Data analysis of overseas chinese literature based on co-occurrence relationship

Análise de dados da literatura chinesa no exterior com base na relação de coocorrência

Danfeng Huang¹ (iD) , Chunpei Lin¹ (iD) , Fang Zhong² (iD) , Hailin Li³ (iD) , Fan Yang³ (iD)

¹ School of Business Administration, Huaqiao University, Quanzhou 362021, China
² College of Management, Jinan University, Guangzhou 510000, China Correspondence to: F. ZHONG. E-mail: <yuzhaoweigo112@163.com>.
³ Business Management Research Center, Huaqiao University, Quanzhou 362021, China.

Abstract

In the past ten years, new progress has been made in the study of Chinese overseas, which provides important scientific basis for the work of Chinese overseas and the formulation of relevant policies. Through data mining and analysis of overseas Chinese literature from 2008 to 2021, this work studied the co-occurrence relationships among keywords, authors, research institutions, publication sources of target articles, references and citations. The results suggest a few things. In terms of keyword analysis, there is a high probability that keywords coexist among overseas Chinese, Chinese society and Chinese history. In addition, the degree of cooperation between authors and research institutions in overseas research in China is not high. In terms of target journals, references and citations, resources such as Journal of Overseas Chinese Historical Studies, Overseas Chinese Publishing Company and People's Publishing House are the main publications published by most overseas Chinese research institutions. In addition, academic articles cite fewer of the journal's journals. Based on the above problems, this study puts forward relevant recommendations for decision-making, in order to further promote the exchange, sharing and development of China's overseas research results.

Keywords: Overseas Chines. Co-occurrence relation. Literature analysis. Source of publication. Network Structure.

Resumo

Nos últimos dez anos, novos avanços foram alcançados no estudo da diáspora chinesa, proporcionando uma base científica crucial para o trabalho relacionado à comunidade chinesa no exterior e a formulação de políticas pertinentes. Por meio da mineração de dados e análise da literatura sobre os chineses no exterior de 2008 a 2021, este trabalho investigou as relações de coocorrência entre palavras-chave, autores, instituições de pesquisa, fontes de publicação dos artigos-alvo, referências e citações. Os resultados indicam algumas observações importantes. No que diz respeito à análise de palavras-chave, há uma alta probabilidade de coexistência de termos

relacionados à "chineses no exterior", à "sociedade chinesa" e à "história chinesa". Além disso, o grau de cooperação entre autores e instituições de pesquisa em estudos sobre os chineses no exterior não é elevado. Em relação aos periódicos alvo, referências e citações, periódicos como "Journal of Overseas Chinese Historical Studies", "Overseas Chinese Publishing Company" e "People's Publishing House" destacam-se como as principais publicações da maioria das instituições de pesquisa sobre os chineses no exterior. Adicionalmente, observa-se uma quantidade limitada de citações a esses periódicos nas publicações acadêmicas. Com base nas questões identificadas, este estudo apresenta recomendações pertinentes para a tomada de decisões, visando promover ainda mais a troca, compartilhamento e desenvolvimento dos resultados de pesquisa sobre a diáspora chinesa na China.

Palavras-chave: *Chinês no exterior. Relação de coocorrência. Análise de Lieratura. Fonte de Publicação. Estrutura de rede.*

Introduction

In recent years, with the accelerating process of globalization, overseas Chinese, as an important group, have played the role of a bridge and a link, connecting the exchange and cooperation between their countries and China (Cheng; Wang, 2023; Larsen; Oehler, 2022; Yu; Xiao; Li, 2021). However, despite a large amount of literature on overseas Chinese, there are relatively few systematic and in-depth data analysis studies on these documents. We can reveal hidden patterns and trends through data analysis of overseas Chinese literature and promote cross-institutional, cross-country, and cross-cultural exchanges and cooperation. Meanwhile, Literature and data related to overseas Chinese are of great value to the contribution of overseas Chinese to the social and economic development of China and the world. It is necessary to sort out and analyze them, to enhance and deepen the attention and understanding of the existence and development of overseas Chinese from all walks of life at home and abroad (Pan, 2015; Zhou, 2014; Larsen; Oehler, 2022; Jin, 2022). Therefore, it is significant to carry out a data analysis study on overseas Chinese literature. Currently, the collection, collation and analysis of overseas Chinese literature and data are mainly through traditional methods such as questionnaire survey, manual collection, probability statistics and bibliometrics (Lu, 2014; Wang, 2016; Wu; Lin, 2016). To some extent, the conclusion has realized the analysis and research of the problem, and obtained some valuable research results, which support the decision-making work of the Overseas Chinese Federation.

In the related academic research on the subject of overseas Chinese, Xu Yun of the Jinan University Library has carried out statistical analysis on the literature related to overseas Chinese in different periods for many times, and made statistical analysis on the literature content from the perspectives of sample size, author, institution, geographical distribution, chronological distribution and citation analysis, to evaluate the research status of overseas Chinese in China at that time (Xu, 2007; Xu, 2010). However, due to the early age of the research data and limited conditions, the data range and objects used are not rich enough, so the research on overseas Chinese needs further in-depth analysis. Deng Sanhong and Xu Xin studied the academic status of Overseas Chinese in China from 2000 to 2009 through bibliometric analysis and subject analysis by extracting documents related to overseas Chinese studies published in CSSCI (Chinese Social Sciences Citation Index) journals (Deng; Xu, 2011). Wu Yuan and Lin Yong made a statistical analysis from author, institution, journal distribution and research content based on the literature on overseas Chinese economic research in the past 30 years collected by CNKI (China National Knowledge Network) (Wu; Lin, 2016). In addition, Cheng Xi discussed the research status of the works of overseas Chinese in China from 2001 to 2006 (Cheng, 2007). Xiang Jun expounded some problems and shortages in the studies of overseas Chinese in his doctoral dissertation (Xiang, 2007).

Traditional research mainly adopts bibliometrics as the research method (Wang; Tan, 2017; Zhu *et al.*, 2016), and statistical and analysis results are usually presented in tables and numbers, lacking data visualization analysis. In particular, the network structure between authors and institutions is not well expressed due to the limitation of conditions, and the in-depth analysis of the cooperative relationship is lacking. In addition, the traditional literature analysis on the related topics of overseas Chinese lacks the analysis and co-occurrence of the reference and citation source publications respectively. With the development of the era of big data, it has become a new trend to use data mining and visualization methods to analyze overseas Chinese literature, which makes it possible to conduct in-depth analysis and research on relevant literature data better. In this paper, we extracted relevant literature on overseas Chinese studies in the recent ten years (2008-2021) from CNKI, and used data analysis techniques and mining methods such as text analysis, database theory, statistical measurement, PageRank (Chen; Shi, 2017; Li; Wang, 2017; Bowater; Stefanakis, 2023; Wang; Yao; Gong, 2023) and community network analysis (Xiong, 2016) to conduct statistical analysis and co-occurrence study on the keywords, authors, research institutions, source journals such as target literature, reference literature and citation literature. The information and knowledge in these fields can provide decision-making basis for overseas Chinese research and overseas Chinese affairs.

Methodology

Data Source

China National Knowledge Network (CNKI) is one of China's largest literature database sources, which accurately and timely collects academic results and data documents of various types of research. Meanwhile, CNKI provides some data resources for academic researchers to analyze and study the literature. By inputting the professional search syntax expression "KY= Overseas Chinese or KY % ethnic Chinese" from the professional search input box in the literature search TAB, the literature search was realized, that is, the search keyword was "overseas Chinese" or the relevant literature containing "overseas Chinese". At the same time, we selected the literature data published from 2008 to 2021 (the deadline for literature download is December 21, 2021) as the research object. The self-defined export format is selected, and the field information of title, author, institution, keywords, publication, annual volume, reference and citation were selected respectively.

