Journal of Informetrics 13 (2019) 100978

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Bibliographically coupled patents: Their temporal pattern and combined relevance

Chung-Huei Kuan[a], Dar-Zen Chen[b,d,*], Mu-Hsuan Huang[c]

[a] National Taiwan University of Science and Technology, Graduate Institute of Patent, No. 43, Sec. 4, Keelung Rd., Taipei, Taiwan, ROC
[b] National Taiwan University, Department of Mechanical Engineering, No. 1, Sec. 4, Roosevelt Rd., Taipei, Taiwan, ROC
[c] National Taiwan University, Department of Library and Information Science, No. 1, Sec. 4, Roosevelt Rd., Taipei, Taiwan, ROC
[d] National Taiwan University, Center for Research in Econometric Theory and Applications (CRETA), No. 1, Sec. 4, Roosevelt Rd., Taipei, Taiwan, ROC

ARTICLE INFO

Article history:
Received 23 July 2019
Received in revised form 2 October 2019
Accepted 4 October 2019
Available online 9 November 2019

Keywords:
Bibliography coupling
Temporal pattern
Patent expansion
Reference expansion
Combined relevance

ABSTRACT

Bibliographically coupled patents reveal a temporal pattern associated with their ages (how long ago they have been issued) and spans (their distance in time). The coupling strength of the aged or long-spanned patent pairs may be inherently limited, especially for fields that reveal patent and reference expansions. This study proposes a simple measure, *combined relevance* (CR), to counter such impact. Unlike traditional measures, CR provides more balanced treatment to short- and long-spanned but favors more aged patents pairs. A fixed CR threshold may be more safely applied with a reduced possibility of erroneously removing patent pairs that are truly related. For observing long-term knowledge dissemination or tracing overall development trajectory, CR may be an alternative.

1. Introduction

Patent bibliometric works often involve the detection and measurement of relatedness between patents. Based on their relatedness, structures may be derived from thousands of seemingly disorganized patents. Then, collections of related patents may be modeled and monitored as a unit; evolving trends may be detected by observing these collections in their chronological orders. The cooperation/competition relationship and knowledge exchange between firms, institutions, counties, and fields may be examined and inferred based on the relatedness between their patents.

Researchers have been using terms "relatedness," "similarity," and "proximity" along with terms "technology," "patent," and "knowledge" in different contexts to refer to similar concepts. It seems to us that, for specific features of patents, "similarity" is more often used, such as "keyword similarity" (cf. Giudice, Musarella, Sofo, & Ursino, 2019) and "classification similarity" (cf. Kuan, Chiu, Liu, Huang, & Chen, 2018); "proximity" is more often applied in knowledge spillover context and usually at firm or higher aggregate level (cf. Chu, Tian, & Wang, 2018). This study chooses to use "relatedness" as it seems to be a broader and more abstract concept. For example, Makri, Hitt, and Lane (2010) considered that technological relatedness covers not only technological similarity but also technological complementarity.

There are three major categories of approaches in detecting and measuring patent relatedness. The text-based approaches extract textual information from patents' specifications (cf. Arts, Cassiman, & Gomez, 2018; Moehrle & Gerken, 2012; Niemann, Moehrle, & Frischkorn, 2017; Ortiz-de-Urbina-Criado, Nájera-Sánchez, & Mora-Valentín, 2018; Yoon & Magee,

* Corresponding author at: National Taiwan University, Department of Mechanical Engineering, No. 1, Sec. 4, Roosevelt Rd., Taipei, Taiwan, ROC.
E-mail addresses: maxkuan@mail.ntust.edu.tw (C.-H. Kuan), dzchen@ntu.edu.tw (D.-Z. Chen), mhhuang@ntu.edu.tw (M.-H. Huang).

