



Inventor bibliographic-patent-coupling analysis and inventor-patent-classification-coupling analysis: a comparative analysis based on NPE

Song Yanhui¹ · Lei Lixin¹

Received: 22 March 2021 / Accepted: 19 April 2023 / Published online: 5 August 2023
© Akadémiai Kiadó, Budapest, Hungary 2023

Abstract

The patent literature is an important scientific and technological literature, which integrates technical information, market information, and legal information. It is of great significance to expand the bibliometric methods to the measurement of patent literature. This paper takes 4624 NPE (Non Practicing Entities) patents as samples and establishes an inventor coupling network based on two types of feature items, patent literature and Derwent classification codes. We have explored the technical structure of NPE patents. Through centrality analysis, correlation analysis, factor analysis, and visualization analysis, the two coupling analysis methods of Inventor Bibliographic-Patent-Coupling (IBPCA) and Inventor Patent Classification-Coupling (IPPCA) are compared. It is found that inventor centrality analysis, frequency correlation analysis, and cosine similarity measurement all show that IBPCA is correlated with IPPCA; The core technical topics of NPE patents discovered by IBPCA and IPCCA are digital computers, digital telecommunication transmission, and data storage and transmission. However, the two methods differ in factor models fitting analysis and intellectual structure detection. The factor fitting analysis of IPPCA is better than that of IBPCA; IBPCA can detect more topics than IPCCA, and has more advantages in small-scale topic detection; IPCCA is more sensitive to traditional and more stable research topics. Therefore, The combination of the two methods for intellectual structure detection and analysis will be more effective, then more comprehensive and specific conclusions will be obtained.

Keywords Inventor bibliographic-patent-coupling analysis · Inventor patent classification-coupling analysis · Patent analysis

✉ Song Yanhui
syh687@163.com

¹ School of Management, Hangzhou Dianzi University, Hangzhou 310018, China

Introduction

The bibliographic coupling analysis method has been fully applied and developed in the field of bibliometrics. However, the patent is also a kind of literature. It is meaningful work to expand the bibliographic coupling analysis method to patent literature and explore the patent bibliometrics method. The patent bibliometrics research was proposed by Narin (1994). Inventor bibliographic-patent-coupling analysis (IBPCA) originated from the author bibliographic coupling in bibliometrics, and the author bibliographic coupling analysis was based on bibliographic coupling. The bibliographic coupling means that two papers cite the same literature to form a kind of the same citation, which is called a coupling relationship. It was first proposed by Kessler (1963). The author bibliographic coupling promotes the bibliographic coupling analysis to the author level, not just stay in the paper level, with the author of the paper as the main research object (Zhao & Strotmann, 2014). After expanding to the author level, the bibliographic coupling overcomes the original shortcoming of fixed quantity and becomes a dynamic relationship. Therefore, the author bibliographic coupling analysis becomes more analytically meaningful (Huang & Chang, 2014). IBPCA indicates that two inventors form a relationship because they cite the same patent technology or the same research article in the invention patent. Inventors in IBPCA are equivalent to the authors of author bibliographic coupling, the research of a patent inventor is relatively fixed. Of course, an inventor may be involved in multiple research fields, and the research fields may also migrate. However, the possibility of significant changes or frequent changes in a short time is not great. Therefore, the IBPCA is also of great metrological value.

In addition, researchers have extended the author bibliographic coupling analysis and explored the author keyword coupling analysis method with keywords as the research object. In research articles, authors summarize the article, thus refining the keywords reflecting the content of the document, and the author keyword coupling analysis method is generated with the keywords as the research object. However, in patent documents, patent inventors and patentees do not give the declared patent keyword fields, and often reflect the technical field and technical direction of patent documents in the form of classification codes. Patent classification codes provide another coupling medium besides the citation literature. The establishment of a coupling relationship with the help of Derwent classification codes can open up a new dimension of patent coupling analysis. In the Derwent database, Derwent classification codes are extracted and summarized by more than 1000 technical experts of Thomson Reuters based on patent content, novelty, application, and other aspects, and then the technical classification index is given (Shen et al., 2012). Each patent contains one or more classification codes, and a classification code can actually represent a technical field. Therefore, the patent inventor, the citation literature, and the Derwent classification codes constitute the important analysis unit and correlation relationship of patent coupling. Inventor patent classification-coupling analysis (IPPCA) refers to the coupling relationship established by two inventors using the same classification codes.

The current research on the patent coupling is far from mature, and there are still many perspectives worth studying. For example (1) Most researches focus on the application of methods (Daim et al., 2006; Noh et al., 2016; Zhang et al., 2016), and the discussion of the method itself is slightly insufficient, that is, the basic theoretical research is insufficient. (2) Researchers mostly conduct coupling studies from the perspective of patentees (Chen & Fang, 2011; Wen, 2017). The patent inventor is the developer of the patent, and only the person who has put creative brain work into the invention can be considered as

the inventor. The patentee is the collective name of the owner and holder of the patent right, i.e. the actual owner of the patent, and the patent owner can be either a unit or an individual. The inventor of a patent is the creator of the invention, but not necessarily the actual owner of the patent. Patent inventors are like the authors of research articles. The literature they cite is more representative of the content of the patent. Therefore, for the specific study of the technical content of the patent, the patent inventor will be more valuable for analysis and will be able to obtain more research conclusions than the patentee. (3) At present, there are only a few discussions on the coupling of patent classification codes (Wen, 2017), but these studies are far from enough and need to continue to move forward. wen et al. only applied the patentee classification code coupling to the exploration of cooperative relationships between samples and did not involve both patent inventors and patent references, which are highly susceptible to coupling relationships. Distinguishing from their study, the study in this paper is an exploration of the method, rather than applying the method to a particular domain.

Based on the above research deficiencies, this paper proposes to study the patent coupling from the perspective of the patent inventor and conduct a comparative study of IBPCA and IPPCA from the perspective of the exploration of the method itself. Previous studies have shown that there is a correlation between author bibliographic coupling analysis and author keyword coupling analysis in research articles (Song & Wu, 2014), so is there also a correlation between IBPCA and IPPCA? This paper tries to explore. Since bibliographic coupling methods are widely used for intellectual structure detection, we further explore the similarities and differences between the two patent coupling methods in exploring the intellectual structure of technical fields. The research questions of this paper are as follows:

- (1) Is there any correlation between IBPCA and IPPCA? Where, if any, is the correlation?
- (2) What are the similarities and differences between the two coupling methods in exploring the intellectual structure of the technological field?

Most previous studies on bibliographic coupling analysis have focused on the application of the method (Liu et al., 2017; Park et al., 2014; Yeh et al., 2012). On the contrary, the analysis of the method itself has been neglected so far. This study aims to fill this gap. The current analysis of bibliographic coupling methods mostly uses papers as data sources and rarely involves the analysis of patent data. In addition, this study takes NPE (Non Practicing Entities) patents as an example, which can reveal the intellectual structure of the literature in the field of NPE to a certain extent, and facilitate to draw the attention and understanding of political circles, industries and academics to the patent giants.

Related work

Patent scientometrics analysis

According to the World Intellectual Property Organization (WIPO), patent information is one of the world's largest sources of open technology information, it contains 90–95% of the world's technical information, and the disclosure of technical information is 1–2 years earlier than other carriers. Several previous studies have also shown that a large part of the information in patents is relatively new (Bonino et al., 2010; Noh et al., 2015). Therefore,

the maximum research and utilization of patent information have become an important guarantee for countries and enterprises to gain a competitive advantage.

According to the characteristics of patent information, the researchers analyze patent information from the shallow to the deep: statistical analysis of the external characteristics of patent literature, that is, analysis of individual items (states, industries, institutions, individuals, etc.) in patent literature to obtain “point” information (Hasner et al., 2019; Huang et al., 2003a, 2003b; Lee & Lee, 2020); The “point” information obtained is sorted in a certain way, to obtain the ordered “line” information. That is, the process of analyzing the number of patents, patentees, and patent application date in the preliminary statistics of patents from the perspectives of time, space, and classification (Chang & Fan, 2016; Kang et al., 2007; Leydesdorff et al., 2012); By combining the above information, we can get a variety of interrelated “face” information about technological development. That is to carry out citation analysis and cluster analysis on patent literature, and deeply excavate the technical information in the patent literature and the relationship between various technologies (Jun et al., 2012; Kim & Bae, 2017; Von Wartburg et al., 2005).

