An improved solution for partner selection of industry-university cooperation

Congjing Ran, Kai Song, Le Yang (correspondence: yanglegd@gmail.com)

ABSTRACT

Industry-University Cooperation (IUC) is recognized as an effective model for technological innovation, helping small and medium-sized enterprises (SMEs) to seek opportunities from technological development and to achieve sustainable competitiveness. The study proposes an improved solution to resolve the 4W questions commonly faced by SMEs, including identifying cooperative topics, comparing technology competitors, evaluating cooperative universities, and selecting research teams. Firstly, the Latent Dirichlet Allocation model is applied to extract sub-technology topics in a specific technology domain, and the patent clustering is analyzed via K-means algorithm. Secondly, competitive and uncompetitive sub-technologies of the enterprise, as well as the geographical distribution of competitors in sub-technologies are analyzed. Thirdly, the study proposes a new topic matching index (TMI) to help determine the best cooperative universities. Finally, the cooperation network is visualized to help identify the specific core research teams for each specific technology. The research team conducted an experimental study using real patent data on a "Blockchain" technology enterprise to verify the proposed solution.

KEYWORDS: Patent analysis; clustering analysis; data visualization; partner selection; industry-university cooperation

1.Introduction

In the modern technology economy, the ability of an enterprise to improve its technological innovation is a key factor for its success. This is true for enterprises in technology-intensive industries, which cannot only rely solely on internal knowledge creation, but also on external information to achieve innovation (Chesbrough and Crowther 2006). The theory of learning organizations explains why enterprises need cooperation and provides a practical path to enable enterprises on sustainable development (Gil, Francisco, and Carrillo 2016). Therefore, over the last few decades, a large number of enterprises have established cooperation with external partners, including other enterprises and universities (Perkmann, Neely, and Walsh 2011).

Industry-University Cooperation (IUC) has received growing attentions (Hagedoorn, Link, and Vonortas 2000; Lehrer, Nell, and Garber 2009). With the help of IUC, knowledge flow and technology transfer can be realized between enterprises and universities, and technological innovation can be achieved through complementary advantages (Wang 2016). Furthermore, IUC can realize the application and publication of technological patents through joint R&D, promoting a win-win development for both sides (Mueller 2006). IUC unites market demand and common interests, and adopts various ways to carry out strategic cooperation activities such as the integration of upstream, midstream and downstream of technological innovation (Steinmo and Rasmussen 2018); besides, IUC allows academia for access to a broader area of resources and knowledge, bringing the practical knowledge from industries to universities, which, in return, mitigating risks in industrial activities (Lin 2017). Empirical research also finds that universities act as important means for private sectors to acquire technological opportunities (Klerovick et al. 1995).

However, these potential benefits are not always realized (Estrada et al. 2016), mainly because enterprises often find it challenging to collaborate effectively in the IUC relationship (Howells, Ramlogan, and Cheng 2012). When carrying out R&D collaboration, suitable selection on partners is a core factor affecting the relationship between enterprises and universities (Ireland, Hitt, and Vaidyanath 2002). SMEs in industrial sectors always face the following 4W questions when making partner selection: “What are the technology domains for cooperation,” “Who are the top competitors,” “Where to find the best cooperative universities,” and “Which research teams to collaborate with.”

To answer the 4W questions, this study uses patent data for quantitative analysis on the SMEs and proposes an improved systematic framework for partner selection. In doing so, the study makes three contributions to the current literature in the field of IUC. Firstly, the paper moves deeper into sub-technologies to unveil both competitive and uncompetitive technologies based on an enterprise’s patent information. For competitive technologies, enterprises can choose to deepen the relevant research and consolidate competitive advantages through IUC relationship. For uncompetitive technologies, enterprises can switch to IUC to save R&D costs avoid patent infringement. The first contribution provides a solution to the first question “What are the technology domains for cooperation.”

Secondly, SMEs often lack the resources needed to compete with large corporations in research activities. Thus, SMEs are more likely to rely on external sources of R&D to help sustain their research efforts (Kirchhoff et al. 2007) and overcome barriers of research innovation (Lee et al. 2010). Through the patent analysis, SMEs can better understand the current technology portfolios of the top competitors and their R&D directions. More importantly, the study proposes a new topic matching index (TMI) for partner selection in IUC at the level of sub-technology, addressing the other two questions “Who are the top competitors” and “Where to find the best cooperative universities”.

Finally, current studies on IUC relationships stop at the level of cooperative universities. The relationships of IUC, however, usually end up in the cooperation between the enterprise and a specific research team. Consequently, the study moves a level deeper to establish the inventor networks by identifying core research teams in universities, answering the question “Which research teams to collaborate with.”

The paper is structured as follows: Section 2 discusses the current literature and in the field of IUC; Section 3 explains the purpose of the study; Section 4 details the methodology and theoretical frameworks; Section 5 describes the experimental study; and Section 6 concludes the paper and discusses the main implications of the research.

2. Literature review

One of the directions in IUC research has attempted to identify and categorize factors that are playing significant roles in performance. Kyung et al. (2016) found that governmental funding in R&D projects have mediation and promotion effects in IUC. Szucs (2018) analyzed that in large-scale research subsidy programs, the number of project participants and university participation positively affect IUC performance. Wong et al. (2018) found that a successful IUC requires the following factors: review methods for the mechanisms, policy needs in regulations, and plan incentives in operations. Sjoo and Hellstrom (2019) identified seven main central factors stimulating collaborative innovation between industrials and universities, including resources, university organization, boundary-spanning functions, collaborative experience, culture, status centrality, and environmental context.

In addition to the analysis of influential factors, scholars also studied the evolution of the IUC network through the social network analysis (SNA). Zhang et al. (2016) depicted the evolution process and spatial distribution of the patent cooperation network using SNA. Xu et al. (2018) summarized characteristic features of spatial distribution and network profile on patent cooperation among 71 universities and enterprises. Lyu et al. (2019) outlined the cooperation trend of enterprises, universities and research institutions in China based on SNA and spatial analysis. Such research demonstrates the progressive development of IUC from a perspective of geographical distribution model and provides a reference portfolio for the government to better customize development policies.

Partner selection has been identified as one of the important determinants on the success of IUC (Ireland, Hitt, and Vaidyanath 2002). The current literature has revealed several criteria for selecting partners. Calvo et al. (1998) studied attitudes and decisions of research groups towards the collaboration with firms in R&D joint projects and suggested that gender can be a criterion for partner selection. Cappelli et al. (1998) pointed out factors such as enterprise size and corporate cultures affect the decision of enterprises to participate in IUC. In addition, Luo et al. (2019) indicated that work experience, monetary benefit, and HR policies are considered as important factors by university research teams to determine participation in IUC programs. The criteria discussed in the literature is helpful in partner selection during the process of IUC development; however, evaluating and comparing the overall attributes of potential partners is still difficult and time-consuming. Moreover, the lack of an efficient and comprehensive methodology for partner selection remains the limitation.

Realizing the limitation, some studies have proposed patent analysis for partner selection. Kang et al. (2019) proposed a systematic methodology that combines topic model and clustering algorithm to classify sub-technologies of a particular technology domain and identify the best university partners in each category. However, it only includes the highest-ranked universities in the candidate pool based on the number of patents without considering each individual enterprise's strength and weakness, nor their competitors. This is not promising to enterprises to formulate a clear development strategy, choose the right partner, or compete with the competitors. That is to say, questions of "What are the technology domains for cooperation" and "Who are the top competitors" are not answered.

