novelty for another party (Liu, Bu, et al., 2022; Liu, Jaiswal, et al., 2022; Zeng et al., 2021).

In addition, there is a certain correlation between several characteristics of scholars when they first collaborate and their subsequent development. If there is a significant difference in academic age or academic impact between scholars who are collaborating for the first time, their academic cliques generally do not intersect in the future, which corresponds to the phenomenon of age homogeneity mentioned in previous research on scientific collaboration (Wang, Yu, et al., 2017). There is also a certain connection between the specific research points of scholars in the future and several attributes of their first-time collaborators. If there is a significant difference in academic impact and social networks of scholars who collaborate for the first time, and they have similar academic age, publication output, and research fields, they may generate more common specific research points in the future, which aligns with the fact that influential scholars are more inclined to involve collaborators in new topics (Zeng et al., 2022). The evolution of scholars' research fields exhibits strong stability, and some characteristics of their first-time collaborators do not have a significant impact on this. This regular phenomenon has implications for the research about the evolution of scholars' interests and academic social networks, providing references for speculating on scholars' future research points and potential collaborators.

6 | CONCLUSIONS AND FUTURE WORK

The first scientific research cooperation is conducive to increasing the innovation ability of the team and is a prerequisite for the subsequent cooperation. Thus, it is essential to clarify the drivers, characteristic and influence of the first collaboration for precise collaborative recommendation, and furtherly promotes a creative and inclusive atmosphere of scientific research cooperation. With the application of data mining approaches including PSM, differential expression analysis, and correlation analysis, we utilize larger-scale data from CSSCI to conduct the proposed clarification, which not only provides a clue to why the first cooperation happens, but also provides suggestions for recommending new collaborators to scholars.

Based on academic cooperation and non-cooperation in the field of LIS collected from CSSCI, our understandings of the entire collaborative patterns obtained from this study can be summarized as follows: (1) There is a higher likelihood of the first collaboration between two scholars with more common collaborators, similar number of published articles, greater difference in academic influence, consistency of the general research direction and less common specific research points. Additionally, neither the career age of the research nor the difference in whether the researchers belonged to the same clique significantly affected how they came to their first cooperation. (2) If the first collaboration of scholars is through the first paper authored by one or more of them, it is likely that their subsequent collaboration will become more frequent, long-lasting, and stable. Moreover, there is a higher likelihood that their academic interpersonal relationships, specific research points, and research themes will develop in a similar orientation in the future. (3) When scholars collaborate for the first time, the likelihood of sustained collaboration in the future increases if they have greater differences in scientific productivity and research theme, fewer common collaborators, and more overlapping specific research points. Additionally, there is a modest correlation between the first-time collaborators' interpersonal differences and future similarities in academic accomplishments, academic networks, and research direction development.

Based on the above understandings, if scholars have more common collaborators, similar research areas, and publication volume, while having significant differences in academic impact and specific research points, it is more favorable for both parties to initiate first research collaborations. Specifically, the first collaborative relationship formed when scholars publish their first paper is often more stable, frequent, and enduring, leading to a high degree of similarity in future research directions and social network relationships for the collaborating parties. Furthermore, although the likelihood of first collaboration is low for scholars with significantly different academic backgrounds, once they initiate first research collaboration, the probability of subsequent collaborations is often higher due to the novelty each party brings to the other. Additionally, there is a certain correlation between the future development of scholars' academic social networks and specific research points and the degree of difference in their academic backgrounds at the time of their first collaboration.

This research analyzes the driving factors of first scientific collaboration, which can be used to optimize collaboration recommendation services of those existing academic service platforms by facilitating new collaborations with unfamiliar scholars who have similar research productivity, overall research direction, and non-overlapping specific research areas. In addition, these findings reveal that first scientific collaborations are correlated with subsequent collaborations and future academic development. This correlation suggests that first scientific collaborations may influence scholars' future academic trajectories, including the evolution of academic social networks, shifts in research directions, and changes in academic capabilities. This has implications for predicting scholars' future academic behavior or performance. To summary, this study only investigated the collaboration patterns among scholars in the field of library and information science in China. Considering other factors such as discipline and nationality, the generalizability of these conclusions remains to be furtherly verified by using a more extensive dataset in our future work.