Through literature data collection and sorting, we collected 1115 relevant articles. Among 893 journal articles, 233 were included in core journals, including 1 journal indexed by SCI, 1 journal indexed by EI, by CSSCI, and 4 by CSCD. Most of the literature related to overseas Chinese studies mainly came from journals and CSSCI journals account for the majority. However, from the comparison of master's and doctoral dissertations, there were relatively few doctoral research dissertations, indicating that the doctoral training with overseas Chinese as the academic research direction needs to be strengthened in the aspect of postgraduate training. At the same time, compared with other disciplines and fields, the number of papers in domestic and foreign academic conferences on overseas Chinese was insufficient, with less than 1% of the total academic achievements. It showed that the number of domestic and foreign academic conferences on overseas Chinese studies should be increased to better promote the cultural exchanges and academic research among overseas Chinese.

Data Processing

The original literature downloaded from CNKI was semi-structured text data, which except for reference and citation, other fields were separated by "$$" and", ". Although there was a certain regularity, the author, unit and keyword values were also separated by commas, which made it impossible to directly use conventional text processing software for field cutting and information extraction. At the same time, it was difficult to extract the author's information, title, source journal and publication date of the reference. It cited literature due to the different format requirements of the reference and cited literature. In addition, because a target document covered multiple references and citation documents, as well as multiple authors, units and keywords, it was impossible to use the same data table to analyze such literature data. To facilitate the analysis of literature data, it was necessary to transform the semi-structured text into structured data type. This paper used data mining R language and combined regular expression of character processing to convert it into multi-table data. It also established the relation between tables according to ER model in database theory.

Combining with study on the relationship between the co-occurrence data object, converting semi-structured documents data to 6 pieces of structured data tables, respectively, the literature data table (paper ID, title, source, time), keywords data table (paper ID, keywords), the author tables (paper ID, the author), unit of data table (paper ID, Unit), reference journal Data sheet (paper ID, source journal) and reference journal data sheet (paper ID, source journal). These data tables must satisfy at least the first paradigm in database design so that cells in each table do not have child tables. At the same time, the paper ID field of the target literature was used as the key field to realize the connection between the tables, and then completed the design and implementation of the database (Table 1).

The statistical results after data processing are shown in Table 1. Take the keyword variable for example, in the context of this study, "number" refers to the count of unique and non-repetitive keywords present in the analyzed literature. It represents the quantity or diversity of keywords within the field of overseas Chinese studies. This metric provides insights into the range of topics or themes the literature covers. On the other hand, "frequency" measures how frequently a specific keyword co-occurs with other keywords within the same document. It represents the total count of occurrences where a given keyword appears alongside other keywords in the analyzed literature. This metric helps us understand the interrelationships or associations between keywords and reveals the level of relevance or interconnectedness among them. The "number" and "frequency" metrics were derived from statistical analysis conducted using the R programming language. By applying these statistical techniques in R, we could quantify the number of unique keywords and calculate their co-occurrence frequencies within the corpus of overseas Chinese literature. The results obtained through these analyses provide valuable quantitative information about overseas Chinese literature, contributing to a better understanding of the landscape of overseas Chinese studies.

Table 1 — Number of research fields and co-occurrence of research subjects in literature related to the overseas Chinese.

Target literature source journal		Keyword	Author	Unit	Reference Source Journal	Citation Source Journal
Number	598	3316	1213	452	2714	1415
Frequency	598	23017	1984	1146	156960	89715

Source: China National Knowledge Network (2008-2021).

The number of source journals in the research literature with Overseas Chinese as the keyword between 2008 and 2021 was 598, indicating that these papers were published in 598 journals. It should be noted that the co-occurrence frequency of the target literature source journal refers to the co-occurrence relationship of the journal itself, so the co-occurrence frequency was the same as the number of the source journal. In addition, the number of keywords, authors, units, reference source journals and citation source journals were 3316, 1213, 452, 2714 and 1415, respectively, corresponding co-occurrence times were 23017, 1984, 1146, 156960 and 89715, respectively. Based on the analysis of the number and frequency, it can be seen that the co-occurrence degree among authors and units was relatively small. In contrast, the co-occurrence degree among key words, reference source journals and citation source journals was relatively large.

Keyword Co-Occurrence Analysis

A statistical analysis was carried out on a total of 3316 keywords found in 1115 target literature sources to identify commonly used keywords associated with overseas Chinese studies. The results of this analysis, including the identified keywords and their respective frequencies, are presented in Table 2. This statistic involved counting the number of occurrences of each keyword in all the literature. The most frequently used or important keyword can be identified through word frequency analysis, allowing for a deeper understanding of the themes, or thematic trends, present in the overseas Chinese literature.

Table 2 — Frequency of the top 45 keywords.

N°	Keyword	Frequency	N°	Keyword	Frequency	No.	Keyword	Frequency
1	Overseas Chinese	652	16	Chinese dream	15	31	International migration	11
2	Chinese citizens residing abroad	128	17	Chinese society	14	32	Myanmar	11
3	One Belt One Road	59	18	Chinese association	14	33	Overseas policy	11
4	Chinese education	49	19	Confucius Institute	14	34	Thailand	11
5	Southeast Asia	39	20	Dissemination	13	35	Development	10
6	Overseas Chinese affairs	24	21	Chinese abroad	13	36	Reform and opening	10
7	Maritime Silk Road	23	22	Economic development	13	37	International communication	10
8	Chinese media	22	23	Soft power	13	38	Overseas Chinese studies	10
9	Public diplomacy	21	24	Chinese citizen	13	39	Chinese newspaper	10
10	Culture Identity	21	25	Overseas Chinese Affairs Office of the State Council	12	40	culture	10
11	Xi Jinping	21	26	Overseas Chinese literature	12	41	Cultural exchange	10
12	China	21	27	Chinese	12	42	Influence	10
13	Chinese culture	19	28	Hometown of overseas Chinese	12	43	Chinese story	10
14	United front	16	29	Cultural dissemination	12	44	Overseas dissemination	9
15	Dual citizenship	15	30	The Philippines	11	45	Chinese community	9

Source: China National Knowledge Network (2008-2021).