https://doi.org/10.1016/j.joi.2019.100978
1751-1577/

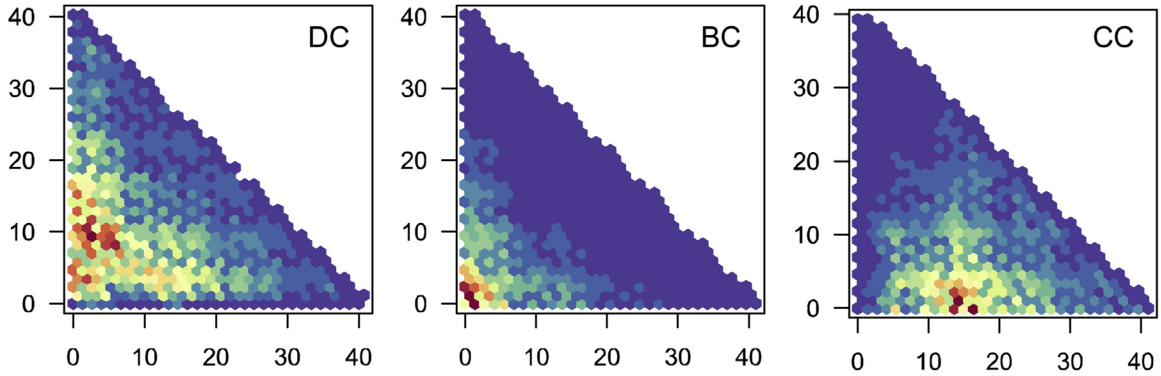

Fig. 1. Frequency distribution of DC, BC, and CC pairs according to ages (x) and spans (y).

2018). The classification-based approaches evaluate the overlapping of classification symbols assigned to patents by examiners (cf. Angue, Ayerbe, & Mitkova, 2014; Chang, 2012; Jaffe, 1986; Kuan, Chiu et al., 2018; Kuan, Huang, & Chen, 2018; Petruzzelli, 2011; Wang, Hou, & Hung, 2018). The third category, citation-based approaches, generally involve three common mechanisms: direct citation (DC) (Trajtenberg, 1990), bibliographic coupling (BC) (Kessler, 1963), and co-citation (CC) (Small, 1973). These mechanisms may be used individually or together. Kuusi and Meyer (2007) employed BC alone to cluster related patents and identify an emerging technological paradigm. Lo (2007) also employed only BC to identify technological connections between major research organizations. Von Wartburg, Teichert, and Rost (2005) combined DC and BC in a multistage analysis to reveal technological change. Chen, Huang, Hsieh, and Lin (2011), Kuan, Huang et al. (2018), and Yeh, Sung, Yang, Tsai, and Chen (2013) used both DC and BC to construct more comprehensive citation networks. Citation-based approaches may also be applied together with the other two approaches. Kuan, Chiu et al. (2018), Kuan, Huang et al. (2018) and Leydesdorff, Kushnir, and Rafols (2014) integrated DC with patent classification codes, Nakamura, Suzuki, Sakata, and Kajikawa (2015) combined DC and co-word analysis, and Park, Jeong, Yoon, and Mortara (2015) used BC and patent text semantic analysis to locate potential R&D collaboration partners.

DC, BC, and CC are all valid means, as evidenced by a large body of works where the aforementioned are only a few samples. These mechanisms, however, may conflict when one suggests relatedness whereas another indicates otherwise. In a previous study investigating one such conflict involving patents that do not cite each other but are strongly bibliographically coupled, we found that this phenomenon, referred to as *missing link*, are not coincidental (Kuan, Chiu et al., 2018; Kuan, Huang et al., 2018). For DC to occur, the cited patent has to be published earlier so that it is visible and citable to the applicant or examiner of the citing patent. Therefore, DC rarely occurs between patents whose application processes are highly overlapped, as their applicants or examiners may be blind to each other, even though the two patents are indeed related. BC is not handicapped as such and may effectively reveal the patents' relatedness. We then proposed and verified through empirical analysis that BC is useful in revealing concurrent patents embodying cotemporaneous development that would be overlooked by using DC alone.

During the previous study, in addition to discovering BC's supplemental function to DC in developing a more comprehensive trajectory of patent evolvement, we also noticed that bibliographically coupled patents reveal a temporal pattern involving their ages (how long ago they have been issued) and spans (their distance in time) and this pattern would place an inherent bound on their coupling strength. This pattern and its implication are described and discussed in the following section. Then, an improved measure to the coupling strength is proposed to counter the age and span effects. Finally, the new measure is applied to real case data, and the result is compared to that by a conventional measure.

2. Temporal pattern and its cause

2.1. Phenomenon

To demonstrate this temporal pattern, a same dataset from the previous study (Kuan, Chiu et al., 2018; Kuan, Huang et al., 2018) is borrowed here, so that observation made in this study may be compared against the previous study. This dataset includes 34,083 US utility patents issued between 1976/1/1 and 2017/3/31 in the field of carbon dioxide capture, storage, recovery, delivery, and regeneration, which is one of the major solutions to global warming. These patents are retrieved if they contain at least one specific keyword in at least one relevant field (i.e., Title, Abstract, Specification, or Claims) and at least one specific Cooperative Patent Classification (CPC) symbol prefix.[1] These keywords and CPC prefixes are compiled

[1] The keyword search command is '(carbon or dioxide$ or co2) AND (storage$ or captur$ or recover$ or deliver$ or regenerat$),' and the CPC symbol prefixes include B63B 35$, C01B 3$, C01B31/20, C01B 21/22, C02F 1$, C07C 7/10, F01N 3/10, F25J 3/02, B01J 20$, B01D 53$, and B01D 11, where '$' is the wildcard character.