The complexity and sheer volume of patent information pose a huge challenge to its effective and timely utilization. Various tools and techniques are available to analyze the patent data, such as text mining, visualization, citation analysis. Text mining is a process of using analytic tools to extract patterns of interest or usefulness to users from the natural language text. The visual outputs of the structured patent data such as the inventor of the patent, assignee of the patent, and citation information are represented in the form of graphs and networks, whereas the results from the unstructured patent data such as patent title, abstract, claims, and description are represented as patent maps (Abbas et al., 2014; Tseng et al., 2007). These mechanisms can also be used together. Zitt and Bassecoulard (2006) combine lexical query and citation analysis to delineate the complex scientific field of nanoscience. Park et al., (2018a, 2018b) combined text mining and visualization methods to identify thematic similarities and differences between the three online mental health communities. Yoon and Magee (2018) used text mining and patent map visualization tools in exploring technology opportunities.

Coupling analysis

The citation analysis method is a mature quantitative research method in bibliometrics, and it is widely used to detect the research frontier of discipline (Hsiao & Chen, 2020; Small & Griffith, 1974; Wang & Hsu, 2020), explore the relationship among academic groups (Gmür, 2003; Jiang et al., 2017), research discipline evolution path (Chang et al., 2015; Liu et al., 2017), measure technology similarity (Lai & Wu, 2005; Rodriguez et al., 2015), identify the intellectual structure of subject (Ferreira, 2018; Nerur et al., 2008). It is an important aspect of citation analysis to study the author’s citation behavior to explore the domain intellectual structure, which mainly includes three aspects: co-citation (Small, 1973), bibliographic coupling, and direct citation (Swanson, 1971). Among them, the accuracy of direct citation is the lowest (Boyack & Klavans, 2010). Although both bibliographic coupling and co-citation refer to the relationship between two papers through one or more other papers, they are different in terms of the object of study. The results of co-citation analysis reflect the relationship between two cited papers (references to research articles), while the results of bibliographic coupling analysis reflect the relationship between two citing papers (the literature data itself). They have natural differences in content and time in the analysis of research frontiers, which also lead to differences in the analysis results.

Although co-citation analysis is currently the dominant method in the field of intellectual structure research (Barirani et al., 2013; Hou et al., 2018), it is not the only method that can be used to analyze the author relationships using citation information. The resurgence of the use of bibliographic coupling is challenging the historical preference for co-citation analysis. Many studies have shown that bibliographic coupling analysis has stronger timeliness, stability, and integrity (Boyack & Klavans, 2010; Huang & Chang, 2014; Kuan et al., 2019). The bibliographic coupling can identify more important research frontiers and detect the emergence of research frontiers earlier than co-citation, showing a better performance in discovering research frontiers.

Small and Koenig (1977) used a bibliographic coupling method to journal clustering. It's pointed out that the strength of bibliographic coupling is the number of same documents cited by the two journals. Liu, Nanetti, and Cheong (2017) built bibliographic coupling networks to study the knowledge evolution in physics. Yang et al. (2016) introduced the author keyword coupling analysis method for analyzing the intellectual structure of the field of information science and compared it with the author bibliographic coupling analysis method, particularly based on first- and all-author coupling.

With the characteristics and advantages of the coupling analysis method becoming more and more prominent, researchers have gradually introduced it into the field of patent metrology to investigate and reveal the relationship between patent literature and patentees. Park et al. (2015) used bibliographic coupling analysis to explore potential Research and Development collaboration partners based on patent information; Lo (2007) took only a patent coupling analysis approach to reveal the technological connection among the primary organizations. The results show the feasibility of utilization of "Patent Coupling" on correlation analysis; In addition, Kuusi and Meyer (2007) used bibliographic coupling to anticipate technological breakthroughs in the field of nanotechnology. Kuan et al. (2018) and Chen et al. (2012) used bibliographic coupling and direct citation to construct comprehensive citation networks.

Methodology

Data source

The data source of this article is Derwent Innovations Index (DII). As an authoritative patent information retrieval system in the world, DII integrates the World Patent Index (WPI) and Patent Citation Index (PCI), providing a wealth of patent sources and patent citation data (Shen et al., 2012). The data samples are NPE patents. NPE are the subject of non-patent implementation. Its patent acquisition is not for a specific implementation. NPE regards patents as commodities that can be circulated and makes profits by relying on patent transaction activities. Its patent operation behavior may be either legitimate business behavior or patent speculation behavior of abusing patent rights. The reasons for choosing NPE patents for study are as follows: (1) This is our last project of National Social Science Foundation of China, so we are more familiar with the NPE, and it is convenient for us to start a full comparison between the two methods of IBPCA and IPPCA. (2) The non-patent enforcing subject is a relatively new thing, and we also want to draw the attention of the academic community through the research of this paper. (3) The patents of non-patent enforcing subjects are often of higher value and the data will be better, which is conducive to the analysis of our research process and the making of research conclusions. The data retrieval in this paper is an advanced retrieval, and the retrieval item is the patentee.

Since many enterprises share the same non-standard code, the non-uniqueness of the non-standard code can be avoided by using the patentee's name to retrieve the data. The data in this paper are based on the definition, morphology, and characterization of NPEs by Pénin (2012) and Zhang et al. (2015). We got NPEs from these papers and retrieved their patents through DII. Finally, we obtained 4624 NPE patents data such as Eolas, Wi-Lan, NorthPeak, Stragent, etc. We found that there are a large number of relevant studies with datasets of around 4000 as well (Huang et al., 2003a, 2003b; Wang & Hsu, 2020; Zhao & Strotmann, 2008a, 2014). Therefore the dataset in this paper is appropriate.

Extraction of core inventors

In this paper, price's law is used as the selection standard of the core patent inventors. We make statistics on inventors in NPE patents data twice: only consider the first patent inventor and consider all inventors. If only the first inventor is considered, WARE, FA has the largest number of patented inventions, and the number of patents is 144, then $m=8.99$ is calculated according to the formula. If all authors are considered, WARE, FA still has the most patent inventions, with 296 patents, that is, $m=12.89$. A statistical comparison of inventors with more than 9 and 13 patents reveals that the two methods have 68 co-authors. Only 6 of the 74 core patent inventors obtained by the first inventor are not included. Although the number of inventors is much larger when all inventors are considered, and the number of core inventors is increased along with it, the core patent inventors identified through the first patent inventor are also valid. Therefore, the 68 authors we further screened out can basically be regarded as outstanding representatives of NPE patents. In addition, another similarity between the two methods is that the proportion of the number of patents owned by the determined core patent inventors to the total number of patents is equivalent, the proportion considering only the first author is 34.0568%, while the proportion considering all authors is 35.7052%. Therefore, patent inventors and their patent volume present a good centralized and discrete distribution. The two calculation methods confirm that our results are credible.

Methods and tools

Inventor bibliographic-patent-coupling analysis

Similar to the author bibliographic coupling analysis, if two inventors cite the same patented technology or the same research article, a coupling relationship is considered to exist between them. It should be pointed out that patent citation includes both research articles and patent documents. In terms of coupling calculation, Zhao and Strotmann (2008a) established a data set for each author, which contains all references of the author. The same number of references in the data set of the two authors is the coupling frequency between them. In the calculation of IBPCA, we have used this same method. The inventor and the author are corresponding, and the references in the paper are corresponding to the research articles and patent documents in the patent. The bibliographic-patent-coupling strength between two inventors is the degree of overlap of all their two literature sets. The coupling strength is calculated for each literature in the literature sets of the two inventors and accumulated one by one, and the final value obtained is taken as the inventor bibliographic-patent-coupling strength of these two inventors. Examples are shown in Table 1. Inventors WARE and ZERBE, have two identical references (B, C), the coupling strength is 2

Table 1 An example of IBPCA and IPCCA

Inventor	Reference	Derwent classification codes
WARE	ABCDE...	T01, U12, S01, W01, T03, W03...
ZERBE	BCFHK.	T01, U12, W03, X11, U14, W03...
SHAEFFER	ADFHK	T03, T05, U14, W03, S01, W03, W03...
...

in IBPCA. Similarly, inventor WARE has a coupling strength of 2 with SHAEFFER, and inventor ZERBE has a coupling strength of 2 with SHAEFFER.