AUTHOR CONTRIBUTIONS

Xianzhe Peng: Collection and analysis of data; paper writing; revision. **Jin Shi**: Providing ideas; paper reviewing; revision.

ACKNOWLEDGMENT

This research is supported by the National Social Science Foundation, People's Republic of China (No. 21BTQ012).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

REFERENCES

Abramo, G., D'Angelo, A. C., & Murgia, G. (2017). The relationship among research productivity, research collaboration, and their determinants. *Journal of Informetrics*, *11*(4), 1016–1030. https://doi.org/10.1016/j.joi.2017.09.007

Abramo, G., D'Angelo, C. A., & Di Costa, F. (2011). Research productivity: Are higher academic ranks more productive than lower ones? *Scientometrics*, *88*(3), 915–928. https://doi.org/10.1007/s11192-011-0426-6

Adams, J. (2013). The fourth age of research. *Nature*, *497*(7451), 557–560. https://doi.org/10.1038/497557a

Adams, J. D., Black, G. C., Clemmons, J. R., & Stephan, P. E. (2005). Scientific teams and institutional collaborations: Evidence from U.S. universities, 1981–1999. *Research Policy*, *34*(3), 259–285. https://doi.org/10.1016/j.respol.2005.01.014

Austin, P. C. (2011). An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behavioral Research*, *46*(3), 399–424. https://doi.org/10.1080/00273171.2011.568786

Barabási, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, *311*(3), 590–614. https://doi.org/10.1016/S0378-4371(02)00736-7

Beaver, D. D. (2001). Reflections on scientific collaboration (and its study): Past, present, and future. *Scientometrics*, *52*(3), 365–377. https://doi.org/10.1023/A:1014254214337

Beaver, D. D., & Rosen, R. (1978). Studies in scientific collaboration—part I. The professional origins of scientific co-authorship. *Scientometrics*, *1*(1), 65–84. https://doi.org/10.1007/BF02016840

Birnholtz, J. P. (2007). When do researchers collaborate? Toward a model of collaboration propensity. *Journal of the American Society for Information Science and Technology*, *58*(14), 2226–2239. https://doi.org/10.1002/asi.20684

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993–1022. https://doi.org/10.5555/944919.944937

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Boardman, P. C., & Corley, E. A. (2008). University research centers and the composition of research collaborations. *Research Policy*, *37*(5), 900–913. https://doi.org/10.1016/j.respol.2008.01.012

Boardman, P. C., & Ponomariov, B. L. (2009). University researchers working with private companies. *Technovation*, *29*(2), 142–153. https://doi.org/10.1016/j.technovation.2008.03.008

Bu, Y., Ding, Y., Liang, X., & Murray, D. S. (2018). Understanding persistent scientific collaboration. *Journal of the Association for Information Science and Technology*, *69*(3), 438–448. https://doi.org/10.1002/asi.23966

Burton, R. E., & Kebler, R. W. (1960). The "half-life" of some scientific and technical literatures. *American Documentation*, *11*(1), 18–22. https://doi.org/10.1002/asi.5090110105

Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, *22*(1), 31–72. https://doi.org/10.1111/j.1467-6419.2007.00527.x

Chen, J., Wang, X., Zhao, S., & Zhang, Y. (2021). Content-enhanced network embedding for academic collaborator recommendation. *Complexity*, *2021*, 7035467. https://doi.org/10.1155/2021/7035467

Coccia, M., & Wang, L. (2016). Evolution and convergence of the patterns of international scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(8), 2057–2061. https://doi.org/10.1073/pnas.1510820113

Cohen, S., & Ebel, L. (2013). Recommending collaborators using keywords. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 959–962). ACM. https://doi.org/10.1145/2487788.2488091

Corrêa, E. A., Jr., Silva, F. N., Costa, L. F., & Amancio, D. R. (2017). Patterns of authors contribution in scientific manuscripts. *Journal of Informetrics*, *11*(2), 498–510. https://doi.org/10.1016/j.joi.2017.03.003