Alternatively, Jeon et al. (2016) proposed a systematic approach for searching potential technology partners using patent information via similarity indicators. Song et al. (2016) proposed a patent portfolio-based approach for assessing potential R&D partners. Lee et al. (2016) suggested a Bayesian network model to select R&D collaboration partners for enterprises. Although these proposed methods in the literature are applicable for practitioners to implement, the methods do not include a ranking system, nor do they reach the specific level of research teams. The existing approaches cannot answer questions such as "Where to find the best cooperative universities" and "Which research teams to collaborate with." In other words, an advanced methodology is still needed to help with identifying enterprises' competitive and uncompetitive sub-technologies, their top competitors, the best cooperative universities, and specific research teams for each unique field.

3.Purpose of study

Therefore, the purpose of this study is to propose an improved framework to help SMEs answer the following 4W questions when making partner selection: "What are the technology domains for cooperation," "Who are the top competitors," "Where to find the best cooperative universities," and

"Which research teams to collaborate with." The study also proposed a new mathematical function to calculate TMI to aid partner selection.

At the level of sub-technologies, competitive and uncompetitive domains can be divided according to the number of patents and enterprise ranking, which plays an important role in guiding the enterprise to determine the technology domain for cooperation. Furthermore, the exploration of technology competitors in different types of sub-technologies can help the enterprises strategize targeted cooperative development. Moreover, according to the patent portfolio of top competitors, the universities with highest technology matching index can be located and core research teams and researchers of relevant research fields can be identified. The proposed improved methodology in the study can be of great significance for enterprises to accelerate the improvement of technology efficiency through IUC.

4.Methodology

4.1. Research framework

As shown in Figure 1, the complete approach framework consists of four steps: data collection & data pre-processing, topic identification, patent clustering, and partner exploration. The first step is to collect patent documents related to the research domain and to prepare the collected documents for analysis via a topic-modeling algorithm. The second step is to identify the main topics of research patents. The third step is to cluster the patent document into every sub-technology through a combined LDA model and K-means algorithm. In the final step, cooperative topics, technology competitors, cooperative universities and research teams are further explored in each technology topic.

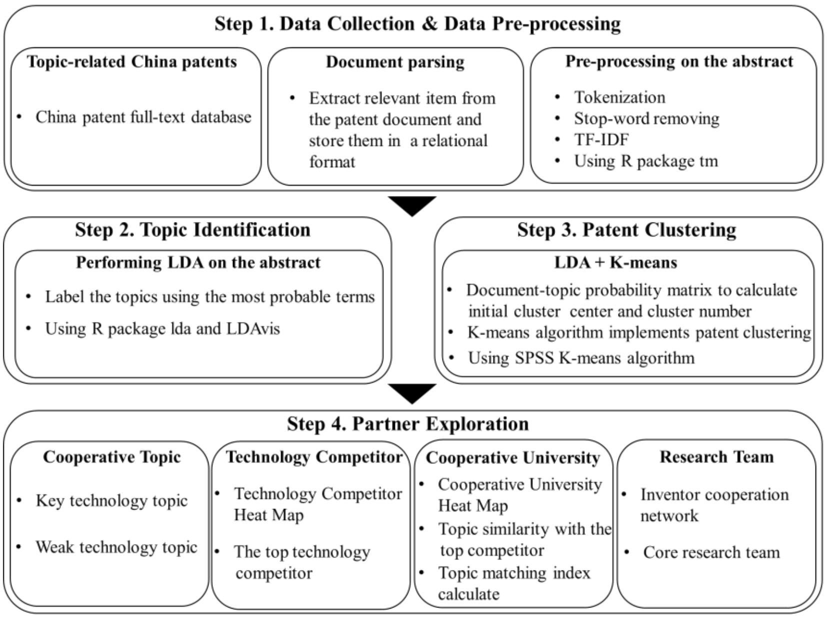


Figure 1. Overall process of the proposed approach

4.2. Step 1: Data collection and data pre-processing

The patent documents related to the research domain can be collected from patent database. Because the collected patent documents are in an unstructured text format, they should be pre-processed and transformed into a structured format for further analysis. Firstly, using the document parsing techniques, the fields of title, abstract, assignees, inventors are extracted from documents. In order for the abstract information to be used in the LDA model for identifying patent topics, the abstract in a free-text format requires natural language processing techniques, including professional dictionary construction, tokenization, stop-word removing, and vector-space representation.

With the help of a professional dictionary about technological nouns, texts in abstracts can be cleaned through tokenization, a process that converts text streams into processing units known as tokens, which are character strings without delimiters. Tokens are then filtered using stop words, which are common but unnecessary words for the patent analysis. Finally, the patent abstract is represented as a term frequency-inverse document frequency (TF-IDF) vector. The TF-IDF weighting scheme has been widely used to measure the importance of a given term in a document (Zhang, Yoshida, and Tang 2011). By incorporating the TF-IDF vector for every patent abstract, the TF-IDF matrix is generated, which is the input of the LDA in the next step.

4.3. Step 2: Topic identification

In order to identify multiple sub-technology topics covered by a particular technology domain, we chose the LDA model for implementation. The LDA model presumes that words are generated from a mixture of topics, and each topic is a polynomial distribution on a fixed vocabulary (Blei, Ng, and Jordan 2003). These topics are shared by all documents in the collection, and each document has a specific topic probability, which is sampled from the Dirichlet distribution. As a generative model, its structure model is complete and clear, and it uses an efficient probability inference algorithm to process large-scale data. Thus, LDA has become a widely used topic recognition model (Song and Suh 2019). At this step, the R package LDAvis is used to present the topic similarity. In LDAvis, multidimensional scale analysis is used to clarify and visualize the similarity between each topic. Through adjusting the number of topics, alpha value, and beta value, the corresponding number of topics can reach optimal when the topics are independent of each other. The LDA on patent abstracts are expected to generate two outputs: the probability that each patent document is related to each topic and the term distribution over each topic. The document-topic probability matrix is used for text clustering. The top n words that are most likely to occur in each topic are used to label the identified topics in the technology domain-related patents.

4.4. Step 3: Patent clustering

After identifying the topics covered by technology domains, each patent document needs to be classified into extracted sub-technology by applying LDA to topic identification. The most used clustering algorithms, K-means clustering, is one of the top ten classic algorithms for data mining presented by J. MacQueen in 1967 and has been widely used in the research of text clustering (Kelaiaia and Merouani 2016). However, K-means can be very sensitive to the selected initial centroids (Velmurugan, 2012). In the application of the K-means algorithm, there is a phenomenon that the initial clustering center is randomly selected, and the clustering result tends to fall into local optimal, resulting in poor clustering results. How to obtain the appropriate initial clustering center and keep the accuracy of the algorithm while ensuring the stability of the result are particularly important for the realization of patent clustering in this paper.

Therefore, the paper proposes to improve this situation by merging LDA and K-means, theoretically ensuring that the selected initial clustering center is determined based on probability. Considering that the technology problem solved by an invention patent has the characteristics of singleness and in-depth, it is reasonable that a patent only appears in one sub-technology topic cluster during patent clustering in this paper. The algorithm is described as follows:

(1) The LDA model is used to extract the topics of N patent documents and generated a document-topic matrix (D-T matrix) with K topics. In the D-T matrix, each row represents the matching degree of K sub-technology topics in a patent document, and each column represents the matching degree of a sub-technology topic in N patent documents.

(2) For each topic, the research team calculates the average matching degree M_i ($0< i<=K$) of the topic to N patent documents. Furthermore, one patent document with a matching degree greater than M_i is regarded as matching document, regarded as MP, while the number of matching documents is regarded as MD_j ($0< j<=K$). Through the above process, the patent document with high relevance to each sub-technology topic is screened out.