From the statistical ladder and distribution perspective, Chinese education, Southeast Asia, overseas Chinese affairs, Chinese media, public diplomacy and cultural identity were important entry points for studying the work of overseas Chinese. Chinese culture, dual citizenship, Chinese society, Chinese community, Confucius Institute and overseas Chinese literature were also closely studied. Regarding geographical distribution, the research on overseas Chinese was carried out in Southeast Asia, the Philippines, Myanmar and Thailand. Among the 45 keywords with the highest frequency of occurrence, the Overseas Chinese Affairs Office of the State Council ranked 25th as a keyword with the highest frequency of occurrence, which was sufficient to show that the Overseas Chinese Affairs Office of the State Council plays a pivotal role in the field of overseas Chinese studies (Table 2).
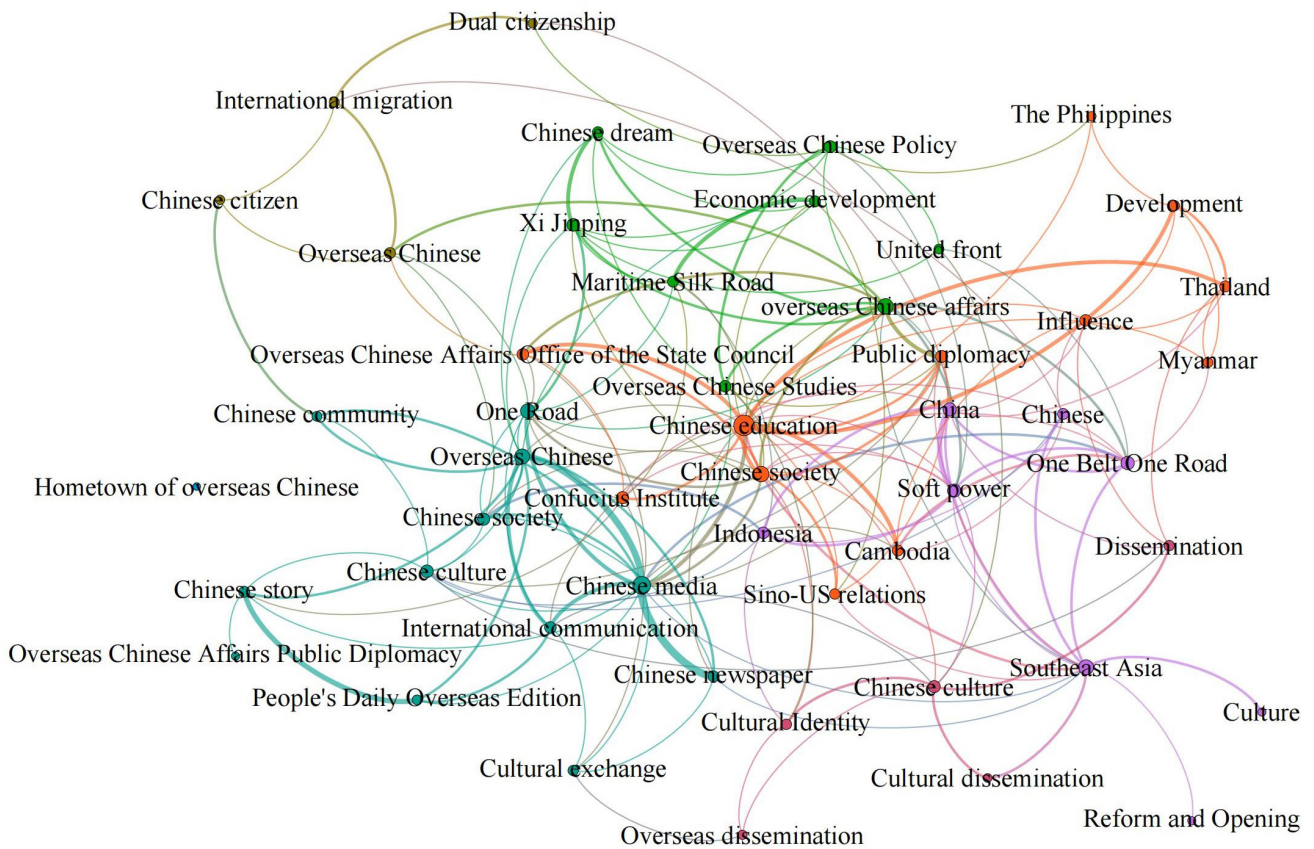
The statistics and analysis of 3316 keywords appeared together in the same target literature, observed which keywords are usually used together to describe the same research topic, and then explained that these keywords strongly correlate with the study of a certain problem. To better display the co-occurrence relationship of keywords, the co-occurrence relationship of all keywords was not presented, and only the top 50 keywords with strong co-occurrence relationship were displayed in this paper. At the same time, due to the overseas Chinese and Chinese citizens residing abroad appeared more frequently than other keywords, and they usually appeared at the same time with other top 50 keywords that appear frequently, the co-occurrence relationship between overseas Chinese and Chinese citizens residing abroad and other keywords was omitted. The analysis results were shown in Figure 1. The larger the node means the more times a keyword co-occurs with other keywords, and the thicker the edge means the higher the co-occurrence probability of keywords corresponding to nodes at both ends of the edge. For example, keywords of Chinese media were highly likely to co-occur with Chinese newspapers and overseas Chinese literature, indicating that current research on Chinese media was closely linked with those of Chinese newspapers and overseas Chinese literature. Similarly, overseas Chinese literature was closely related to Chinese story and international communication. The research focus of overseas Chinese affairs was public diplomacy, One Belt One Road and overseas Chinese policy. At the same time, Chinese education mainly focused on studying Confucius Institute and Chinese society, mainly in Southeast Asia, Thailand and Cambodia.

The joint analysis of Table 2 and Fig. 1 showed that the occurrence frequency of keywords is proportional to the co-occurrence degree to some extent, and the higher the occurrence frequency, the greater the co-occurrence degree with other keywords. However, there were also some keywords with high frequency, such as cultural identity and united front. Still, their co-occurrence degree was small, indicating that the related research of this part needs to be integrated with other research content to improve, and expand the research perspective of these contents. Through keyword co-occurrence analysis, it can be seen that relevant researches on keywords with large co-occurrence can maintain normal research intensity and vision. At the same time, it was necessary to add research directions corresponding to other boundary keywords, so that there was better connectivity between research topics about overseas Chinese.
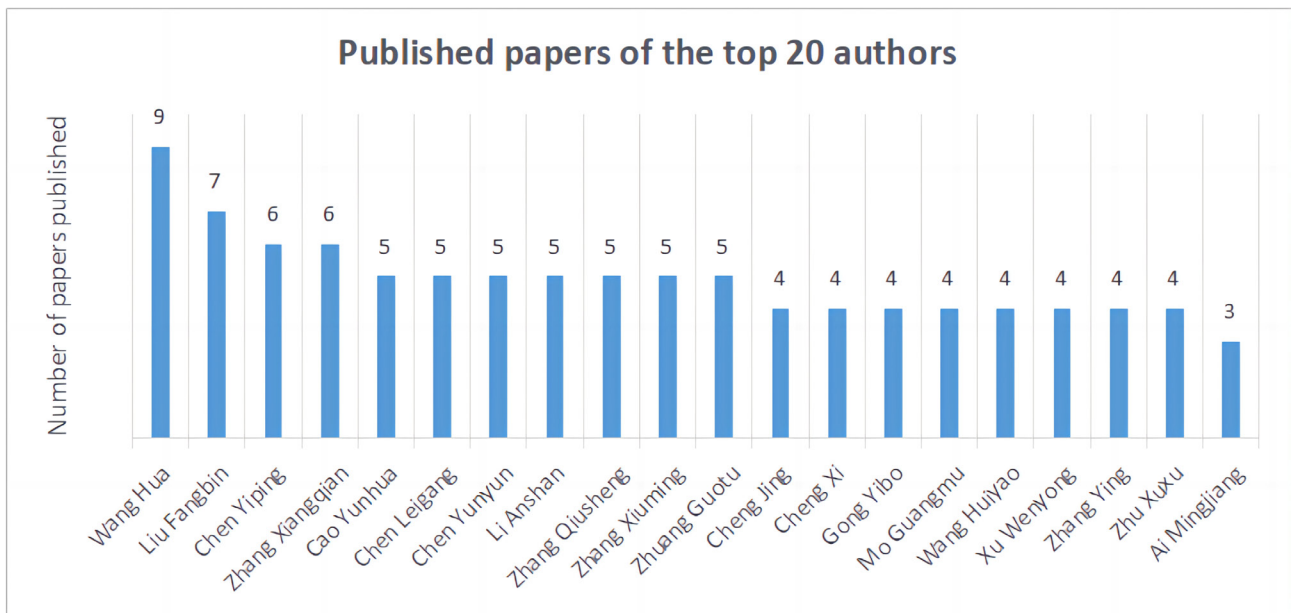
Author Co-Occurrence Analysis

Author co-occurrence analysis was a collaborative analysis conducted by academic research authors to discover which authors have a group cooperation to complete the relevant research of overseas Chinese. Before the co-occurrence analysis, through the statistical analysis of 1213

Figure 1 — The co-occurrence relationship of the top 50 keywords with co-occurrence degree.
Source: China National Knowledge Network (2008-2021).

