A Novel Indicator for Measuring Science-technology Linkage Based on Paper-patent Co-cited

Yiru Liu[1,a], Ruizhe Zhou[2,b], Hanyu Zhang[2,c], Huijun Sun[2,d*]

[1]Southwestern University of Finance and Economics Library, Chengdu, 611130, China
[2]China Agricultural University Library, Beijing, 100193, China

[a]liuyiru@swufe.edu.cn, [b]readchow@163.com
[c]19510175328@163.com, [d*]tsg003@cau.edu.cn

Abstract. This paper proposes a novel method for measuring the linkage between science and technology (S&T) based on the co-citation of patents and academic papers. This method differs significantly from the Science Linkage (SL) index, a science-based index first proposed by Narin. By analyzing the co-citation of patents and papers in references, we can detect the relationship of S&T. This method uses the co-citation strength of papers and patents in references to characterize the strength of S&T linkage. The citations of patent databases in some countries and regions are incomplete, making it difficult to calculate the relevance of S&T by the SL method based on direct citation. However, the method based on patent-paper co-citation proposed in this paper is effective. The author also applied the approach to patent-citing documents registered between 2001 and 2015 from the United States, Germany, France, Australia, South Korea, and Canada (hereinafter referred to as the five countries) and checked the performance. Results showed that the proposed approach and the indirect S&T linkages (indSL), a new indicator, are valuable for detecting the relationship of S&T.

Keywords: patent, paper, co-citation, science-technology linkage.

1 Introduction

The importance of the interaction between science and technology (S&T) for economic growth and progress is indisputable[1].Toynbee compared science and technology to "a pair of dancers" to illustrate the close linkage between them. What is the nature of this linkage? Many scholars have offered their perspectives, including Carpenter, who viewed it as the dependence of technology on science, and Rip, who considered it to be the interaction between science and technology. Narin defined it as the knowledge transfer relationship.

Patents carry technical information and provide new ideas and methods in technology R&D[2]. Academic papers present basic research. As early as the 1960s, Price studied the S&T linkage. He pointed out that S&T have structural tightness and ac-

K. Zhang et al. (eds.), *Proceedings of the 4th International Conference on Management Science and Software Engineering (ICMSSE 2024)*, Advances in Engineering Research 244,
https://doi.org/10.2991/978-94-6463-552-2_5

cumulation, showing that new scientific achievements and technological progress are reflected in academic papers and patents[3].Since then, numerous scholars have conducted research on the S&T linkage via academic papers and patents.

2 Literature Reviews

In order to ascertain the relationships between S&T, several quantitative approaches have been proposed in the literature. These approaches are based on the premise that patent literature knowledge linkages can be identified through citation relations, author-inventor links, international patent classification (IPC)-journal classification mapping, and topic linkages. In essence, these approaches regard scientific publications and patents as proxies for the science and technology systems, respectively.

2.1 Utilising Citation Relations to Identify Science-technology Linkages

Citation-based linkages, such as non-patent references (NPRs) and patents cited in the literature, have been extensively studied in order to measure the increasing interdependencies and interactions between S&T[4-6]. As a direct and straightforward form of linkage, citations confirm and illustrate the social character of knowledge diffusion and transfer between S&T[7]. Yuhang Kang employed an NPR-based approach to identify the internal structure of heterogeneous knowledge flow. He posited that the fluidity of heterogeneous knowledge in this field is limited, and that the contribution of scientific knowledge to technological innovation is minimal[8]. Popp demonstrated that extensively cited academic literature is valuable for the creation of applied technology[9]. This was based on a large-scale patent dataset and NPRs. Meyer and Chen L. conducted a comprehensive analysis of citation patterns between S&T from multiple perspectives. Their findings revealed that the correlation degree of S&T in different subject areas varies considerably[10,11].

The above one-way analyses of S&T citation linkages offer limited insight into the interactions of S&T from a perspective of mutuality[12]. Specifically, they can only observe patent-to-paper citations or paper-to-patent citations, and they are unable to investigate two-way relationships between S&T. To this end, some scholars propose a cross-citation analytical method for investigating mutual relations between them. Jiping Gao implemented hybrid document co-citation analysis (HDCCA) to construct the patent citation data set, generating a co-occurrence matrix and co-citation network of paper citations and patent citations. These can reflect the interaction between science and technology and the impact of the interaction of science on future science and technology[13]. Furthermore, citation analysis can be employed to uncover the interconnections between science and technology, and to identify pivotal areas of scientific research that drive technological advancement. Despite the objective, accessible, and standardised nature of citation analysis, some researchers have questioned the reliability of this approach due to the inherent complexities and ambiguities associated with citation practices. It is suspected that the citations ultimately decided by the patent examiner may not truly represent the actual references of inventors[14].

2.2 Utilising Author-inventor Links to Identify Science-technology Linkages

Some scholars have observed the dual role of authors or inventors in the S&T systems and have matched inventors with authors in order to identify points of exchange between basic science and technology development[15,16]. For example, Li Rui incorporated the citations of different patent examiners, inventors' citations, and inventors' self-citations into the revised science-technology according to the different degrees of scientific-technical connection revealed by them through differential empowerment[17]. Meyer conducted an exploratory comparison of inventor-authors with their non-inventing peers in non-science and technology and concluded that inventor-authors were more productive and more highly cited than their peers who concentrate on scholarly publication[18].