(3) The LDA optimal topic number is taken as the clustering number of K-means algorithm, which is regarded as W, and the average matching degree of matching documents in each topic is calculated to generate the initial clustering center as ($C_1, C_2, C_3 \ldots C_w$).

(4) The D-T matrix generated by the LDA model is imported into SPSS, and the K-means algorithm is used to perform patent clustering. The number of clusters and initial cluster centers is set. The cluster member division is yielded to realize the topic identification of patent documents.

4.5. Step 4: Partner exploration

After determining technology topic of each patent through patent clustering, the panoramic analysis reveals the self-development status of enterprises who intend to conduct joint research. Based on this, the analysis continues to identify competitors under different sub-technology topics and shows the best cooperation universities and their core research teams.

4.5.1. Cooperative topic

Under each sub-technology topic, different enterprises are ranked according to the number of distributed patents. The target enterprise can clarify the competitive and uncompetitive technology topics by aggregating the distribution of existing patents and rankings. Based on the results, the target enterprise can develop a strategy and define the cooperative research topics.

4.5.2. Technology competitor

The number of different companies' patents under each sub-technology topic is counted to determine the technology competitors for the target enterprise. The heat map can also be drawn to better visualize the distribution of technology competitors. The data of the heat map include the latitude, longitude, and the number of companies' patents. The transformation of the company name and latitude/longitude is achieved by using the map (Google, Baidu, or Leaflet) API interface.

4.5.3. Cooperative university

The information about universities' patents under each sub-technology topic is visualized on the heat map to present the distribution of cooperative universities. Firstly, the top technology competitors of the target enterprise under each sub-technology are identified. Secondly, the average patent topic similarity between the university and the top technology competitor under each sub-technology is calculated separately through Equation (1) - (3) shown below. Among them, $Sim(UP_A, IP_B)$ represents the cosine similarity between the patent A of the university and the patent B of the top technology competitor of the target enterprise. $avg\text{-}Sim(IU_A)$ represents the average topic similarity between university's patent A and top competitor's all patents, and NIP represents the number of top competitor patents. $avg\text{-}Sim(IU_{All})$ represents the average topic similarity of university's all patents and top competitor's all patents, and *NUP* represents the number of university patents. Finally, the topic matching index (TMI), proposed by the study, is calculated in Equation (4) by multiplying the number of university patents ($Univ_{num}$) and the average topic similarity, where the best cooperative university under each sub-technology can be determined.

$$Sim(UP_A, IP_B) = \frac{\sum_{i=1}^{n}(M_{Ai} * M_{Bi})}{\sum_{i=1}^{n}(\sqrt{(M_{Ai})^2} * \sqrt{(M_{Bi})^2})} \qquad \text{eq. (1)}$$

$$\text{avg}\_Sim(IU_A) = \frac{\sum_{j=1}^{NIP} Sim(UP_A, IP_j)}{NIP} \qquad \text{eq. (2)}$$

$$\text{avg}\_Sim(IU_{All}) = \frac{\sum_{z=1}^{NUP} Sim(IU_z)}{NUP} \qquad \text{eq. (3)}$$

$$TMI = Univ_{num} * avg\_Sim(IU_{All}) \qquad \text{eq. (4)}$$

4.5.4. Research team

After determining the best cooperative university under each sub-technology, the inventor cooperation network is drawn via the computer application Vosviewer. The target enterprise can determine the core research team based on the intensity of inventor cooperation. The node size in the network indicates the number of patents applied by the inventor, and the core inventors are determined according to the node size.

5. Experimental study

In order to verify the applicability of the proposed methodology in Section 4, an experimental study is conducted using actual patent data. The technology domain to be tested is "Blockchain". Specific information regarding the data is presented in Section 5.1. Section 5.2 - 5.6 present the experimental results according to the procedure of the proposed methodology.

5.1. Data description

The research team performed the keyword search and collected 8,799 documents from Chinese Patent Full-text Database (CPFD) without time limiters. Among the collected patent documents there were 1639 companies and 152 universities. These patent documents were parsed to extract the fields from relevant items, including the title, abstract, assignees, and inventors. All patents' abstracts were pre-processed by tokenization, stop-word removing, and vector-space representation. These pre-processing on the abstracts were performed with an R-script. Specifically, tokenization and stop-word removing were performed using the tm package of R. In order to reduce the dimension, the research team removed the words appearing fewer than five times in the entire corpus. In addition, the weights of the TF-IDF were calculated for every remaining term in order to exclude words that were in the lower 15% quintile. With the remaining 1891 terms, consequently, an 8799

× 1891 document-term matrix was obtained.

5.2. Topic identification and patent clustering

The LDA topic model was implemented using the lda package of R. In addition, the appropriate number of topics is determined using the LDAvis package of R. Figure 2 presents the optimal number of topics using parameters such as the number of topics being 10, the alpha value 0.2, and the beta value 0.05. Based on the visualization results of topic similarity presented by LDAvis, each topic is independent of each other, indicating that the similarity between topics is small and the number of topics is optimal. Thus, by applying the LDA model using 10 topics, the research team obtained the vectors consisting of the appearance probability of each topic corresponding to each document. That is, the probability vector corresponding to each patent document is positioned in the topic-based vector space of 10 dimensions, labeling the 10 identified topics in the blockchain-related patents.

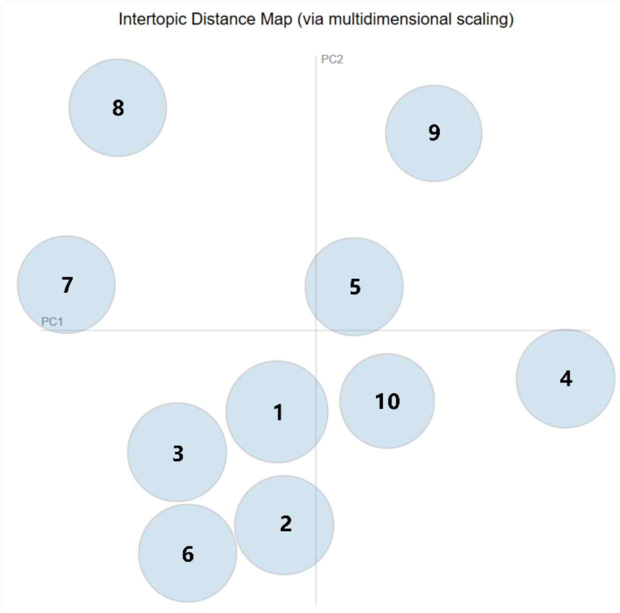


Figure 2. The optimal number of topics in LDA model

The 8799 × 10 document-topic probability matrix was obtained by LDA. According to the patent clustering process described in Section 4, the initial cluster center results are calculated as (0.355433559, 0.373444966, 0.337785209, 0.342380289, 0.313298697, 0.335688124, 0.357835905, 0.329235873, 0.369186285, 0.354366802). Table 1 shows topic definitions and numbers of patents under each sub-technology.

Table 1. Result of topic definition considering their frequent terms

Topic No.	Topic Definition	Number of Patents Included
1	Supply Chain Management	949
2	Consensus Method	968
3	Internet of Things Application	917
4	Authentication	996
5	Smart Contract	830
6	Privacy Protection	992
7	Safety Monitoring	830

8	Blockchain Management System	748
9	Digital Asset	664
10	Data Storage	905

5.3. Cooperative topic

The research team selected the patents held by Sinochain Technology Co., Ltd. (Sinochain-Tech), which focuses on the R&D of blockchain infrastructure and innovation. The enterprise currently has 42 employees, belonging to the category of SMEs. Table 2 shows the number of patents and enterprise ranking included in each different topic clustering for Sinochain-Tech.