authors, combined with the publication information of 1115 target literature, about 1 paper was published per person. Then we could preliminarily understand that the co-occurrence of authors was not high. According to the comparison between the number of authors 1213 and the number of co-occurrence of authors 1984 in Table 1, it could also be seen that the degree of co-creation and research of authors in related literature on overseas Chinese studies was relatively weak, and only a small number of authors were likely to carry out relevant collaborative research. In addition, from the statistics of published papers, as shown in Fig. 2, Wang Hua, who has the highest number of published papers, published less than 1 paper every year in 14 years, regardless of the order of authors. Among 1213 authors, the top 20 authors published a total of 98 papers, about 5 papers per capita, while the average number of all authors published about 1 paper per capita, which to some extent also indicates that the authors were not enthusiastic about cooperation.
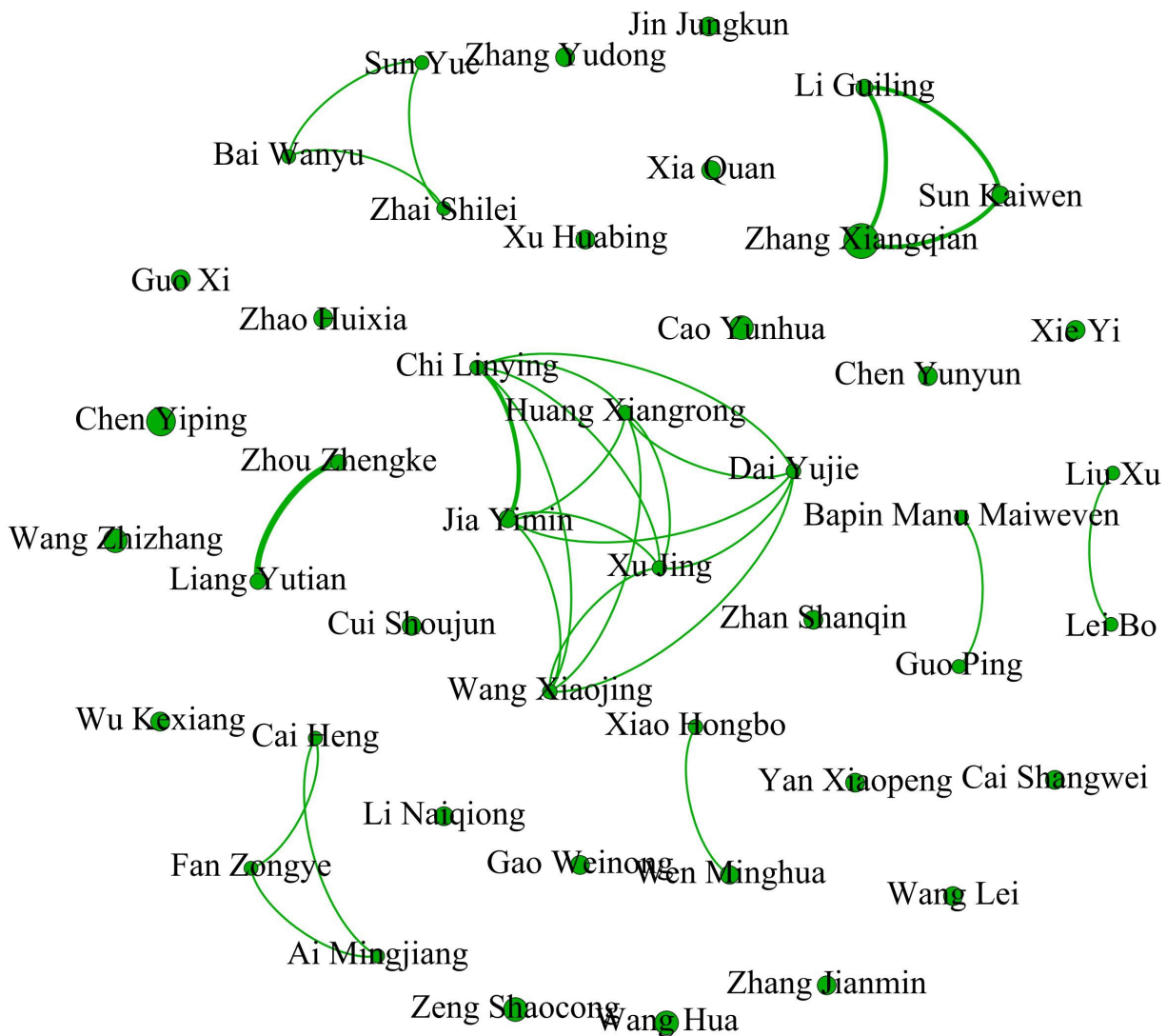
To better analyze the cooperation among the main authors of overseas Chinese studies, the PageRank algorithm commonly used in data mining for information and knowledge search ranking was used to calculate the PR value of each author and observe its importance in the cooperative network. In the community network, if the PR value is larger, it indicates that the author represented by this node has an important position in the cooperative network community, and it can better play the role of author cooperation in the community network. In other words, if an author has the largest PR value in the author co-occurrence network, it indicated that the author has the strongest cooperation and relatively stable in the co-occurrence community, and also has strong academic influence.

Figure 2 — Published papers by the main author.
Source: China National Knowledge Network (2008-2021).

Analyzed which other authors each author collaborates with by analyzing all author collaborations. If cooperation occurs once, set the weight of cooperation between the two authors to 1. In the case of multiple cooperation without regard to the order of cooperation, the weight of cooperation between authors was calculated by the cumulative number of cooperation. Then the network distribution map between authors could be constructed. Meanwhile, combined with PageRank algorithm, PageRank calculation was performed for each node in author cooperation network, and the top 45 network nodes with high PR values were extracted for comprehensive analysis, as shown in Figure 3. In the figure, each node represented the corresponding author, the node size represented the PR value of its node, and the edge represented the strength of cooperation. The thickness of the edge reflected the strength of cooperation between authors. As can be seen from the co-occurrence of authors in Figure 3, the nodes with large PR values included Zhang Xiangqian, Chen Yiping, Cao Yunhua, Zeng Shaocong, Wang Zhizhang and Wang Hua, etc. Still, there were almost only nodes with small PR values around them. Due to the large amount of data, if all co-authors were displayed, the complexity of the graph made it impossible to visually observe and analyze the cooperative relationship in the network structure. Therefore, nodes with smaller PR values were filtered and not displayed in the co-occurrence network according to the forward segmentation display to ensure the visual effect of network structure.

Combined with Figure 2, it can be found that most of these authors in Figure 3 did not publish many papers, indicating that they are likely to appear in the target papers with many collaborators. As shown in Figure 3, the online community where Xu Jing and Wang Xiaojing belong had cooperation among community members, which made the single author have a large PR value and stable cooperative relationship. Still, they only published a small number of papers. On the contrary, Wang Hua, who had the largest number of publications in Figure 2, had a smaller node in Figure 3, that is, a lower PR value, indicating that the author Wang Hua published literature as an independent author with a relatively low degree of cooperation. Therefore, through the analysis

Figure 3 — Top 45 Author Collaboration Networks by PageRank.
Source: China National Knowledge Network (2008-2021).

of author co-occurrence relationship of overseas Chinese studies, it was found that the number of papers published by the author was not related to the degree of cooperation. Still, the degree of author co-occurrence may be related to the number of papers published, mainly depending on the number of collaborators in the academic research. In addition, by analyzing the PageRank value of nodes in the author co-occurrence network, we could find the importance of authors in their communities, which is conducive to the promotion and dissemination of the main research results of overseas Chinese. In particular, identifying these key authors was conducive to the matching and academic cooperation of overseas Chinese research team members on relevant topics.