A fundamental requirement for the author-inventor matching approach is that there must be a sample comprising a significant number of researchers who fulfil a dual role in the S&T system[19]. However, the number of author-inventors in practice is relatively small, which naturally places restrictions on the application of the approach to exploring the S&T overlap. Furthermore, the approach depends heavily on the accuracy of the name disambiguation process, which can be labour-intensive work for large data analysis. In addition, the approach is also unable to recognise semantic associations between S&T at the micro-content level.

2.3 Utilising IPC-ISI Journal Classification Mapping to Identify Science-technology Linkages

The matching of IPC with ISI journal classification can reveal similarities in developing trends and their directions between S&T at the domain level[20]. For example, Verbeek, Debackere and Luwel matched IPC 4-digit classes with ISI journal classifications to trace the IPC class of a given patent and the science-domain classification of the journal in which the "source" publication appeared[21]. Li Rui refined the S&T correlation model designed by Verbeek by building a two-way citation model and performing a validation analysis of the science and technology correlation in the field of catalysis. However, further investigation is required to ascertain whether an accurate subject correspondence between S&T, based on IPC-ISI journal classification linkage, can be achieved[17].

2.4 Paper-patent Co-citation in Science and Technology

The term 'co-citation' refers to the simultaneous citing of two documents by other documents. the co-citation of literature can be regarded as a certain knowledge association between them[22]. The greater the co-citation intensity, the closer their relationship will be[23].Paper-patent co-citation is a method used to measure the linkage between science and technology based on the co-occurrence of patents and papers in patent references. The greater the number of papers and patents that appear together, the closer the relationship between science and technology. Ji Ping Gao constructed the patent citation data by setting out a method for generating the co-occurrence matrix and co-citation network of paper citations and patent citations. This method was used to reflect the interaction between science and technology.

However, the patent citation information of some countries, including China, is incomplete. Furthermore, it is challenging to obtain citation data directly from patent references. The co-citation of papers and patents based on paper datasets effectively addresses this practical issue. Currently, the majority of references in papers are described in a standardized manner, which also facilitates the study of the relationship between science and technology.

Despite the numerous limitations, criticisms and disputations, the method of citation analysis is a feasible approach to studying the correlation between science and technology in different countries and regions, as well as across different disciplines. A significant body of research also supports the scientific validity of measuring the degree of science and technology correlation through patent analysis. The citation method is a valuable tool for studying the relationship between science and technology. However, there is a need for more innovative research methods in this field. Currently, most research is quantitative, focusing on specific fields. There is a lack of research on the relationship between science and technology based on the co-citation of papers and patents. In particular, many national patent databases, including those in China, lack standardized management, lack patent citation information and non-standard descriptions, which makes it challenging for researchers to obtain direct scientific relevance through patent citation information.

The paper-patent hybrid co-citation method enables the study of the correlation between science and technology[24]. This study employs a co-citation network of papers and patents to measure the co-citation frequency of patents and papers and to quantify the relationship between them. This approach enables the characterization of the relationship between science and technology. It is hypothesized that the paper-patent co-citation can reduce the noise impact caused by direct citation and that the linkage between science and technology can be detected. The empirical research seeks to gather supporting evidence for this hypothesis.

3 Research Framwork and Methodology

3.1 Research Framwork

This article intends to start from the existing problems in the science-technology linkage research methods. And on the basis of empirical testing, we find a method to detect the science-technology linkage from the view of co-citation. The technical route of this research is shown in Fig. 1.

