Pre-grant knowledge burst Patent: Identification and Evolutionary Trajectory

Yang Xiucai　Tang Shiqi　Song Haoyang Hou Jianhua[1]
（School of Information Management of Sun Yat -sen University　Guangdong Guangzhou 510006）

Abstract: In order to explore the impact of the implementation of the "pre-grant disclosure system" on the diffusion of pre-grant knowledge. This study proposes the "pre-grant knowledge burst patent"(P-GKBP) and constructs a knowledge burst patent detection model(KBPDM) by citation, assignment and licensing lying in the process of patent knowledge diffusion. We retrieved 23,626 patents in the field of "Wireless Communication" as the basic data set Data A. Based on the KBPDM, it is calculated that there were 12,608 patents for pre-grant knowledge burst in Data A. Via the knowledge diffusion chain of Data A, we further collected 36031 citing patents (Data B) and 69,110 cited patents (Data C) of these 12,608 patents. Through KBPDM method, we analyze the intensity of knowledge burst before patent grant and the correlation of patent knowledge burst between different data sets. Additionally, we revealed the characteristics of P-GKBP and its impact on knowledge diffusion from several dimensions. Compared with the current advanced BHDDT method, the results show that the method this paper proposed has better applicability and effectiveness in knowledge diffusion trajectory burst detection. The findings recovered that the implementation of the "pre-grant disclosure system" made the development of the polarization of patent technology knowledge diffusion more obvious, and promoted the knowledge diffusion and earlier visibility of important patent technologies. In Data A, 53.36% of the patents had P-GKBP, while the excellence P-GKBP only accounts for 0.74% in all patents. When exploring the relationship between P-GKBP and citing patents and cited patents, it is found that P-GKBP reflected the characteristics of celebrity effect and Matthew effect; these effects were entirely different when an ordinary P-GKBP was cited, or an

[1] 作者简介：侯剑华(1980-)(ORCID：0000-0001-7080-7131)，男，辽宁北票人，教授，博士生导师，博士，研究方向为科学学、科学计量、专利计量、科技政策管理；E-mail：houjh5@mail.sysu.edu.cn；
杨秀财(1996-)(ORCID：0000-0002-1127-3288)，男，辽宁丹东人，博士研究生，研究方向为科学学、科学计量、专利计量、科技政策管理，E-mail：yangxc5@mail2.sysu.edu.cn；
唐诗琪(2000-)(ORCID:0000-0001-9955-9178)，女，浙江衢州人，硕士研究生，研究方向为科学学、专利计量，E-mail：tangshq7@mail2.sysu.edu.cn；
宋昊阳(1988-)(ORCID：0000-0003-4893-5676)，男，辽宁朝阳人，副教授，硕士生导师，博士，研究方向为科学计量、专利计量、科技管理 E-mail：songhy29@mail.sysu.edu.cn。
基金：本研究受到广东省自然科学基金一般项目支持[2021A1515012291]。

excellence P-GKBP was cited. When a patent cited a P-GKBP, citing an excellence P-GKBP was more likely to produce a pre-grant knowledge burst into an ordinary P-GKBP and an excellence P-GKBP than citing an ordinary P-GKBP. This provided a new perspective for researchers to track emerging technologies and estimate future trends in technology. In addition, identifying P-GKBP could aid in shortening technology innovation cycles and identify disruptive technologies.

Keywords: pre-grant knowledge burst patent; patent diffusion trajectory; knowledge burst; technological uncertainty

1. Introduction

Since the Dutch Patent Law formally recognized the “early disclosure and delayed examination system” in 1964 (i.e., the pre-grant disclosure system), the “pre-grant disclosure system” has been adopted by an increasing number of countries, including China, the United States, France, and Germany. Scholars have investigated the impact of the “pre-grant disclosure system” on technology, economy, and knowledge diffusion before and after the implementation, revealing that early disclosure of useful technical information to the public could facilitate the diffusion of technical knowledge and avoid duplication of R&D activities (Johnson & Popp, 2001; Okada & Nagaoka, 2020; Plečnik, 2021). At the same time, many studies on the diffusion of knowledge after patent authorization have reported that early discovery of patents is crucial for improving and expanding patent protection, stimulating technological change, and accelerating the evolution of technological innovation (Coccia & Watts, 2020; Goldsby et al., 2016; Coccia & Watts, 2020). However, there exists a paucity of studies investigating the impact of the “pre-grant disclosure system” on the diffusion of knowledge about patented technologies, as well as revealing the characteristics of pre-grant knowledge diffusion in the context of this system. It remains unclear whether the implementation of a “pre-grant disclosure system” will exert a high impact on patents from the application to pre-grant. No study focused specifically on the impact and characteristics of knowledge diffusion during the period from the publication of the application to the authorization (before authorization). Therefore, this study aims to address the following two questions: (1) What impact will the implementation of the "pre-grant disclosure system" have on the diffusion of pre-grant knowledge? (2) What are the characteristics of P-GKBP in the knowledge diffusion chain and how to promote the development of technological innovation?

In order to address these two problems, we construct a knowledge burst patent detection model(KBPDM) based on patent citations and transfer, and attempts to measure the influence of the implementation of the “pre-grant disclosure system” on the pre-grant knowledge diffusion of patent technology by identifying the burst of knowledge before patent grant. In addition, we calculate the intensity and correlation of the citing and cited patents of the “pre-grant knowledge burst patent”(P-GKBP). Besideswe figured out the impact of the P-GKBP influenced by the implementation of the “pre-grant disclosure system” on shortening technological innovation cycle and identifying disruptive technologies. In summary, the main contributions of this research are as follows:

(1) We identified P-GKBP through the KBPDM we constructed, and revealed the specific impact of the implementation of the “pre-grant disclosure system” on the diffusion of pre-patent knowledge. In this study, KBPDM is constructed to identify P-GKBP using patent citation and transfer indicators. In addition, based on the 23626 patent data (Data A) in the “Wireless Communication” technical field published in the IncoPat patent database from 1996 to 2015, the accuracy and applicability of the model are measured. We further reveal the specific impact of the implementation of the “pre-grant disclosure system” on the proliferation of patented technology. Besides, we reveal the ordinary characteristic laws of P-GKBP; this is the supplement and expansion of the knowledge diffusion theory under the existing intellectual property system.

(2) We find that P-GKBPs have “Celebrity effect” and “Matthew effect” in the process of knowledge diffusion and are more likely to promote technological innovation. Based on the 36031 citing patents (Data B) and 69,110 cited patents (Data C) of P-GKBP in Data A, the relationship and influence between P-GKBP, citing patents and cited patents are revealed. We find that the P-GKBP exerts apparent “celebrity effect” and “Matthew effect” for its citing and cited patents. And compared with patents issued by unlicensed knowledge, patents issued by pre-grant knowledge are easier to promote technological innovation, reflecting their high value in several aspects such as revealing hotspot technology, identifying disruptive technologies, recognizing cutting-edge technology, and speculating the future technology trend.

This paper is organized as follows. Section 2 reviews the related literature of “pre-grant disclosure system” on the impact of knowledge diffusion, patent knowledge diffusion and knowledge diffusion trajectory types as well as methods, and points out the similarities and differences between this research and the existing research. Section 3 introduces the data acquisition and processing process in details. It also explains how this study constructs a patent knowledge burst detection model through patent citation, transfer and other indicators.

Section 4 reports research results, mainly from three aspects: （1） Identification and distribution of P-GKBP.（2）Do patents get high impact when the P-GKBP cite them?（3） Is it easier to occur knowledge burst if a patent cites a P-GKBP? Theoretical implications and practical implications are discussed in Section 5. Section 6 concludes with limitations and points out research space for future study.

2. Literature Review

2.1 The impact of “Pre-grant disclosure system” on knowledge diffusion

The patent system exerts a critical impact on technological innovation and technological knowledge diffusion (Okada & Nagaoka, 2020). On the one hand, some studies examined the overall innovation and diffusion of patents since the implementation of the “pre-grant disclosure system” from the perspective of law and economics. Reportedly, pre-grant disclosure would increase the number of patent citations, promote the dissemination of technical knowledge, and avoid repeated R&D activities (Aoki & Spiegel, 2009; Baruffaldi & Simeth, 2020; Johnson & Popp, 2001; Okada & Nagaoka, 2020; Plečnik, 2021). On the other hand, many studies on the diffusion of knowledge after patent authorization have reported that early discovery of patents is crucial for improving and expanding patent protection (Hutchins, 2003), stimulating technological change (Hekkert & Negro, 2009), shortening technology lifecycle (Tsang et al., 2016) and accelerating the evolution of technological innovation (Coccia & Watts, 2020; Goldsby et al., 2016).

In an empirical study, Johnson and Popp (2001) used the patent citation data of the United States Patent Office (USPTO) from 1976 to 1996 to suggest that public publication before grant would promote the flow of patent knowledge. However, as the US “pre-grant disclosure system” was established in 1999 and implemented in 2000, all the data used for analysis were before the pre-grant disclosure system took effect; thus, Johnson and Popp (2001) could not directly determine whether the knowledge diffusion started from the publication date of the application or the “American Inventor’s Protection Act (AIPA)” effective date (Okada & Nagaoka, 2020). Thus, Okada and Nagaoka (2020) compared the citation of authorized patents in different fields a year before and a year after AIPA came into force in the United States, revealing that the implementation of the patent system accelerated the knowledge diffusion in different technical fields. In the same year, Baruffaldi and Simeth (2020) used the patent data of the “European Patent Office” (EPO) to demonstrate that pre-authorization publication would increase the number of patent citations; they measured the geographic distance of the patent citations and found that the geographic distance of citations remained unchanged, but the technological proximity increased moderately. However, significant uncertainty exists as to whether the patented technology will be cited, transferred, authorized, and have an impact and economic value after application (Czarnitzki & Toole, 2011; Higham, 2021; Zhang & Xiong, 2020). Obviously, the implementation of the “pre-grant disclosure system” of patents promotes the diffusion of patented technology knowledge. However, it remains unclear whether the implementation of a “pre-grant disclosure system” will exert a high impact on patents from the application to pre-grant. Particularly, the research on the changing characteristics and laws of the citation and assignment of patents before being granted has not received sufficient attention.