Inventor patent classification-coupling analysis

Inventor patent classification-coupling is similar to the author subject coupling and the author keyword coupling, which refers to the coupling relationship established by two inventors using the same classification codes. The classification code here is the Derwent classification code, and a Derwent subcategory is similar to a subject or a keyword in the research article. In the calculation of IPCCA, each Derwent classification code is equivalent to a patent citation, and the classification code can also be extracted to establish a data set, which is basically consistent with the establishment process of the patent citation data set. Specifically, we create a set of classification codes for each patent inventor (S1, S2..... SN), and then compare the sets of classification codes between inventors, and then count the number of identical classification codes in different sets of classification codes. We consider this value as the inventor patent classification-coupling strength. Exceptionally, if a classification code appears multiple times in two sets of classification codes, a further calculation is required. For example, if the classification code appears M times and N times in the two sets, then its classification code coupling strength is min (M, N). Examples are shown in Table 1. Inventors WARE and ZERBE, have three identical classification codes (T01, U12, W03), IPCCA coupling strength of 3. In particular, W03 appears two times in inventor ZERBE, and W03 appears three times in inventor SHAEFFER, and we take its minimum value of 2, then the IPCCA coupling strength of inventor ZERBE and SHAEFFER is 3.

Factor analysis

The basic principle of factor analysis is to reduce the dimension of the measured variables with a complex relationship. The common variables are classified into one category, each category has a factor, and finally, a few factors are used to show the relationship between the variables and the factors. In this article, the factor analysis method is used to analyze the sample data, and the inventors with similarities are gathered together to form a factor, and then form a technical theme. Firstly, the IBPCA matrix and IPCCA matrix of the patent inventor are constructed respectively. The diagonal adopts the maximum coupling frequency plus 1 to reflect the self-coupling of the inventor (White, 2003). Then, the IBPCA matrix and the IPCCA matrix are similarly transformed to eliminate the difference in magnitude and dimension of the data. The similarity matrix is imported into SPSS for factor analysis. The principal component analysis is used for factor extraction. The rotation method of factor analysis is the direct Oblimin method (McCain, 1990; White & Griffith, 1981; Zhao & Strotmann, 2008b).

Visual analysis

In this paper, we choose NERDRAW to visualize the factor matrix. The factors are represented by square nodes, and the inventors are represented by circular nodes. The connecting line between the circular node and the square node indicates that the inventor has a load on the factor, and the load value must be greater than 0.2 to appear. The thickness of the line represents the size of the factor loading value. The factors are represented by a uniform color. The color of the circular node represents different degree centrality. Red represents the degree centrality of 1, military green represents the degree centrality of 2, pink represents the degree centrality of 3, the orange node's degree centrality is 4, the gold node's degree centrality is 5, and the size of the node represents the betweenness centrality. Figure 1 illustrates the overall analysis process of this paper.

Results

Analysis of core inventors

Centrality describes the position and importance of the inventor in the patent coupling network. This article analyzes the degree centrality, betweenness centrality, and eigenvector centrality of IBCCA and IPCCCA to explore the core inventors of the two coupling methods.

Table 2 shows that in the degree centrality of IBPCA and IPCCCA, five of the top ten inventors are the same. The first and third inventors in IBPCA are also the first and third inventors in IPCCCA. The inventor's high degree of centrality reflects the inventor's high

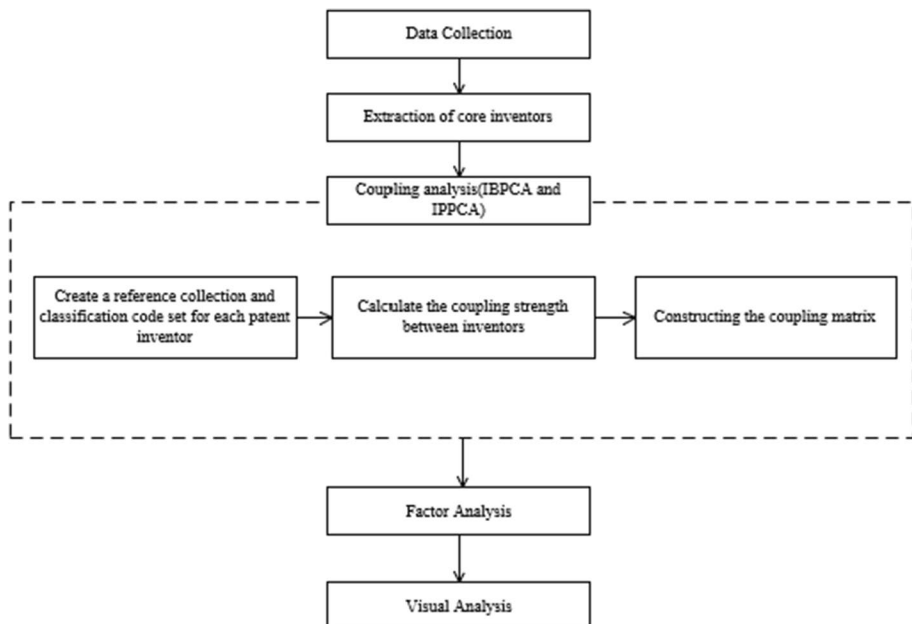


Fig. 1 The overall analysis process

Table 2 Inventor centrality ranking (top 10)

IBPCA				IPCCA			
Rank	Degree	Betweenness	Eigenvec	Rank	Degree	Betweenness	Eigenvec
1	WARE, FA	WARE, FA	WARE, FA	1	WARE, FA	PEREGO, RE	WARE, FA
2	HAMPEL, CE	STOCKHAM, MA	HAMPEL, CE	2	BEST, SC	PARKER, JR	HAMPEL, CE
3	ZERBE, JL	PARKER, JR	ZERBE, JL	3	ZERBE, JL	STOCKHAM, MA	ZERBE, JL
4	SHAEFFER, I	MANSFIELD, P	PEREGO, RE	4	WANG, Y	SMITH, JAS	PEREGO, RE
5	BARTH, RM	ZERBE, JL	TSEHN, EK	5	NGUYEN, HM	MONRO, DM	TSEHN, EK
6	PEREGO, RE	ABBASFAR, A	HOROWITZ, MA	6	HAMPEL, CE	OH, KS	HOROWITZ, MA
7	BEST, SC	BEST, SC	BARTH, RM	7	ABBASFAR, A	NGUYEN, HM	BARTH, RM
8	HOROWITZ, MA	SHLOMOT, E	BEST, SC	8	GAO, Y	STOJANOVIC, VM	BEST, SC
9	NGUYEN, HM	SHAEFFER, I	STANWOOD, KL	9	STOJANOVIC, VM	WARE, FA	STANWOOD, KL
10	SPINAR, B	BOURLAS, Y	STARK, DC	10	MONRO, DM	STARK, DC	STARK, DC

research vitality, enabling more relationships with other inventors can be established. Therefore, the inventors of WARE, FA; HAMPEL, CE; ZERBE, JL; BEST, SC are relatively active researchers in NPE patents technology. In IBPCA, the top three inventors with the highest betweenness centrality are WARE, FA; STOCKHAM, MA; PARKER, JR. At the same time, their betweenness centrality in IPCCCA is also high. This shows that the inventors with an “intermediary” nature detected by the two coupling methods are similar. We also find that inventors rank roughly the same in the centrality of eigenvector in IBPCA and IPCCCA. In the inventor coupling network, if the centrality of an inventor’s eigenvector is high, it usually indicates that the centrality of other inventors who have a coupling relationship with the inventor is also high.