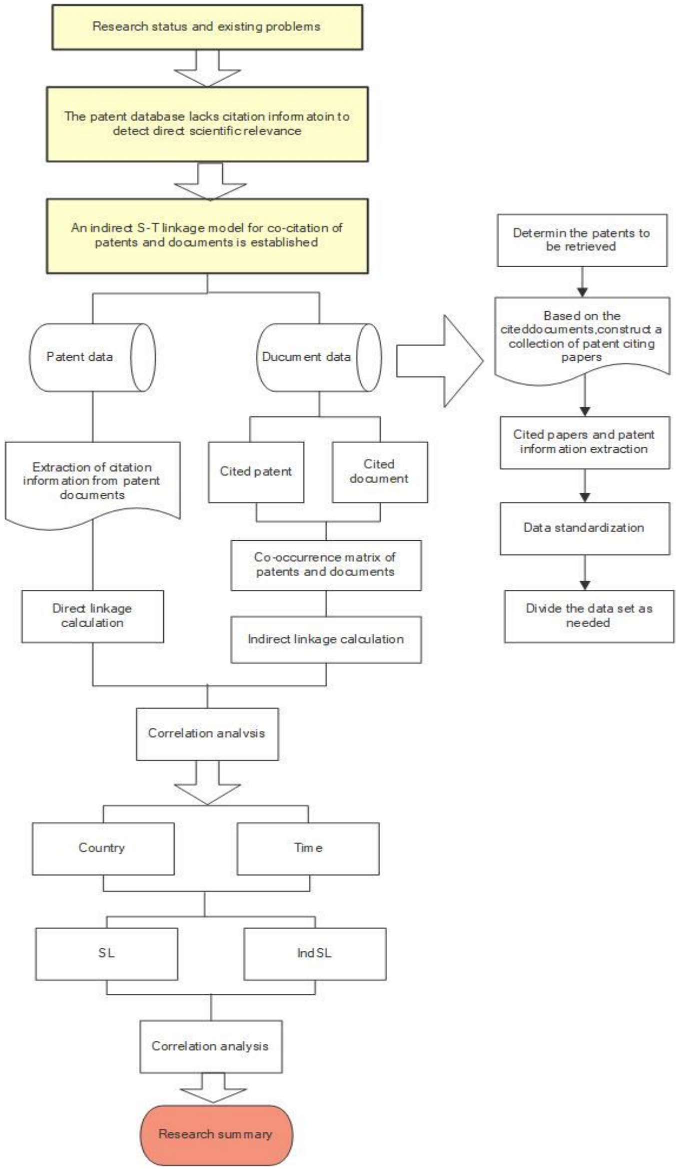

Fig. 1. Schematic diagram of science-technology linkage based on co-cited

3.2 Data Sources

The scientific document data in this study was sourced from the core collection of Web of Science (WOS). In WOS, the "Cited Reference Search" function was employed to identify documents citing patents in the United States and five countries. The search was limited to documents citing patents from 2001 to 2015. A total of 6922 American documents citing patents and 1362 documents citing patents from other five countries were retrieved. The following data, which has been subjected to cleaning, is presented in Fig. 2.

NEMETH L. T., 2001, [No title captured], Patent No. [US 6.288.281, 6288281]	6288281	US6288281
Pakulski M. K., 2001, US Patent, Patent No. [6 331 508, 6331508]	6331508	US6331508
Perrier E., 2001, Patent No. [6235294 B1, 6235294, WO 6,235,294 B1.]	6235294B1	
Rongved P. I., 2001, US, Patent No. 6180012	6180012	US6180012
Rudolf C., 2001, EP patent, Patent No. [6,218,000, 6218000]	6218000	
SALEH R. Y., 2001, [No title captured], Patent No. [US 6.320.083, 6320083]	6320083	US6320083
Siemens Automotive Corporation, 2001, US, Patent No. [6758408B2, 6758408]	6758408B2	US6758408B2
Silenius P., 2001, US, Patent No. [6251222, 6251222US]	6251222US	US6251222US
Sun D.-C., 2001, US, Patent No. [6174934 B1, 6174934] Non-American Patent Information	6174934B1	US6174934B1
Tanaka Y., 2001, US, Patent No. 6613298B2	6613298B2	US6613298B2
Tomka I., 2001, US, Patent No. [6,242102, 6242102]	6242102	US6242102
Yuying T, 2001, Patent No. [US 6231854, 6231854]	US6231854	US6231854
Zabinski J, 2001, MAGNETRON SPUTTER PU, Patent No. US 2001H1933	US2001H1933	US2001H1933
Zavattari C., 2001, US Patent, Patent No. 6231628	6231628	US6231628

Fig. 2. Schematic diagram of document citation patent information cleaning

The patent data for this study was derived from Derwent Innovation (DI). The reference information was extracted from Web of Science scientific documents to obtain patent numbers, and then the patent data was obtained from the Derwent Patent Database. A total of 8,216 American patents and 1,734 patents from five countries were obtained.

The patent information was then standardised. At the same time, the detected patents were cleaned again. Any patents that fail to meet the requisite criteria are excluded, and the requisite fields are added to the retrieved patent information, which is then saved in the Derwent Data Analyzer format, as shown in Table 1.

Table 1. Patent format standardization

Original patent format	Standardized patent format
Velocys Inc, 2006, US, Patent No. [7084180, US 7,084,180]	US7084180
Yu W. C., 1997, US, Patent No. 5691077	US5691077
Aiello M. F., 2001, United States Patent, Patent No. [US 006236580 B1, 006236580]	US006236580B1

Following preprocessing, the saved patent data is imported into Derwent Data Analyzer format within DDA. The document information in patent citations is then extracted directly using the "Cited Refs-Non patent" function.

3.3 Calculation of Scientific-technological Linkage Degree

The concept of Science Linkage (SL), as proposed by the American scholar Narin, is defined as the ratio of the number of documents cited by a patent to the number of patents, expressed as:

$$SL = PT_{num} / P_{num} \tag{1}$$

Where SL is the degree of science linkage, PT_{num} is the number of scientific documents cited by the patent, and P_{num} is the number of patents.

The patent citation analysis method is employed to identify the scientific-technological relationship, ascertain its quantitative characteristics and internal laws, and determine the relationship between technology and science based on this[20]. The principle of direct science linkage is illustrated in Fig. 3.

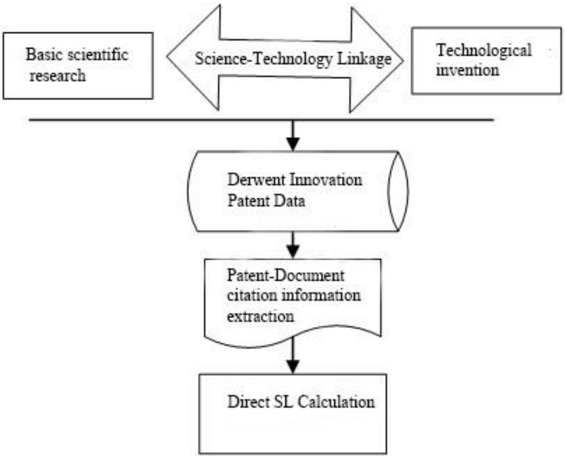

Fig. 3. Principle of direct scientific linkage

Document co-citation can be used to characterise the interrelationship between different knowledge domains. This article proposes a hypothesis that the science-technology linkage can be detected through the co-occurrence of patents and documents. The method uses the absolute co-citation strength of the documents and patents in the references to characterise the strength of the science-technology linkage. The science-technology linkage based on co-citation is referred to as the "ind Science Linkage" (indSL), expressed as:

$$indSL_1 = \frac{1}{mn} \sum_{1}^{mn} \frac{U_i^2}{CP_m \bullet A_n} \tag{2}$$

Where Ui is the number of times the document and patent are cited in the literature, CP_m and A_n are the number of patents and documents cited in the literature, and *indSL₁* is the degree of indirect science linkage.