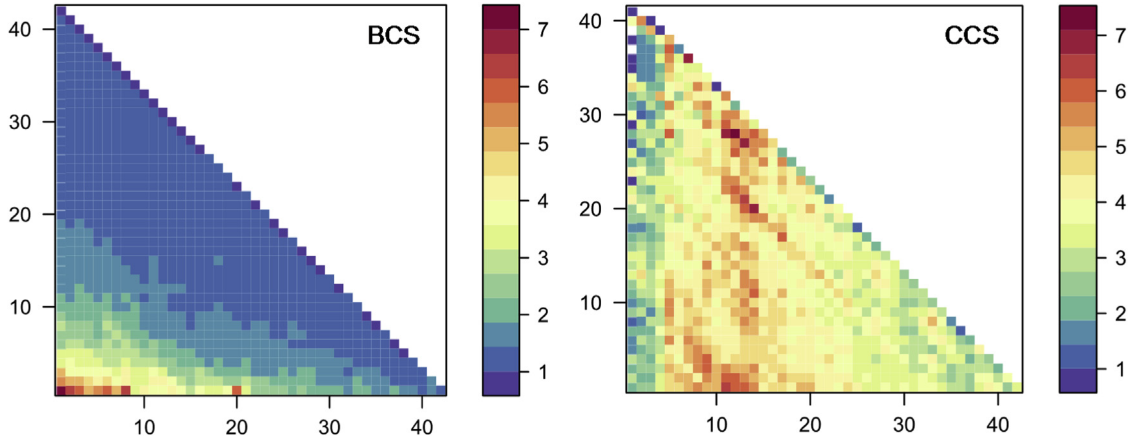

Fig. 2. Average BCS and CCS according to ages (x) and spans (y).

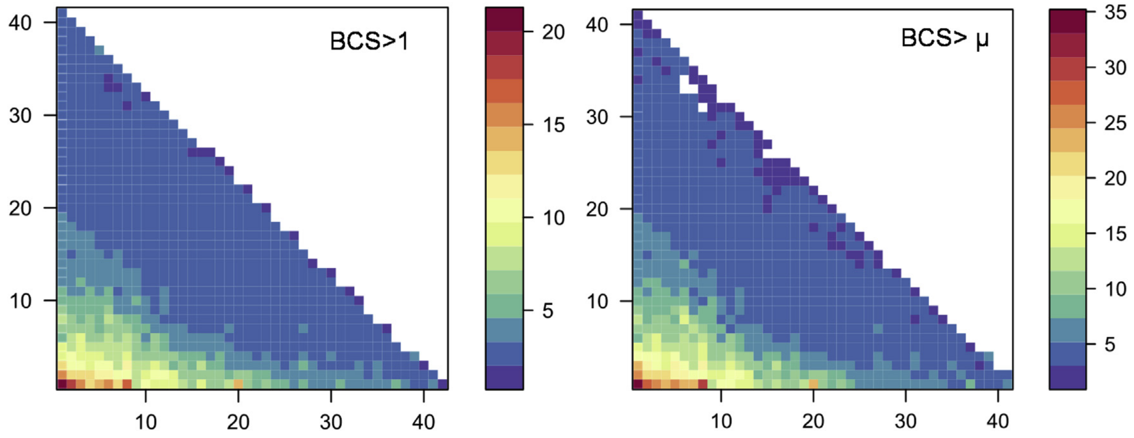

Fig. 3. Average BCS for pairs with BCS > 1 and BCS > μ according to ages (x) and spans (y).

based on the search strategies published in an official technology analysis report (World Intellectual Property Organization, 2009) and in a prior study (Gray et al., 2009).

Among the 34,083 patents, there are 154,505 pairs of cited and citing patents, 1,609,549 pairs of bibliographically coupled patents, and 644,376 pairs of co-cited patents, hereinafter respectively referred to as DC pairs, BC pairs, and CC pairs. Each of the DC, BC, and CC pairs includes patents P_E and P_L respectively issued at an earlier date t_E and a later date t_L ($t_E \leq t_L$). Then, the *span* and *age* of a DC, BC, or CC pair are defined as $t_L - t_E$ and $t_{NOW} - t_L$, where t_{NOW} denotes the cut-off date of the patent data collection (e.g., 2017/03/31 for this dataset).