In conclusion, IBPCA and IPCCCA have a certain similarity in core inventors based on the centrality analysis of the above three different properties. The phenomenon can be observed, at least among the top-ranked inventors, So does the overall analysis still show a good similarity? This will be analyzed in the following.

Coupling correlation analysis

Frequency correlation analysis

In order to further analyze the correlation level of IBPCA and IPCCCA as a whole, we conducted a correlation analysis on the coupling frequency and coupling frequency ranking of 68 inventors in IBPCA and IPCCCA, as shown in Table 3. All the correlation levels in this article are measured at the 0.01 level and the Sig. value is far less than 0.01, which means that the correlation is significant.

The average coupling frequency is calculated on the basis of the total coupling frequency. Therefore, the correlation between the total coupling frequency and the average coupling frequency is 1, which means a complete correlation. In IBPCA, there is an obvious correlation between the total coupling frequency, the average coupling frequency, and the maximum coupling frequency. The correlation coefficients are all 0.817, indicating that the higher the total coupling frequency, the higher the maximum coupling frequency; The

Table 3 Coupling frequency and ranking correlation analysis

IBPCA				IPCCCA			
Frequency correlation analysis	A	B	C	Frequency correlation analysis	A	B	C
A	1	1	0.817	A	1	1	0.75
B	1	1	0.8	B	1	1	0.751
C	0.817	0.817	1	C	0.751	0.751	1
Frequency ranking correlation analysis	A Rank	B Rank	C Rank	Frequency ranking correlation analysis	A Rank	B Rank	C Rank
A Rank	1	1	0.927	A Rank	1	1	0.749
B Rank	1	1	0.927	B Rank	1	1	0.749
C Rank	0.927	0.927	1	C Rank	0.749	0.749	1

A represents the total coupling frequency, B represents the average coupling frequency, and C represents the maximum coupling frequency

total coupling frequency ranking, the average coupling frequency ranking, and the maximum coupling frequency ranking also have a high correlation coefficient of 0.927, which indicates that inventors having a higher ranking in the total coupling frequency ranking tend to have a higher ranking in the maximum coupling frequency; In addition, the correlation of the inventor’s coupling frequency ranking is slightly higher than the correlation of the frequency value. In IPCCA, the correlation of the inventor’s coupling frequency ranking is equivalent to the correlation of the frequency value. In IPCCA, the correlation coefficient between the total coupling frequency and the maximum coupling frequency is 0.751, and the correlation coefficient between the total coupling frequency ranking and the maximum coupling frequency ranking is 0.749, these two values are extremely close. Whether it is IBPCA or IPCCA, the coupling frequency (total coupling frequency, average coupling frequency) and the maximum coupling frequency are highly correlated. The coupling frequency ranking (total coupling frequency ranking, average coupling frequency ranking) and the maximum coupling frequency ranking are also highly correlated. But in IPCCA, this correlation level is slightly lower than IBPCA. In other words, compared to IPCCA, if the inventor has a higher coupling frequency or a higher ranking in IBPCA, it is easier to obtain a higher maximum coupling frequency and its ranking.

In general, inventors with more patents are more likely to establish higher coupling frequency. In order to analyze the correlation between them, we add two variables of patent quantity and patent quantity ranking to conduct correlation analysis. Because the average coupling frequency and total coupling frequency are completely correlated, here we only take the total coupling frequency and the maximum coupling frequency. The results are shown in Table 4. In IBPCA, the correlation between the number of patents and the total frequency of patent coupling is greater than that of IPCCA. The correlation coefficient in IBPCA is 0.779, while the correlation coefficient in IPCCA is 0.570. However, there is no significant difference between the patent quantity ranking and the total coupling frequency ranking, which are 0.479 and 0.579 respectively. This shows that in IBPCA, the direct correlation between the number of patents and the total frequency of patent coupling is greater. The inventor has more patented inventions, and it is easier to obtain a higher coupling frequency. And with regard to ranking, compared to IPCCA, IBPCA is not obvious.

Table 4 Correlation analysis of total coupling frequency, maximum coupling frequency and patent amount

IBPCA				IPCCA			
Frequency correlation analysis	A	B	C	Frequency correlation analysis	A	B	C
A	1	0.817	0.779	A	1	0.751	0.570
B	0.817	1	0.591	B	0.751	1	0.711
C	0.779	0.591	1	C	0.570	0.711	1
Frequency ranking correlation analysis	A Rank	B Rank	C Rank	Frequency ranking correlation analysis	A Rank	B Rank	C Rank
A Rank	1	0.927	0.479	A Rank	1	0.749	0.579
B Rank	0.927	1	0.415	B Rank	0.749	1	0.780
C Rank	0.479	0.415	1	C Rank	0.579	0.780	1

A represents the total coupling frequency, B represents the maximum coupling frequency, and C represents the amount of patents

In terms of the correlation analysis coefficients (0.591, 0.415; 0.711, 0.780) between the number of patents and their rankings and the maximum coupling frequency and their rankings, IPCCA is more advantageous than IBPCA. That is to say, compared with IBPCA, Inventors with more patents are more likely to obtain a higher maximum coupling frequency in IPCCA. Similarly, if an inventor ranks higher in patents, his maximum coupling frequency ranking also tends to have a higher ranking. Table 4 shows that the inventor's patent ownership does affect the coupling between the inventor and other inventors.

The above correlation analysis does not distinguish the similarities and differences of the same inventor in the two methods. The analysis of the performance of the same inventor in different methods can better reflect the similarities and differences of the two methods (Zhang et al., 2016). To further explore the similarities between IBPCA and IPCCA, we analyze the co-inventors of IBPCA and IPCCA, as shown in Table 5. The correlation coefficient of the total coupling frequency of the inventors between IBPCA and IPCCA is the highest, which is 0.618. The correlation coefficient of the total coupling frequency ranking of the inventors between IBPCA and IPCCA is 0.568. This means that the inventor's coupling behavior in IBPCA and IPCCA is basically consistent. The total coupling frequency of the inventor in IBPCA is related to the total coupling frequency in IPCCA to a certain extent, and the coupling frequency ranking of inventors in IBPCA is also related to the frequency ranking of the same inventor in IPCCA. For example, the rankings of MONRO, DM; ABBASFAR, A; ARYANFAR, F; COK, RS; HOSSAIN, M; WENTINK, M; SPINAR, B in IBPCA and IPCCA are the same, being 1, 3, 22, 65, 52, 55, 68. PALMER, RE; NGUYEN, H; BEST, SC; KIZER, JM only differ by 1 place in the ranking of IBPCA and IPCCA.

The average coupling frequency is calculated according to the total coupling frequency, therefore, the correlation analysis results of the average coupling frequency and the total coupling frequency in IBPCA and IPCCA are consistent. The correlation analysis result of the inventor's maximum coupling frequency in IBPCA and IPCCA is 0.455; the correlation analysis result of the inventor's maximum coupling frequency ranking in IBPCA and IPCCA is 0.467. These two values are very close, and the correlation coefficient is not high. This shows that IBPCA and IPCCA are not strongly correlated in terms of maximum coupling frequency. It is found that many inventors have good performance in IBPCA, but poor performance in IPCCA. For example, SPINAR, B ranked 3 in IBPCA, but ranked 63 in IPCCA, with a difference of 60. These inventors will weaken the correlation between the inventor's maximum coupling frequency in IBPCA and IPCCA. Of course, the maximum

Table 5 Correlation analysis of IBPCA and IPCCA

		IBPCA					
		A	A Rank	B	B Rank	C	C Rank
IPCCA	A	0.618					
	A Rank	0.568					
	B	0.618					
	B Rank	0.568					
	C	0.455					
	C Rank	0.467					

A represents the total coupling frequency, B represents the average coupling frequency, and C represents the maximum coupling frequency

coupling frequency and its ranking of most inventors in IBPCA and IPCCA still have a certain correlation.