Cugmas, M., Ferligoj, A., & Kronegger, L. (2016). The stability of co-authorship structures. *Scientometrics*, *106*(1), 163–186. https://doi.org/10.1007/s11192-015-1790-4

Cugmas, M., Mali, F., & Žiberna, A. (2020). Scientific collaboration of researchers and organizations: A two-level blockmodeling approach. *Scientometrics*, *125*(3), 2471–2489. https://doi.org/10.1007/s11192-020-03708-x

de Oliveira, S. C., Cobre, J., & Pereira, D. F. (2021). A measure of reliability for scientific co-authorship networks using fuzzy logic. *Scientometrics*, *126*(6), 4551–4563. https://doi.org/10.1007/s11192-021-03915-0

Dehdarirad, T., & Nasini, S. (2017). Research impact in co-authorship networks: A two-mode analysis. *Journal of Informetrics*, *11*(2), 371–388. https://doi.org/10.1016/j.joi.2017.02.002

Fell, C. B., & König, C. J. (2016). Is there a gender difference in scientific collaboration? A scientometric examination of co-authorships among industrial–organizational psychologists. *Scientometrics*, *108*(1), 113–141. https://doi.org/10.1007/s11192-016-1967-5

Floyd, S. W., Schroeder, D. M., & Finn, D. M. (1994). Only if I'm first author: Conflict over credit in management scholarship. *Academy of Management Journal*, *37*(3), 734–747. https://doi.org/10.2307/256709

Gao, J., Yin, Y., Myers, K. R., Lakhani, K. R., & Wang, D. (2021). Potentially long-lasting effects of the pandemic on scientists. *Nature Communications*, *12*(1), 6188. https://doi.org/10.1038/s41467-021-26428-z

Gaughan, M., & Bozeman, B. (2016). Using the prisms of gender and rank to interpret research collaboration power dynamics. *Social Studies of Science*, *46*(4), 536–558. https://doi.org/10.1177/0306312716652249

Glänzel, W., Schubert, A., & Czerwon, H. J. (1999). A bibliometric analysis of international scientific cooperation of the European Union (1985–1995). *Scientometrics*, *45*(2), 185–202. https://doi.org/10.1007/BF02458432

Gómez, I., Teresa Fernández, M., & Sebastián, J. (1999). Analysis of the structure of international scientific cooperation networks through bibliometric indicators. *Scientometrics*, *44*(3), 441–457. https://doi.org/10.1007/BF02458489

Gong, K., & Cheng, Y. (2022). Patterns and impact of collaboration in China's social sciences: Cross-database comparisons between CSSCI and SSCI. *Scientometrics*, *127*(10), 5947–5964. https://doi.org/10.1007/s11192-022-04483-7

Guan, J., Yan, Y., & Zhang, J. J. (2017). The impact of collaboration and knowledge networks on citations. *Journal of Informetrics*, *11*(2), 407–422. https://doi.org/10.1016/j.joi.2017.02.007

He, C., Wu, J., & Zhang, Q. (2021). Characterizing research leadership on geographically weighted collaboration network. *Scientometrics*, *126*(5), 4005–4037. https://doi.org/10.1007/s11192-021-03943-w

He, Z.-L., Geng, X.-S., & Campbell-Hunt, C. (2009). Research collaboration and research output: A longitudinal study of 65 biomedical scientists in a New Zealand university. *Research Policy*, *38*(2), 306–317. https://doi.org/10.1016/j.respol.2008.11.011

Hilário, C. M., & Grácio, M. C. (2017). Scientific collaboration in Brazilian researches: A comparative study in the information science, mathematics and dentistry fields. *Scientometrics*, *113*(2), 929–950. https://doi.org/10.1007/s11192-017-2498-4

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(46), 16569–16572. https://doi.org/10.1073/pnas.0507655102

Huang, L., Chen, X., Zhang, Y., Zhu, Y., Li, S., & Ni, X. (2021). Dynamic network analytics for recommending scientific collaborators. *Scientometrics*, *126*(11), 8789–8814. https://doi.org/10.1007/s11192-021-04164-x