In order to obtain a scientific and logical indirect science linkage index, two indirect science linkage algorithms are added for comparative research, and a more reasonable indirect science linkage algorithm is obtained, expressed as:

$$indSL_2 = \frac{1}{mn}\sum_{1}^{mn}\frac{U_i}{\sqrt{CP_m * A_n}} \tag{3}$$

Where U_i is the number of times the document and patent are cited in the literature, CP_m and A_n are the number of patents and documents cited in the literature, and $indSL_2$ is the degree of indirect science linkage.

The third indirect science linkage algorithm formula is:

$$indSL_3 = \sum CA_a / \sum CP_b \tag{4}$$

Where CA_a is the frequency of cited documents, CP_b is the frequency of cited patents, and $indSL_3$ is the degree of indirect science linkage.

The hypothesis's validity is then assessed through a linkage analysis between the indirect and direct scientific relevance. If the degree of linkage is high, the hypothesis is deemed to be true. The indirect science linkage degree based on document-patent co-citation can therefore be regarded as an important measurement indicator for the research of science-technology linkage. The principle of indirect science-technology linkage is illustrated in Fig. 4.

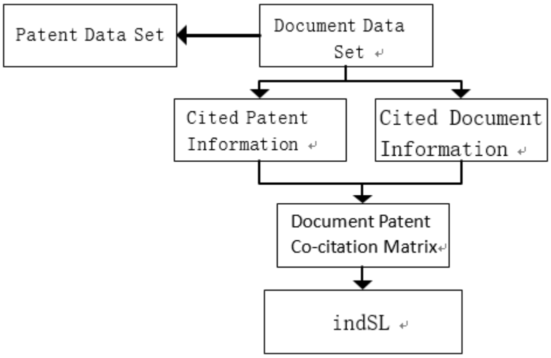

Fig. 4. Principle of indirect science linkage

3.4 Calculation of Linkage Coefficient

The linkage coefficient is a frequently employed metric for describing the degree of linkage between two variables. In this study, the Pearson product-moment coefficient is employed to determine the linkage. The product-moment coefficient (r) is capable of accurately reflecting the degree of linear linkage between two variables in numerical form.

If r is greater than 0.8, there is a strong linear linkage between the two variables; If r is greater than 0.2, there is a weak linear linkage between the two variables; If r is less than 0, there is a negative linear linkage between the two variables, expressed as:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2 \sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{5}$$

4 Results

4.1 Indirect Science Linkage Analysis between the United States and the Five Countries

The co-occurrence frequency of patents and documents, and their respective citations, are considered in indSL in order to explore the correlation between science and technology. The document data set is used to measure the linkage between basic research and technological innovation in different countries and in different time dimensions. Fig. 5 describes the knowledge process in the citation network[20]. D1~D5 respectively represent a document (patent). The connection between the two indicates that there is a co-citation relationship. The number (1-5) on the connection indicates the number of co-citations.

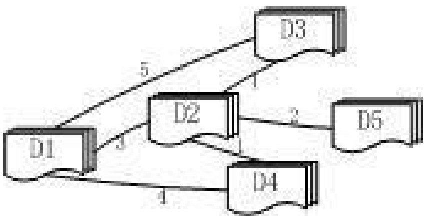

Fig. 5. Document-Patent Co-citation Network

In accordance with the aforementioned principles, the fundamental aspect of document-patent co-citation is the delineation of the interrelationship between disparate bodies of knowledge. It is commonly accepted that if two articles are cited by other articles concurrently, it is probable that their themes are interrelated. This suggests the existence of a knowledge connection between the co-cited documents[12].Co-citation analysis is conducted on the basis of patent citing documents. The patents and documents referenced by citing documents are employed for co-citation analysis, with the objective of exploring the correlation between science and technology.

4.1.1 Overview of American Patent Citations and Analysis of indSL.

Following the processing and cleansing of the retrieved document data, and the removal of incomplete and erroneous data, a total of 6,687 documents that cited US patents from 2001 to 2015 were obtained.

4.1.1.1 Structure a Co-citation Matrix.

The obtained patent and document co-citation datasets are grouped into different years in which the patent was cited, and a document-patent co-citation matrix is created.

Using American patents as a sample, a co-citation matrix of documents and patents is structured. Fig. 6 shows the co-occurrence matrix of documents referenced by documents and American patents from 2001 to 2015. The horizontal axis represents the cited document information, and the vertical axis represents the cited patent information. The number of records in the matrix indicates the total number of appearances of the patents or documents, and the number in the middle part represents the frequency of co-citation of the patents or documents. In the co-citation matrix, any patent or document that is co-cited will appear in the matrix, while cases where they are not co-cited with other documents or patents will be left blank. As depicted in Fig. 6.