Cosine similarity comparison

Cosine similarity uses the cosine value of the angle between two vectors in the vector space as a measure of the difference between two individuals. The important consideration is the difference in the direction of the vector rather than the difference in distance or length. We run SPSS to measure the cosine similarity of the IBPCA matrix and the IPCCA matrix, and the results show that the IBPCA matrix and the IPCCA matrix are similar matrices, and the similarity is 0.396. This is the cosine similarity calculation result of the original data matrix. We believe that the original data has large data differences, which will affect the expression of the results to a certain extent. To reduce the differences between the data, the matrix after the similarity conversion is calculated for the cosine similarity. The calculation result is much more ideal than the calculation result of the original matrix, and the similarity is greatly enhanced. Therefore, it can be said that by calculating the cosine similarity of the IBPCA and IPCCA matrices, we can basically conclude that IBPCA and IPCCA have a certain similarity.

In the above, we have demonstrated that IBPCA and IPCCA are related from the correlation analysis of total coupling frequency, average coupling frequency, maximum coupling frequency, etc., which can be said to be a macro-level argument. The cosine similarity goes deep into the comparison and calculation of each data, which can be said to demonstrate the similarity between IBPCA and IPCCA from the micro-level. They all proved that IBPCA and IPCCA have a certain correlation.

Comparison of factor fitting analysis results

Factor model fitting

Factor analysis of the IBPCA matrix and IPCCA matrix is carried out, and the scree plot is shown in Fig. 2. From the scree plot, the fitting result of IPCCA is better than that of IBPCA. Specifically, IBPCA extracted 12 factors, accounting for 90.138% of the total variance, while IPCCA only used 5 factors, accounting for 97.327% of the total variance.

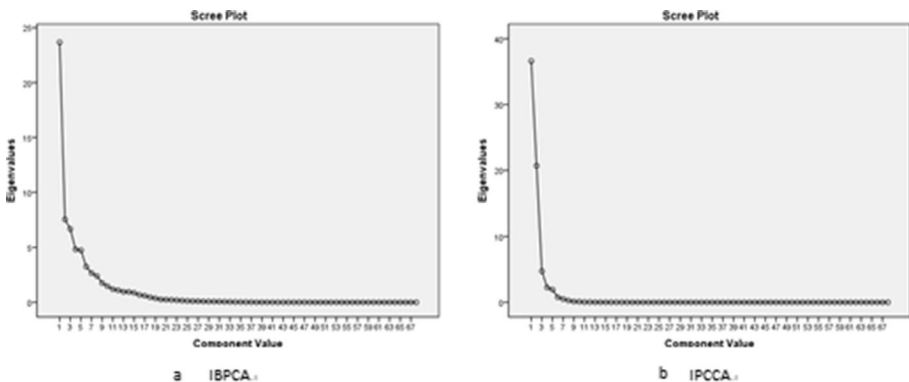


Fig. 2 Factor fitting scree plot

Compared with IBPCA, IPCCCA can explain more total variance with fewer factors, and the explanation of variance is better.

There are 95 (4%) non-redundant residuals with an absolute value greater than 0.05 between the observed correlations and the regenerated correlations calculated by IBPCA; there are 20 (0%) non-redundant residuals with absolute values greater than 0.05 between the observed correlation and the regenerated correlation calculated by IPCCCA; therefore, the residuals also show that the fitting results of IPCCCA are better than IBPCA. The common factor of IPCCCA varies from 0.716 to 0.999, the highest common factor is 0.999, and the lowest common factor is 0.716. The common factor of IBPCA varies from 0.482 to 0.992, the highest common factor 0.992 is less than 0.999, and the lowest common factor 0.482 is also much less than 0.716. Therefore, IPCCCA is better than IBPCA from the variation range of common factor, highest common factor, and lowest common factor.

Factor loading analysis

After performing factor analysis on the IBPCA matrix transformed by similarity, a total of 12 factors were detected. We examine the highest load inventors, investigate the high-frequency coupling pairs between the highest load inventors and other inventors (especially the high load inventors), analyze the commonalities among these inventors, especially study the patent literature contents that lead to the coupling of these high load inventors, and consult experts or scholars in related fields to determine the factor labels.

Factor 1: the maximum load inventor is CONNORS, DP, the coupling frequency with WENTING, M is 76, the coupling frequency with DALLY, WJ is 46, the coupling frequency with MAENG, J is 40, and the coupling frequency with WARE FA (95) is the maximum coupling frequency. The main research field of this factor is “Digital Computers”. Factor 2: The maximum load inventor is GARLEPP, BW, the coupling frequency with ZERBE, J is 276, the coupling frequency with STOJANOVIC, VM is 160, and the coupling frequency with WERNER, CW is 64. The most involved is “Signal Generation and Distribution”. Factor 3: The maximum load inventor is SPINAR, B. The coupling frequency with STANWOOD, KL is 1689, which is also the maximum coupling frequency. The coupling frequency with ONG, AE is 181, and the coupling frequency with VOGELANG, T is 16. A large amount of content involved is digital information transmission, which is the field of communication. To avoid confusion, different from factor 11 (Data Storage and Transmission, with emphasis on computer and semiconductor fields), we define this factor as “Digital Telecommunications Transmission”. There are no high-load inventors for factors 4 and factor 5. The largest inventors are LEE, W, and BENYASSINE, A. Factor 6: The inventor of the largest load is HYNECEK, J. The factor is determined as “Semiconductors and Integrated Circuits” mainly through the joint technical research of HYNECEK, J and COK, RS, HOSSAIN, M. Factor 7, there are no high-load inventors, focusing on the joint patent inventions of SU, H, and GAO, Y. Factor 8, factor 9, factor 10, and factor 12 are generally lower than 0.4, and there is no high-load inventor. Factor 11, the inventor of the highest load is WEBSTER, Ma, which involves data storage and memory, interconnection, data transmission, and data static storage. It can be seen that although the two belong to different categories, there are still many overlapping contents. The content of the factor can be summarized as “Data Storage and Transmission”. For factor 4, factor 5, factor 7, factor 8, factor 9, factor 10, and factor 12, we label the factor content uniformly, because we found that these 7 factors are derived from the 5 inventors of SHLOMOT, E; GAO, Y; SU, H; THYSSEN, J; BENYASSINE, A., these factors are relatively independent, and the factor

load distribution is relatively uniform. To determine the factor label, we first mainly consider the commonality research between the inventor with the highest load and the inventor with the highest coupling frequency. When there is a conflict between factors, such as the inventor with the highest load of factor 10 and factor 12, and the highest coupling pair may be SHLOMOT, E, and GAO, Y, we will consider the study of the inventor with the second or third load. In this way, we determine the factor labels as: factor 4 is “Computer Speech Processing”, factor 5 is “General Speech Processing”, and factor 7 has too few authors and too low a load value to determine the research content. We use “Undetermined” to express it. Factor 8 is “Portable Phone”, factor 9 is “Noise Treatment”, factor 10 is “Coding and Information Theory”, and factor 12 is “Data Conversion and Transmission”.

After performing factor analysis on the IPPCA matrix transformed by similarity, a total of five topics (factors) were detected, which are mainly some traditional and more stable research topics. Factor 1: LAU, BC is the inventor of the largest load, the maximum coupling frequency is 22, the coupling frequency of LAU, BC, and KIZER, JM is 19, the coupling frequency with STARK, DC is 21, and the coupling frequency with BEST, SC is 22. The factor name is recorded as “Digital Computers”. Factor 2: GAO, Y is the inventor of the largest load, the maximum coupling frequency is 51, the coupling frequency with CONNORS, DP is 27, the coupling frequency with MONRO, DM is 51, and the coupling frequency with SHLOMOT, E is 30. This factor is marked as “Digital Telecommunications Transmission”. Factor 3: We analyze the common technical research of highly coupled inventors, especially BELL, M and ARMSTRONG, BA. The study found that data storage and memory, interconnection, data transmission, and data static storage are the most coupled. The factor content can be summarized as “Data Storage and Transmission”, which involves not only the field of digital computers but also semiconductors and electronic circuits. It is the intersection of the two fields. Factor 4: There are only 2 inventors, namely STOCKHAM, MA, and HIDER, RC. STOCKHAM, MA is the inventor of the maximum load, and the maximum coupling frequency object also happens to be HIDER, RC, with a frequency value of 4. The studies on the analysis of the interaction between Stockham, MA, and Hider, RC are mostly “Cyclic Compounds Research” in the B category. Factor 5: The technical fields with more coupling are: liquid crystal displays and optics. The factor content can be “LED Optical Research”.