Hückstädt, M. (2023). Ten reasons why research collaborations succeed—A random forest approach. *Scientometrics*, *128*(3), 1923–1950. https://doi.org/10.1007/s11192-022-04629-7

Jie, Z., Xinning, S., & Sanhong, D. (2008). The academic impact of Chinese humanities and social science research. *ASLIB Proceedings*, *60*(1), 55–74. https://doi.org/10.1108/00012530810847371

Katz, J. S. (1994). Geographical proximity and scientific collaboration. *Scientometrics*, *31*(1), 31–43. https://doi.org/10.1007/BF02018100

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*(1), 1–18. https://doi.org/10.1016/S0048-7333(96)00917-1

Kim, S., & Park, K. C. (2021). Government funded R&D collaboration and it's impact on SME's business performance. *Journal of Informetrics*, *15*(3), 101197. https://doi.org/10.1016/j.joi.2021.101197

Kong, X., Jiang, H., Wang, W., Bekele, T. M., Xu, Z., & Wang, M. (2017). Exploring dynamic research interest and academic influence for scientific collaborator recommendation. *Scientometrics*, *113*(1), 369–385. https://doi.org/10.1007/s11192-017-2485-9

Kwiek, M., & Roszka, W. (2021). Gender-based homophily in research: A large-scale study of man-woman collaboration. *Journal of Informetrics*, *15*(3), 101171. https://doi.org/10.1016/j.joi.2021.101171

Laband, D. N., & Tollison, R. D. (2000). Intellectual collaboration. *Journal of Political Economy*, *108*(3), 632–662. https://doi.org/10.1086/262132

Laudel, G. (2002). What do we measure by co-authorships? *Research Evaluation*, *11*(1), 3–15. https://doi.org/10.3152/147154402781776961

Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, *15*(2), R29. https://doi.org/10.1186/gb-2014-15-2-r29

Leclerc, M., & Gagné, J. (1994). International scientific cooperation: The continentalization of science. *Scientometrics*, *31*(3), 261–292. https://doi.org/10.1007/BF02016876

Lee, S., & Bozeman, B. (2005). The impact of research collaboration on scientific productivity. *Social Studies of Science*, *35*(5), 673–702. https://doi.org/10.1177/0306312705052359

Li, F., Miao, Y., & Yang, C. (2015). How do alumni faculty behave in research collaboration? An analysis of Chang Jiang scholars in China. *Research Policy*, *44*(2), 438–450. https://doi.org/10.1016/j.respol.2014.09.002

Li, J., Xia, F., Wang, W., Chen, Z., Asabere, N. Y., & Jiang, H. (2014). ACRec: A co-authorship based random walk model for academic collaboration recommendation. In *Proceedings of the 23rd international conference on World Wide Web* (pp. 1209–1214). ACM. https://doi.org/10.1145/2567948.2579034

Li, W., Aste, T., Caccioli, F., & Livan, G. (2019). Early coauthorship with top scientists predicts success in academic careers. *Nature Communications*, *10*(1), 5170. https://doi.org/10.1038/s41467-019-13130-4

Li, Y., Wang, K., Xiao, Y., & Froyd, J. E. (2020). Research and trends in STEM education: A systematic review of journal publications. *International Journal of STEM Education*, *7*(1), 11. https://doi.org/10.1186/s40594-020-00207-6

Liang, L., Kretschmer, H., Guo, Y., & Beaver, D. (2001). Age structures of scientific collaboration in Chinese computer science. *Scientometrics*, *52*(3), 471–486. https://doi.org/10.1023/A:1014252017971

Liu, M., Bu, Y., Chen, C., Xu, J., Li, D., Leng, Y., Freeman, R. B., Meyer, E. T., Yoon, W., Sung, M., Jeong, M., Lee, J., Kang, J., Min, C., Song, M., Zhai, Y., & Ding, Y. (2022). Pandemics are catalysts of scientific novelty: Evidence from COVID-19. *Journal of the Association for Information Science and Technology*, *73*(8), 1065–1078. https://doi.org/10.1002/asi.24612