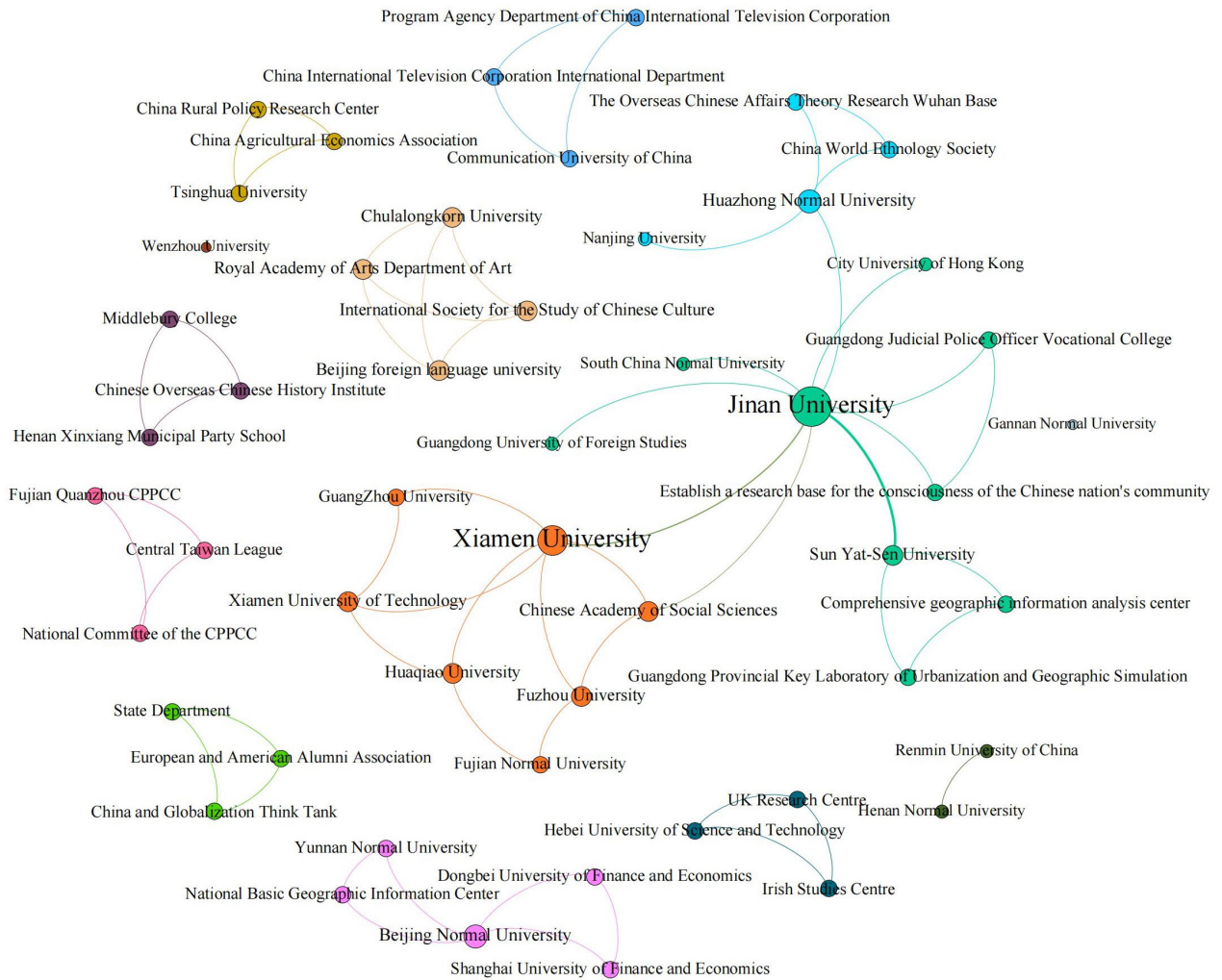## Co-Occurrence Analysis of Research Institutions

According to the statistical analysis of research institutions, from 2008 to 2021, Jinan University as the first research unit or cooperative unit published a total of 114 papers related to overseas Chinese, ranking first, Huaqiao University, Xiamen University and Sun Yat-sen University

ranked second, third and fourth with 54, 36 and 28 papers, respectively. Similarly, according to the frequency of research institutions appearing in relevant target literature, word clouds of the top 100 frequency of research institutions were constructed.

It is easy to find that Jinan University and Huaqiao University are the major contributing institutions to the research of overseas Chinese, which is related to the nature of running schools of these two universities. Jinan University and Huaqiao University are national key universities under the supervision of Overseas Chinese Affairs Office of the State Council. Jinan University is also a 211 key university with rich resources related to overseas Chinese. These two universities many overseas Chinese students, and set up relevant departments and institutions to conduct research overseas Chinese subjects. Compared with other universities, they also have more research topics on overseas Chinese studies. Therefore, the research results of Jinan University and Huaqiao University on overseas Chinese are relatively intensive. Due to their geographical location and historical background, Xiamen University and Sun Yat-sen University are the ancestral home of many Overseas Chinese, and the research on their survival and development has become the focus of their universities and colleges. As the capital of China and the political and cultural center of China, Beijing also competes courageously in the field of overseas Chinese research, with numerous research institutions, such as Renmin University of China, Peking University, Chinese Academy of Social Sciences, Minzu University of China, and The Institute of Overseas Chinese History in China. In addition, overseas Chinese and native distribution are mainly distributed in Fujian, Guangdong and Zhejiang provinces, which makes the Huaqiao university, Xiamen university and Fujian normal university in Fujian, Jinan university and Sun yat-sen university in Guangdong, Wenzhou university and Zhejiang normal university in Zhejiang and other research institutions relevant studies of overseas Chinese has become the main unit.

The cooperation between institutions is analyzed from the perspective of co-occurrence relationship of research institutions. As shown in Figure 4, the node represents the research institution related to overseas Chinese, and its size represents the cooperation intensity between the institution and other research institutions, the edge represents the cooperation relationship between the mechanisms, and the thickness of the edge represents the cooperation intensity of the corresponding mechanisms at both ends. The figure shows the co-occurrence network of research institutions with cooperation degree greater than or equal to 2, while nodes with cooperation degree less than 2 are filtered and not presented. It is not difficult to find that Jinan University and Huaqiao University, which have many achievements and similar educational properties, do not have cooperative relationship in the research of overseas Chinese subjects, which is not conducive to researching and developing relevant themes and important subjects. At the same time, from the perspective of co-occurrence network community, research institutions are more inclined to cooperate between research institutions located in close geographical locations. For example, Xiamen university, Huaqiao university, Fuzhou university, institute of technology in Xiamen and Fujian normal university in Fujian Province, South China normal University, Jinan university, Sun yat-sen University and Guangdong University of Foreign Studies in Guangdong Province, Tsinghua University and the research centers such as Agricultural Economic Association of China and China Rural Policy Research Center in Beijing have strong cooperation between each other. However, the cooperation between Wenzhou University, Ningbo University and Zhejiang Normal University in Zhejiang Province and have many achievements, is relatively weak. At the same time, there is a lack of cooperation between research institutions in provinces or larger regions.

Based on the comprehensive analysis of the published achievements, geographical distribution and cooperation of overseas Chinese research institutions, there are the following

Figure 4 — Co-occurrence network between major research institutions.
Source: China National Knowledge Network (2008-2021).

characteristics: (1) most of the research achievements about overseas Chinese are produced by Jinan University and Huaqiao University, but there is no cooperative relationship between them. (2) The main research results come from the research institutions in Guangdong, Fujian, Zhejiang and other provinces, and there is close cooperation between the research institutions in Guangdong and Fujian provinces. Still, the cooperation between the research institutions in the two provinces is scarce. In addition, the research institutions in Zhejiang province, whether within or outside the province, do not have strong research cooperation with overseas Chinese. (3) Except for the cooperative relationship between Jinan University and Sun Yat-sen University, cooperation among research institutions is weak. (4) From the perspective of the division of research institutions' network community, institutions lack trans-regional research cooperation.

Co-Occurrence Analysis of Source Publications

Literature source publication refers to the source of a document, that is, the document is published by which press, journal or newspaper and other media. According to the different research objects, it can be divided into target literature source publication, reference literature

source publication and citation literature source publication. 1155 articles on overseas Chinese were collected from 598 source publications. However, the same target literature contains multiple references or is cited by multiple references from different journals. Among these target documents, references from 2714 different journals were cited, and 1415 different journals also cited references. Therefore, it is beneficial to recommend the target literature and spread the relevant knowledge to study which journals have published papers related to overseas Chinese, which journals are usually cited by papers related to overseas Chinese, and which journals prefer to cite papers related to overseas Chinese. In addition, by studying the co-occurrence relationship between different journals as sources of different journals, we can find the degree of association between journals when studying topics related to overseas Chinese, that is, which journals are likely to be cited by the target literature and the cited literature. The occurrence frequency statistics of target literature source journals were conducted to construct the word cloud of the top 100 occurrence frequency of target literature source journals.