Fig. 1 shows the frequency distributions of all DC, BC, and CC pairs according to their ages (horizontal axis) and spans (vertical axis) in years, where more reddish or bluish points reflect higher or lower counts. BC pairs are particularly concentrated in the lower left corner, meaning, more BC pairs have shorter spans and younger ages, or BC tends to occur between patents issued close to each other and more recently. One may speculate that this phenomenon is inevitable as there are more patents issued more recently (which is partially true, as will be discussed in the next section). However, once DC, BC, and CC pairs are observed alongside each other relative to their respective ages and spans, it is interesting to see that DC and CC pairs do not reveal such propensity, even though they are all from the same set of patents revealing the same more-patents-in-more-recent-years phenomenon.

Not only that, the bibliographic coupling strength (BCS) of the BC pairs also reveals a similar pattern. Fig. 2 shows the average BCS and average co-citation strength (CCS) for BC and CC pairs across ages and spans, where BCS and CCS here are measured as the number of cited and citing patents in common. Again, BC pairs having higher BCS (i.e., more reddish points) are particularly concentrated in the lower left corner, and CCS does not show any significant pattern.

Jarneving (2007a) and Swanson (1971) indicated that only BC pairs having BCS above a threshold are truly related. Indeed, among the 1,609,549 BC pairs, up to 1,167,794 (72.55%) of them have the smallest BCS of 1. To avoid that the pattern shown in Fig. 2 is resulted from a large volume of noises, Fig. 3 shows the distributions of average BCS after those BC pairs having BCS not greater than 1 (left) and having BCS not greater than the overall average BCS (2,74 or μ) (right) are filtered out. The same pattern still prevails.

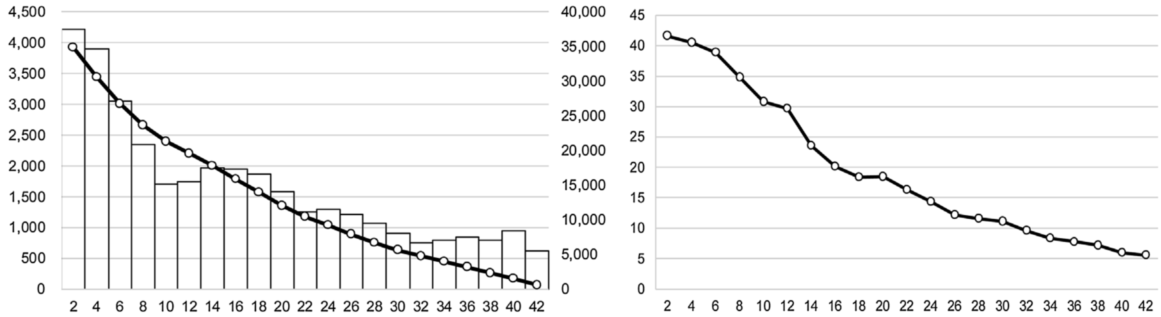

Fig. 4. A field's expanding numbers of total patents (left) and cited references (right).

Using Pearson correlation test on BC pairs having BCS > 1, BC's frequency is significantly correlated with age and span with respective correlation coefficients of −0.83 (p-value 7.098e-12) and -0.79 (p-value 5.557e-10). BCS is also significantly correlated with age and span with respective correlation coefficients of −0.91 (p-value of 2.2e-16) and −0.76 (p-value of 6.27e-09).

2.2. Patent and reference expansion

Bibliographically coupled patents' unique temporal pattern is related to the increasing number of total patents in a field, referred to as the field's *patent expansion*. The left diagram of Fig. 4 depicts this growing trend using the case data, where patents are distributed along the horizontal axe according to their ages (i.e., how long ago they are issued or the difference between their issue dates and t_{NOW}) in 2-year intervals. Therefore, older patents issued earlier in the past are plotted farther to the right, whereas younger patents issued recently are drawn closer to the left. Then, the bars show the numbers of patents issued within each age interval against the left scale; the curve depicts the numbers of total patents accumulated up to each age interval against the right scale. As illustrated in the left diagram, the total patent curve rises monotonically from right to left as patents of the field are issued and accumulated over time.

Due to a field's patent expansion, later patents of the field have more prior patents available to cite and, therefore, may have more references than earlier patents do. This growing numbers of references over time, referred to as the field's *reference expansion*, is depicted in the right diagram of Fig. 4, where the curve shows the average numbers of references cited by patents issued within each age interval, and the curve also rises continuously from right to left.