2.2 Patent knowledge diffusion

Technology knowledge diffusion is crucial in technological innovation and social progress(Liu & Qiao, 2021). Schumpeter (1934) regarded the great imitation of technological innovation as the diffusion of technological innovation in 1912. Stoneman (1989) believed that technology knowledge diffusion is a process in which new technology is transferred (Komada, 1986), broadly applied, and popularized. Thus, innovation can hardly exert an economic impact without technology knowledge diffusion (Schultz, 1990). Patents are the main carrier of technical knowledge. The generation of new patents and the diffusion of technical knowledge are often accompanied by enhancements or applications of original innovations (Dosi, 1982, 1991; Pezzoni, 2022). Thus, by measuring the trajectory of patent technology knowledge diffusion (Narin et al., 1997; Sharma & Tripathi, 2017), it is possible to identify major innovations and transformative technologies in patents (Chang et al., 2009; Hou, 2019) and to examine the influencing factors and law of patent knowledge diffusion (Alkemade & Castaldi, 2005; Delre et al., 2007; Delre et al., 2010).

In previous studies, scholars have often cited patent documents and other indicators to perform quantitative measurement of technology knowledge diffusion (Narin et al., 1997; Meyer, 2000; Breitzman & Mogee, 2002; Chang et al., 2009; Liu, 2012; Hou & Yang, 2019; Veugelers & Wang, 2019), examine influencing factors of technology knowledge diffusion (Alkemade & Castaldi, 2005; Bodo, 2016; Cainelli et al., 2007; Delre et al., 2007, 2010; Mac Garvie, 2005), and discuss patent knowledge overflow (Jaffe et al., 1993; Verspagen & Loo, 1999; Jaffe & Trajtenberg, 1999; Hu & Jaffe, 2003; Aoki & Spiegel, 2009; Sharma & Tripathi, 2017).Besides, some scholars have used patent citation networks to investigate patent diffusion and reveal the network structure characteristics of various patent technologies in the process of knowledge diffusion (Cainelli et al., 2007; Li et al., 2007; Shih & Chang, 2009; Hsueh & Wang, 2009; Kim et al., 2016; Zheng et al., 2017; Yang et al., 2021).

However, on the one hand, the studies mentioned above usually explored the cited situations of patents in an extended timeframe, which neglected the influence characteristics of patents in a short timeframe. Meanwhile, more attention was paid to the factors other than the patent text and their influence on the technology in an extended timeframe, such as the influence of scientific publications on the patented technology (Narin et al., 1997; de moya-anegon et al., 2019; Veugelers & Wang, 2019), and the influence of regional productivity and market factors on the diffusion of patented technologies (Anderson & Wincoop, 2001; Fung & Chow, 2002; Jiang et al., 2015; Bodo, 2016). On the other hand, the existing studies on the trajectory of patent technology knowledge diffusion have primarily investigated the impact and characteristics of the patent technology knowledge diffusion trajectory after authorization. Correspondingly, limited studies have explored the characteristics of pre-grant knowledge diffusion and its evolution trajectory under the "pre-grant disclosure system," as well as the impact of the pre-grant knowledge burst on the patent forward citation (citing) and backward citation (cited). Therefore, exploring the influence of the patent text property on the diffusion trajectory of pre-grant patent technological knowledge from the patent text property warrants further attention, which is conducive for inventors and patentees to grasp the innovation and social value of technology from the source of technology and promote the progress and innovation of science and technology.

2.3 Knowledge diffusion trajectory and method

In the field of knowledge diffusion trajectory, scientific papers have investigated extensively. Some studies measured the citation trajectory of scientific papers by using the citation counts per year or within a specific timeframe and then discussed papers' impact or the change rules and characteristics of their life cycle. Based on the differences of citation trajectories, "sleeping beauty" (Glanzel et al., 2003; Glanzel & Garfield, 2004; Lachance & Larivière, 2014), "total factor – sleeping beauty" (Li, 2014), "one-hit wonder" (Dalen & Henkens, 2005; Li et al., 2014), "Smart Girl" (Ye & Bornmann, 2018), and other special citation track types are defined through subjective judgments (Garfield, 1980, 1989a and b, 1990; Stent, 1972), curve fitting (Avramescu, 1979), three-index method (Van Raan, 2004; Van Raan, 2021), B index (Ke et al., 2015), quartile method (Costas et al., 2010), heart rate index (Li, 2014; Li et al., 2014), and the method of citation angle of exploitation (Ye & Bornmann, 2018)(Table 1). Other scholars, in order to reveal the burst situation of the subject terms in the time series during the development of the subject field, proposed "the burst feature detection algorithm"(Kleinberg, 2003;Chen, 2006; Takahashi, Tomioka & Yamanishi, 2014), "K-state automaton burst detection model"(Wu & Zhang, 2014), "BurstFuseX"(Liang,& de Rijke, 2015), "association rule mining between keywords and burst terms(ARM-KB)"(Li & Chu, 2016), "Co-occurrence matrix Detection"(Lee, 2019), "Burst hotspots dynamic detection and tracking(BHDDT)"(Huang, et al, 2019) and other methods. It should be noted that these different methods have different applicable objects and characteristics of applications. Since these methods examine the changes in a period of time, they have certain limitations in measuring the continuity of knowledge bursts.

Table 1. Some main methods of identifying knowledge burst based on the trajectory of knowledge diffusion

Study	Method	Features
van Raan(2004)	Three-index method	(1) For a single indicator trajectory, it has a fixed threshold parameter and depends on the time interval and citation count, which is suitable for small or single data sets;
Costas et al（2010）	Quartile method	(2) Only the first knowledge burst node is measured, which cannot be applied to the situation of multiple knowledge bursts.
Li（2014）	Heart rate index	For the trajectory of a single indicator, the situation of multiple bursts of knowledge is not considered, and the time of knowledge burst is affected by time.
Ke et al(2015)	B index	
Ye & Bornmann(2018)	The method of citation angle of exploitation	For a single index, the selection of the angle threshold is subjective.
Hou(2019)	PA index	Subjectivity is strong, since the Subjectivity sets the threshold, suitable for small sample data set.
Kleinberg(2003)	The burst feature detection algorithm	These Burst detection methods are mainly for the time series detection of subject terms, which inspects the transformation in a period of time.
Wu & Zhang(2014)	K-state automaton burst detection model	
Liang,& de Rijke (2015)	BurstFuseX	
Li & Chu(2016)	Association rule mining between keywords and burst terms(ARM-KB)	

Lee（2019）	Co-occurrence matrix Detection
Huang, et al (2019)	Burst hotspots dynamic detection and tracking(BHDDT)

However, limited attention has been paid to the patent citation at different times and legal status (transfer, license) to determine the influence diffusion trajectory of patent technology and assess the characteristics of different trajectories. Based on the two indices, patent citation and legal status (transfer and license), Hou and Yang (2019) explored the knowledge diffusion trajectory of granted patents and identified four types of knowledge diffusion trajectories—early sudden awakening (the "flash in the pan"), early gradual awakening (the "pea princess"), delay gradual awakening (the "ugly duckling"), delay sudden awakening (the "sleeping beauty"); they also ardently discussed the trajectory characteristics of "ugly duckling" and "sleeping beauty." Among them, the patent with "flash in the pan" denotes that a knowledge burst of the granted patent merges only in the early stage of the application and no more knowledge burst occurs in the remaining time. Nevertheless, their study did not highlight whether the patent with "flash in the pan" and its property characteristics merit attention when it bursts out. Thus, this study adapted the same indices to elucidate the knowledge diffusion trajectories of patents, by constructing a non-parameter dynamic knowledge burst detection model. And we identified P-GKBP based on the knowledge bursts,. In addition, we explored the property characteristics of the P-GKBP knowledge diffusion and the reasons for the knowledge burst, as well as the differences between P-GKBP and "flash in the pan" patent, which is of great practical value for developers, scientists, and managers.

3. Methodology and Data Sources

3.1 Data source and processing

In this study, the patent data were obtained from IncoPat database, a commercial database covering the vast amount of patent information worldwide (https://www.incopat.com/), to search P-GKBP. Owing to its comprehensive and reliable data, professional functions, high retrieval efficiency, and friendly user interface, IncoPat database has become a robust tool for technological research, competitive analysis, and early warning of legal risks. Currently, IncoPat is home to >140 million patent documents in 120 countries, organizations, or regions around the world. Its data are purchased from official and commercial data providers, and its patent description information, law (transfer and license), operation, family, in-depth processing and integration of citation and other information can understand the 24-hour dynamic update of data (Hou & Yang, 2019). Through the present research, we have constructed three patent data sets: Data A, Data B, and Data C(Table 2).

Table 2. Base data set information table

Category	Data A	Data B	Data C
Search Method	Title abstract	Patent Public Notice No.	Patent Public Notice No.
Search terms	wireless communication	Data A Citation Patent Publication No.	Data A Cited Patent Publication No.
Search time	19960101-20151231	All	All
Number of patents	23,626	36,031	69,110
Whether data is de-duplicated	Yes	Yes	Yes

For Data A, we took the field of “Wireless Communication” as an example and retrieved the patent documents granted in China in the IncoPat database, based on the following rules: “title abstract = wireless communication” and “patent application date = 19960101–20151231”. We retrieved a total of 23,626 patent documents (Fig. 1), of which 3835 were transferred or licensed, and 2 were declassified for national defense.

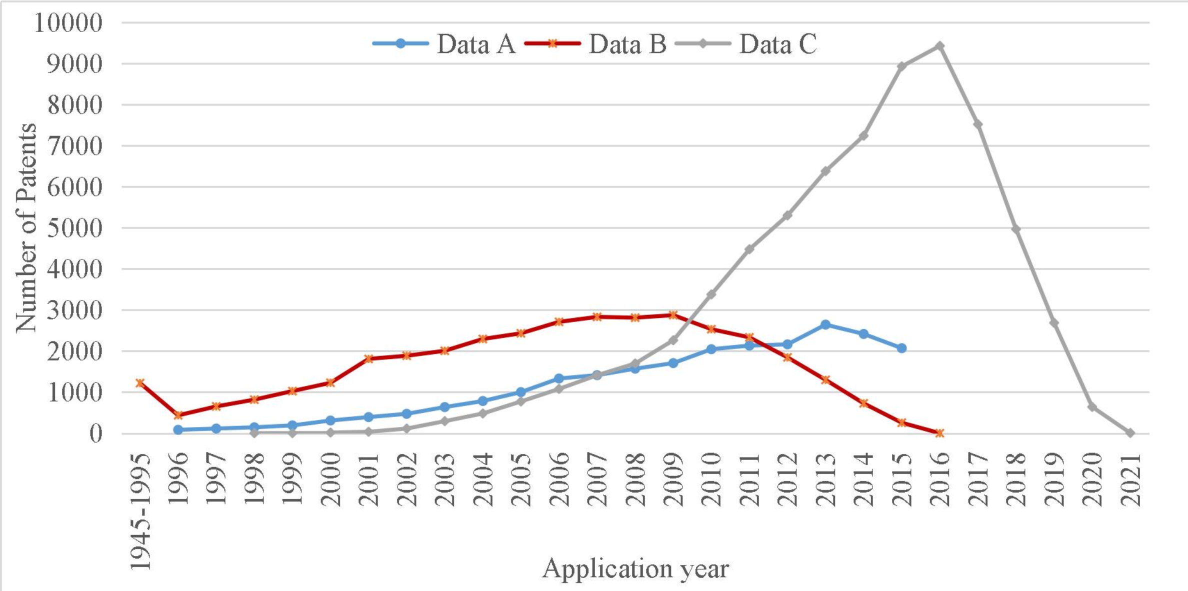

Fig 1. Time distribution of patent applications in Data A, Data B, and Data C

In Data A, the top 10 international patent classification (IPC) subcategories of these patents included H04W, H04L, H04B, H04Q, H04J, H06F, H04M, H01Q, H04N, and G08C (Fig. 2). To precisely evaluate the influence of each patent in Data A at different times, we performed mathematical statistics on each patent's citation, transfer, and license in each year. The results revealed that by 2019, these 23,626 patents were cited 104,169 times and cumulatively transferred 5015 times.