Visual analysis

Importing each factor and the nodes with load values greater than 0.2 on these factors into Netdraw for processing. The visualization results are shown in the following figures. In the visualization map of IBPCA (Fig. 3), through the analysis of betweenness centrality, three important factors can be found: digital computers, digital telecommunications transmission, and data storage and transmission. Through the K-core analysis below, we will also find that these three factors are the most important. These three factors interact and interweave. The inventors of “Digital Telecommunications Transmission” and “Data Storage and Transmission” are VOGELSANG, T, and ARYANFAR, F. There are more inventors who contact “Digital Computers” and “Data Storage and Transmission”. The inventor of “Digital Computers” and “Digital Telecommunications Transmission” is WENTINK, M. “Digital Computers” and “Semiconductors and Integrated Circuits” are also relatively close, and there are many contacts between inventors.

In the IPCCCA visualization map (Fig. 4), the most important factors are digital computers and digital telecommunications transmission. These are also the two factors with the

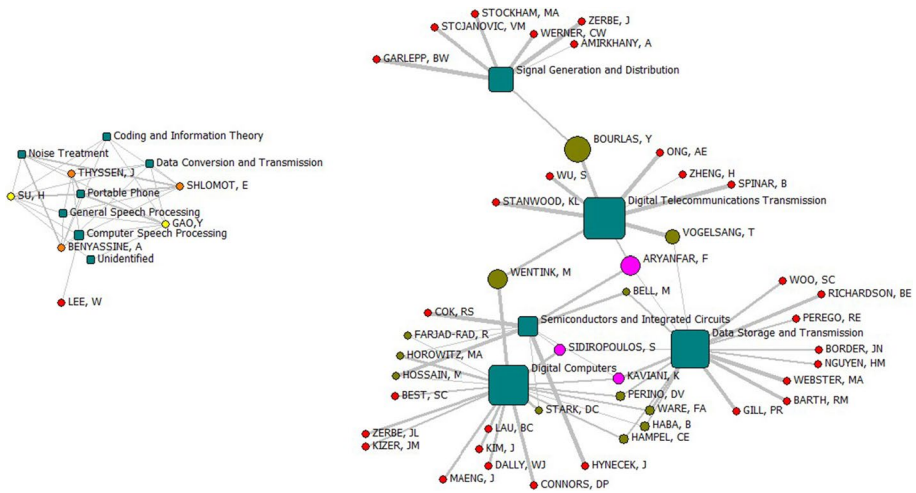


Fig. 3 IBPCA visualization map

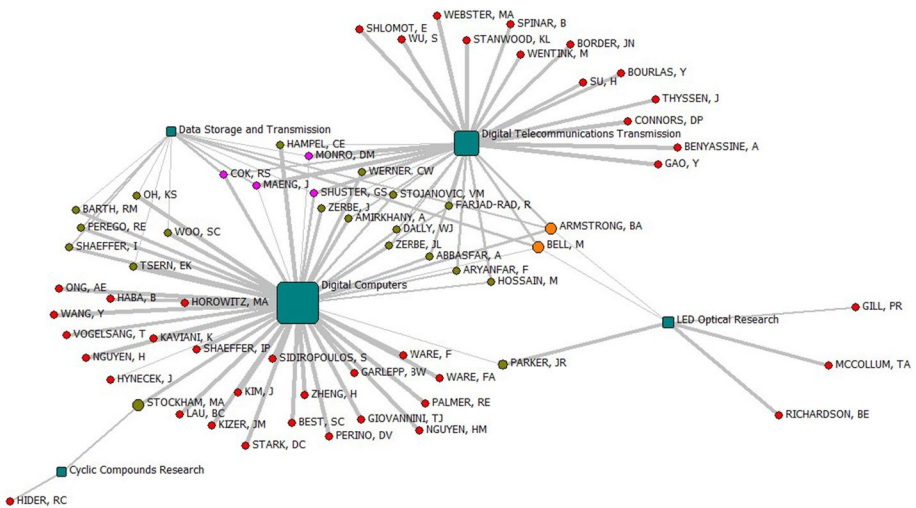


Fig. 4 IPPCA visualization map

strongest interaction. There are many inventors connected, which is different from IBPCA. IBPCA has only one inventor contact, and the interaction between the two factors is obviously weaker. The interaction between “Data Storage and Transmission” and “Digital Computers” is relatively strong, and the inventors with intermediate connections are W00, SC; TSERN, EK; SHAEFFER, I; OH, KS; PEREGO, RE; BARTH, RM. The interaction between “Data Storage and Transmission” and “Digital Telecommunications Transmission” is also relatively strong, and the inventors with intermediate contact are MONRO, DM; SHUSTER, GS; COK, RS; MAENG, J. At the same time, these inventors are also

important inventors who contact “Digital Computers” and “Digital Telecommunications Transmission”.

It can be seen that the most important factors detected by IPCCCA are: digital computers, digital telecommunications transmission, which are all detected in IBPCA. The data storage and transmission detected by IPCCCA is also detected in IBPCA, and this factor is an important factor in IBPCA. The LED optical research detected by IPCCCA was not detected in IBPCA. IBPCA detects many small factors, such as portable phones, general speech processing, computer speech processing, coding and information theory, data conversion and transmission, noise treatment, etc., which are also not available in IPCCCA. Therefore, it can be said that IBPCA can detect more factors than IPCCCA, especially small factors. The interaction between factors is also different. The strong interaction in IPCCCA may not necessarily show strong force in IBPCA; the strong force in IBPCA may not necessarily show strong force in IPCCCA.

We can get more core research fields and inventors by further k -core analysis. By calculating $k=1$, $k=2$, and $k=5$, we can see that there are four core research fields (factors) of IBPCA: digital computers, digital telecommunications transmission, data storage and transmission, semiconductor, and integrated circuits. For the calculation of $k=3$, $k=1$, and $k=2$, IPCCCA also found three core research areas: digital computers, digital telecommunications transmission, data storage, and transmission. Therefore, we can see that the core areas of IBPCA and IPCCCA are roughly equivalent.

Conclusion

Taking NPE patents as samples, this paper uses centrality analysis, correlation analysis, factor analysis, visual analysis, and other analysis methods to compare the two analysis methods of IBPCA and IPCCCA, and analyzes the NPE patents technological structure. The main research findings are as follows:

(1) First of all, it can be seen from the inventor centrality ranking that IBPCA and IPCCCA have certain similarities in core inventors. Secondly, IBPCA and IPCCCA show similar characteristics in the correlation analysis results of patent quantity and its ranking, total coupling frequency and its ranking, average coupling frequency and its ranking, maximum coupling frequency, and its ranking. Coupling frequency is highly correlated with maximum coupling frequency; coupling frequency ranking is highly correlated with maximum coupling frequency ranking; inventor’s patent ownership does affect the inventor coupling with other inventors. Finally, the cosine similarity measurement of the similarity matrix of IBPCA and IPCCCA demonstrates the similarity between IBPCA and IPCCCA from the micro-level.

(2) Through factor analysis and visualization analysis, it is found that the fitting result of IPCCCA is better than that of IBPCA in terms of scree plot, variance explanation degree, common factor variation range, and common factor size. In terms of the intellectual structure exploration in the technological field, IBPCA can detect more topics than IPCCCA, and has advantages in the discovery of smaller topics. The intellectual structure of IPCCCA detection is mainly a traditional and more stable research topic. The core areas of NPE discovered by IBPCA and IPCCCA are roughly the same. The core research areas of IBPCA are digital computers, digital telecommunications transmission, and data storage, and transmission, semiconductors, and integrated circuits; and IPCCCA also found three core

research areas: digital computers, digital telecommunications transmission, data storage, and transmission.