What is depicted in Fig. 4 conforms to the observations made by Hall, Jaffe, and Trajtenberg (2001) and Zhang, Huang, and Chen (2018). Both works noticed that the number of references made per U.S. patent and the number of U.S. patents issued both increase over time. However, we speculates that, while the patent expansion phenomenon should be common to all fields and to all counties, reference expansion should be particularly applicable to U.S. patents. U.S. regulation obligates applicants to disclose and cite all information (e.g., prior patents) known to be relevant to the applications' patentability (Bicknell, 2008). U.S. applicants, therefore, tend to cite more comprehensively when there are more prior patents to cite.

Facing the patent and reference expansions, Hall et al. (2001) and Zhang et al. (2018) were concerned about the "devaluation" of citations (i.e., a patent's "later citations are less significant than earlier ones"). This study, however, is concerned about their implication on BC and patent relatedness based on BC.

2.3. Implication for BC and BCS

Fig. 5, using the case data's yearly issued patents as backdrop, provides four scenarios for a pair of patents P_E and P_L from a same field. P_E and P_L are issued on t_E and t_L ($t_E \le t_L$) with references REF_E and REF_L respectively drawn from total patents TP_{t_E} and TP_{t_L} accumulated up to t_E and t_L. In the figure, TP_{t_E} is represented by dark bars, and TP_{t_L} by both light and dark bars. The light bars, therefore, reflect the size difference between TP_{t_E} and TP_{t_L}.

The four scenarios are designated respectively as young or old age, and short or long span, based on the age and span of the patent pair in a relative manner. For scenarios (A) and (B), the patent pair is said to have young age as t_L is more recent (closer to the left) than that in scenarios (C) and (D). Similarly, the patent pair in scenarios (A) and (C) is of short span as t_E and t_L are closer to each other than those of scenarios (B) and (D).

P_E and P_L then should satisfy the following Eqs. (1) and (2):

$$REF_E \subseteq TP_{t_E}, \quad \left|REF_E\right| \le \left|TP_{t_E}\right|, \text{ and} \tag{1}$$

$$REF_L \subseteq TP_{t_L}, \quad \left|REF_L\right| \le \left|TP_{t_L}\right|. \tag{2}$$

Then, Eq. (3) may be derived according to the patent expansion:

$$TP_{t_E} \subseteq TP_{t_L}, \quad \left|TP_{t_E}\right| \le \left|TP_{t_L}\right| \tag{3}$$

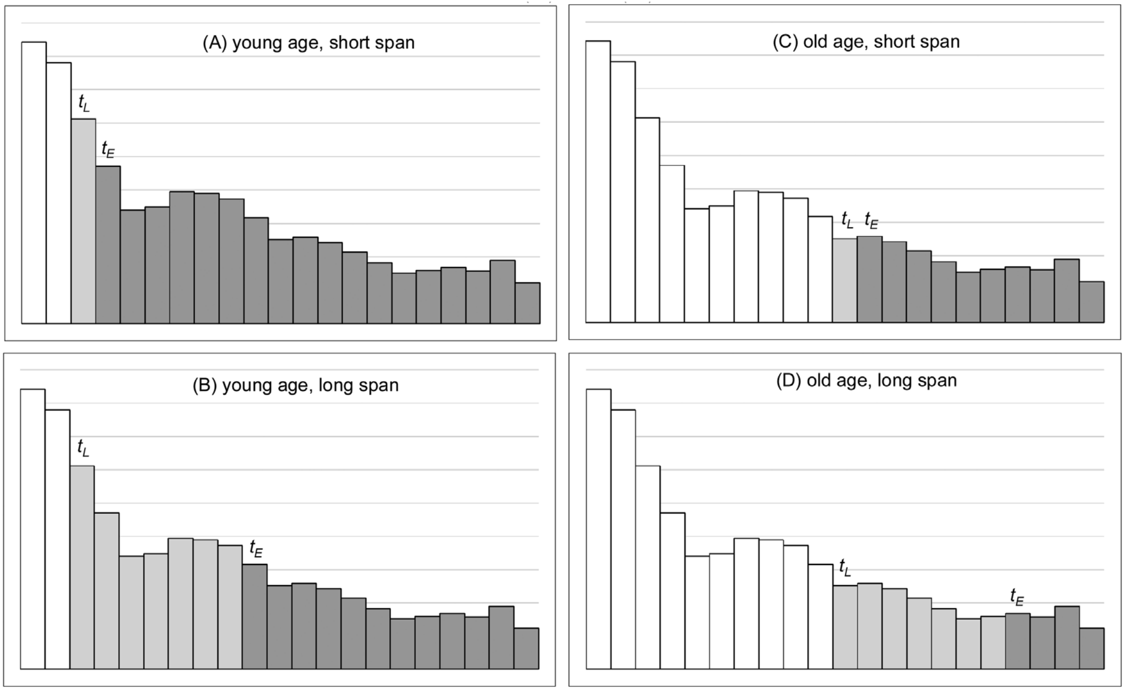

Fig. 5. Four scenarios for a pair of patents P_E and P_L from a same field.