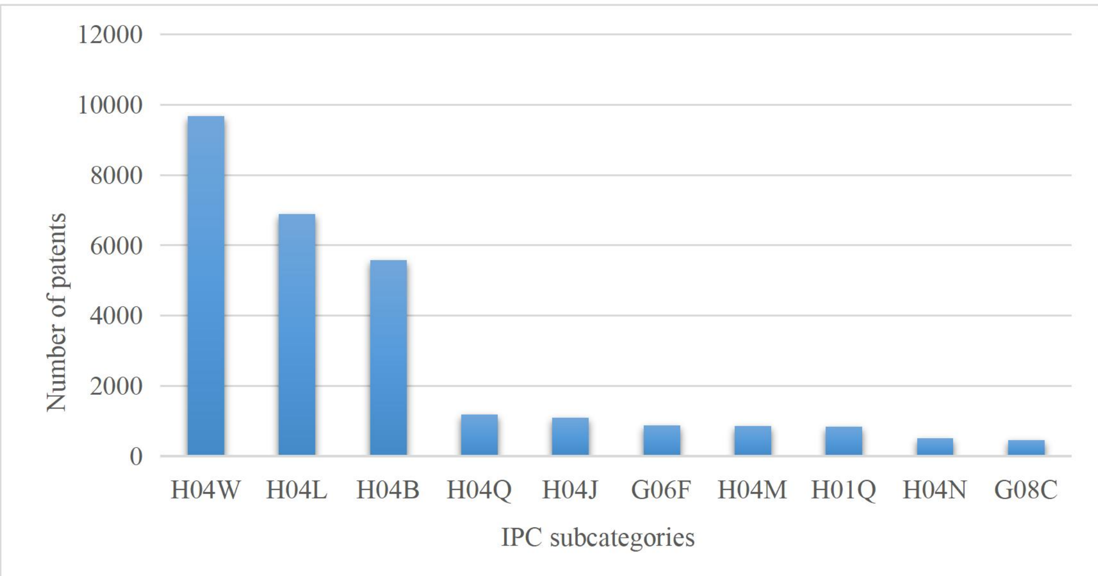

Fig 2. The top 10 international patent classification (IPC) subcategories of the patents in the field of wireless communication technology during 1996–2015.

Data B data comes from the citation data of P-GKBP in Data A. After extracting and de-duplicating information such as the public announcements of the cited patents of the P-GKBP in Data A, a total of 36,031 patents were obtained. The basic information of the 36,031 patents, such as application number, applicant, application date, inventor, cited patent, times of citation, cited patent, cited frequency, IPC, whether transfer occurs, transfer time and other information, was retrieved in IncoPat database as Data B(Fig 1).

Data C data are obtained from the cited patent data of P-GKBP in Data A. 69,110 patents in total were obtained after extracting and reprocessing the public announcements of cited patents of P-GKBP in Data A. The basic information of the 69,110 patents, such as application number, applicant, application date, inventor, cited patent, times of citation, cited patent, cited frequency, IPC, whether transfer occurs, transfer time and other information, was retrieved in IncoPat database as Data C(Fig 1).

3.2 Methodology

This study has constructed a knowledge burst patent detection model(KBPDM) to allow us to effectively and stably identify the knowledge bursts that are generated during the diffusion of proprietary knowledge and calculate the intensity of different knowledge burst intervals. The model mainly comprises of the following processes:

First, the weights of each indicator in the process of patent knowledge diffusion are calculated to determine the influence that the patent documents generate at different time points. Based on Hou and Yang (2019), to describe the technology knowledge diffusion of patents under the joint action of citation (Ci), transfer (Tr), and license (Li), we proposed the following Patent Affect(PA) function:

$$\mathrm{PA_i} = \mathrm{Wci} \cdot \mathrm{Ci} + \mathrm{Wtr} \cdot \mathrm{Tr} + \mathrm{Wli} \cdot \mathrm{Li} \tag{1}$$

where Ci is the annual citation frequency of patents; Tr denotes the annual transfer frequency of patents; Li denotes the annual licensing frequency of patents; and Wci, Wtr, and Wli are the weight of the three indices. Moreover, $\mathrm{PA_i}$ is the annual influence of technological knowledge diffusion created by the combined action of the three indices.

Assuming that patent transfer and license exert the same influence on patents, we constructed the structural matrix based on the correlation between the three indexes

Table 3. Model matrix construction

Xij	Ci	Tr	Li
Ci	1	1/9	1/9
Tr	9	1	1

Li	9	1	1

The results obtained were tested for consistency; PA passes the consistency check. After normalizing the eigenvalue vector $(0.0783 \quad 0.7049 \quad 0.7049)^T$, the weight vector $(0.0526 \quad 0.4737 \quad 0.4737)^T$ was derived to obtain. Evaluated the weight of the three indexes as Wci = 0.0526, Wtr = Wli = 0.4737 (Hou & Yang, 2019). In addition, we adopted this function and considered the influence of the three indices as the measure index of the patent knowledge diffusion trajectory. Accordingly, we drew the knowledge diffusion trajectory of each patent.

Second, the Z-Score formula was used to identify the nodes and durations of knowledge burst generated in the knowledge diffusion process of patent documents before licensing. The Z-Score is a means to notice the relative position of a value in a distribution and provides a true picture of the relative standard distance of a score from the mean. A score that is above the mean gets a positive standard score and a score below the mean gets a negative standard score. The underlying formula is $Z = (x_i - \bar{x}) / \sigma$, The $\bar{x}$ is the mean value of a set of data, σ is the standard deviation of a set of data.

After calculating the PA values of different time nodes of the patent, the PA values of different time nodes from the time of patent application announcement to the time of patent grant announcement are formed into a time series as $PA_1, PA_2, PA_3, \cdots, PA_i$, the values in the time series denote the influence values obtained in year *i* after the patent application was disclosed, respectively. Thus, the formula for the identification of knowledge burst in the process of patent knowledge diffusion is:

$$Z(i) = \frac{PA_i - \overline{PA}}{\sigma\,(PA)} \qquad i = 1,2,3,\cdots,n \tag{2}$$

PA_i is the impact of patent b in year *i* of the application disclosure, $\overline{PA}$ is the average value of the influence generated by the patent from the date of application publication to the date of grant publication, $\sigma\,(PA)$ is the standard deviation of the influence generated by the patent from the filing notice date to the grant notice date. Z(i) is the knowledge burst of the patent in year i. If Z(i) > 0, suggesting that the patent undergoes a knowledge burst in year *i*. IF $Z(i) \leqslant 0$,indicates that no knowledge burst occurred in year i of the patent.

So,$\overline{PA}$、$\sigma\,(PA)$、Z(i) is：

$$\overline{\mathrm{PA}} = \left(\sum_{i=1}^{n} \mathrm{PA_i}\right)/n \tag{3}$$

$$\sigma\,(\mathrm{PA}) = \sqrt{\frac{\sum_{i=1}^{n} (\mathrm{PA_i} - \overline{\mathrm{PA}})^2}{n}} \tag{4}$$

$$Z(i) = \frac{\mathrm{PA_i} - \left(\left(\sum_{i=1}^{n} \mathrm{PA_i}\right)/n\right)}{\sqrt{\frac{\sum_{i=1}^{n} (\mathrm{PA_i} - \overline{\mathrm{PA}})^2}{n}}} \qquad i=1,2,3,\cdots,n \tag{5}$$

Third, based on the identifying of knowledge of burst nodes, we calculate the knowledge of burst intensity in different knowledge burst time intervals. The main calculation was performed using the exponential density function. Since the simplest randomized model for generating information arrival time series is based on the exponential distribution, the information is sent out probabilistically, and the time gap t between information i and i+1 is distributed according to the "memoryless" exponential density function $f(x) = \alpha e^{-\alpha x}$ (Kleinberg, 2002) for the parameter $\alpha > 0$. Thus, we extended the exponential density function to the following formula for calculating the strength of knowledge mutation, and the patent knowledge has an exponential density function in the diffusion process as

$$f(t) = \overline{PA} \cdot e^{-\overline{PA} \cdot t} \quad (\overline{PA} > 0) \tag{6}$$

Therefore, the intensity of the knowledge burst of a patent with different knowledge burst intervals before grant is:

$$Pbs = \int_0^{t_1} \overline{PA} \cdot e^{-\overline{PA} \cdot t} dt = 1 - e^{-\overline{PA} \cdot t_1} \qquad (\overline{PA} \geq 0) \tag{7}$$

where Pbs indicates the intensity of knowledge burst of the patent, $0 \leqslant \mathrm{Pbs} \leqslant 1$. More the value of Pbs tends to be closer to 1, the stronger is the intensity of knowledge burst of the patent. In addition, t denotes the interval of duration of patent knowledge burst, and t_1 denotes the duration interval of the first burst of patent knowledge. This mutation intensity calculation formula not only considers the influence of the patent at the time of knowledge burst, but also the duration of the patent knowledge burst.

During the actual research process, to better reveal the characteristics of pre-grant knowledge burst patents, we divide P-GKBP into ordinary P-GKBP and excellence P-GKBP based on the variation of the total intensity of knowledge burst before patent grant. Among them, excellence P-GKBP are those with pre-grant knowledge burst total intensity $\sum \mathrm{Pbs} \geqslant 0.5$, and pre-grant knowledge burst patents total intensity $\sum \mathrm{Pbs} < 0.5$ are ordinary P-GKBP.