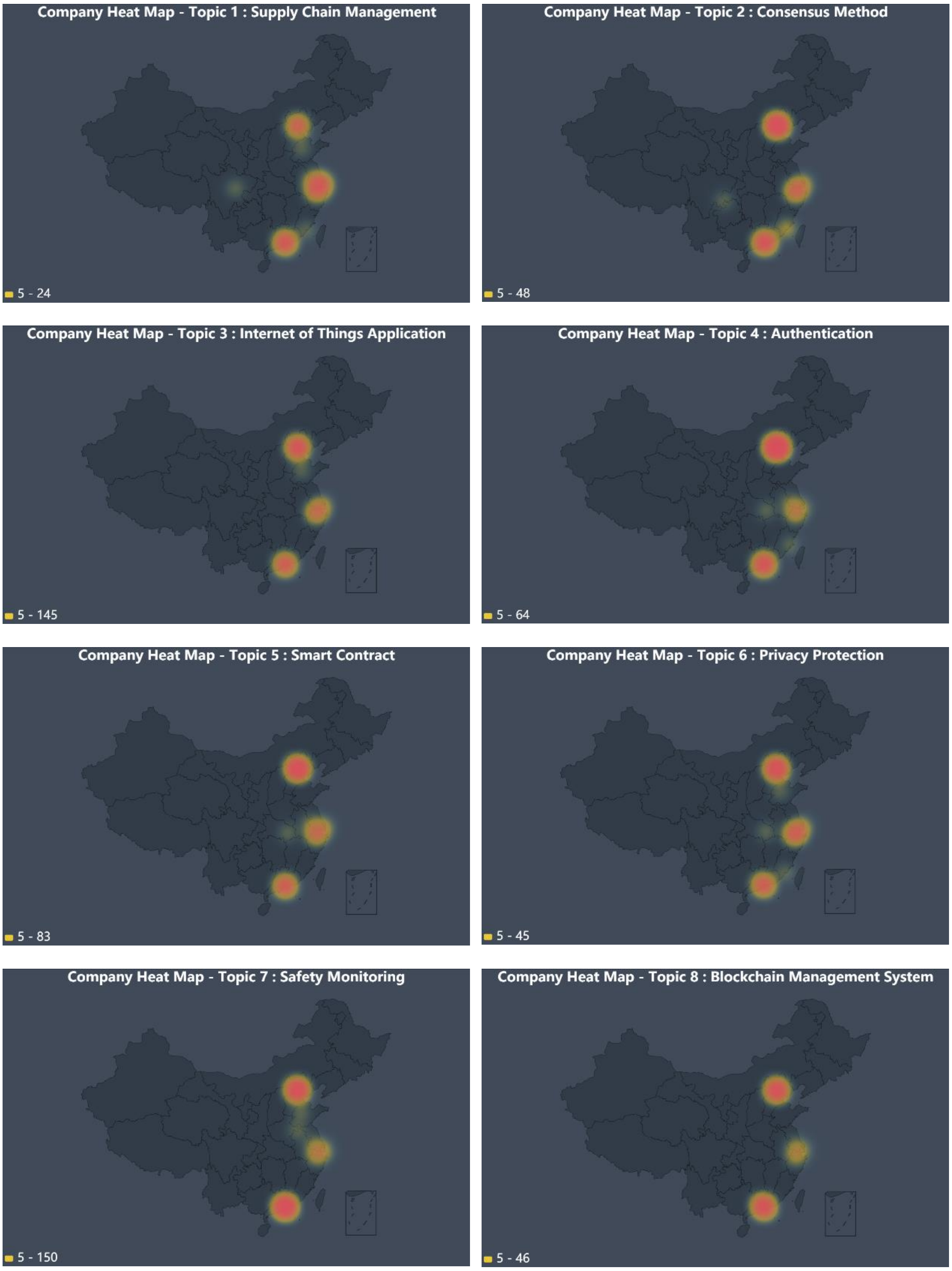
Table 2. Result of Sinochain-Tech's patents' number and ranking in each different topic

Topic No.	Number of Patents Included	Ranking
1 Supply Chain Management	13	5
2 Consensus Method	8	14
3 Internet of Things Application	20	4
4 Authentication	9	12
5 Smart Contract	4	14
6 Privacy Protection	8	9
7 Safety Monitoring	10	10
8 Blockchain Management System	6	14
9 Digital Asset	8	11
10 Data Storage	8	8

Impacted by development strategy, enterprise size, and R&D capacity, SMEs' research competencies cannot be equally covering every sub-technology. Therefore, there will be differences in the distribution of patents' number under sub-technologies. The competitive technology topics and uncompetitive technology topics can be scaled out by setting thresholds for the patents' number and the enterprise ranking based on the specific technology domain and the specific situation of the enterprise. In this case, according to the patent distribution of Sinochain-Tech, the threshold of patents' number and enterprise ranking in the sub-technology were set to 10. Table 2 shows that the first, the third, and the seventh can be considered as its competitive technologies for Sinochain-Tech; whereas the remaining are deemed as uncompetitive ones. The Sinochain-Tech should seek IUC according to the fit between corporate development strategy and benefits. Seeking cooperation in competitive technologies can further expand technological advantages and research boundaries. Seeking cooperation in uncompetitive technologies can help open a technological vision, develop patent portfolio through university's patent transfer, save R&D costs, and avoid the infringement of competitors in the future development process (Lee and Kang 2017). The question "what are the technology domains for cooperation" is answered in this step.

5.4. Technology competitor

After Sinochain-Tech's competitive and uncompetitive technology topics were identified, statistical analytics was also performed on other companies under each sub-technology. Companies with more than five patents were selected and visualized in Figure 3 to directly present the distribution of Sinochain-Tech's competitors. The maximum value on the map represents the number of patents owned the top technology competitor. In addition, Sinochain-Tech's top technology competitors under each sub-technology are presented in Table 3.



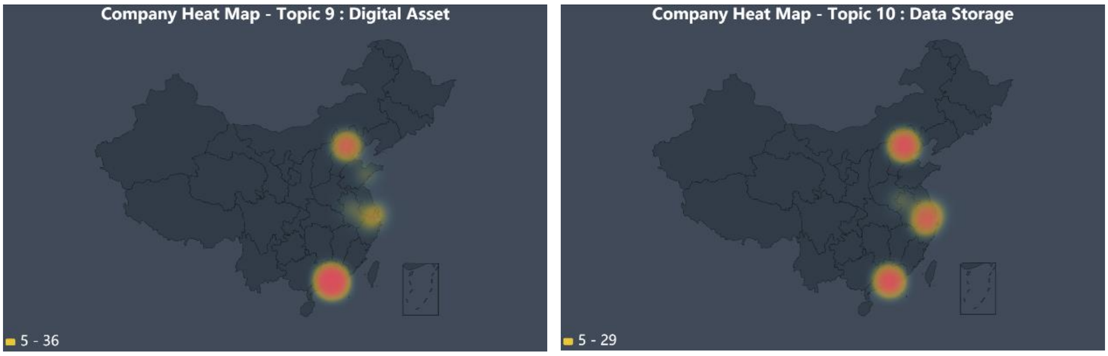


Figure 3. Heat map of technology competitors' distribution

Figure 3 shows that under the 10 sub-technology topics, Sinochain-Tech's technology competitors are mainly distributed in three regions of China: Beijing-Tianjin-Hebei district, Yangtze River Delta district, and Pearl River Delta district. These three regions are also the primary locations representing active scientific and technological innovations in China.