It is easy to find that Bagui Overseas Chinese Journal, Overseas Chinese History Studies, Southeast Asian Studies and Jinan University are the main publications of overseas Chinese studies, with 55, 26, 24 and 18 targeted documents respectively. Huaqiao University, which ranked second in publication volume, ranked eighth in the statistics of source publication frequency with 12 publications. The appearance of Jinan University and Huaqiao University as journals implies that these target documents are master's or doctor's dissertations, and also indicates that these two schools have the postgraduate training system of researching related topics of overseas Chinese, and have obtained many achievements. Through the statistical analysis of target source journals, we can find out which journals are willing to receive and publish academic research results on overseas Chinese studies to facilitate relevant scholars to inquire, read and publish research results on topics related to overseas Chinese in these journals.

A total of 15,413 articles from 2714 journals are cited in the target literature of overseas Chinese research, among which 50 journals of pre-overseas Chinese history are cited as shown in Table 3. Overseas Chinese History Studies has the most cited papers in the target literature with 523, far ahead of The Overseas Chinese Publishing Company in second place. This shows that most of the research results on Overseas Chinese read and cite the literature published by Overseas Chinese History Studies. Overseas Chinese History Studies has become a major publication for collecting and studying Overseas Chinese literature and publishing relevant results. At the same time, it is not difficult to find from Table 3 that most of the literature cited on overseas Chinese studies come from publishing houses. Compared with other fields and other disciplines, there are not many kinds of periodicals and magazines. In addition, from the perspective of frequency distribution, most of the literature cited in the target literature are from the first 27 journals, which enables overseas Chinese researchers to focus their literature search work on these 27 journals, or subscribe to these publications appropriately, and pay attention to the dynamics and progress of relevant research (Table 3).

According to the number of times different journals appear in the same target literature to establish the co-occurrence network of journals with different reference sources. A co-occurrence network with complex relationships including 2714 nodes and 156960 edges can be obtained. Sufficient computer memory space is needed for data analysis and co-occurrence relationship construction of arge nodes and edge weights. Based on the constraints of operating environment and the effectiveness of mining knowledge, this paper presents only the co-occurrence networks
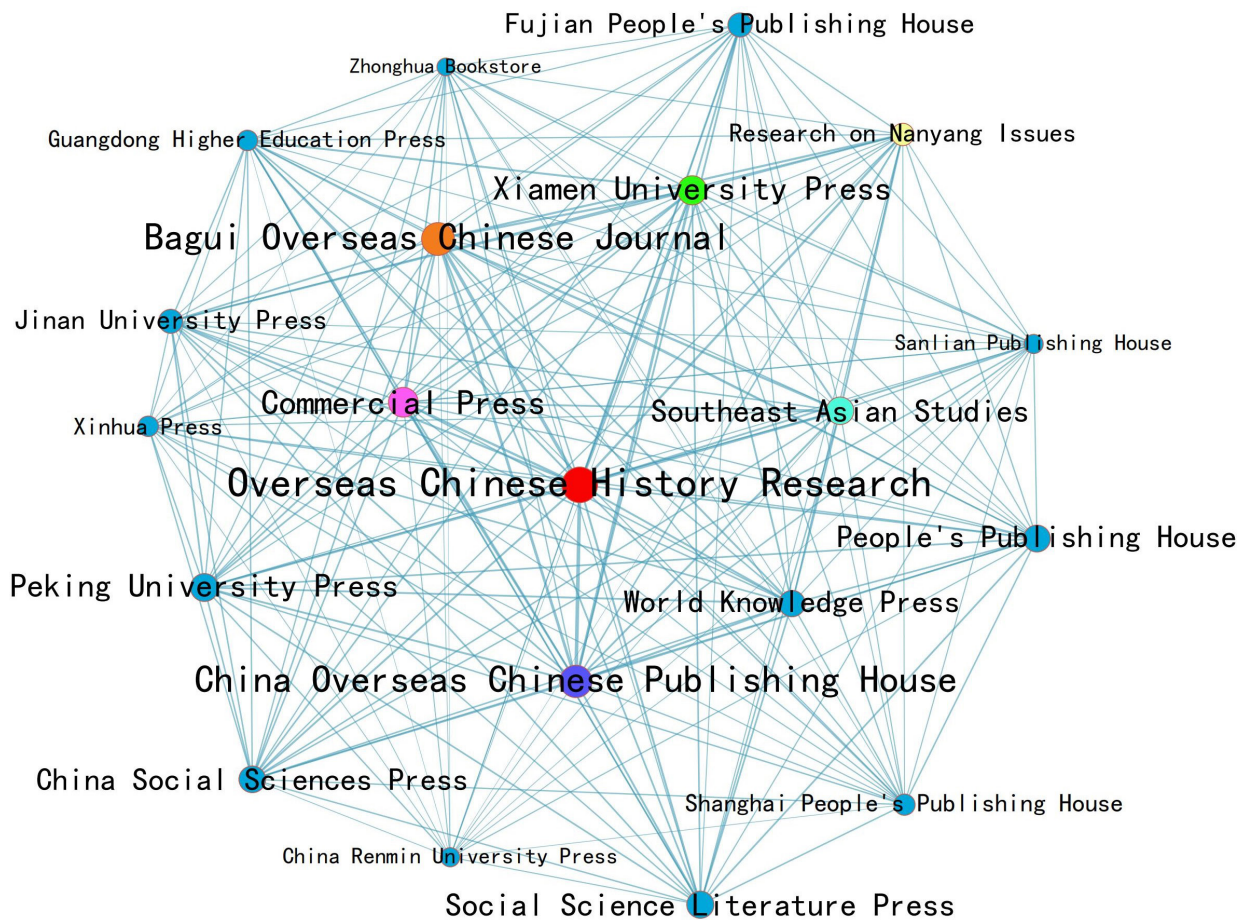
constructed by the top 50 source journals with the highest frequency in references. However, there will be 84,872 times co-occurrence between these 50 reference source journals and 2553 other reference source journals. The connection between network nodes and sideline makes the co-occurrence network diagram unable to display its structure intuitively and effectively. Therefore, nodes (source journals) with top 20 co-occurrence degree are selected to generate the co-occurrence network of the main source journals of references, as shown in Figure 5.

Table 3 — Top 50 major reference source publications.

No.	Source Publication	Frequency	No.	Source Publication	Frequency
1	Overseas Chinese History Studies	523	26	China Renmin University Press	112
2	China Overseas Chinese Publishing House	333	27	Sanlian Publishing House	106
3	People's Publishing House	325	28	Research on Overseas Chinese Affairs	81
4	Xiamen University Press	325	29	Kaiming Bulletin	75
5	Bagui Overseas Chinese Journal	322	30	Central Party Literature Press	75
6	Southeast Asian Studies	283	31	Journal of Xiamen University	74
7	World Knowledge Press	246	32	Fudan University Press	73
8	Peking University Press	219	33	Zhejiang People's Publishing House	73
9	Zhonghua Bookstore	217	34	Asia-Pacific Research Center of Peking University	69
10	People's Daily	205	35	Daily News	64
11	Social Sciences Literature Press	197	36	Huaxia Press	63
12	China Social Sciences Press	183	37	Guangxi People's Publishing House	59
13	Commercial Press	172	38	Jinan University	59
14	Guangdong People's Publishing House	167	39	Xiamen University	56
15	Research on Nanyang Issues	165	40	Journal of Huaqiao University	54
16	Fujian People's Publishing House	162	41	Jinan Journal	52
17	Fujian People's Publishing House	162	42	Jiuzhou Press	50
18	Shanghai People's Publishing House	161	43	Guangming Daily	49
19	Chinese Business Daily	160	44	Shun Pao	49
20	Jinan University Press	160	45	Southeast Asian Studies: A Quarterly Journal of Translation	43
21	Xinhua Press	159	46	Journal of Yunnan Normal University	42
22	People's Daily Overseas	152	47	Guangxi Normal University Press	41
23	World Nation	126	48	Fujian Education Publishing House	40
24	Crossroads: Southeast Asian Studies	117	49	Shanghai Translation Publishing House	39
25	Guangzhou Press	113	50	QIUSHI	38