4. Results

The function of KBPDM is to detect the P-GKBP in Data A in the field of wireless communication technology, identify the P-GKBP with knowledge burst before grant, and distinguish between the excellence P-GKBP and ordinary P-GKBP based on the intensity of knowledge burst of the P-GKBP. After that, based on the identification results, the data in Data B and Data C are analyzed respectively to explore the celebrity effect and the Matthew effect of P-GKBP.

4.1 Identification and distribution of P-GKBP

Using the constructed KBPDM, we performed the pre-grant knowledge Burst Detection on 23,626 patents in Data A in the field of wireless communication technology. We found that among the 23,626 patents, there were 12,608 P-GKBP with a P-GKBP rate of 53.36%, and these 12,608 P-GKBP occurred with different degrees of knowledge burst. Based on the change in burst intensity knowledge of P-GKBP, this study found 175 excellence P-GKBP and 12,433 ordinary P-GKBP (Table 4). In terms of the number of pre-grant knowledge bursts of P-GKBP, the maximum number of P-GKBP was four. Among the 12,608 P-GKBP, cases comprising one knowledge burst before the patent grant were 10,722, accounting for 85.04% of the total number of knowledge burst patents. A total of 1825 patent knowledge bursts occurred twice before patent grant, which accounted for 14.47% of the total number of patent knowledge bursts. There were 60 patent knowledge bursts that occurred three times before patent grant, thereby accounting for 0.48% of the total number of patent knowledge bursts (Table 4). There was one case of four patent knowledge bursts before the patent was granted, accounting for 0.01% of the total number of patent knowledge bursts. The patent application number "CN01103432.7" (Fig 3), for which four bursts of knowledge occurred before the patent was granted, was a patent for an invention filed by Lucent Technology Corp, and the patent application cycle was 14 years long. "CN01103432.7" was filed with the State Intellectual Property Office of China in February 2001, first filed for disclosure in September 2001, granted in June 2014, and expired in March 2021 owing to the expiration of the patent right. "CN01103432.7" was cited 27 times before the grant, all of which were cited by the examiner, and no citation or assignment by the inventor occurred.

Table 4. Information on the distribution of the number and intensity of knowledge bursts of different P-GKBP before granting

Total intensity range of pre-licensing knowledge bursts	A burst of knowledge occurs before authorization	Secondary knowledge bursts occur before authorization	Three knowledge bursts occur before authorization	Four knowledge bursts occur before authorization
[0.0,0.1)	7030	757	13	0
[0.1,0.2)	2375	619	23	0
[0.2,0.3)	768	220	11	0
[0.3,0.4)	289	118	8	0
[0.4,0.5)	136	64	2	0
[0.5,0.6)	55	24	1	1
[0.6,0.7)	29	15	1	0
[0.7,0.8)	25	7	1	0
[0.8,0.9)	10	1	0	0
[0.9,+∞)	5	0	0	0

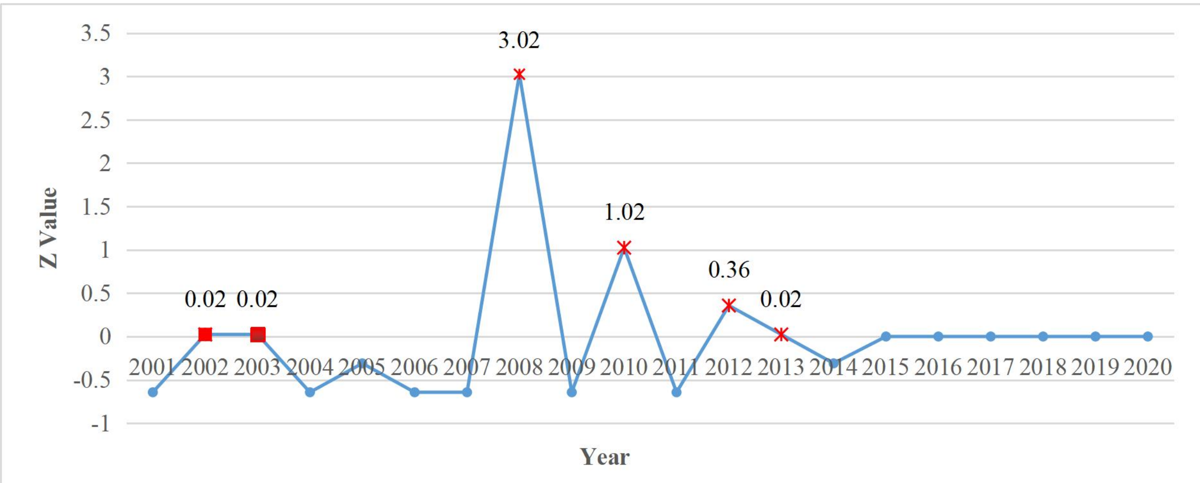

Fig 3. Z-value distribution trajectory of the P-GKBP with application number "CN01103432.7"

To further explore the relationship between the application intervals, the burst frequency and burst intensity of pre-grant knowledge of P-GKBP, we conducted a correlation analysis using Pearson correlation and significance (two-sided) tests. We analyzed the correlation between these variables using the Pearson correlation and significance (two-sided) tests. We established that there was a significant positive correlation between the time interval of P-GKBP application, the number of pre-grant knowledge bursts and the intensity of pre-grant knowledge bursts. This means that the longer the time interval of P-GKBP application, the higher the number of pre-grant knowledge bursts, and similarly, the longer the time interval of P-GKBP application, the stronger the pre-grant knowledge burst intensity. The correlation between the time interval of P-GKBP application and the number of pre-grant knowledge bursts was 0.323**, and the correlation between the time interval of P-GKBP application and the intensity of pre-grant knowledge bursts was 0.190**. However, we did not find significant correlation between the number of pre-grant knowledge bursts of P-GKBP and the intensity of pre-grant knowledge bursts (Fig 4). On the other hand, from the viewpoint of the time interval of P-GKBP applications, the time interval of patent applications in which knowledge bursts occurred was mainly concentrated in 3–5 years, and there were always 10,576 patents, occupying 83.88% of the total number of P-GKBP, among which the maximum number of P-GKBP with a four-year application interval was 4857, occupying 38.52% of the total number of P-GKBP (Fig 5).

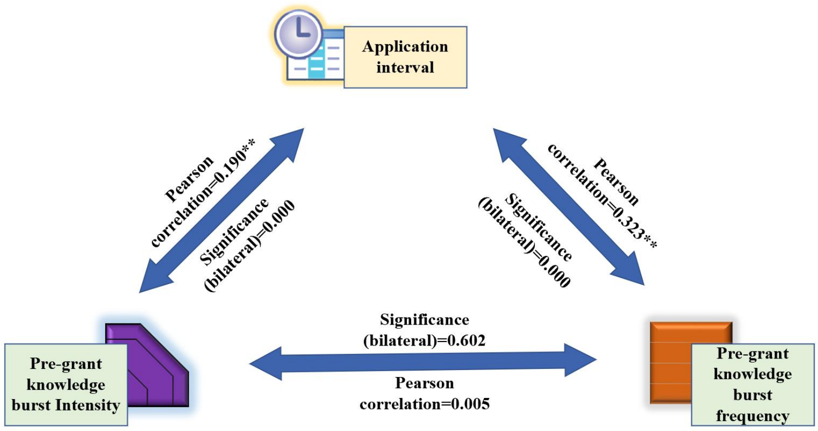

Fig 4. Correlation between the lengths of time between P-GKBP applications, the number of pre-grant knowledge bursts, and the intensity of pre-grant knowledge burst

Note: **. Significantly correlated at the 0.01 level (bilateral)

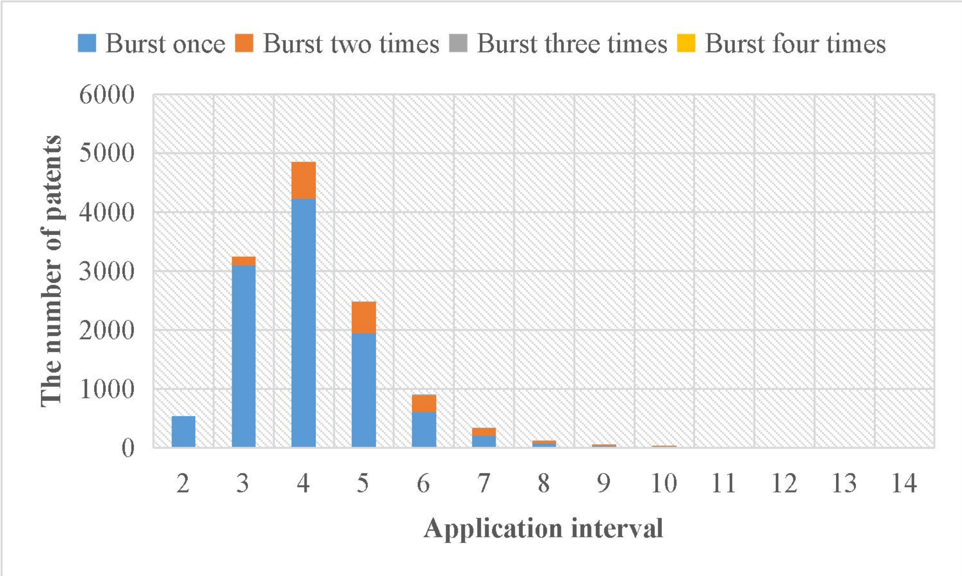

Fig 5. Distribution of the number of P-GKBP with different application time intervals

4.2 Do patents get high impact when the P-GKBP cite them?

Based on the P-GKBP in Data A that have been identified, we analyzed the data in Data B, the cited patent set of P-GKBP, to explore whether being cited by a P-GKBP leads to a high impact (PA). As of May 2021, the patents in Data B were cited 9,82,942 times, and there was an occurrence of 26,872 assignments. We matched the patents in Data B with the P-GKBP in Data A for the knowledge diffusion relationship.