This paper is a positive attempt to expand bibliometric methods to the field of patent literature and hopes that patent coupling will eventually become a mature method like bibliographic coupling and be widely used. We believe that the inventor bibliographic-patent-coupling and inventor patent classification-coupling proposed in this paper can play an important role in patent intellectual structure detection in the future. Moreover, these two methods have their characteristics and advantages. If they can be used together, a better research conclusion will be obtained. Although the study provides a set of interesting findings and shows that the IPCCCA method can be used to support the exploration of the intellectual structure, some questions must be addressed in future research. First, we only choose the Derwent database for empirical research, future studies should analyze other patent databases (e.g., USPTO) similarly, to examine thoroughly the effectiveness of IBPCA and IPCCCA; second, the data source for this study can be expanded, as we only examined NPE patents.

Acknowledgements This study was funded in part by major project of National Social Science Foundation of China (19ZDA348), and Supported by the Fundamental Research Funds for the Provincial Universities of Zhejiang (GK209907299001-201).

Funding Major project of National Social Science Foundation of China, 19ZDA348, Song Yanhui, Fundamental Research Funds for the Provincial Universities of Zhejiang, GK209907299001-201, Song Yanhui.

References

- Abbas, A., Zhang, L., & Khan, S. U. (2014). A literature review on the state-of-the-art in patent analysis. *World Patent Information*, 37, 3–13. <https://doi.org/10.1016/j.wpi.2013.12.006>
- Barirani, A., Agard, B., & Beaudry, C. (2013). Discovering and assessing fields of expertise in nanomedicine: A patent co-citation network perspective. *Scientometrics*, 94(3), 1111–1136. <https://doi.org/10.1007/s11192-012-0891-6>
- Bonino, D., Caramella, A., & Corno, F. (2010). Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. *World Patent Information*, 32(1), 30–38. <https://doi.org/10.1016/j.wpi.2009.05.008>
- Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology*, 61(12), 2389–2404. <https://doi.org/10.1002/asi.21419>
- Chang, S. H., & Fan, C. Y. (2016). Identification of the technology life cycle of telematics: A patent-based analytical perspective. *Technological Forecasting and Social Change*, 105, 1–10. <https://doi.org/10.1016/j.techfore.2016.01.023>
- Chang, Y. W., Huang, M. H., & Lin, C. W. (2015). Evolution of research subjects in library and information science based on keyword, bibliographical coupling, and co-citation analyses. *Scientometrics*, 105(3), 2071–2087. <https://doi.org/10.1007/s11192-015-1762-8>
- Chen, S. H., Huang, M. H., Chen, D. Z., & Lin, S. Z. (2012). Detecting the temporal gaps of technology fronts: A case study of smart grid field. *Technological Forecasting and Social Change*, 79(9), 1705–1719. <https://doi.org/10.1016/j.techfore.2012.06.005>
- Chen, Y., & Fang, S. (2011). Methods of social network analysis on patent assignees' correlation networks. *Documentation, Information & Knowledge*, 3, 58–66. <https://doi.org/10.13366/j.dik.2011.03.015>
- Daim, T. U., Rueda, G., Martin, H., & Gerdtsri, P. (2006). Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technological Forecasting and Social Change*, 73(8), 981–1012. <https://doi.org/10.1016/j.techfore.2006.04.004>
- Ferreira, F. A. F. (2018). Mapping the field of arts-based management: Bibliographic coupling and co-citation analyses. *Journal of Business Research*, 85, 348–357. <https://doi.org/10.1016/j.jbusres.2017.03.026>

- Gmür, M. (2003). Co-citation analysis and the search for invisible colleges: A methodological evaluation. *Scientometrics*, 57(1), 27–57. <https://doi.org/10.1023/A:1023619503005>
- Hasner, C., de Lima, A. A., & Winter, E. (2019). Technology advances in sugarcane propagation: A patent citation study. *World Patent Information*, 56(9), 16. <https://doi.org/10.1016/j.wpi.2018.09.001>
- Hou, J., Yang, X., & Chen, C. (2018). Emerging trends and new developments in information science: A document co-citation analysis (2009–2016). *Scientometrics*, 115(2), 869–892. <https://doi.org/10.1007/s11192-018-2695-9>
- Hsiao, T. M., & Chen, K. H. (2020). The dynamics of research subfields for library and information science: An investigation based on word bibliographic coupling. *Scientometrics*, 125(1), 717–737. <https://doi.org/10.1007/s11192-020-03645-9>
- Huang, M. H., & Chang, C. P. (2014). A comparative study on detecting research fronts in the organic light-emitting diode (OLED) field using bibliographic coupling and co-citation. *Scientometrics*, 102(3), 2041–2057. <https://doi.org/10.1007/s11192-014-1494-1>
- Huang, M. H., Chiang, L. Y., & Chen, D. Z. (2003a). Constructing a patent citation map using bibliographic coupling: A study of Taiwan's high-tech companies. *Scientometrics*, 58(3), 489–506. <https://doi.org/10.1023/B:SCIE.000006876.29052.bf>
- Huang, Z., Chen, H., Yip, A., Ng, G., Guo, F., Chen, Z. K., & Roco, M. C. (2003b). Longitudinal patent analysis for nanoscale science and engineering: country, institution and technology field. *Journal of Nanoparticle Research*, 5(3), 333–363. <https://doi.org/10.1023/A:1025556800994>
- Jiang, J., Shi, P., An, B., Yu, J., & Wang, C. (2017). Measuring the social influences of scientist groups based on multiple types of collaboration relations. *Information Processing & Management*, 53(1), 1–20. <https://doi.org/10.1016/j.ipm.2016.06.003>
- Jun, S., Sung Park, S., & Sik Jang, D. (2012). Technology forecasting using matrix map and patent clustering. *Industrial Management & Data Systems*, 112(5), 786–807. <https://doi.org/10.1108/02635571211232352>
- Kang, I. S., Na, S. H., Kim, J., & Lee, J. H. (2007). Cluster-based patent retrieval. *Information Processing & Management*, 43(5), 1173–1182. <https://doi.org/10.1016/j.ipm.2006.11.006>
- Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14(1), 10–25. <https://doi.org/10.1002/asi.5090140103>
- Kim, G., & Bae, J. (2017). A novel approach to forecast promising technology through patent analysis. *Technological Forecasting and Social Change*, 117, 228–237. <https://doi.org/10.1016/j.techfore.2016.11.023>
- Kuan, C. H., Chen, D. Z., & Huang, M. H. (2019). Bibliographically coupled patents: Their temporal pattern and combined relevance. *Journal of Informetrics*. <https://doi.org/10.1016/j.joi.2019.100978>
- Kuan, C. H., Huang, M. H., & Chen, D. Z. (2018). Missing links: Timing characteristics and their implications for capturing contemporaneous technological developments. *Journal of Informetrics*, 12(1), 259–270. <https://doi.org/10.1016/j.joi.2018.01.005>
- Kuusi, O., & Meyer, M. (2007). Anticipating technological breakthroughs: Using bibliographic coupling to explore the nanotubes paradigm. *Scientometrics*, 70(3), 759–777. <https://doi.org/10.1007/s11192-007-0311-5>
- Lai, K. K., & Wu, S. J. (2005). Using the patent co-citation approach to establish a new patent classification system. *Information Processing & Management*, 41(2), 313–330. <https://doi.org/10.1016/j.ipm.2003.11.004>
- Lee, K., & Lee, J. (2020). National innovation systems, economic complexity, and economic growth: Country panel analysis using the US patent data. *Journal of Evolutionary Economics*, 30(4), 897–928. <https://doi.org/10.1007/s00191-019-00612-3>
- Leydesdorff, L., Kushnir, D., & Rafols, I. (2012). Interactive overlay maps for US patent (USPTO) data based on International Patent Classification (IPC). *Scientometrics*, 98(3), 1583–1599. <https://doi.org/10.1007/s11192-012-0923-2>
- Liu, W., Nanetti, A., & Cheong, S. A. (2017). Knowledge evolution in physics research: An analysis of bibliographic coupling networks. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0184821>
- Lo, S. C. (2007). Patent coupling analysis of primary organizations in genetic engineering research. *Scientometrics*, 74(1), 143–151. <https://doi.org/10.1007/s11192-008-0110-7>
- McCain, K. W. (1990). Mapping authors in intellectual space: A technical overview. *Journal of the American Society for Information Science*, 41(6), 433–443. [https://doi.org/10.1002/\(SICI\)1097-4571\(199009\)41:6%3c433::AID-ASII1%3e3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1097-4571(199009)41:6%3c433::AID-ASII1%3e3.0.CO;2-Q)
- Narin, F. (1994). Patent bibliometrics. *Scientometrics*, 30(1), 147–155. <https://doi.org/10.1007/BF02017219>
- Nerur, S. P., Rasheed, A. A., & Natarajan, V. (2008). The intellectual structure of the strategic management field: An author co-citation analysis. *Strategic Management Journal*, 29(3), 319–336. <https://doi.org/10.1002/smj.659>