Table 3 shows Sinochain-Tech's top technology competitors under each sub-technology. Alibaba, Rechain, and Tencent have appeared twice, and they are also the three companies with the largest number of patent applications in China. Alibaba and Tencent are the most representative Internet giants, while Rechain is a unique Chinese company in the blockchain field. Fuzamei, Dianrong, Qulian, and Blockchain Holdings are emerging blockchain R&D companies. Through the above analysis, the geographical distribution, patent portfolio, and strength of the competitors of Sinochain-Tech under the sub-technologies were profiled, which resolves the question "Who are the top competitors."

Table 3. Result of the top technology competitors

Topic No.	The Top Technology Competitor	Number of Patents Included
1 Supply Chain Management	Hangzhou Fuzamei Technology Co., Ltd.	24
2 Consensus Method	Tencent Tech Shenzhen Co., Ltd.	48
3 Internet of Things Application	Alibaba Group Holding Ltd.	145
4 Authentication	Tencent Tech Shenzhen Co., Ltd.	64
5 Smart Contract	Beijing Rechain Technology Co., Ltd.	83
6 Privacy Protection	Blockchain Holdings Ltd.	45
7 Safety Monitoring	Alibaba Group Holding Ltd.	150
8 Blockchain Management System	Beijing Rechain Technology Co., Ltd.	46
9 Digital Asset	Shanghai Dianrong Information Technology Co., Ltd.	36
10 Data Storage	Hangzhou Qulian Technology Co., Ltd.	29

5.5. Cooperative university

In this part of study, the question "Where to find the best cooperative universities" will be resolved. Universities with more than two invention patents were selected to draw the heat map, as depicted in Figure 4, to show the distribution of Sinochain-Tech's possible cooperative universities. The maximum values on the map represent the largest number of patents owned by universities.

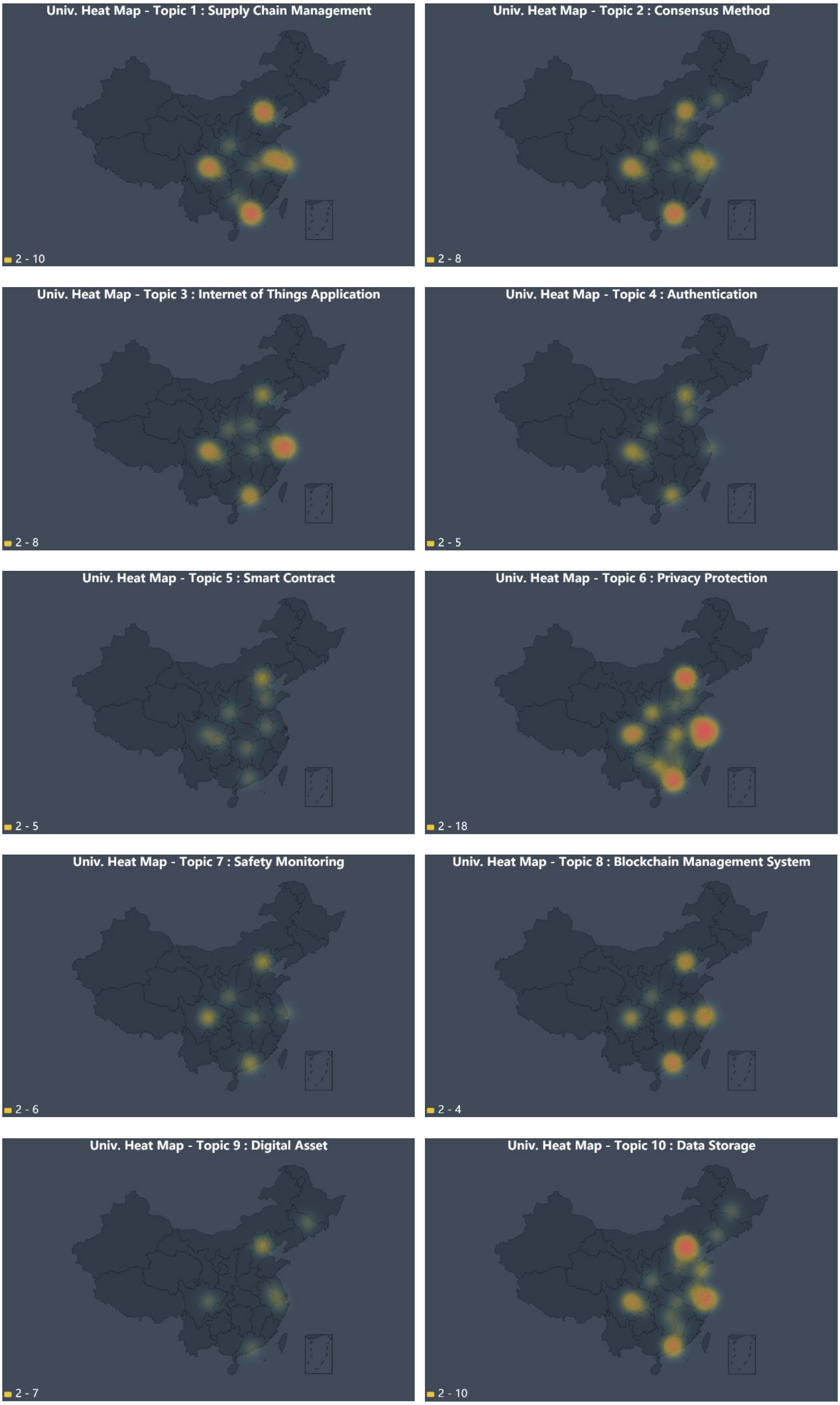

Univ. Heat Map - Topic 1 : Supply Chain Management
2 - 10
Univ. Heat Map - Topic 2 : Consensus Method
2 - 8
Univ. Heat Map - Topic 3 : Internet of Things Application
2 - 8
Univ. Heat Map - Topic 4 : Authentication
2 - 5
Univ. Heat Map - Topic 5 : Smart Contract
2 - 5
Univ. Heat Map - Topic 6 : Privacy Protection
2 - 18
Univ. Heat Map - Topic 7 : Safety Monitoring
2 - 6
Univ. Heat Map - Topic 8 : Blockchain Management System
2 - 4
Univ. Heat Map - Topic 9 : Digital Asset
2 - 7
Univ. Heat Map - Topic 10 : Data Storage
2 - 10