It was established that there was a significant positive correlation between the influence (PA value) of patent b in Data B, the frequency of patent b being cited by the P-GKBP in Data A, and the knowledge burst intensity of the P-GKBP in Data A (Fig 6). Therefore, the change in influence of patent b after being cited by P-GKBP is closely related to the intensity of knowledge burst of P-GKBP and the number of pre-grant knowledge burst patent citations. Among them, the correlation between the influence of patent b in Data B and the frequency of patent b being cited by P-GKBP in Data A is 0.087** in the case of no control variables. The correlation between the influence of patent b in Data B and the frequency of patent b being cited by P-GKBP in Data A was 0.072** when the patent application time and the intensity of knowledge burst of P-GKBP were used as control variables (Table 5). When exploring the pattern of quantitative changes between the influence of patent b and the number of times patent b was cited by P-GKBP by controlling the patent application time and the burst of the knowledge intensity of P-GKBP, we found that the more a patent b in Data B and was cited by P-GKBP, the greater is its influence (Table 6). In addition, from the actual citation relationship network between patent b and the P-GKBP (Fig 7), the patent with application number CN200910161920.0 has not changed its status such as assignment, but it is cited by the P-GKBP CN201110080379.8, CN201010129111.4, CN201010540113.2, and CN201110183346.6, thereby resulting in high impact of CN200910161920.0, which was cited 173 times. The patent with application number CN200810217598.4 has not changed its status such as transfer, but it is cited by pre-grant knowledge burst patents CN201110033969.5 and CN201110183346.6, hence resulting in CN200810217598.4 having high influence and being cited 36 times.

In the absence of control variables, the correlation between the influence of patent b in Data B and the intensity of knowledge burst of P-GKBP in Data A was 0.032**. The correlation between the influence of patent b in Data B and the burst of knowledge of P-GKBP in Data A was 0.017** when the time of patent application and the frequency of being cited by P-GKBP were used as control variables (Table 5). Moreover, when exploring the pattern of quantitative changes between the influence of patent b and the burst of the knowledge intensity of the P-GKBP by controlling the patent application time and the frequency of being cited by the P-GKBP, it was found that the stronger the burst of the knowledge intensity of the P-GKBP, the greater was the influence generated by that cited patent b (Table 7). This means that when the application time of patent b and the frequency of being cited by a P-GKBP are taken as control variables, the effect of a patent being cited by a excellence P-GKBP and an ordinary P-GKBP is different, and it will gain higher influence if it is cited by a excellence P-GKBP. Figure 7 shows that patents WODE05000774 and CN200510029798.3 in Data B are only cited by P-GKBP CN200980160549.4 in Data A. Since P-GKBP CN200980160549.4 is an ordinary P-GKBP with a pre-grant knowledge burst intensity of 0.017, the patent WODE05000774 (cited 2 times), CN200510029798.3 (cited 1 time) are less influential. Patent CN201010111879.9 in Data B is cited by P-GKBP in excellence celebrity CN201110183346.6 (pre-grant knowledge burst intensity of 0.7), the CN201010111879.9 was cited 33 times. Patent US09217235 in Data B was cited by excellence celebrity in P-GKBP CN201080017134.4 (pre-grant knowledge burst intensity of 0.938), and US09217235 had an impact of 207 citations and 2 transfers.

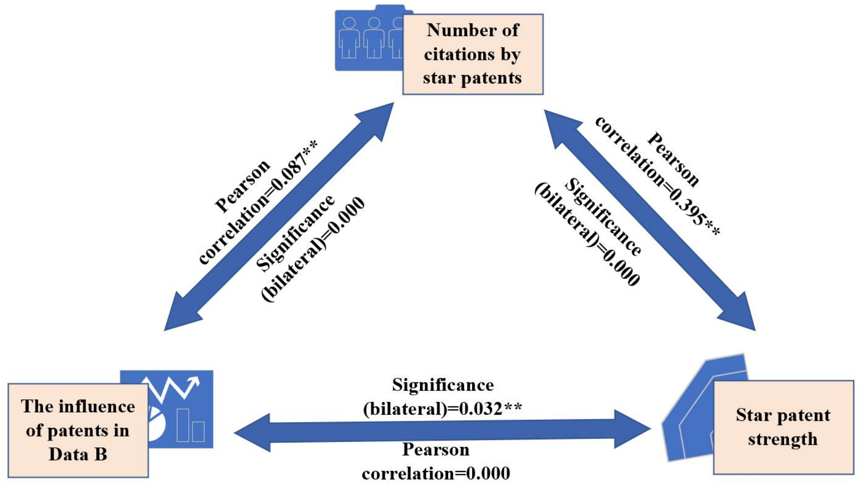

Fig 6. Relationship between the influence of patent b in Data B, the frequency of patent b being cited by P-GKBP in Data A, and the knowledge burst intensity of pre-license knowledge bursts in Data A

Note: **. Significantly correlated at the 0.01 level (bilateral)

Table 5. Results of correlation analysis between different variables

Control variables	Variables	Variable Correlation Value	Significantly
No	Frequency of being cited by P-GKBP Influence with patent b	0.087^{**}	0.000
Patent b's filing time, patent strength of P-GKBP	Frequency of being cited by P-GKBP and the impact of patent b	0.072^{**}	0.000
No	Cited P-GKBP intensity and impact of patent b	0.032^{**}	0.000
Patent b's filing time, frequency of being cited by pre-grant knowledge burst patents	Cited P-GKBP intensity and impact of patent b	0.017^{**}	0.001

Note: *. Significantly correlated at the 0.01 level (bilateral), **. Significantly correlated at the 0.01 level (bilateral)

Table 6. Distribution of the relationship between the frequency of being cited P-GKBP and patent influence under the same patent application time and P-GKBP strength (results of some time periods)

Patent Application Time	P-GKBP strength	Frequency of being cited by P-GKBP	Average value of patent impact	Percentage of patents with patent impact greater than 1
2003	(0-0.5)	1	2.87	52.06%
		2	4.17	60.58%
		≥3	8.12	86.87%
	[0.5,+∞)	1	2.83	69.23%
		2	3.67	71.43%
		≥3	12.26	100%
2009	(0-0.5)	1	1.79	46.66%
		2	2.96	72.33%
		≥3	5.32	83.78%
	[0.5,+∞)	1	2.06	45.83%
		2	3.02	80.00%
		≥3	7.14	95.00%

Table 7. Distribution of the correlation between the intensity of P-GKBP and patent influence under the consistency of patent application time and frequency of being cited by P-GKBP(results of some time periods)

Patent Application Time	Frequency of being cited by P-GKBP	P-GKBP strength	Average value of patent impact	Percentage of patents with patent impact greater than 1
2003	1	(0-0.5)	2.87	52.06%
		[0.5,+∞)	2.83	69.23%
	2	(0-0.5)	4.17	60.58%
		[0.5,+∞)	3.67	71.43%
	≥3	(0-0.5)	8.12	86.87%
		[0.5,+∞)	12.26	100%

2009	1	(0-0.5)	1.79	46.66%
		[0.5,+∞)	2.06	45.83%
	2	(0-0.5)	2.96	72.33%
		[0.5,+∞)	3.02	80.00%
	≥3	(0-0.5)	5.32	83.78%
		[0.5,+∞)	7.14	95.00%

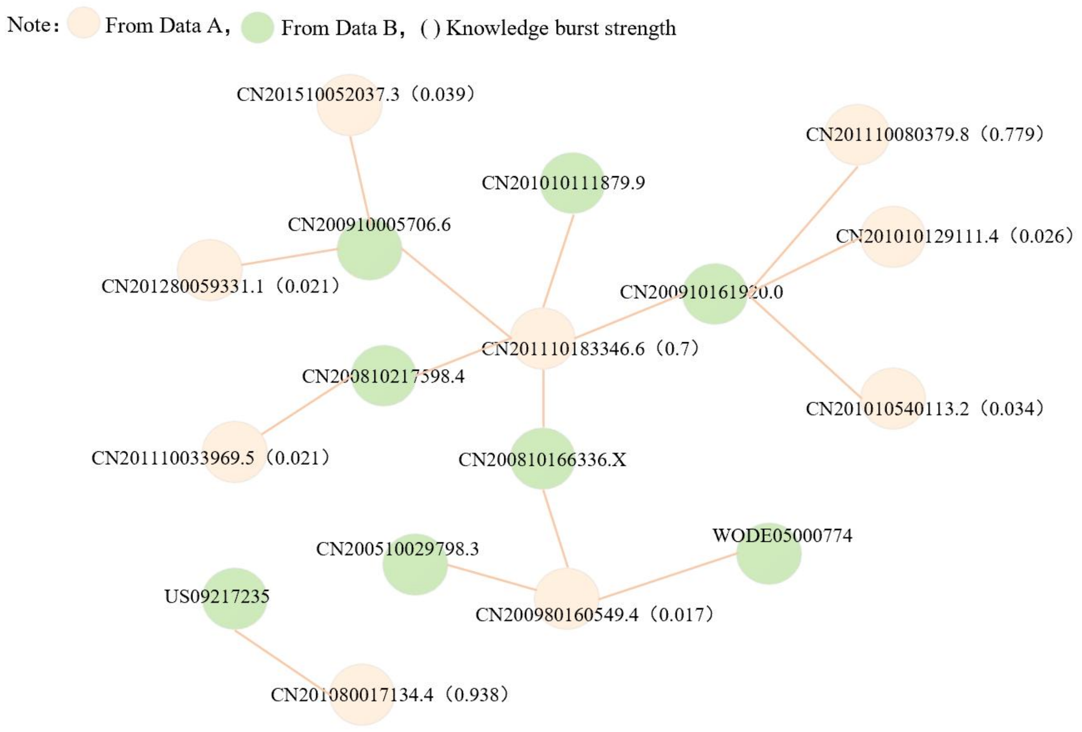

Fig 7. Relationship between some patents between Data B and Data A

4.3 Is it easier to occur knowledge burst if a patent cites a P-GKBP?

Based on the P-GKBP that have been identified in Data A, to explore whether patents that cite P-GKBP are more likely to be P-GKBP, we analyzed the data in Data C, the cited patent set of P-GKBP. As of May 2021, the patents in Data C were cited 2,33,688 times and 20,843 assignments occurred. We matched the patents in Data C with the P-GKBP in Data A in terms of knowledge diffusion relationship.

We established a significant correlation between the knowledge burst intensity of a P-GKBP and the knowledge burst intensity of a patent citing a P-GKBP, with a correlation of 0.106** (Table 8). This means that if a patent cites a P-GKBP, it is more likely to be called a

P-GKBP. When a patent cites a P-GKBP, does it have the same impact on itself when citing a ordinary P-GKBP and citing a excellence P-GKBP? We have classified and calculated the patents in Data C that cite ordinary P-GKBP and those that cite excellence P-GKBP. It was found that when a patent cites a P-GKBP, citing a excellence P-GKBP is more likely to produce a pre-grant knowledge burst into an ordinary P-GKBP and a excellence P-GKBP than citing an ordinary P-GKBP. Among the 69,110 P-GKBP in Data C, 64,251 patents cited ordinary P-GKBP in Data A. The number of P-GKBP in these 64,251 patents was 35,585, and the number of excellence P-GKBP was 998. A total of 55.38% of the 64,251 patents were P-GKBP, and 2.8% were excellence P-GKBP. Among 69,110 P-GKBP in Data C, 4859 patents cited excellence P-GKBP in Data A. The number of P-GKBP in these 4859 patents was 3493, and the number of excellence P-GKBP was 470. A total of 71.89% of the 4859 patents were P-GKBP, and 13.46% were excellence P-GKBP (Table 9).