- Noh, H., Jo, Y., & Lee, S. (2015). Keyword selection and processing strategy for applying text mining to patent analysis. *Expert Systems with Applications*, 42(9), 4348–4360. <https://doi.org/10.1016/j.eswa.2015.01.050>
- Noh, H., Song, Y. K., & Lee, S. (2016). Identifying emerging core technologies for the future: Case study of patents published by leading telecommunication organizations. *Telecommunications Policy*, 40(10), 956–970. <https://doi.org/10.1016/j.telpol.2016.04.003>
- Park, A., Conway, M., & Chen, A. T. (2018a). Examining thematic similarity, difference, and membership in three online mental health communities from reddit: A text mining and visualization approach. *Computers in Human Behavior*, 78, 98–112. <https://doi.org/10.1016/j.chb.2017.09.001>
- Park, I., Jeong, Y., Yoon, B., & Mortara, L. (2014). Exploring potential R&D collaboration partners through patent analysis based on bibliographic coupling and latent semantic analysis. *Technology Analysis & Strategic Management*, 27(7), 759–781. <https://doi.org/10.1080/09537325.2014.971004>
- Park, I., Jeong, Y., Yoon, B., & 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(7), 759–781. <https://doi.org/10.1080/09537325.2014.971004>
- Park, T. Y., Lim, H., & Ji, I. (2018b). Identifying potential users of technology for technology transfer using patent citation analysis: A case analysis of a Korean research institute. *Scientometrics*, 116(3), 1541–1558. <https://doi.org/10.1007/s11192-018-2792-9>
- Pénin, J. (2012). Strategic uses of patents in markets for technology: A story of fabless firms, brokers and trolls. *Journal of Economic Behavior & Organization*, 84(2), 633–641. <https://doi.org/10.1016/j.jebo.2012.09.007>
- Rodriguez, A., Kim, B., Turkoz, M., Lee, J. M., Coh, B. Y., & Jeong, M. K. (2015). New multi-stage similarity measure for calculation of pairwise patent similarity in a patent citation network. *Scientometrics*, 103(2), 565–581. <https://doi.org/10.1007/s11192-015-1531-8>
- Shen, J., Gao, J., & Teng, L. (2012). Derwent manual code co-occurrence: A practical method in patent map. *Science of Science and Management of S & T*, 33(1), 12–16.
- Small, H. (1973). Co-citation in the scientific literature a new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(24), 265–269. <https://doi.org/10.1002/asi.4630240406>
- Small, H., & Griffith, B. C. (1974). The structure of scientific literatures I: Identifying and graphing specialties. *Science Studies*, 4(1), 17–40.
- Small, H. G., & Koenig, M. E. D. (1977). Journal clustering using a bibliographic coupling method. *Information Processing & Management*, 13(5), 277–288. [https://doi.org/10.1016/0306-4573\(77\)90017-6](https://doi.org/10.1016/0306-4573(77)90017-6)
- Song, Y., & Wu, Y. (2014). A comparative study on author bibliographic-coupling analysis and author keyword-coupling analysis based on scientometrics. *The Journal of the Library Science in China*, 40(1), 25–38.
- Swanson, D. R. (1971). Some unexplained aspects of the cranfield tests of indexing performance factors. *The Library Quarterly*, 41(3), 223–228. <https://doi.org/10.1086/619959>
- Tseng, Y. H., Lin, C. J., & Lin, Y. I. (2007). Text mining techniques for patent analysis. *Information Processing & Management*, 43(5), 1216–1247. <https://doi.org/10.1016/j.ipm.2006.11.011>
- Von Wartburg, I., Teichert, T., & Rost, K. (2005). Inventive progress measured by multi-stage patent citation analysis. *Research Policy*, 34(10), 1591–1607. <https://doi.org/10.1016/j.respol.2005.08.001>
- Wang, J., & Hsu, C. C. (2020). A topic-based patent analytics approach for exploring technological trends in smart manufacturing. *Journal of Manufacturing Technology Management*, 32(1), 110–135. <https://doi.org/10.1108/jmtm-03-2020-0106>
- Wen, F. (2017). Research on the technology diversity and similarity based on the coupling of derwent patent classification codes. *Information Studies: Theory & Application*, 40(8), 87–92. <https://doi.org/10.16353/j.cnki.1000-7490.2017.08.016>
- White, H. D. (2003). Author cocitation analysis and Pearson's r. *Journal of the American Society for Information Science and Technology*, 54(13), 1250–1259. <https://doi.org/10.1002/asi.10325>
- White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32(3), 163–171. <https://doi.org/10.1002/asi.4630320302>
- Yang, S., Han, R., Wolfram, D., & Zhao, Y. (2016). Visualizing the intellectual structure of information science (2006–2015): Introducing author keyword coupling analysis. *Journal of Informetrics*, 10(1), 132–150. <https://doi.org/10.1016/j.joi.2015.12.003>
- Yeh, H. Y., Sung, Y. S., Yang, H. W., Tsai, W. C., & Chen, D. Z. (2012). The bibliographic coupling approach to filter the cited and uncited patent citations: A case of electric vehicle technology. *Scientometrics*, 94(1), 75–93. <https://doi.org/10.1007/s11192-012-0820-8>

- Yoon, B., & Magee, C. L. (2018). Exploring technology opportunities by visualizing patent information based on generative topographic mapping and link prediction. *Technological Forecasting and Social Change*, *132*, 105–117. <https://doi.org/10.1016/j.techfore.2018.01.019>
- Zhang, K., Xia, W., Yuan, J., Chen, J., & Geng, Y. (2015). Study on the definition, types and characteristics of NPEs. *Science and Technology Management Research*, *35*(15), 141–146+151.
- Zhang, Y., Shang, L., Huang, L., Porter, A. L., Zhang, G., Lu, J., & Zhu, D. (2016). A hybrid similarity measure method for patent portfolio analysis. *Journal of Informetrics*, *10*(4), 1108–1130. <https://doi.org/10.1016/j.joi.2016.09.006>
- Zhao, D., & Strotmann, A. (2008a). Evolution of research activities and intellectual influences in information science 1996–2005: Introducing author bibliographic-coupling analysis. *Journal of the American Society for Information Science and Technology*, *59*(13), 2070–2086. <https://doi.org/10.1002/asi.20910>
- Zhao, D., & Strotmann, A. (2008b). Information science during the first decade of the web: An enriched author cocitation analysis. *Journal of the American Society for Information Science and Technology*, *59*(6), 916–937. <https://doi.org/10.1002/asi.20799>
- Zhao, D., & Strotmann, A. (2014). The knowledge base and research front of information science 2006–2010: An author cocitation and bibliographic coupling analysis. *Journal of the Association for Information Science and Technology*, *65*(5), 995–1006. <https://doi.org/10.1002/asi.23027>
- Zitt, M., & Bassecoulard, E. (2006). Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences. *Information Processing & Management*, *42*(6), 1513–1531. <https://doi.org/10.1016/j.ipm.2006.03.016>

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.