Liu, M., Jaiswal, A., Bu, Y., Min, C., Yang, S., Liu, Z., Acuña, D., & Ding, Y. (2022). Team formation and team impact: The balance between team freshness and repeat collaboration. *Journal of Informetrics*, *16*(4), 101337. https://doi.org/10.1016/j.joi.2022.101337

Liu, X., Wu, K., Liu, B., & Qian, R. (2023). HNERec: Scientific collaborator recommendation model based on heterogeneous network embedding. *Information Processing & Management*, *60*(2), 103253. https://doi.org/10.1016/j.ipm.2022.103253

Lopaciuk-Gonczaryk, B. (2016). Collaboration strategies for publishing articles in international journals—A study of Polish scientists in economics. *Social Networks*, *44*, 50–63. https://doi.org/10.1016/j.socnet.2015.07.001

Lopes, G. R., Moro, M. M., Wives, L. K., & De Oliveira, J. P. M. (2010). Collaboration recommendation on academic social networks. In J. Trujillo, G. Dobbie, H. Kangassalo, S. Hartmann, M. Kirchberg, M. Rossi, I. Reinhartz-Berger, E. Zimányi, & F. Frasincar (Eds.), *Advances in conceptual modeling: Applications and challenges* (pp. 190–199). Springer. https://doi.org/10.1007/978-3-642-16385-2_24

Lu, C., Zhang, Y., Ahn, Y.-Y., Ding, Y., Zhang, C., & Ma, D. (2020). Co-contributorship network and division of labor in individual scientific collaborations. *Journal of the Association for Information Science and Technology*, *71*(10), 1162–1178. https://doi.org/10.1002/asi.24321

Luellen, J. K., Shadish, W. R., & Clark, M. H. (2005). Propensity scores: An introduction and experimental test. *Evaluation Review*, *29*(6), 530–558. https://doi.org/10.1177/0193841X05275596

Lundberg, J., Tomson, G., Lundkvist, I., Skr, J., & Brommels, M. (2006). Collaborationuncovered: Exploring the adequacy of measuring university-industrycollaboration through co-authorship and funding. *Scientometrics*, *69*(3), 575–589. https://doi.org/10.1007/s11192-006-0170-5

Matveeva, N., & Ferligoj, A. (2020). Scientific collaboration in Russian universities before and after the excellence initiative Project 5-100. *Scientometrics*, *124*(3), 2383–2407. https://doi.org/10.1007/s11192-020-03602-6

Milojević, S. (2010). Modes of collaboration in modern science: Beyond power laws and preferential attachment. *Journal of the American Society for Information Science and Technology*, *61*(7), 1410–1423. https://doi.org/10.1002/asi.21331

Misra, S., Hall, K., Feng, A., Stipelman, B., & Stokols, D. (2011). Collaborative processes in transdisciplinary research. In M. Kirst, N. Schaefer-McDaniel, S. Hwang, & P. O'Campo (Eds.), *Converging disciplines: A transdisciplinary research approach to urban health problems* (pp. 97–110). Springer. https://doi.org/10.1007/978-1-4419-6330-7_8

Moody, J. (2004). The structure of a social science collaboration network: Disciplinary cohesion from 1963 to 1999. *American Sociological Review*, *69*(2), 213–238. https://doi.org/10.1177/000312240406900204

Morgan, C. J. (2018). Reducing bias using propensity score matching. *Journal of Nuclear Cardiology*, *25*(2), 404–406. https://doi.org/10.1007/s12350-017-1012-y

Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(2), 404–409. https://doi.org/10.1073/pnas.98.2.404

Newman, M. E. J. (2004). Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, *101*, 5200–5205. https://doi.org/10.1073/pnas.0307545100

Perc, M. (2010). Growth and structure of Slovenia's scientific collaboration network. *Journal of Informetrics*, *4*(4), 475–482. https://doi.org/10.1016/j.joi.2010.04.003

Petersen, A. M. (2015). Quantifying the impact of weak, strong, and super ties in scientific careers. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(34), E4671–E4680. https://doi.org/10.1073/pnas.1501444112

Petersen, A. M., Fortunato, S., Pan, R. K., Kaski, K., Penner, O., Rungi, A., Riccaboni, M., Stanley, H. E., & Pammolli, F. (2014). Reputation and impact in academic careers. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(43), 15316–15321. https://doi.org/10.1073/pnas.1323111111