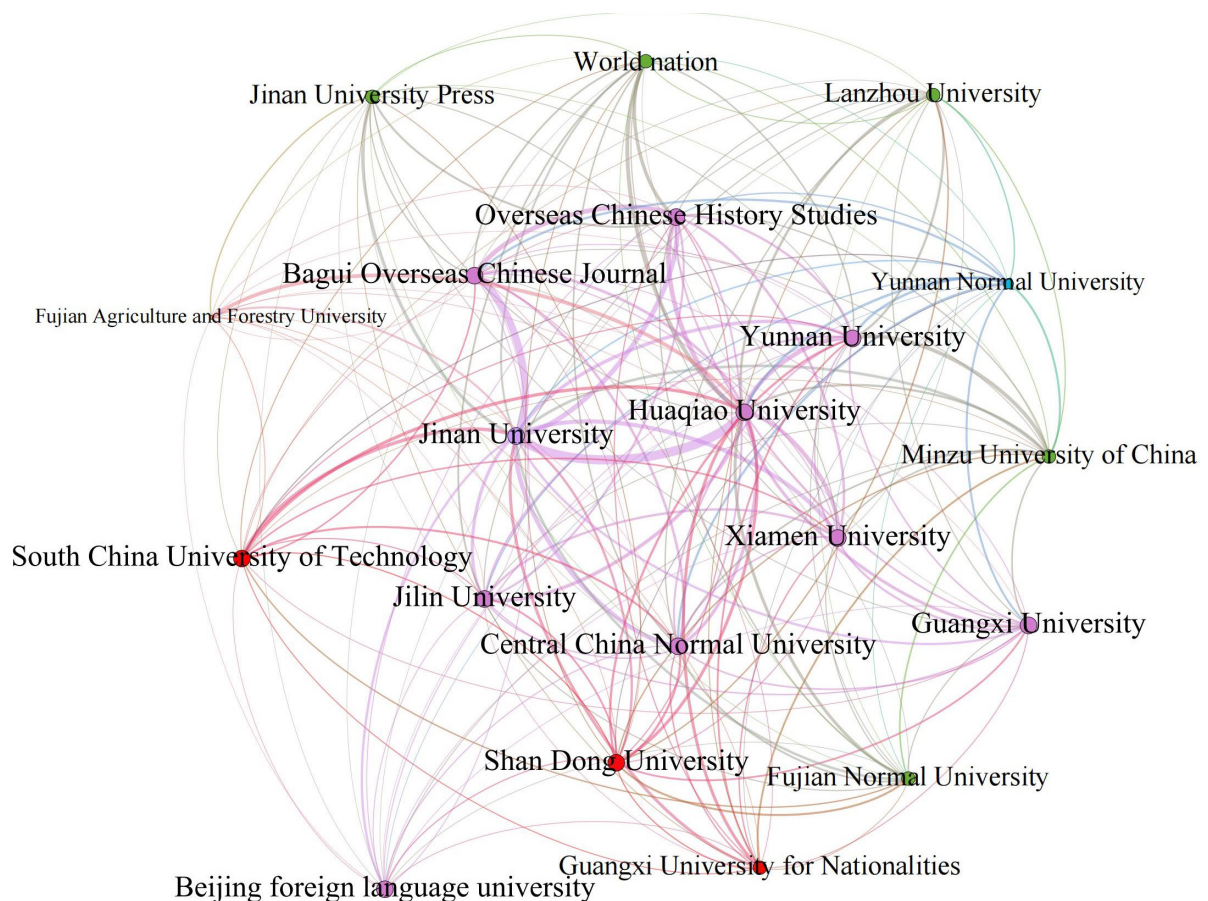Source: China National Knowledge Network (2008-2021).

Figure 5 — The co-occurrence network of reference sources.
Source: China National Knowledge Network (2008-2021).

From the perspective of the co-occurrence network of reference sources, the co-occurrence relationship of Overseas Chinese History Studies, China Overseas Chinese Publishing House, Bagui Overseas Chinese Journal, Xiamen University Press, Southeast Asian Studies, Commercial Press and Southeast Asian Affairs is closely related. In contrast, the co-occurrence relationship between publications such as China Renmin University Press, Xinhua Publishing House, Zhonghua Bookstore and Shanghai People's Publishing House and Overseas Chinese History Studies is not obvious. This shows that the target literature that cites papers from Overseas Chinese History Studies usually prefers to cite publications with high co-occurrence relationship, such as The Chinese Overseas Publishing House and Bagui Overseas Chinese Journal, but is less likely to cite some publications with high target literature citation, such as Zhonghua Bookstore and Shanghai People's Publishing House. Similarly, through the co-occurrence network of reference source journals, can find the co-occurrence feature of a source journal and other source journals. When find an important literature in the publication, the co-occurrence relation can be used to find other publication whether also have similar documents related to the study, at the same time of improving the quality of references, researchers can purposefully search and retrieve relevant literature in other journals with high co-occurrence, thus improving the efficiency and quality of data search and overseas Chinese research.

Through the analysis of literature sources, we can find out which publications prefer to cite the relevant achievements of overseas Chinese studies, which is conducive to promoting achievements and knowledge dissemination. Through the analysis of the cited literature, it is easy to know that the target literature of overseas Chinese studies have been cited for a total of 5562 times, and the co-occurrence of these cited literature has reached 89715 times. The journals of Jinan University (master and doctoral papers) cited the target literature 270 times, with the highest number of citations. The journal of Huaqiao University followed with 183 citations. It is not difficult to find from the top 20 source journals with the most cited target documents that most of the research results on overseas Chinese are cited by master's theses and doctoral theses. Only three journals, namely Overseas Chinese History Studies, Bagui Overseas Chinese Journal and Journal of World Peoples Studies, cite the research results of overseas Chinese. Therefore, given the situation that the research achievements of overseas Chinese are cited, it is necessary to promote the citation of newspapers, magazines and other publications, and promote and accelerate the sharing and research of academic achievements.

By constructing the co-occurrence network of journals with different cited sources that cite the same target literature, we can determine which journals will often focus on the same target literature of overseas Chinese related research. As shown in Fig. 6, the co-occurrence network of the main source journals cited by literature is described. It can be found that the same target literature

Figure 6 — Co-occurrence network of main citation source publications.
Source: China National Knowledge Network (2008-2021).

is highly likely to be cited by the master and doctoral papers of Jinan University and Huaqiao University. In addition, the same target literature is often cited in the master's and doctor's theses of Jinan University, Bagui Overseas Chinese Journal, Overseas Chinese History Studies and Central China Normal University. The co-occurrence relationship of source publications indicates that if a research achievement related to overseas Chinese studies is cited in the master's and doctor's dissertations of Jinan University, it is likely to be also favored by scholars of Overseas Chinese History Studies and master and doctoral papers of Huaqiao University and Central China Normal University. It is conducive to accurately recommending overseas Chinese research results to these university research institutions or publications.

Results Discussion and Decision-Making Suggestions

The research results can provide meaningful guidance and decision-making suggestions for overseas Chinese research to a certain extent, and improve the efficiency and quality of overseas Chinese research.

1) From the data source processing and statistical analysis results, there is a lack of doctoral training in the research results on overseas Chinese, and the related research results lack a certain systematization and organization. In addition, the lack of academic conferences on overseas Chinese studies at home and abroad makes the research results unable to be timely discussed between experts and scholars, thus lacking academic exchanges and cooperation between relevant schools and units. In addition, from the analysis of the number of authors and units and the number of co-occurrence, the degree of cooperation between authors and research institutions is not high, leading to the phenomenon of "single soldier fighting" in this field, which is not conducive to the related research with overseas Chinese as the theme. Therefore, in the study of overseas Chinese, it is necessary to promote the research of doctoral students in the field of overseas Chinese to make the research content and results more targeted, in-depth, organized and systematic. In addition, while strengthening the cooperation between authors and research institutions, planning some academic conferences on overseas Chinese studies is necessary, further strengthening the cooperation and exchange among scholars, experts and institutions, and realize the sharing of resources and research results.