Table 8. Correlation between the burst of knowledge intensity of P-GKBP and the burst of knowledge intensity of patents citing P-GKBP

		P-GKBP knowledge bursts into intensity	Cited knowledge bursts patent strength
P-GKBP knowledge bursts into intensity	Pearson	1	.106**
	Sig.(bilateral)		0
	Sum of squares and product of forks	2849.93	221.888
	covariance	0.032	0.002
	Number of cases	88,904	88,904
Cited knowledge bursts patent strength	Pearson	.106**	1
	Sig.(bilateral)	0	
	Sum of squares and product of forks	221.888	1,528.275
	covariance	0.002	0.017
	Number of cases	88,904	88,904

**. Significantly correlated at the 0.01 level (bilateral)

Table 9. Distribution of the types of patents citing ordinary P-GKBP and excellence P-GKBP

Reference to patent types in Data A	Ordinary P-GKBP	Excellence P-GKBP
Number of cited patents	64,251	4,859
Number of P-GKBP among the cited patents	35,585	3,493
Number of excellence P-GKBP among cited patents	998	470
Percentage of P-GKBP among cited patents	55.38%	71.89%
The proportion of excellence P-GKBP in P-GKBP	2.80%	13.46%

In addition, when 11,018 patents without pre-grant knowledge burst in Data A were studied, we found that these 11,018 patents were cited by 2,972 patents. Using the Burst Detection model to detect the pre-grant knowledge burst of these 2,972 patents, it was found that 2,512 patents generated knowledge burst before grant because of citation and assignment. The pre-grant patent knowledge burst rate was 84.52%. Since this figure was unusually high, we calculated the 2,512 P-GKBP to explore why 84.52% of them still became P-GKBP after citing non-P-GKBP. It was found that these 2,512 P-GKBP not only cited the non-P-GKBP in Data A but also cited a large number of P-GKBP. The cited P-GKBP and the fact that assignments occurred before they were granted were key factors in their becoming P-GKBP. This further indicates that P-GKBP have a significant Matthew effect.

Hence, we observed an apparent Matthew effect in the process of technological knowledge diffusion of P-GKBP. In practice, through the existing P-GKBP knowledge diffusion path, the enterprise could use this to find and buy patents that are currently in the stage of the application and significant for technology layout, which can effectively decrease and save significant technology capital and time cost, optimize the enterprise's existing technology patent layout and core patent control, and promote the high-quality development.

5. Discussion and Implication

5.1 Applicability and effectiveness of the method

In order to verify the applicability and effectiveness of the proposed KBPDM method in detecting the knowledge burst nodes in the trajectory of patent knowledge diffusion, KBPDM is compared with "Burst hotspots dynamic detection and tracking(BHDDT)"(Huang, et al, 2019). According to the BHDDT method proposed by Huang, et al (2019), the time series composed of the knowledge diffusion trajectory of each patent is:

$$\mathrm{PA} = \{\mathrm{PA_1}, \mathrm{PA_2}, \mathrm{PA_3}, \cdots, \mathrm{PA_i},\} \tag{8}$$

PA is patent affect sequence, which refers to the influence produced in the i^{th} year after the publication of the patent application.

The time window before patent authorization is defined as $(t_0, t_0+p]$:

$$\mathrm{SW_t} = \mathrm{PA}_{t_0}^{t_0+p} \tag{9}$$

SW_t is Sliding Windows Patent set, t_0 is the year of patent application disclosure, t_0+P is the patent grant year, and P is the time interval between the year of patent application disclosure and the time before authorization.

Suppose v is the speed of patent knowledge diffusion, S_t is the Total amount of SWt, S_t can be calculated by:

$$\mathrm{S_t} = \mathrm{v} * \mathrm{p} \tag{10}$$

If p is a constant, SWt is a steady sliding time window. If p is mutable, SWt is a dynamic window with higher flexibility and complexity (Huang, et al, 2019). In this study, p is a variable quantity. And it should be noted that Huang, et al (2019) defined Burst as "peak popularity above 20% of its total popularity" according to Crane & Sornette, et al (2008). In other words, if the influence of the patent in a certain year before the grant exceeds 20% of the total influence, it is considered that a Burst occurred in that year.

We used the BHDDT method to perform pre-grant knowledge burst tests on 23,626 patents in Data A. The study found that the results identified by the KBPDM method we proposed were significantly more than those identified by the BHDDT method. Our KBPDM method identified a total of 12,608 P-GKBP, and the BHDDT method identified a total of 10,720 P-GKBP. The patents identified by the BHDDT method were all included in the identification results of our proposed method (Table 10).

Table 10. Comparison of identification results between KBPDM method and BHDDT method

Method	Total	A burst of knowledge occurs before	Secondary knowledge bursts occur before	Three knowledge bursts occur before	Four knowledge bursts occur before

		authorization	authorization	authorization	authorization
KBPDM method identification results	12608	10722	1825	60	1
BHDDT identification results (Huang, et al, 2019)	10720	9253	1442	25	0

Furthermore, it is found that the number of knowledge bursts identified by the KBPDM method before patent grant is significantly more than the number of patent bursts identified by the BHDDT method. This is mainly because the burst threshold of the BHDDT method is affected by the total influence of the patent. If the total influence of a patent is higher, the threshold for a burst to occur before the patent is granted is higher, which cannot objectively measure the point of knowledge burst in the trajectory of patent knowledge diffusion. On the other hand, the BHDDT method could not detect some P-GKBP with sustained high impact and patent actual knowledge burst time in the process of identifying patent knowledge burst. For example, the patent application number CN201310565001.6 was published in 2014 and authorized in 2016. According to the BHDDT method, its burst threshold is 0.5684, and the time of burst occurrence is 2018. This patent does not belong to P-GKBP. However, according to our KBPDM method, the first knowledge burst of the patent was in 2016, and continuous knowledge bursts occurred during 2016-2019. The patent belongs to P-GKBP (Figure 8). Therefore, compared with advanced burst detection methods such as BHDDT, our method is more applicable and effective in detecting the burst of knowledge in the trajectory of knowledge diffusion before patent grant.

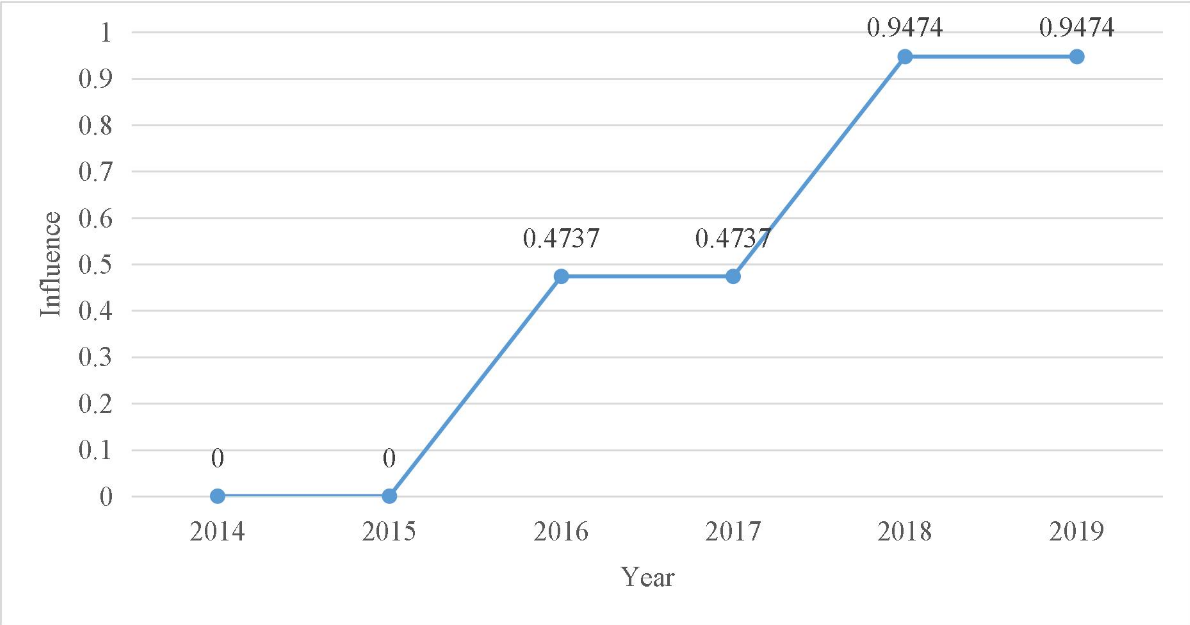

Fig 8. The knowledge diffusion trajectory of the patent application number CN201310565001.6

5.2 Theoretical Implications

(1) This research proposes a new method —KBPDM, which can better identify P-GKBP. Compared with the past methods for detecting bursts based on the frequency of documents in the entire corpus, "the burst feature detection algorithm", "K-state automaton burst detection model", "BurstFuseX", "ARM-KB", "BHDDT", etc., we identify patent knowledge bursts time combined with the characteristics of stability and continuity of patent knowledge diffusion trajectory in the process of knowledge burst, so that the recognition result is more accurate. Therefore, our method is more suitable for the detection of bursts of patent knowledge compared with the latest detection method of identifying the subject term Burst. On the other hand, in the process of identifying the pre-grant knowledge burst of 23,626 patents in the field of wireless communication technology, we measure the knowledge burst in the trajectory of knowledge diffusion by KBPDM we proposed. Compared with the methods proposed by Hou (2019), Li (2019), the model we proposed is automated, simple to calculate, reliable and objective in detecting patent knowledge burst and is not affected by the threshold settings. The model established that there are P-GKBP in the field of wireless communication technology.