Table 1
Shares of BC pairs having $|\boldsymbol{REF_E}| \leq |\boldsymbol{REF_L}|$.

BCS>	0	2	4	6	8	10	12	14	16	18
# of BC pairs	1,609,549	214,324	94,090	59,260	43,381	34,989	29,357	25,882	23,283	21,284
% of $\lvert REF_E\rvert \leq \lvert REF_L\rvert$	65.46%	66.44%	67.70%	68.46%	69.54%	70.62%	71.70%	72.49%	73.16%	73.59%

If the reference expansion is valid for the field, then we have Eqs. (4) and (5):

$$\left|REF_E\right| \leq \left|REF_L\right|, \text{ and} \tag{4}$$

$$\left|REF_E \cap REF_L\right| \leq \min\left(\left|REF_E\right|, \left|REF_L\right|\right) = \left|REF_E\right| \leq \left|TP_{t_E}\right|. \tag{5}$$

If P_E and P_L are to form a BC pair, they need to have non-empty reference intersection or $REF_E \cap REF_L \neq \emptyset$. According to Eq. (5), $\left|REF_E \cap REF_L\right|$ is bounded by $\left|TP_{t_E}\right|$, which is the number of patents commonly available for citation to both P_E and P_L (i.e., the dark grey bars in Fig. 5). Therefore, larger TP_{t_E} implies a greater chance for P_E and P_L to achieve non-empty intersection and thus forming a BC pair.

For the four scenarios, whether t_L is more recent as in scenarios (A) and (B) or earlier in the past as in (C) and (D), $\left|TP_{t_E}\right|$ is greater when P_E and P_L have a shorter span. On the other hand, whether P_E and P_L have a shorter span in scenarios (A) and (C) or a longer span as in (B) and (D), $\left|TP_{t_E}\right|$ is greater when t_L is more recent. This is why more BC pairs have shorter spans and younger ages as illustrated in Fig. 1. Similarly, for a BC pair involving P_E and P_L, its BCS, $\left|REF_E \cap REF_L\right|$, is also bounded by $\left|TP_{t_E}\right|$, and larger TP_{t_E} provides a greater chance in achieving higher BCS. This is why BC pairs having shorter spans and younger ages tend to have greater BCS as illustrated in Figs. 2 and 3.

For BC pairs to deliver the unique temporal behavior, the reference expansion or Eq. (4) should hold. This is indeed true for a major portion of the case's BC pairs. Table 1 lists the shares of BC pairs satisfying Eq. (4), $\left|REF_E\right| \leq \left|REF_L\right|$, among those whose BCS is above ten different thresholds. For all BC pairs (BCS > 0), despite a large number of BC pairs that may be noises, there is still more than 65% of the BC pairs satisfying $\left|REF_E\right| \leq \left|REF_L\right|$. At a higher threshold, there are even greater portions satisfying $\left|REF_E\right| \leq \left|REF_L\right|$.

3. Adjustment for temporal pattern

3.1. Problem with conventional measure and threshold

Using a threshold to filter BC pairs, despite a common practice, may not be appropriate for a field revealing the temporal pattern. For example, if a threshold for BCS is set to the overall average (3) of the case's BC pairs, pretty much all BC pairs denoted by the greenish or bluish points in Fig. 2 are filtered out, even though some of these aged or long-spanned BC pairs may reflect true relatedness.

This filtering-without-distinction problem is not only due to the fixed threshold use, but also due to the BCS measure's overlooking the age and span factors of BC pairs. The frequently used BCS measures may be classified into two broad categories: *intersection-based* and *vector-based* measures with Jaccard coefficient (Jaccard, 1901) and coupling angle (Glänzel & Czerwon, 1996) as representatives. If Jaccard coefficient or coupling angle is applied to a field with the temporal pattern, all BCS would be bounded by $\left|REF_E\right|$, the earlier patents' reference size, as derived in the following Eqs. (6) and (7):

$$\frac{\left|REF_E \cap REF_L\right|}{\left|REF_E \cup REF_L\right|} \le \frac{\left|REF_E\right|}{\left|REF_E \cup REF_L\right|}, \text{ and} \tag{6}$$

$$\frac{\vec{REF_E} \cdot \vec{REF_L}}{\left|\vec{REF_E}\right| \left|\vec{REF_L}\right|} = \frac{\left|REF_E \cap REF_L\right|}{\left|\vec{REF_E}\right| \left|\vec{REF_L}\right|} \le \frac{\left|REF_E\right|}{\left|\vec{REF_E}\right| \left|\vec{REF_L}\right|}, \tag{7}$$

where $\vec{REF_E}$ and $\vec{REF_E}$ are REF_E and REF_L expressed in binary vectors of equal dimension.