2) According to the statistics of authors' publications and the results of co-occurrence relationship, it was found that the number of authors' publications per capita was very low, resulting in a low co-occurrence relationship among authors. At the same time, authors with a high number of publications often conduct academic research and publish their results as independent authors, which makes them have a low PageRank value and lack the influence of team cooperation, which is not conducive to the expansion of the research scope of overseas Chinese and the cultivation of research talents. In improving per capita number of papers at the same time, therefore, need to fully consider the author collaboration network key authors in the community recognition, to train team of overseas Chinese and related research elite and study leader to provide back-up talent, at the same time is conducive to the members of the project research and academic cooperation collocation, to better promote the overseas and ethnic Chinese studies related to the promotion and spread of information and knowledge.

3) Research achievements on overseas Chinese are mainly concentrated in Jinan University and Huaqiao University, two key universities under the supervision of Overseas Chinese Affairs Office of the State Council. Unfortunately, these two universities lack strong research cooperation in studying overseas Chinese. At the same time, most research institutions are located in

Guangdong, Fujian and Zhejiang, the ancestral home of most overseas Chinese, which is conducive to studying the life, roots, economy, culture and history of overseas Chinese. However, through the co-occurrence network of research institutions, it is found that although there is a certain degree of cooperation between research institutions in Guangdong and Fujian provinces, there is a lack of regional cooperation between different provinces, which has a low spatial connection. Therefore, while improving university-university cooperation represented by Jinan University and Huaqiao University, we should make full use of the resources and backgrounds of each region, strengthen regional cooperation, enhance the sharing of research resources and results, and achieve a comprehensive understanding of overseas Chinese studies in the whole country and even the world.

4) Through statistical analysis and cooccurrence relationship study of target journal, reference and citation, the research results can provide the source and destination of results for overseas Chinese researchers, and provide reference information and resources for data collection, reference and publication of results. (1) Regarding source and journal analysis of target journals, most of the research results of overseas Chinese are published in three journals, Namely, Bagui Overseas Chinese Journal, Overseas Chinese History Studies and Southeast Asian Studies. Jinan University and Huaqiao University promote the relevant research of overseas Chinese employing postgraduate training. (2) The analysis of the co-occurrence of reference sources and periodicals shows that most of the literature on overseas Chinese studies cite works of publishing houses, but lack of references to periodicals and newspapers. Therefore, major journals and mainstream media of Overseas Chinese studies need to hold academic conferences to enhance the academic influence of journals, promote the sharing, communication and teamwork among researchers, and improve the citation rate of journals and newspapers. (3) For literature source journals cited concurrence relationship, which can be used to find journal focusing on the common interests of overseas Chinese and study a particular subject, makes the target literature can be recommended to these publications for accuracy at the same time, the overseas Chinese and related research results to improve their reading and reference, speed up the transmission of information and knowledge. In addition, given the situation that overseas Chinese research achievements are cited, it is necessary to promote the citation of newspapers, magazines and other publications, and promote and accelerate the sharing and research of academic achievements.

Conclusion

Through the selection of overseas Chinese or including overseas Chinese as the main keywords of the target literature as the research object, focusing on the statistical analysis and co-occurrence study of the information characteristics such as authors, research institutions, keywords, target literature source journals, reference literature source journals and cited literature source journals in the target literature, to grasp the topic positioning, author relationship, institutional cooperation, quality of literature inquiry and reading, and the current situation of the use of the results of research on overseas Chinese.

In conclusion, this study has highlighted the importance of systematic organization, collaboration, and knowledge sharing in overseas Chinese research. Promoting doctoral training, enhancing cooperation between authors and institutions, and planning academic conferences can significantly improve the efficiency and quality of research in this field. Moreover, recognizing the significance of author collaboration networks, encouraging team-based research, and strengthening regional cooperation will contribute to expanding research scope and talent cultivation. Lastly, efforts should be made to enhance the academic influence of journals,

improve citation rates, and facilitate the dissemination of information and knowledge in overseas Chinese studies. These actions can achieve a more comprehensive and impactful understanding of overseas Chinese.

References

Bowater, D.; Stefanakis, E. Extending the Adapted PageRank Algorithm centrality model for urban street networks using non-local random walks. *Applied Mathematics and Computation*, v. 446, n. 127888, 2023.

Chen, X. W.; Shi, Y. T. Identifying key nodes in social network with improved pagerank algorithm. *Data Analysis and Knowledge Discovery*, v. 1, n. 8, p. 68-75, 2017.

Cheng, X. The research developments on southeast asian Chinese of china: reviews on recently published book. *Southeast Asian Studies*, n. 5, p. 64-71, 2007.

Cheng, S.; Wang, B. Impact of the Belt and Road Initiative on China's overseas renewable energy development finance: Effects and features. *Renewable Energy*, v. 206, p. 1036-1048, 2023.

Deng, S. H.; Xu, X. The status of Chinese overseas Chinese research in china in the last 10 years: an analysis based on CSSCI. *Dongyue Tribune*, v. 32, n. 11, p. 74-78, 2011.

Larsen, M. L.; Oehler, L. Clean at home, polluting abroad: the role of the Chinese financial system's differential treatment of state-owned and private enterprises. *Climate Policy*, v. 23, n. 1, p. 57-70, 2022.

Li, T.; Wang, G. Y. PageRank in measuring the importance of standard literature. *Journal of Suzhou University of Science and Technology (Natural Science Edition)*, v. 34, n. 2, p. 59-62, 2017.

Lu, Y. Analysis on the status quo of overseas Chinese studies in the perspective of national fund of philosophy and social science (NPSS): quantitative analysis on the programs of overseas Chinese studies funded by NPSS during 1991 to 2013. *Southeast Asian and South Asian Studies*, n. 2, p. 94-100, 2014.

Jin, H. L. An analysis of the return and the reasons of overseas Chinese students to their home country. *Journal of North-east Asian Cultures*, v. 1, n. 62, p. 145-158, 2022.

Pan, Y. N. Challenges of Guangdong enterprises in the process of "going out to Southeast Asia" and the role of ethnic Chinese. *Overseas Chinese History Studies*, n. 1, p. 11-20, 2015.

Wang, S.; Yao, X.; Gong, D. Overlapping community detection in software ecosystem based on pheromone guided personalized PageRank algorithm. *Information and Software Technology*, v. 163, n. 107283, 2023.

Wang, H. Formation, collection status and consolidation measures of genealogical literature of overseas Chinese in Southeast Asia. *Library Theory and Practice*, n. 4, p. 103-107, 2016.

Wang, J. H.; Tan, Z. Y. A review of bibliometric evaluation research of scientific creativity. *Library and Information Service*, v. 61, n. 3, p. 131-139, 2017.

Wu, Y.; Lin, Y. A Review of 30 years' research on overseas Chinese economy based on essays included in CNKI. *Asia-pacific Economic Review*, n. 5, p. 143-148, 2016.

Xiang, J. A study on the overseas Chinese and china's economic modernization. PhD thesis, Jinan University, 2007.

Xiong, Z. L. An inquiry of returned overseas Chinese'community networks and identity construction in beihai Qiaogang town. *Bagui Overseas Chinese Journal*, n. 3, p. 64-68, 2016.

Xu, Y. Ethnic Chinese studies in mainland China: A preliminary analysis based on the Chinese Social Science Citation Index (CSSCI). *Overseas Chinese History Studies*, n. 1, p. 59-69, 2007.

Xu, Y. Ethnic Chinese studies in mainland China: An analysis on the quotations cited in articles published in the Journal of Overseas Chinese History Studies between 1999 and 2008. *Overseas Chinese History Studies*, n. 2, p. 1-13, 2010.

Yu, Z.; Xiao, Y.; Li J. How does geopolitical uncertainty affect Chinese overseas investment in the energy sector? Evidence from the South China Sea Dispute. *Energy Economics*, v. 100, n. 105361, 2021.

Zhou, H. J. The Living Condition of Overseas Chinese in Africa and Their Relations with the Local Ethnicities. *Southeast Asian Studies*, n. 1, p. 79-84, 2014.

Zhu, S. L. *et al.* The status and countermeasures of domestic bibliometrics application research in the age of big data. *Information Science*, v. 34, n. 8, p. 116-121, 2016.

Acknowledgments