In the process of granting of a patent, the P-GKBP with knowledge burst before granting occupy a large proportion, and the P-GKBP in Data A account for 53.36%, while the excellence P-GKBP only account for 0.74% of the total number of patents. From the perspective of the number of pre-grant knowledge bursts of P-GKBP, the numbers of P-GKBP are gradually decreasing as the number of pre-grant knowledge bursts of patents increase. The longer the time interval of the P-GKBP application, the more the number of pre-grant knowledge bursts of the patent. In addition, the longer the time interval of P-GKBP applications, the stronger the intensity of knowledge bursts obtained by the patents before granting. Therefore, compared with the research of Johnson and Popp (2001)， Okada and Nagaoka (2020) and other scholars, we have not only revealed that the implementation of the “pre-grnat disclosure system” has accelerated the overall knowledge diffusion in the technical field, and more accurately revealed the impact of the implementation of the system on the diffusion of knowledge in the pre-grant stage of a single patent. The “pre-grant disclosure system” does not speed up the diffusion of all patent knowledge, but makes the diffusion of patented technology knowledge more polarized. It has brought earlier and faster knowledge diffusion for some important patented technologies, accelerated the discovery of important patented technologies by enterprises and inventors, making some unimportant and infringing patents discovered and invalid earlier.

(2) P-GKBP have a significant celebrity effect and Matthew effect. There has always been an objective existence of the celebrity effect and the Matthew effect of P-GKBP and this phenomenon is proved for the first time in this study by quantitative means. The celebrity effect and the Matthew effect in patents belong to one of the phenomena in the accumulation of advantage theory proposed by Merton in 1942 (Cole, 1970, 1973; Dannefer, 2003; Merton, 1973). Price proposed an advantageous accumulation distribution, arguing that "success breeds success" and that "a millionaire earns extra income faster and easier than a beggar" (Price, 1965, 1976). In this study, we not only figured out that the consequence of patent advantage accumulation has the Matthew effect but also found that the antecedent of patent advantage accumulation has the celebrity effect. Among them, when studying the celebrity effect of P-GKBP, it was found that the effect of a patent being cited by a excellence P-GKBP and an ordinary P-GKBP is different, and that it will gain more influence if it is cited by a excellence P-GKBP.

In addition, in an in-depth investigation, it was found that the coupling behavior of P-GKBP would intensify the increase in influence of target patents, and those patents with high level of technology played a more vital role after coupling (Gui et al., 2015), and the coupling behavior of P-GKBP made the celebrity effect of patents more prominent. Similarly, while exploring the Matthew effect of P-GKBP, it was found that the patents that cite P-GKBPare more likely to become P-GKBP. In fact, citing ordinary P-GKBP and citing excellence P-GKBP have different effects on themselves. When a patent cites a P-GKBP, citing a excellence P-GKBP is more likely to produce a pre-grant knowledge burst to become an ordinary P-GKBP and a excellence P-GKBP than citing an ordinary P-GKBP. Notably, in the process of citing P-GKBP, some of the P-GKBP cited by the target patent have co-citation relationship, and the stronger the co-citation intensity of the P-GKBP, then the higher will be the influence of the target patent, and the higher the technical level of the representative target patent. Thus, in terms of impact on the celebrity effect of the target patent, P-GKBP coupling behavior > excellence P-GKBP > ordinary P-GKBP on improving the influence of the target patent. Concerning the Matthew effect of P-GKBP, the citation of ordinary P-GKBP and the citation of excellence P-GKBP have different effects on themselves, and P-GKBP are cited altogether > by excellence P-GKBP > by ordinary P-GKBP. Compared with Okada and Nagaoka et al. (2020), this study not only confirms that pre-grant disclosure will increase the diffusion of knowledge such as patent citation and transfer, but more importantly, reveals how the patent in “pre-grant disclosure system” influences the innovation development and knowledge diffusion of other patents through citing and being cited in the process of knowledge diffusion.

5.3 Practical Implications

(1) The P-GKBP has a Matthew effect by guiding how to enhance patent visibility (becoming a P-GKBP, early licensing). The P-GKBP is discovered early in the patent application and this can help shorten the technological innovation cycle and facilitate the identification of disruptive technologies.P-GKBP and their knowledge diffusion trajectories could optimize the layout of enterprise patent technology and save significant technical capital and time cost, which promotes the high-quality development of the enterprise. Besides, scholars can investigate hot technologies in different fields, identify core patent technologies, and estimate future technology trends through P-GKBP and their characteristics of knowledge diffusion. Furthermore, it can provide decision-making reference for technological innovation, technological management, and other strategic planning and policy-making.

(2) An automated identification path is offered to reduce patent sleeping beauty and discover valuable patents at the earliest. In this study, from the pre-grant knowledge diffusion trajectory of P-GKBP, it is found that the pre-grant knowledge burst trajectory of P-GKBP has various types, including "flash in the pan", "pea princess", "ugly duckling", "sleeping beauty" and other types of trajectories (Hou, 2019). For example, the P-GKBP trajectories in Data A with application numbers CN200480014463.8 and CN200680039085.8 underwent knowledge burst only in their sixth year after filing and became the type of Sleeping Beauty patent trajectories proposed by Hou (2019). Many types of knowledge burst trajectories of P-GKBP appear before granting because some P-GKBP have a long time interval from

application to grant owing to special reasons such as delayed examination and re-appeal examination by rejection when conducting the examination of the application for grant. Thus, when knowledge Burst Detection is conducted, the knowledge burst node becomes a sleeping beauty patent owing to the long time interval from the application disclosure date and is cited or transferred during the examination period. Hence, to decrease the number of sleeping beauties and the time of patent application, relevant institutions can improve the examination cycle of patent applications by expanding the scale of examiners, strengthening the professional training of patent examiners, and establishing a linkage mechanism between industry and patent examination institutions (examiners communicate with industry experts for joint examination) when conducting patent examination.

6. Limitations and future study

The study comes across with some limitations that need to be addressed. The type of subject of patent application has an impact on P-GKBP; however, this factor has not been considered in this study. Regarding the types of P-GKBP applicants in Data A, in the field of wireless communication technology, enterprises are located in the main position of science and technology innovation. In the Data A dataset, there are 10,088 patents with the applicant type containing enterprises, which account for 80% of the total number of P-GKBP. This was followed by 2416 patents with patent applicant types containing universities and research institutes that accounted for 19.16% of the total number of P-GKBP. Among the 175 excellence P-GKBP in the Data A dataset, 131 patents had the applicant type of enterprise, accounting for 74.86% of the total number of P-GKBP. Next, there were 43 patents with the applicant type of universities and research institutes, accounting for 24.57% of the total number of P-GKBP. Therefore, future should focus on exploring the factors influencing the applicant's subject type on P-GKBP, as well as to explore whether the same phenomenon exists in other technical fields. In addition, the importance of patents in the network is crucial to identify the potential/impact of patents (Park, 2018; You, 2017); however, this study does not consider the important characteristics of P-GKBP in the relationship network. Future research should focus on the location characteristics and structural characteristics of P-GKBP in the patent relationship network.

References

Alkemade, F., Castaldi, C. (2005). Strategies for the diffusion of innovations on social networks. *Computational Economics*, *25*(1/2), 3-23.

Anderson, J., Wincoop, E. V. (2001). Cravity with gravitas: A solution to the border puzzles. *American Economic Review*, *93*,170-192.

Aoki, R., & Spiegel, Y. (2009). Pre-grant patent publication and cumulative innovation. *International Journal of Industrial Organization*, *27*(3), 333-345.

Avramescu, A. (1979). Actuality and obsolescence of scientific literature. *Journal of the American Society for Information Science*, *30*(5), 296-303.

Baruffaldi, S. H., & Simeth, M. (2020). Patents and knowledge diffusion: The effect of early disclosure. *Research Policy*, *49*(4), 103927.

Bodo, P. (2016). MADness in the method: On the volatility and irregularity of technology diffusion. *Technological Forecasting and social Change*, *111*, 2-11.

Breitzman, A. F., Mogee, M. E. (2002). The many applications of patent analysis. *Journal of Information Science*, *28*(3), 187-205.

Cainelli, G., Mancinelli, S., Mazzanti, M. (2007). Social capital and innovation dynamics in district-based local systems. *The Journal of Socio-Economics*, 932−948.

Chang, S. B., Lai, K. K., Chang, S. M. (2009). Exploring technology diffusion and classification of business methods: Using the patent citation network. *Technological Forecasting and Social Change*, *76*(1), 107-117.

Chen, CM. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology,* 57(3): 359-377.

Coccia, M., & Watts, J. (2020). A theory of the evolution of technology: Technological parasitism and the implications for innovation magement. *Journal of Engineering and Technology Management*, *55*, 101552.

Cole, J. R., Cole, S. (1973). Social Stratification in Science. Chicago：University of Chicago Press.

Cole, S. (1970). Professional Standing and The Reception of Scientific Discoveries. *American Journal of Sociology*, *76*(2), 286-306.

Costas, R., Leeuwen, T. N. V., & Raan, A. F. J. V. (2010). Is scientific literature subject to a 'sell-by-date'? a ordinary methodology to analyze the 'durability' of scientific documents. *Journal of the American Society for Information Science & Technology*, *61*(2), 329-339.

Crane, R., & Sornette, D. (2008). Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences, 105*(41), 15649-15653.

Czarnitzki, D., & Toole, A. A. (2011). Patent Protection, Market Uncertainty, and R&D Investment. *Review of Economics and Statistics*, *93*(1), 147-159.

Dalen, H. P. V., & Henkens, K. (2005). Signals in science - on the importance of signaling in gaining attention in science. *Scientometrics*, *64*(2), 209-233.

Dannefer, D. (2003). Cumulative Advantage/Disadvantage and the Life Course: Cross-Fertilizing Age and Social Science Theory. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *58*(6), S327–S337.

de Moya-Anegon, F., Lopez-Illescas, C., Guerrero-Bote, V., Moed, H. F. (2019). Exploring the impact of scholarly journals in social sciences and humanities upon patentable technology. 17th International Conference of the International-Society-for-Scientometrics-and-Informetrics (ISSI) on Scientometrics and Informetrics, Sapienza Univ Rome, Rome, ITALY. 744-755.

Delre, S. A., Jager, W., Bijmolt, T. H. A., Janssen, M. A. (2007). Targeting and timing promotional activities: An agent-based model for the take off of new products. *Journal of Business Research*, *60*(8), 826-835.

Delre, S. A., Jager, W., Bijmolt, T. H. A., Janssen, M. A. (2010). Will it spread or not the effects of social influences and network topology on innovation diffusion. *Journal of Product Innovation Management*, *27*(2), 267-282.

Dosi, G . (1982). Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research Policy*, *11*, 147-162.

Dosi, G. (1991). The research on innovation diffusion: An assessment. *Diffusion of technologies and social behavior*,179-208.

Fung, M. M., Chow, W. W. (2002). Measuring intensity of knowledge flow with patent statistics. *Economic Letter*, *74*, 353-358.

Garfield, E. (1980). Premature Discovery or Delayed Recognition-Why?. *Current Contents*, (21), 5-10.

Garfield, E. (1989a). Delayed recognition in scientific discovery: Citation frequency analysis aids the search for case histories. *Current Contents*, *23*, 3-9.