Figure 4. Heat map of corporative universities' distribution

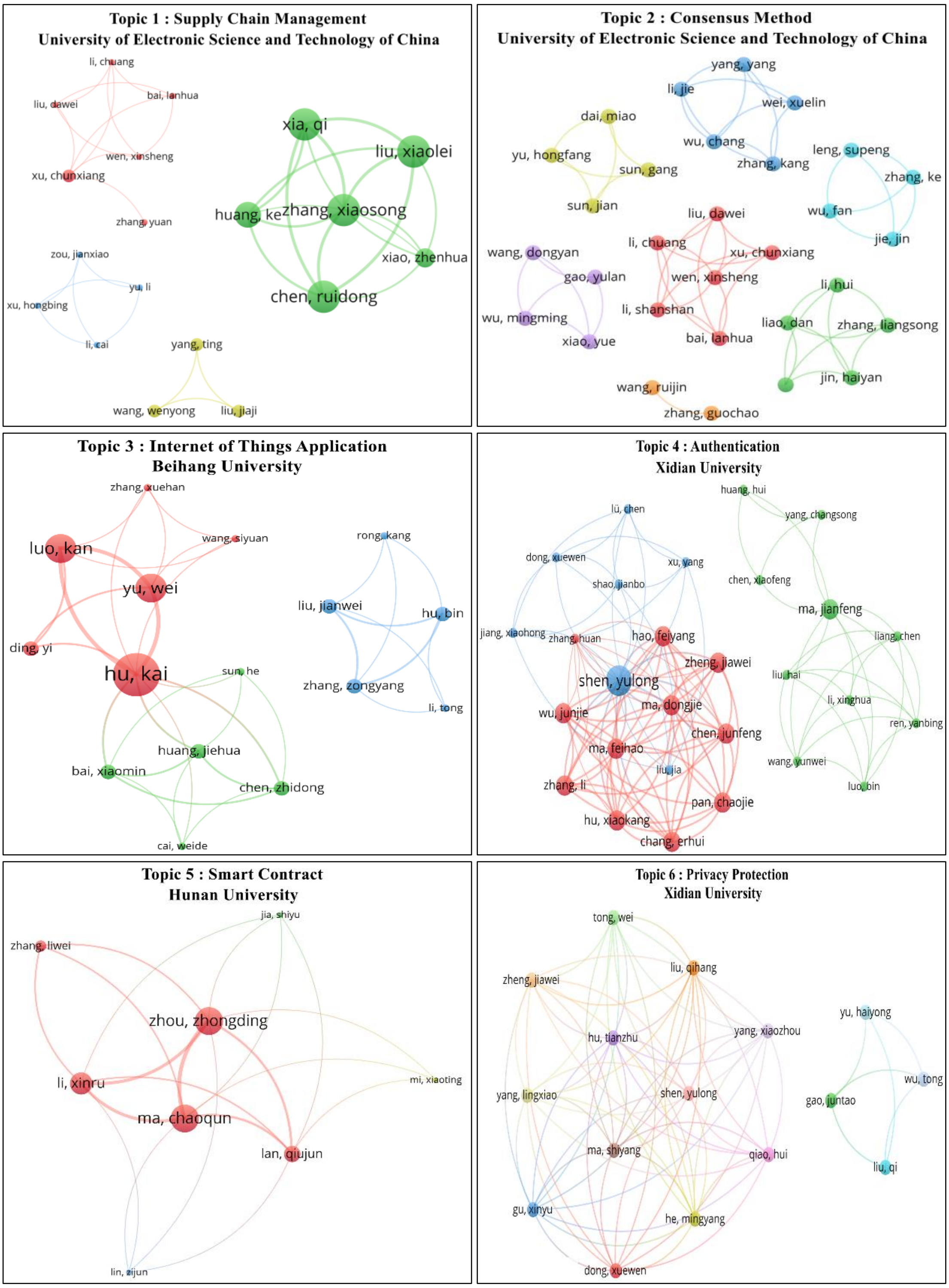
Figure 4 suggests that under the 10 sub-technology topics, the possible cooperative universities for Sinochain-Tech are mainly distributed in the following cities: Beijing, Nanjing, Guangzhou, Chengdu, and Wuhan. These cities are the primary locations of Chinese top universities with strong scientific research capabilities. According to the selection method of the best cooperative universities described in Section 4, the ranking was conducted by the topic matching index. Table 4 profiles the status of cooperative universities with Sinochain-Tech's competitors under each sub-technology. Universities are ranked based on TMI value, the university with the greatest TMI value can be regarded as the Sinochain-Tech's best R&D partner for each sub-technology.

Table 4. Result of the best cooperative universities with competitors

Topic No.	The Top Technology Competitor	Cooperative University (Number of Patents)	Topic Similarity	Topic Matching Index
1	Hangzhou Fuzamei Technology Co., Ltd. (24)	University of Electronic Science and Technology of China（10）	0.8302	8.302
2	Tencent Tech Shenzhen Co., Ltd. (48)	University of Electronic Science and Technology of China（8）	0.8049	6.4392
3	Alibaba Group Holding Ltd. (145)	Beihang University（8）	0.8233	6.5864
4	Tencent Tech Shenzhen Co., Ltd.(60)	Xidian University（5）	0.8566	4.283
5	Beijing Rechain Technology co., Ltd. (80)	Hunan University（5）	0.7740	3.87
6	Blockchain holdings Ltd. (44)	Xidian University（18）	0.8496	15.2928
7	Alibaba Group Holding Ltd. (150)	China University of Geosciences，Wuhan（6）	0.8980	5.388
8	Beijing Rechain Technology Co., Ltd. (46)	Tsinghua University（4）	0.8460	3.384
9	Shanghai Dianrong Information Technology Co., Ltd. (36)	Nanjing University of Posts and Telecommunications（7）	0.8558	5.9906
10	Hangzhou Qulian Technology Co., Ltd. (29)	Guangdong University of Technology（10）	0.8069	8.069

5.6. Research team

At the final step, the research team used Vosviewer software to visualize the inventor cooperation network to resolve the question "Which core research team to collaborate with" (Figure 5). Based on its own competitive and uncompetitive technologies and strategies, Sinochain-Tech can select the core research teams with large nodes and close cooperative relationships among the best cooperative universities. For example, on the fifth topic "Smart Contract," Sinochain-Tech has only 4 patents, ranking 14th in the companies. To seek university cooperation on this topic, Sinochain-Tech may consider the research team of Hunan University with Zhou and Ma as the core research partner.



Topic 1 : Supply Chain Management
University of Electronic Science and Technology of China
li, chuang
liu, dawei
bai, lanhua
wen, xinsheng
xu, chunxiang
zhang, yuan
xia, qi
liu, xiaolei
huang, ke
zhang, xiaosong
xiao, zhenhua
chen, ruidong
zou, jianxiao
yu, li
xu, hongbing
li, cai
yang, ting
wang, wenyong
liu, jiaji
Topic 2 : Consensus Method
University of Electronic Science and Technology of China
yang, yang
li, jie
wei, xuelin
dai, miao
wu, chang
yu, hongfang
zhang, kang
sun, gang
leng, supeng
zhang, ke
sun, jian
liu, dawei
wu, fan
jie, jin
wang, dongyan
li, chuang
xu, chunxiang
gao, yulan
wen, xinsheng
li, shanshan
li, hui
wu, mingming
liao, dan
zhang, liangsong
xiao, yue
bai, lanhua
jin, haiyan
wang, ruijin
zhang, guochao
Topic 3 : Internet of Things Application
Beihang University
zhang, xuehan
luo, kan
wang, siyuan
rong, kang
yu, wei
liu, jianwei
hu, bin
ding, yi
hu, kai
sun, he
zhang, zongyang
li, tong
huang, jiehua
bai, xiaomin
chen, zhidong
cai, weide
Topic 4 : Authentication
Xidian University
huang, hui
lü, chen
yang, changsong
dong, xuewen
xu, yang
shao, jianbo
chen, xiaofeng
ma, jianfeng
jiang, xiaohong
zhang, huan
hao, feiyang
liang, chen
zheng, jiawei
liu, hai
shen, yulong
li, xinghua
ma, dongjie
wu, junjie
chen, junfeng
ren, yanbing
ma, feihao
wang, yunwei
liu, jia
zhang, li
luo, bin
pan, chaojie
hu, xiaokang
chang, erhui
Topic 5 : Smart Contract
Hunan University
jia, shiyu
zhang, liwei
zhou, zhongding
li, xinru
mi, xiaoting
ma, chaoqun
lan, qiujun
lin, zijun
Topic 6 : Privacy Protection
Xidian University
tong, wei
liu, qihang
zheng, jiawei
yu, haiyong
yang, xiaozhou
hu, tianzhu
wu, tong
yang, lingxiao
shen, yulong
gao, juntao
ma, shiyang
qiao, hui
liu, qi
gu, xinyu
he, mingyang
dong, xuewen