Phipson, B., Lee, S. C., Majewski, I. J., Alexander, W. S., & Smyth, G. K. (2016). Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. *The Annals of Applied Statistics*, *10*(2), 946–963. https://doi.org/10.1214/16-AOAS920

Ponomariov, B. L., & Boardman, P. C. (2008). The effect of informal industry contacts on the time university scientists allocate to collaborative research with industry. *Journal of Technology Transfer*, *33*(3), 301–313. https://doi.org/10.1007/s10961-007-9029-z

Ponomariov, B. L., & Boardman, P. C. (2010). Influencing scientists' collaboration and productivity patterns through new institutions: University research centers and scientific and technical human capital. *Research Policy*, *39*(5), 613–624. https://doi.org/10.1016/j.respol.2010.02.013

Ponomariov, B. L., & Boardman, P. C. (2016). What is co-authorship? *Scientometrics*, *109*(3), 1939–1963. https://doi.org/10.1007/s11192-016-2127-7

Porac, J. F., Wade, J. B., Fischer, H. M., Brown, J., Kanfer, A., & Bowker, G. (2004). Human capital heterogeneity, collaborative relationships, and publication patterns in a multidisciplinary scientific alliance: A comparative case study of two scientific teams. *Research Policy*, *33*(4), 661–678. https://doi.org/10.1016/j.respol.2004.01.007

Pradhan, T., & Pal, S. (2020). A multi-level fusion based decision support system for academic collaborator recommendation. *Knowledge-Based Systems*, *197*, 105784. https://doi.org/10.1016/j.knosys.2020.105784

Prathap, G. (2010). The 100 most prolific economists using the p-index. *Scientometrics*, *84*(1), 167–172. https://doi.org/10.1007/s11192-009-0068-0

Prathap, G. (2011). Qualifying scholarly impact using an iCX (impact-citations-exergy) analysis. *DESIDOC Journal of Library & Information Technology*, *31*(5), 382–386. https://doi.org/10.14429/djlit.31.5.1196

Ribeiro, L. C., Rapini, M. S., Silva, L. A., & Albuquerque, E. M. (2018). Growth patterns of the network of international collaboration in science. *Scientometrics*, *114*(1), 159–179. https://doi.org/10.1007/s11192-017-2573-x

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, *43*(7), e47. https://doi.org/10.1093/nar/gkv007

Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, *79*(387), 516–524. https://doi.org/10.1080/01621459.1984.10478078

Russell, J. M. (1995). The increasing role of international cooperation in science and technology research in Mexico. *Scientometrics*, *34*(1), 45–61. https://doi.org/10.1007/BF02019172

Saad, G. (2010). Applying the h-index in exploring bibliometric properties of elite marketing scholars. *Scientometrics*, *83*(2), 423–433. https://doi.org/10.1007/s11192-009-0069-z

Sattari, M., & Zamanifar, K. (2018). A cascade information diffusion based label propagation algorithm for community detection in dynamic social networks. *Journal of Computational Science*, *25*, 122–133. https://doi.org/10.1016/j.jocs.2018.01.004

Scarazzati, S., & Wang, L. (2019). The effect of collaborations on scientific research output: The case of nanoscience in Chinese regions. *Scientometrics*, *121*(2), 839–868. https://doi.org/10.1007/s11192-019-03220-x

Shen, H., Xie, J., Li, J., & Cheng, Y. (2021). The correlation between scientific collaboration and citation count at the paper level: A meta-analysis. *Scientometrics*, *126*(4), 3443–3470. https://doi.org/10.1007/s11192-021-03888-0

Sinatra, R., Wang, D., DeVillé, P., Ci, S., & Barabási, A. L. (2016). Quantifying the evolution of individual scientific impact. *Science*, *354*, 6312. https://doi.org/10.1126/science.aaf5239

Skilton, P. F., & Dooley, K. J. (2010). The effects of repeat collaboration on creative abrasion. *Academy of Management Review*, *35*(1), 118–134. https://doi.org/10.5465/amr.35.1.zok118