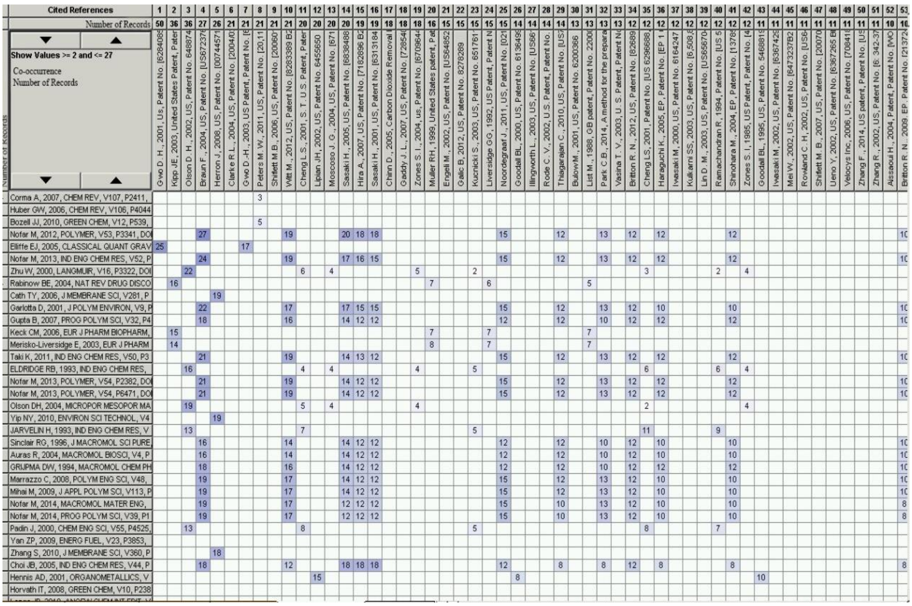

Fig. 6. American Document-Patent Co-citation Matrix

4.1.1.2 indSL in American Documents and Patents.

The frequency of documents and patents in the co-occurrence matrix, along with the quantitative correlation between co-occurrences of the two, indicate that the three mathematical formulas proposed in this study are employed to assess the indirect science-technology linkage under varying co-citation frequencies (the figure is precise to four decimal places). As illustrated in Table 2.

Table 2. indSL between different co-citation frequencies in the United States from 2001 to 2015

Year	Frequency	indSL1	indSL2	indSL3	Year	Frequency	indSL1	indSL2	indSL3
	≧2	0.6364	0.7630	0.8041		≧2	0.8929	0.9359	1.0191
2001	≧3	0.6616	0.7847	0.6909	2009	≧3	0.8985	0.9410	0.9988
	≧4	0.6934	0.8087	0.5930		≧4	0.9278	0.9591	0.9651
	≧2	0.7071	0.8084	0.8246		≧2	0.8657	0.9212	1.0683
2002	≧3	0.6861	0.7900	0.8071	2010	≧3	0.8563	0.9161	1.0763
	≧4	0.6409	0.7587	0.7166		≧4	0.8391	0.9057	1.0058
	≧2	0.7712	0.8535	0.8850		≧2	0.8852	0.9292	1.0404
2003	≧3	0.8458	0.9067	0.8424	2011	≧3	0.8897	0.9333	1.0716
	≧4	0.7712	0.8535	0.7307		≧4	0.8859	0.9312	1.0521
	≧2	0.7442	0.8286	0.7847		≧2	0.8514	0.9083	0.9448
2004	≧3	0.7590	0.8406	0.7644	2012	≧3	0.7844	0.8663	0.9475
	≧4	0.7726	0.8494	0.7165		≧4	0.7844	0.8663	0.9475
	≧2	0.8331	0.8949	1.0498		≧2	0.9037	0.9424	1.0461
2005	≧3	0.8149	0.8843	0.9970	2013	≧3	0.8964	0.9401	1.0924
	≧4	0.8281	0.8931	0.9826		≧4	0.8863	0.9349	1.0857
	≧2	0.8081	0.8811	0.9764		≧2	0.8886	0.9317	0.9986
2006	≧3	0.8529	0.9134	0.9436	2014	≧3	0.9215	0.9550	1.1205
	≧4	0.8500	0.9130	0.8488		≧4	0.9328	0.9613	1.1490
	≧2	0.8929	0.9359	1.0191		≧2	0.9701	0.9830	0.9724
2007	≧3	0.8121	0.8806	0.9683	2015	≧3	0.9712	0.9838	1.0316
	≧4	0.8224	0.8892	0.8103		≧4	0.9709	0.9836	1.0964
	≧2	0.8073	0.8795	0.9966					
2008	≧3	0.7595	0.8518	0.9066					
	≧4	0.8224	0.8938	0.8696					

In order to facilitate a more intuitive comparison of the indirect science linkage of the three algorithms in different years, Fig. 7, 8 and 9 have been constructed based on the frequency of the three co-citations.

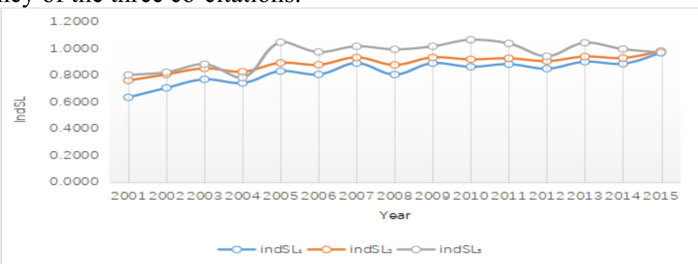

Fig. 7. American document-patent indirect science linkage with citation frequency ≧2 times

The highest values for indSL1 and indSL2 were observed in 2015, at 0.9701 and 0.9830, respectively. Both indices consider the frequency of documents and patents, as well as the number of common citations between the two. The highest value for indSL3 was observed in 2005, at 1.0498. This index assigns a greater weight to the frequency of documents and patents. This method solely considers the frequency of documents and patents in the co-citation matrix.