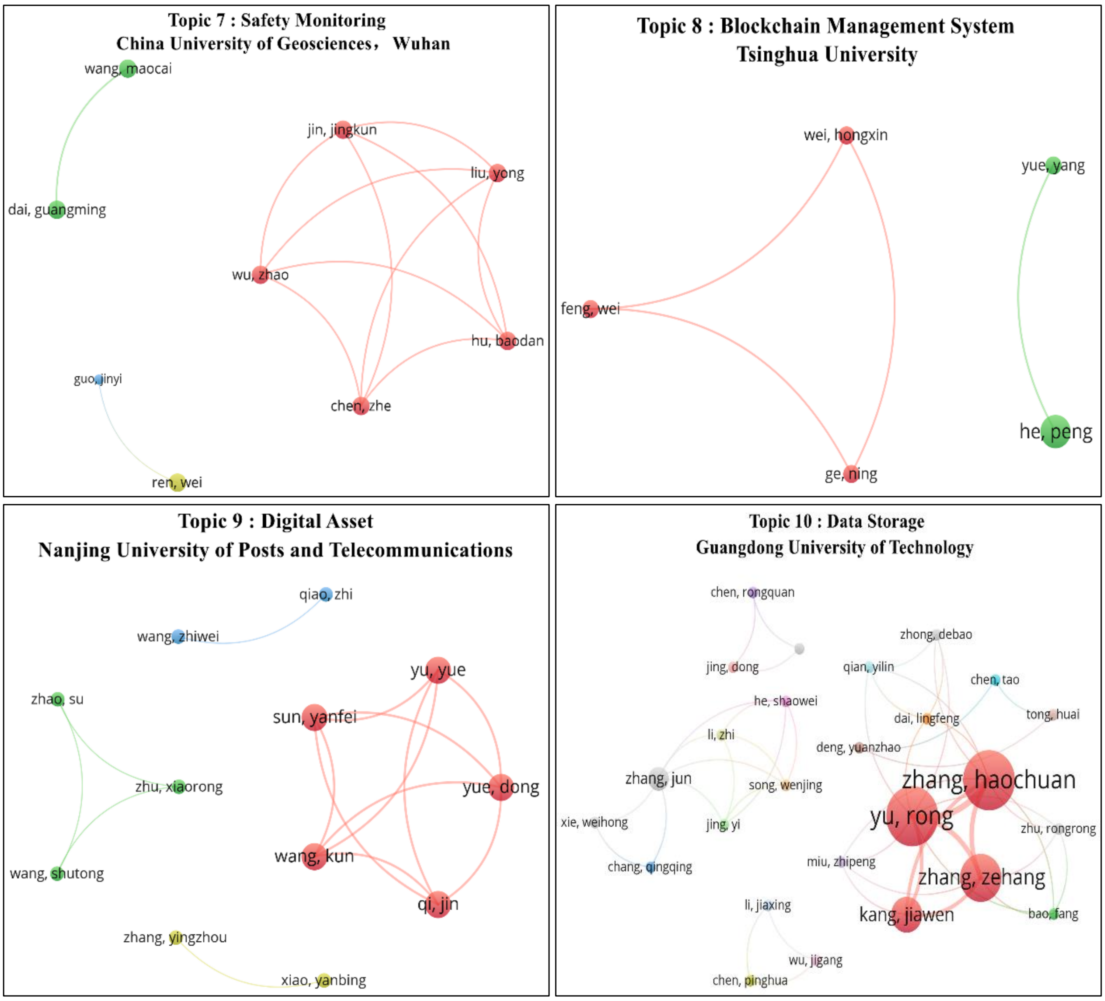


Figure 5. Visualization of inventor cooperation network

6. Discussion and conclusion

SMEs usually need to seek external cooperation to improve their technological capabilities so as to survive the competition with industrial giants. IUC mode has been proved to be an effective way to strengthen SMEs' technological capabilities and suitable partner selection is one of the key factors for the success of IUC. When SMEs are in the process of evaluating and selecting potential partners, they often face the unresolved 4W questions, "What are the technology domains for cooperation," "Who are the top competitors," "Where to find the best cooperative universities," and "Which research teams to collaborate with." The study proposes an effective improved framework to answer the 4W questions.

Current research has found that usually there are several subordinate fields of technologies in a technology domain, while the research capability and concentration of an enterprise vary in those sub-domains (Kim et al. 2016; Lee and Kang 2017; Kang, et al. 2019). Based on the findings in the literature, the research team firstly proposed to use the LDA model to extract topics from a specific technology domain and identify sub-technologies. In order to effectively identify the technology topic of each patent, the LDA model was combined with the K-means algorithm. The document-topic probability matrix extracted from the LDA model was used to determine the initial clustering center of the K-means algorithm, which achieved the classification of each patent document within each sub-technology topic. In this way, it is beneficial for enterprises to understand their own technology profiles in the field and resolve the question “What are the technology domains for cooperation.”

Secondly, through the statistical analysis of the number of companies’ patents under every sub-technology, technological competitors were visualized geographically on the heat map, which clearly show the overall picture of innovation in the technology domain and serves as a guidance for further exploration at the regional level. Moreover, it presents the top technology competitors under each sub-technology, which can clearly outline the patent portfolio of competitors for the enterprises and help them answer the question “Who are the top technology competitor.”

Thirdly, the research team performed statistical analysis on universities’ patents in each sub-technology and visualized the distribution of universities. Compared to previous research that only chose the R&D partners with the largest number of patents or the highest technology similarity in the technology domain of the target enterprise (Jeon et al. 2011; Park et al. 2015; Kang et al. 2019) the paper puts forward a new topic matching index to evaluate the best cooperative universities through the patent topic similarity in the evaluation index, resolving the question of “Where to find the best cooperative universities.”

Finally, previous research only identified partners at the organization level without locating specific research teams (Jeon et al. 2011; Park et al. 2015; Kang et al. 2019), providing insufficient information to effectively guide the target enterprise. Therefore, after determining the best cooperative universities in each sub-technology, the paper continues to visualize the university inventor cooperation network via Vosviewer, through which the target enterprise can select research partners at the level within universities and answer the final question “Which research team to collaborate with.”

In conclusion, there are a few prior studies that used patent data to search potential partners in IUC. The study extends and deepens the current research by dissecting technology sub-fields, identifying technology competitors, matching cooperative universities via improved algorithm, and mining research teams to answer the 4W questions commonly faced by SMEs in the process of IUC. In order to help identify core research teams in universities, the study also proposes a new index called TMI to aid the process and decision making of partner selection, which can be regarded as one unique contribution to the literature. The study then applies the improved methodology to an experimental study to illustrate how the approach is implemented. Although the experimental study is based only on Chinese enterprise data and invention patents from China, this methodology can be applied by SMEs in different countries. The patent database company could also develop the system module accordingly, as the function supplement of the patent database, and further play the decision-making role of the patent analysis assisting the SMEs to build the cooperation development strategy.

The study has certain limitations that require future research to rectify. The method provided in this article requires manual modeling and calculating, which appears to be somewhat inefficient. Therefore, how to implement the methodology logic of this article to a developed system is more appealing in the future studies. In the following research, scholars can also apply and test this methodology to other technology domains or import data from other countries around the world.

References

Alfonso, J. G., and Francisco, J. C. 2016. "Knowledge Transfer and the Learning Process in Spanish Wineries." *Knowledge Management Research & Practice* 14: 60-68.