Smyth, G. K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology*, *3*(1), 1–25. https://doi.org/10.2202/1544-6115.1027

Smyth, G. K. (2005). limma: Linear models for microarray data. In R. Gentleman, V. J. Carey, W. Huber, R. A. Irizarry, & S. Dudoit (Eds.), *Bioinformatics and computational biology solutions using R and bioconductor* (pp. 397–420). Springer. https://doi.org/10.1007/0-387-29362-0_23

Song, Y., Ma, F., & Yang, S. (2015). Comparative study on the obsolescence of humanities and social sciences in China: Under the new situation of web. *Scientometrics*, *102*(1), 365–388. https://doi.org/10.1007/s11192-014-1410-8

Sonnenwald, D. H. (2007). Scientific collaboration. *Annual Review of Information Science and Technology*, *41*(1), 643–681. https://doi.org/10.1002/aris.2007.1440410121

Sørensen, C. (2003). Scandinavian versus UK research: The importance of institutional context. *Scandinavian Journal of Information Systems*, *15*(1), 95–99.

Steyvers, M., Smyth, P., Rosen-Zvi, M., & Griffiths, T. (2004). Probabilistic author-topic models for information discovery. In *Proceedings of the tenth ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 306–315). ACM. https://doi.org/10.1145/1014052.1014087

Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., & Su, Z. (2008). ArnetMiner: Extraction and mining of academic social networks. In *Proceedings of the 14th ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 990–998). ACM. https://doi.org/10.1145/1401890.1402008

Tsai, C.-H., & Lin, Y.-R. (2016). Tracing and predicting collaboration for junior scholars. In *Proceedings of the 25th international conference companion on World Wide Web* (pp. 375–380). ACM. https://doi.org/10.1145/2872518.2890516

Wagner, C. S., & Leydesdorff, L. (2005). Network structure, self-organization, and the growth of international collaboration in science. *Research Policy*, *34*(10), 1608–1618. https://doi.org/10.1016/j.respol.2005.08.002

Wang, H., Deng, S., & Su, X. (2016). A study on construction and analysis of discipline knowledge structure of Chinese LIS based on CSSCI. *Scientometrics*, *109*(3), 1725–1759. https://doi.org/10.1007/s11192-016-2146-4

Wang, M., Zou, Y., Cao, Y., & Xie, B. (2019). Searching software knowledge graph with question. In T. Bhowmik (Ed.), *Reuse in the big data era* (pp. 115–131). Springer. https://doi.org/10.1007/978-3-030-22888-0_9

Wang, W., Cui, Z., Gao, T., Yu, S., Kong, X., & Xia, F. (2017). Is scientific collaboration sustainability predictable? In *Proceedings of the 26th international conference on World Wide Web companion* (pp. 853–854). ACM. https://doi.org/10.1145/3041021.3054230

Wang, W., Liu, J., Tang, T., Tuarob, S., Xia, F., Gong, Z., & King, I. (2020). Attributed collaboration network embedding for academic relationship mining. *ACM Transactions on the Web*, *15*(1), 1–20. https://doi.org/10.1145/3409736

Wang, W., Liu, J., Yang, Z., Kong, X., & Xia, F. (2019). Sustainable collaborator recommendation based on conference closure. *IEEE Transactions on Computational Social Systems*, *6*(2), 311–322. https://doi.org/10.1109/TCSS.2019.2898198