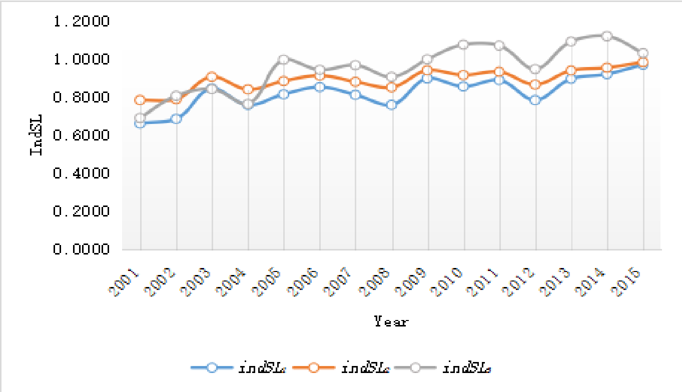

Fig. 8. American document-patent indirect science linkage with citation frequency ≧3 times

indSL1 and *indSL2* reached the highest value in 2015, 0.9712 and 0.9838 respectively; *indSL3* reached the highest value of 1.1205 in 2014.

Fig. 9. American document-patent indirect science linkage with citation frequency ≧4 times

The highest values for indSL1 and indSL2 were observed in 2015, at 0.9709 and 0.9836, respectively. The highest value for the indirect science-technology linkage degree was observed in indSL3 in 2014, at 1.1490.

In the United States, the indirect science linkage (indSL) indices demonstrate a similar pattern of change when documents and patents are cited at least twice, three times, or four times. The indSL1 and indSL2 indices exhibit a consistent trend, whereas the indSL3 index exhibits a different pattern. The indSL1 and indSL2 indices consider the respective frequency of documents and patents and the number of co-citations of the two, while the indSL3 index only considers the total frequency of documents and patents in the co-citation matrix. Which indirect science linkage algorithm under which co-citation frequency is more suitable as a quantitative indicator for researching science-technology linkage requires further analysis and discussion.

4.1.2 Overview of Citations of Documents from Other Five Countries and Analysis of Indirect Science Linkage.

Following the processing and cleansing of the retrieved document data, which involved the removal of incomplete and erroneous data, a total of 1,255 documents citing patents from five countries were obtained between the years 2001 and 2015.

4.1.2.1 Structure a Co-citation Matrix.

The obtained patent and document co-citation data sets are grouped into different years in which the patent was cited by DDA, and a co-citation matrix is created. As shown in Fig. 10.

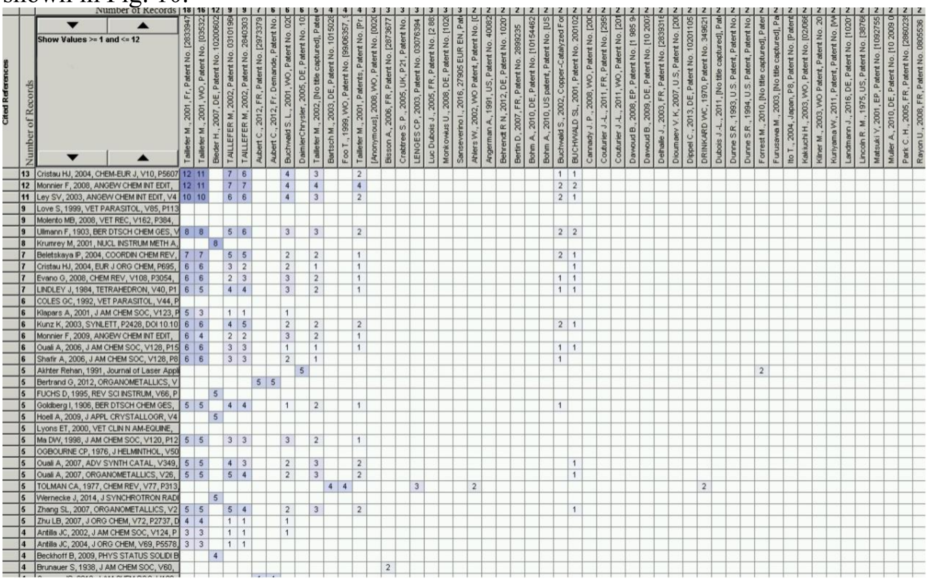

Fig. 10. Five countries' Document-Patent Co-citation Matrix

4.1.2.2 indSL in the Five Countries.

According to the frequency of the documents and patents in the co-occurrence matrix and the quantitative correlation between the co-occurrences of the two, (The figure is precise to four decimal).

The three mathematical formulas proposed in this research are used to measure the indirect science-technology linkages of the three algorithms in which documents and patents from five countries have been cited ≧2 times. As shown in Table 3.

Table 3. Documents and patents cited≧2 times of *3 indSL* in five countries from 2001 to 2005

Year	$indSL_1$	$indSL_2$	$indSL_3$
2001	0.2030	0.4300	0.3864
2002	0.3166	0.5321	0.6617
2003	0.7344	0.8483	0.8905
2004	0.8500	0.9174	1.2250
2005	0.7090	0.8332	0.6571
2006	0.6491	0.8039	0.7018
2007	0.3166	0.5268	0.5533
2008	0.8182	0.8999	0.9767
2009	1.0000	1.0000	1.0000
2010	0.9232	0.9553	1.2333
2011	0.9400	0.9670	1.1286
2012	0.4723	0.6721	0.5185
2013	1.0000	1.0000	1.0000
2014	1.0000	1.0000	1.0000

We create Fig. 11 to compare the indirect science linkage of the three algorithms more intuitively.