Blei, D.M., Ng, A.Y., and Jordan, M.I. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3: 993-1022.

Calvo, N., Fernandez-Lopez, S., and Rodeiro-Pazos, D. 2019. "Is University-industry Collaboration Biased by Sex Criteria?" *Knowledge Management Research and Practice* 17: 408-420.

Cappelli, P., Shapiro, D., and Shumanis, N. 1998. "Employer Participation in School-to-Work Programs." *Annals of The American Academy of Political and Social Science* 559: 109-124.

Chesbrough, H., and Crowther, A. K. 2006. "Beyond High Tech: Early Adopters of Open Innovation in Other Industries." *R& D Management* 36 (3): 229-236.

Estrada, I., Faems, D., Martin Cruz, N., and Perez Santa, P. 2016. "The Role of Interpartner Dissimilarities in Industry-University Alliances: Insights from A Comparative Case Study." *Research Policy* 45: 2008-2022.

Geisler, E. 2003. "Benchmarking Interorganisational Technology Cooperation: The Link Between Infrastructure and Sustained Performance." *International Journal of Technology Management* 25: 675-702.

Hagedoorn, J., Link, A. N., and Vonortas, N. S. 2000. "Research partnerships." *Research Policy* 29: 567-586.

Howells, J., Ramlogan, R., and Cheng, S. L. 2012. "Innovation and University Collaboration: Paradox and Complexity within the Knowledge Economy." *Cambridge Journal of Economics* 36 (3): 703-721.

Ireland, R.D., Hitt, M.A., and Vaidyanath, D. 2002. "Alliance Management as a Source of Competitive Advantage." *Journal of Management* 28 (3): 413-446.

Jeon, J., Lee, C., and Park, Y. 2011. "How to Use Patent Information to Search Potential Technology Partners in Open Innovation." *Journal of Intellectual Property Rights* 16: 385-393.

Kang, J., Lee, J., Jang, D., and Park, S. 2019. "A Methodology of Partner Selection for Sustainable Industry-University Cooperation Based on LDA Topic Model." *Sustainability* 11: 3478.

Kelaiaia, A., and Merouani, H. 2016. "Clustering with Probabilistic Topic Models on Arabic Texts: A Comparative Study of LDA and K-means." *The International Arab Journal of Information Technology* 13: 332-338.

Kim, J., Lee, J., Kim, G., Park, S., and Jang, D. 2016. "A Hybrid Method of Analyzing Patents for Sustainable Technology Management in Humanoid Robot Industry." *Sustainability* 8 (5): 474.

Kirchhoff, B.A., Newbert, S.L., Hasan, I., and Armington, C. 2007. "The Influence of University R&D Expenditures on New Business Formations and Employment Growth." *Entrepreneurship Theory and Practice* 31: 543-559.

Klerovick, A.K., Levin, R.C., Nelson, R.R., and Winter, S.G. 1995. "On the Sources and Significance of Interindustry Differences in Technological Opportunities." *Research Policy* 24: 185-205.

Kyung, L.H., Duk, Y.H., Jeoung, K.S., and Kyo, S.Y. 2016. "Factors Affecting University-Industry Cooperation Performance Study of the Mediating Effects of Government and Enterprise Support." *Journal of Science and Technology Policy Management* 7: 233-254.

Lee, H., and Kang, P. 2018. "Identifying Core Topics in Technology and Innovation Management

Studies: A Topic Model Approach.” *The Journal of Technology Transfer* 43 (5): 1291-1317.

Lee, K., Park, I., and Yoon, B. 2016. “An Approach for R&D Partner Selection in Alliances between Large Companies, and Small and Medium Enterprises (SMEs): Application of Bayesian Network and Patent Analysis.” *Sustainability* 8: 117.

Lee, S., Park, G., Yoon, B., and Park, J. 2010. “Open Innovation in SMEs - An Intermediated Network Model.” *Research Policy* 29: 290-300.

Lehrer, M., Nell, P., and Garber, L. 2009. “A National System View of University Entrepreneurialism: Inferences from Comparison of the German and US Experience.” *Research Policy* 38 (2): 268-280.

Lin, J.Y. 2017. “Balancing Industry Collaboration and Academic Innovation: The Contingent Role of Collaboration-Specific Attributes.” *Technological Forecasting and Social Change* 123: 216-228.

Luo, J.M., and Lam, C.F. 2019. “Qualitative Analysis of Satisfying and Dissatisfying Factors in A University-Industry Cooperation Programme.” *Education Sciences* 9 (1): 56.

Lyu, L.C., Wu W.P., Hu H.P., and Huang R. 2019. “An Evolving Regional Innovation Network: Collaboration Among Industry, University, and Research Institution in China's First Technology Hub.” *The Journal of Technology Transfer* 44: 659-680.

Mueller, P. 2006. “Exploring the Knowledge Filter: How Entrepreneurship and University-Industry Relationships Drive Economic Growth.” *Research Policy* 35: 1499-1508.

Park, I., Jeong, Y., Yoon, B., and Mortara, L. 2015. “Exploring Potential R&D Collaboration Partners Through Patent Analysis Based on Bibliographic Coupling and Latent Semantic Analysis.” *Technology Analysis & Strategic Management* 27: 759-781.

Perkmann, M., Neely, A., and Walsh, K. 2011. “How Should Firms Evaluate Success in University–Industry Alliances? A Performance Measurement System.” R&D Management 41: 202–216.

Sjoo, K., and Hellstrom, T. 2019. “University-Industry Collaboration: A Literature Review and Synthesis.” *Industry and Higher Education* 33: 275-285.

Song, B., and Suh, Y. 2019. “Identifying Convergence Fields and Technologies for Industrial Safety: LDA Based Network Analysis.” *Technological Forecasting and Social Change* 138:115-126.

Song, B., Seol, H., and Park, Y. 2016. “A Patent Portfolio-Based Approach for Assessing Potential R&D Partners: An Application of the Shapley Value.” *Technological Forecasting and Social Change* 103: 156-165.

Steinmo, M., and Rasmussen, E. 2018. “The Interplay of Cognitive and Relational Social Capital Dimensions in University-Industry Collaboration: Overcoming the Experience Barrier.” *Research Policy* 47: 1964-1974.

Szucs, F. 2018. “Research Subsidies, Industry-University Cooperation and Innovation.” *Research Policy* 47: 1256-1266.

Velmurugan, T. 2012. “Efficiency of K-means and K-medoids Algorithms for Clustering Arbitrary Data Points.” *International Journal of Computer Applications in Technology* 3 (5): 1758-1764.

Wang, J. 2016. “Knowledge Creation in Collaboration Networks: Effects of Tie Configuration.” *Research Policy* 45: 68-80.

Wong, J.Y., Wan, T.H., and Chen, H.C. 2018. “The Innovative Grant of University-Industry-Research Cooperation: A Case Study for Taiwan's Technology Development Programs.” *International Journal of Innovation Science* 10: 316-332.

Xu, M., and Wang, H.M. 2018. “Research on the Structure of University-Industry Patent

Cooperation Network Based on Social Network Analysis." *Science and Technology Management Research* 38: 87-95.

Zhang, H., Zhu, Y.M., and Yang, L.S. 2016. "Research on the Evolution of University-Industry Patent Cooperation Network in Jiangsu Province." *Science & Technology Progress and Policy* 33: 50-57.

Zhang, W., Yoshida, T., and Tang, X. 2011. "A Comparative Study of TF*IDF, LSI and Multi-Words for Text Classification." *Expert Systems with Applications* 38 (3): 2758-2765.