Bibliometric researchers had noticed the age and span problem. For example, Jarneving (2007b) indicated that "an increase of the distance in time between bibliographically coupled articles leads to a diminishing pool of shared references as there is a tendency to cite the more current articles" That is why usually an observation window is set up so that bibliographically coupled research articles published closer (i.e., about the same age) within the window (i.e., limited span) are collected and compared together (cf. Glänzel & Czerwon, 1996; Jarneving, 2007b).

3.2. Combined relevance

To observe the knowledge flow or to develop a representative trajectory among patents across an extended period of time, where all relevant BC pairs have to be considered, a BCS measure as much immune to the age and span effects as possible would be desirable.

This study, therefore, proposes a simple BCS measure, referred to as *combined relevance* (CR), in Eq. (8):

$$\left(\frac{\left|REF_E \cap REF_L\right|}{\left|REF_E\right|}\right)\left(\frac{\left|REF_E \cap REF_L\right|}{\left|REF_L\right|}\right) = \frac{\left|REF_E \cap REF_L\right|^2}{\left|REF_E\right| \left|REF_L\right|}. \tag{8}$$

The idea behind Eq. (8) is straightforward. Imaging that REF_E and REF_L respectively represent the information respectively combined within patents P_E and P_L, and that $REF_E \cap REF_L$ is the piece of information shared between them, the two factors in Eq. (8) measure how much this shared information accounts for P_E and P_L's combined information, or this shared information's *individual relevance* to P_E and P_L. Then, P_E and P_L are highly related if their shared information is significant to both P_E and P_L, or they have high CR.

Fig. 6 shows the distributions of the two factors' averages for BC pairs whose BCS is greater than 1 in two separate diagrams. The BC pairs are limited to those having BCS > 1 so that the observation is not impaired by a large volume of noises. As illustrated in the left diagram, P_E's factor tends to be greater for BC pairs that are distributed closer to the diagonal and somewhat more concentrated towards the lower right corner (there are a few red dots on the lower right section of the diagonal). In other words, these BC pairs have higher (age + span) values, and age seems to play a more dominant role. When a pair's age is great, its span is inherently limited and, when its span is great, the pair should have young age. When either age or span is great, the pair's should have an early t_E in the past and, therefore, have a small REF_E. The left factor of Eq. (8), i.e., the individual relevance to P_E, as such would be close to 1 by having $|REF_E|$ as denominator. As to the BC pairs denoted by the red dots, they are aged pairs with limited spans. Their P_E and P_L, as suggested in scenario (C) in Fig. 5, face a similar set of candidates for reference and, therefore, there is a greater chance that whatever cited by P_E is also cited by P_L, meaning $REF_E \cap REF_L = REF_E$ and $\frac{|REF_E \cap REF_L|}{|REF_E|} = 1$.

On the other hand, as illustrated in the right diagram, P_L's factor tends to have higher value for those aged but shorter-spanned BC pairs located closer to the lower right corner. A similar reasoning as applied above would explain why. When a pair's age is great, the pair's should have both early t_E and t_L in the past and, therefore, have limited REF_E and REF_L. The right factor of Eq. (8), i.e., the individual relevance to P_L, as such would be close to 1 by having $|REF_L|$ as denominator.

Fig. 7 shows the distribution of the average CR for BC pairs whose BCS is greater than 1. By comparing Fig. 7 with Fig. 3, one may see that CR is more uniformly distributed than BCS both in terms of ages and spans. CR achieves diminished span effect as the left and right factors of Eq. (8) cancel out each other's favor and disfavor to the long-spanned pairs, as discussed