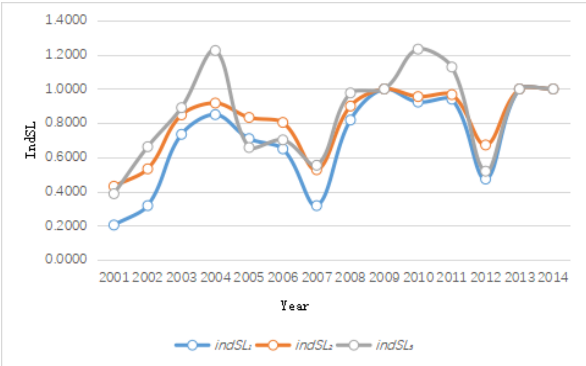

Fig. 11. *indSL* in five countries

4.1.3 Correlation Analysis between Indirect Science Linkage and Direct Science Linkage.

The correlation coefficient is the earliest statistical indicator designed by statistician Carl Pearson. It is the amount of linear correlation between the study variables. The degree of linear correlation between two variables can be accurately reflected in the form of numerical values, generally with letters r means.

In this study, if r>0.8, the two variables have a strong linear linkage;

if r>0.2, the two variables have a weaker linear linkage;

if r<0, the two variables have a negative linear linkage.

Table 4 shows the linkage between the indirect science linkage and the direct science linkage under different co-citation times of American patents and documents.

Table 4. The linkage between the indirect science linkage and the direct science linkage of America.

		SL	$indSL_1$	$indSL_2$	$indSL_3$
SL	*Pearson* Correlation	1	.698**	.680**	.313
	Significance (bilateral)		.004	.005	.255
	N	15	15	15	15

Science-technology linkage of the patents being cited ≥2 times

		SL	$indSL_1$	$indSL_2$	$indSL_3$
SL	*Pearson* Correlation	1	.620**	.597**	.624
	Significance (bilateral)		.014	.019	.013
	N	15	15	15	15

Science-technology linkage of the patents being cited ≥3times

		SL	$indSL_1$	$indSL_2$	$indSL_3$
SL	*Pearson* Correlation	1	.512	.489	.752**
	Significance (bilateral)		.051	.064	.001
	N	15	15	15	15

Science-technology linkage of the patents being cited ≥4 times

In the total number of citations of American patents ≥2 times, the indirect science-technology linkage degree indSL1 and the direct science linkage degree SL exhibit the highest linkage. The linkage coefficient is 0.698. The significance (bilateral) value is the lowest, which is 0.004. The linkage coefficients between the indirect science-technology linkage degrees indSL2 and indSL3 and the direct science linkage degree SL are 0.680 and 0.313, respectively, with significance (bilateral) values of 0.005 and 0.255, respectively. Among American patents cited for at least three times, the indirect science-technology linkage (indSL3) and the direct science linkage (SL) exhibited the highest linkage, with a linkage coefficient of 0.0624 and a significant (bilateral) value of 0.013.

The linkage coefficients of indSL1 and indSL2 and SL are 0.620 and 0.597, respectively, with the significance (bilateral) values being 0.014 and 0.019. Among American patents cited ≥4 times, indSL3 has the highest linkage value with SL, with a linkage coefficient of 0.752 and a significance (bilateral) value of 0.001.

The linkage coefficients of indSL1 and indSL2 and SL are 0.512 and 0.489, respectively, with significance values of 0.051 and 0.064, respectively. Based on these values, it is possible to identify indirect science linkages under different co-citation times of American patents and documents. This study posits that the indirect science-technology linkage, indSL1, with two or more citations of American patents is more suitable for studying the correlation between science and technology.

Table 5 presents the linkage test results of indirect science linkage and direct science linkage in five countries. The linkage coefficients of the indirect science-technology linkages indSL1, indSL2, and indSL3 of the five countries with the direct science linkage (SL) are 0.235, 0.225, and -0.383, respectively. The significance values for the bilateral relationships are 0.235, 0.438, and 0.176, which are comparable to the results of the indirect science linkage of American patents cited twice or more. IndSL1 has the highest linkage with direct science relevance (SL).

Table 5. Correlation analysis of direct science linkage degree and indirect science linkage degree in five countries

		SL	*indSL1*	*indSL2*	*indSL3*
SL	*Pearson Correlation*	*1*	*.235*	*.225*	*-.383*
	Significance (bilateral)		*.418*	*.438*	*.176*
	N	*14*	*14*	*14*	*14*

The findings of this study indicate that the results of the mathematical algorithm indSL1 are highly correlated with those of the traditional direct science linkage at the national level.

5 Conclusion

This article introduces a novel approach to measuring the relationship between science and technology through co-citation analysis. It proposes a new indicator, termed indirect science linkage (indSL), aimed at detecting this relationship. The effectiveness of this indicator is demonstrated through an analysis of data from the United States and five other countries. This approach enables the identification of the impact of technological innovation on basic scientific research, which has been steadily increasing. The results suggest that indSL1, focusing on documents and patents cited at least twice, serves as a more suitable quantitative indicator for studying science-technology linkage.

In terms of the practical implications of the findings, it is challenging to assert that the accuracies are sufficiently high. This challenge stems from the absence of a theoretical standard against which to assess the accuracy of this type of indicator. Moreover, it is always preferable in both theory and practice to employ more accurate methods.