Garfield, E. (1989b). More delayed recognition. Part 1. Examples from the genetics of color blindness, the entropy of short-term memory, phosphoinositides, and polymer Rheology. *Current Content*s, *38*, 3–8.

Garfield, E. (1990). More delayed recognition. Part 2. From inhibin to scanning electron microcopy. *Current Contents*, *9*, 3-9.

Glanzel, W., Garfield, E. (2004). The myth of delayed recognition. *The Scientist*, *18*(11), 8-9.

Glanzel, W., Schlemmer, B., Thijs, B. (2003). Better Late than Never? On the Chance to Become Highly Cited only Beyond the Standard Time Horizon. *Scientometrics*, *58*(3), 571-586.

Goldsby, T. J., & Zinn, W. (2016). Technology Innovation and New Business Models: Can Logistics and Supply Chain Research Accelerate the Evolution?. *Journal of Business Logistics*, *37*(2), 80-81.

Gui, J., Yao, C. Q ., Zeng, W., et al. (2015). How does Basic Research Promote the Innovation for Patented Invention: a Measuring of NPC and Technology Coupling. *International Conference on Management Science and Management Innovation* (MSMI), Guilin, PEOPLES R CHINA, *6*, 261-266.

Huang, J., Li, J., Chen, Y., & Sun, J. (2019). Burst Hotspots Dynamic Detection and Tracking on Large-scale Text Stream. *IEEE Access, 7*, 30913-30924.

Hegde. D., Luo. H. (2013).Imperfect information, Patent Publication, and the Market for Ideas. *Harvard Business School Strategy Unit Working Paper*, 14-19.

Hekkert, M. P., Negro, S. O. (2009). Functions of innovation systems as a framework to understand sustainable technological change: Empirical evidence for earlier claims. *Technological Forecasting and Social Change, 76*(4), 584-594.

Higham, K., de Rassenfosse, G., & Jaffe, A. B. (2021). Patent Quality: Towards a Systematic Framework for Analysis and Measurement. *Research Polic*y, *50*(4), 104215.

Hou, J., & Yang, X. (2019). Patent sleeping beauties: evolutionary trajectories and identification methods. *Scientometric*s, *120*(1), 187-215.

Hsueh. C., Wang, C. (2009). The use of social network analysis in knowledge diffusion research from patent data. 2009 International Conference on Advances in Social Networks Analysis and Mining, Athens, GREECE, 393-398.

Hu, A. G., Jaffe, A. B. (2003). Patent citation and international knowledge flow: The cases of Korea and Taiwan. International Journal of Industrial Organization, *21*,849-880.

Hutchins, M. (2003). Using Interlocking Additional Early Stage Patents to Improve and Extend Patent Protection. Journal of Medical Marketing: Device, Diagnostic, and Pharmaceutical Marketing, *3*, 212-215.

Jaffe, A. B., Trajtenberg, M. (1999). International knowledge flows: Evidence from patent citations. *Economics of Innovation and New Technology*, *8*, 105-136.

Jaffe, A. B., Trajtenberg, M., Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, *108*(3), 577-598.

Jiang, S., Gao, Q., Chen, H, C., Roco, M. C. (2015). The Roles of Sharing, Transfer, and Public Funding in Nanotechnology Knowledge-Diffusion Networks. *Journal of the Association for Information Science and Technology*. *66*(5), 1017-1029.

Johnson, D. K. N., & Popp, D. (2001). Forced Out of the Closet: The Impact of the American Inventors Protection Act on the Timing of Patent Disclosure. *RAND Journal of Economics*, *34* (1), 96-112.

Ke, Q., Ferrara, E., Radicchi, F., & Flammini, A. (2015). Defining and identifying sleeping beauties in science. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(24), 7426.

Kim, D. H., Lee, B. K., Sohn, S. Y. (2016). Quantifying technology-industry spillover effects based on patent citation network analysis of unmanned aerial vehicle (UAV). *Technological Forecasting and social Change*, *105*, 140-157.

Kleinberg, J. (2002). Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery*, *7*(4), 373-397.

Kleinberg, J. (2003). Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery*, 7, 373-397.

Komada, F. (1986). Japanese Studies on Technology Transfer to developing Countries : A Survey. *The Developing Economies*, *24*, 405-420.

Lachance, C., & Larivière, V. (2014). On the citation lifecycle of papers with delayed recognition. *Journal of Informetrics*, *8*(4), 863-872.

Lee, C., Kogler, D. F., & Lee, D. (2019). Capturing information on technology convergence, international collaboration, and knowledge flow from patent documents: A case of information and communication technology. *Information Processing & Management*, *56*(4), 1576-1591.

Li, J. (2014). Citation curves of all-elements-sleeping-beauties: flash in the pan first and then delayed recognition. *Scientometrics*, *100*(2), 595-601.

Li, J., Shi, D. B., Zhao, S. X., Ye, F. Y. (2014). A study of the "heartbeat spectra" for "sleeping beauties". *Journal of Informetrics*, *8*(3), 493-502.

Li, M., & Chu, Y. (2016). Explore the research front of a specific research theme based on a novel technique of enhanced co-word analysis. *Journal of Information Science*, *43*(6), 725-741.

Li, X., Chen, H., Huang, Z., et al. (2007). Patent citation network in nanotechnology(1976-2004). *Journal of Nanoparticle Research*, *9*(3), 337-352.

Liang, S., & de Rijke, M. (2015). Burst-aware data fusion for microblog search. *Information Processing & Management, 51*(2), 89-113.

Liu, W., Qiao, W., Wang, Y., & Xu, Z. (2021). Patent transformation opportunity to realize patent value: Discussion about the conditions to be used or exchanged. *Information Processing & Management*, *58*(4), 102582.

Liu, Y., & Rousseau, R. (2012). Towards a representation of diffusion and interaction of scientific ideas: The case of fiber optics communication. *Information Processing & Management*, *48*(4), 791-801.

Mac Garvie, M. (2005). The determinants of international knowledge diffusion as measured by patent citations. *Economics Letters*, *87*(1), 121-126.

Merton, Robert. K.(1973). The Sociology of Science: Theoretical and Empirical Investigations. Chicago: University of Chicago Press.

Meyer, M. (2000). Does science push technology? Patents citing scientific literature. *Research Policy*, *29*, 409-434.

Narin, F., Hamilton, K. S., Olivastro, D. (1997). The increasing linkage between US technology and public science. *Research Policy*, *26*, 317-330.

Okada, Y., & Nagaoka, S. (2020). Effects of early patent publication on knowledge dissemination: Evidence from U.S. patent law reform. *Information Economics and Policy*, *51,* 100852.

Park, I., Yoon, B. (2018). Technological opportunity discovery for technological convergence based on the prediction of technology knowledge flow in a citation network. *Journal of informetrics*, *12*(4), 1199-1222.

Pezzoni, M, R Veugelers, and F Visentin (2022), How fast is this novel technology going to be a hit? Antecedents predicting follow-on inventions. *Research Policy*, *51*(3), 104454.

Plečnik, J. M., Yang, L. L., & Zhang, J. H. (2021). Corporate innovation and future earnings: does early patent disclosure matter? *Accounting & Finance*. doi:10.1111/acfi.12851

Price, D. D. S. (1976). A ordinary theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, *27*(5), 292-306.

Price, D. J. D. S. (1965). The Scientific Foundations of Science Policy. *Nature*, *206*(4981), 233-238.

Schultz, T. W. (1990). Human Capital Investment and Urban Competitiveness. Beijing：The Commercial Press.

Schumpeter J A. (1934). The theory of economic development. Cambridge, MA: Harvard University Press.

Sharma, P., Tripathi, R, C. (2017). Patent citation: A technique for measuring the knowledge flow of information and innovation. *World Patent Information*, *51*, 31-42.

Shih, H., Chang, T. S. (2009). International diffusion of embodied and disembodied technology: A network analysis approach. *Technological Forecasting and Social Change*, *76*, 821-834.

Stent, G. S. (1972). Prematurity and Uniqueness in Scientific Discovery. *Scientific American*, *227*(6), 84-93.

Stoneman, T. P. (1989). Economic analysis of technological change. Beijing: Machinery Industry Press.

Takahashi, T, Tomioka, R, Yamanishi, K. (2014). Discovering emerging topics in social streams via link-anomaly detection. *IEEE Transactions on Knowledge and Data Engineering*, *26*(1), 120-130.

Tsang, M. P., Sonnemann, G. W., & Bassani, D. M. (2016). Life-cycle assessment of cradle-to-grave opportunities and environmental impacts of organic photovoltaic solar panels compared to conventional technologies. *Solar Energy Materials and Solar Cells*, *156*, 37-48.

Van Raan, A. F. J. (2004). Sleeping beauties in science. *Scientometrics*, *59*(3), 467-472.

Van Raan, A. F. J. (2021). Sleeping beauties gain impact in overdrive mode. *Scientometrics*, *126*(5), 4311-4332.

Verspagen, B., Loo, I. D. (1999). Technology spillovers between sectors and over time. *Technological Forecasting and Social Change*, *60*, 215-235.

Veugelers, R., Wang, J. (2019). Scientific novelty and technological impact. *Research Policy*, *48*(6),1362-1372.

Wu, Q., Zhang, H., & Lan, J. (2014). K-State automaton burst detection model based on KOS: Emerging trends in cancer field. *Journal of Information Science*, *41*(1), 16-26.

Yang, W., Yu, X., Zhang, B., Huang, Z. Y. (2021). Mapping the landscape of international technology diffusion (1994-2017): network analysis of transnational patents. *Journal of Technology Transfer*, *46*, 138-171.

Ye, F. Y., & Bornmann, L. (2018). “Smart girls” versus “sleeping beauties” in the sciences: The identification of instant and delayed recognition by using the citation angle. *Journal of the Association for Information Science and Technology*, *69*(3), 359-367.

You, H. L., Li, M. J., Hipel, K. W., et al. (2017). Development trend forecasting for coherent light generator technology based on patent citation network analysis. *Scientometrics*, *111*(1), 297-315.

Zhang, G., Xiong, L., Duan, H., & Huang, D. (2020). Obtaining certainty vs. creating uncertainty: Does firms’ patent filing strategy work as expected? *Technological Forecasting and Social Change*, *160*, 120234.

Zheng, Q., Huang, L. C., Wu, F. F., Dan, W., Hui, Z. (2017). Analyzing Technological Knowledge Diffusion Among Technological Fields Using Patent Data: The Example of Microfluidics. *International Journal of Innovation and Technology Management*, *14*(1): SI.