Wang, W., Ren, J., Alrashoud, M., Xia, F., Mao, M., & Tolba, A. (2020). Early-stage reciprocity in sustainable scientific collaboration. *Journal of Informetrics*, *14*(3), 101041. https://doi.org/10.1016/j.joi.2020.101041
Wang, W., Yu, S., Bekele, T. M., Kong, X., & Xia, F. (2017). Scientific collaboration patterns vary with scholars' academic ages. *Scientometrics*, *112*(1), 329–343. https://doi.org/10.1007/s11192-017-2388-9
Wu, Y., Venkatramanan, S., & Chiu, D. M. (2015). Research collaboration and topic trends in computer science: An analysis based on UCP authors. In *Proceedings of the 24th international conference on World Wide Web* (pp. 1045–1050). ACM. https://doi.org/10.1145/2740908.2742015
Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in production of knowledge. *Science*, *316*(5827), 1036–1039. https://doi.org/10.1126/science.1136099
Xia, F., Wang, W., Bekele, T. M., & Liu, H. (2017). Big scholarly data: A survey. *IEEE Transactions on Big Data*, *3*(1), 18–35. https://doi.org/10.1109/TBDATA.2016.2641460
Xie, Z., Li, M., Li, J., Duan, X., & Ouyang, Z. (2018). Feature analysis of multidisciplinary scientific collaboration patterns based on PNAS. *EPJ Data Science*, *7*(1), 5. https://doi.org/10.1140/epjds/s13688-018-0134-z
Xu, H., Bu, Y., Liu, M., Zhang, C., Sun, M., Zhang, Y., Meyer, E., Salas, E., & Ding, Y. (2022). Team power dynamics and team impact: New perspectives on scientific collaboration using career age as a proxy for team power. *Journal of the Association for Information Science and Technology*, *73*(10), 1489–1505. https://doi.org/10.1002/asi.24653
Yan, E., Ding, Y., & Zhu, Q. (2010). Mapping library and information science in China: A coauthorship network analysis. *Scientometrics*, *83*(1), 115–131. https://doi.org/10.1007/s11192-009-0027-9
Yong, K., Sauer, S. J., & Mannix, E. A. (2014). Conflict and creativity in interdisciplinary teams. *Small Group Research*, *45*(3), 266–289. https://doi.org/10.1177/1046496414530789
Yoshikane, F., & Kageura, K. (2004). Comparative analysis of coauthorship networks of different domains: The growth and change of networks. *Scientometrics*, *60*(3), 435–446. https://doi.org/10.1023/B:SCIE.0000034385.05897.46
Zeng, A., Fan, Y., Di, Z., Wang, Y., & Havlin, S. (2021). Fresh teams are associated with original and multidisciplinary research. *Nature Human Behaviour*, *5*(10), 1314–1322. https://doi.org/10.1038/s41562-021-01084-x
Zeng, A., Fan, Y., Di, Z., Wang, Y., & Havlin, S. (2022). Impactful scientists have higher tendency to involve collaborators in new topics. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(33), e2207436119. https://doi.org/10.1073/pnas.2207436119
Zhai, L., & Yan, X. (2022). A directed collaboration network for exploring the order of scientific collaboration. *Journal of Informetrics*, *16*(4), 101345. https://doi.org/10.1016/j.joi.2022.101345
Zhai, L., Yan, X., Shibchurn, J., & Song, X. (2014). Evolutionary analysis of international collaboration network of Chinese scholars in management research. *Scientometrics*, *98*(2), 1435–1454. https://doi.org/10.1007/s11192-013-1040-6
Zhang, C., Bu, Y., Ding, Y., & Xu, J. (2018). Understanding scientific collaboration: Homophily, transitivity, and preferential attachment. *Journal of the Association for Information Science and Technology*, *69*(1), 72–86. https://doi.org/10.1002/asi.23916
Zhao, Z., Pan, X., & Hua, W. (2021). Comparative analysis of the research productivity, publication quality, and collaboration patterns of top ranked library and information science schools in China and the United States. *Scientometrics*, *126*(2), 931–950. https://doi.org/10.1007/s11192-020-03796-9
Zong, Q., Xie, Y., & Liang, J. (2020). Does open peer review improve citation count? Evidence from a propensity score matching analysis of PeerJ. *Scientometrics*, *125*(1), 607–623. https://doi.org/10.1007/s11192-020-03545-y
Zuo, Z., & Zhao, K. (2018). The more multidisciplinary the better?—The prevalence and interdisciplinarity of research collaborations in multidisciplinary institutions. *Journal of Informetrics*, *12*(3), 736–756. https://doi.org/10.1016/j.joi.2018.06.006

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Peng, X., & Shi, J. (2024). The drivers, features, and influence of first scientific collaboration among core scholars from Chinese library and information field. *Journal of the Association for Information Science and Technology*, 1–20. https://doi.org/10.1002/asi.24888