Acknowledging the limitations of this research, including constraints on research abilities, knowledge reserves, and time, several shortcomings are evident: Firstly, the use of citation methods for analysis presents certain complexities. It is difficult, for instance, to ascertain the intent behind citing patents and the motivations driving such citations. Secondly, there is a need to develop effective methods for distinguishing between various citing motives and for curbing non-related citing behaviors. Thirdly, the sheer volume of data in this study poses a significant challenge in terms of data cleaning and processing. As a result, this study primarily addresses the scientific merit of the method at the national level, highlighting the need for further validation across various dimensions, including subject areas.

The aforementioned limitations can be regarded as an opportunity to develop related research in the future, intending to make it the main direction of future research. Additionally, the following two aspects can also be studied and discussed: The introduction of the document-patent co-citation analysis method into subject field analysis will facilitate the study of the science-technological linkage in different fields with a novel approach, thereby providing a reference for decision support in certain subjects. Furthermore, the attempt to introduce the document-patent co-citation method into the references of patent data, relying on the patents' data, will enable the exploration of the science linkage of document-patent co-citation. Additionally, differences in document-patent co-citation based on document data can be identified. This approach will provide a novel perspective and facilitate further advancement in research on the correlation between science and technology.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Beijing University Intellectual Property Information Public Service Capability Improvement Project "Beijing Intellectual Property Public Information Service Platform Data Collection, Processing and Training" (NO. BJIP2023130501076).

References

1. Dai Y, Zhang QX, and Zhu Gl. How the degree of scientific correlation and the degree of technical correlation affect the innovation performance of pharmaceutical companies. Scientific and technological progress and countermeasures 2023; 40(19):96-105.
2. Li Y, Ma YX, Zhang Y, and Sun JJ. Prediction of diffusion effect of scientific papers based on knowledge correlation degree: The role of early citation. Modern information 2023; 43(11):73-84.
3. Price, J D, De S. Is technology historically independent of science? A study in statistical historiography. The Political Economy of Science 2000:101-16.
4. Qi Y. Research on the development trend of domain science and technology correlation based on patent citation analysis -- A case study of hcv. Library and information work 2019; 63(23):97-105.

5. Guo Y, Gao JP. The impact of basic science on technological progress is studied from the perspective of patent citation of scientific literature. Chinese basic science 2023; 25(01):60-4.
6. Wang SW, Chen C. A review of potential knowledge discovery methods based on correlation between scientific papers and technical patents. Data analysis and knowledge discovery 2023; 7(07):18-31.
7. Li D, Azoulay P, Sampat BN. The applied value of public investments in biomedical research. Science 2017; 356(6333):78-81.
8. Kang YH. Research on heterogeneous knowledge flow network detection based on the perspective of fusion innovation. Information journal 2016; 35(09):963-70.
9. Popp D. From science to technology: the value of knowledge from different energy research institutions. Res Policy 2017; 46(9):1580-94.
10. Wu QY, Zhao SD. An empirical comparative study of patent citations from different citation subjects -- taking 5g communication technology as an example. Journal of Agricultural Library and Information 2021; 33(03):16-27.
11. Zhang Q. Research on patent-based biotechnology fusion analysis. Academy of Military Sciences; 2023.
12. Xu H, Winnink J, Yue Z, Liu Z, Yuan G. Topic-linked innovation paths in science and technology. J Informetr 2020; 14(2).
13. Gao JP, Yao CQ, and Ma Z et al. The application and deficiency of patent intercitation analysis method -- in the correlation analysis of science and technology. Science and management of science and technology 2014; 35(12):39-44.
14. Zhang AN. The characteristics and evaluation function of patent examiner citation. Chinese Journal of Medical Library and Information 2022; 31(02):36-41.
15. Wu FF, Li YW, Miao H. Prediction of cutting-edge topics in China's key technology fields based on technology R&D and basic research. Information journal 2022; 41(01):60-6.
16. Zhang G, Liu L, Wei F. Key nodes mining in the inventor-author knowledge diffusion network. Scientometrics 2019; 118(3):721-35.
17. Li, Meng LS. On the problems existing in the detection method of science-technology correlation based on patent citation. Information theory and practice 2010; 33(03):87-90.
18. Meyer M. Are patenting scientists the better scholars? An exploratory comparison of inventor-authors with their non-inventing peers in nano-science and technology. Res Policy 2006; 35(10):1646-62.
19. Hao MR. Analysis of Science-technology correlation in Artificial Intelligence based on Patent Citation. Nanjing University; 2019.
20. Han F, Magee CL. Testing the science/technology relationship by analysis of patent citations of scientific papers after decomposition of both science and technology. Scientometrics 2018; 116(2):767-96.
21. Verbeek A, Debackere K, Luwel M, Andries P. Linking science to technology: using bibliographic references in patents to build linkage schemes. Scientometrics 2002; 54(3):399-420.
22. Zhao ZY, Zhong YH, and Wang H et al. A review on the identification methods of potential R&D partners based on technology correlatio. Modern information 2023; 43(10):152-63.
23. Liu YR, Sun HJ. A review of the application and methods of non-patent citation in "science-technology" association research. Agricultural library and information 2019; 31(02):11-9.
24. Gao JP, Ding K, and Teng L et al. An analysis of knowledge flow in patent-paper hybrid co-citation network. Scientific research 2011; 29(08):1184-9.