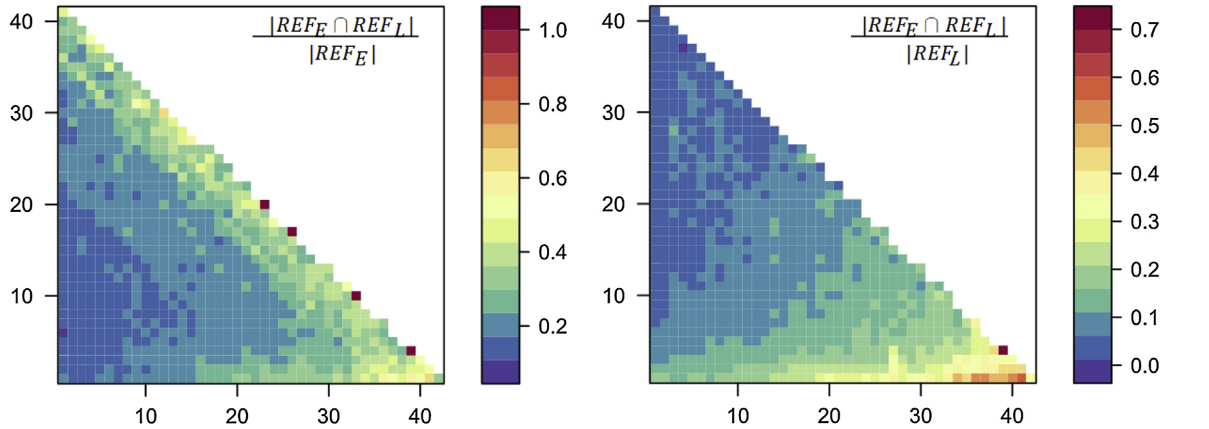

Fig. 6. Distributions of average individual relevance to P_E (left) and P_L (right) of pairs with BCS > 1 according to ages (x) and spans (y). (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

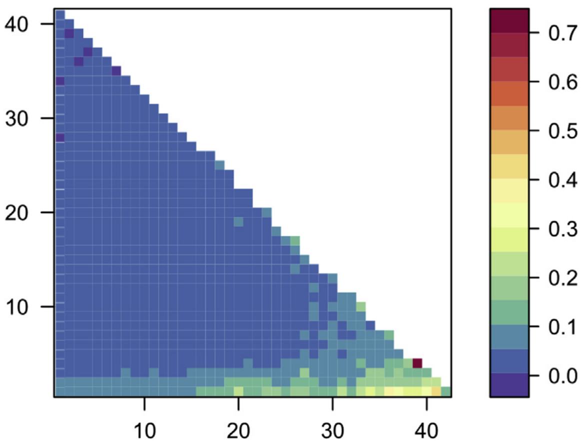

Fig. 7. Distributions of average CR for pairs with BCS > 1 according to ages (x) and spans (y).

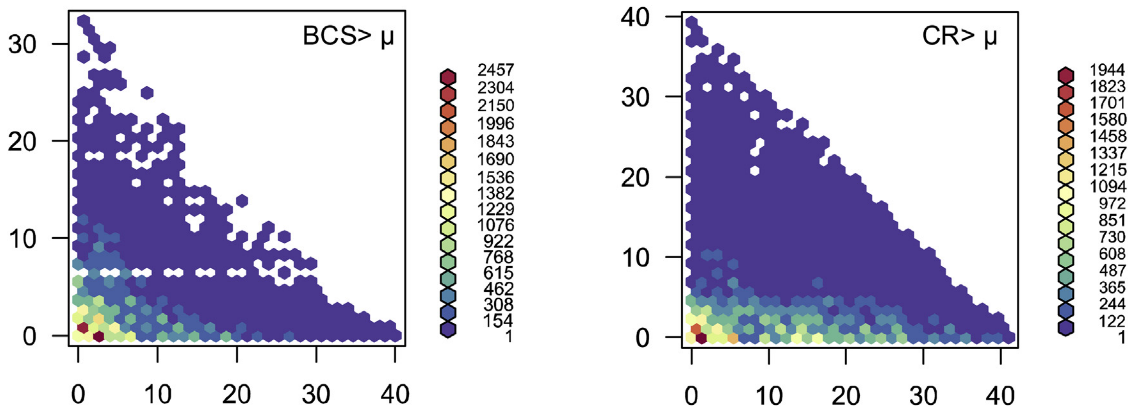

Fig. 8. Frequency distributions of BC pairs having above average BCS (left) and CR (right) according to ages (x) and spans (y).

above. The age effect, however, remains as both left and right factors of Eq. (8) prefer aged pairs. Nonetheless, CR's bias towards aged pairs is not as severe as BCS's overly preference over young or short-spanned pairs.

Due to CR's more uniform distribution, CR retains more long-spanned pairs. Fig. 8 provides two diagrams, the left one showing the frequency distributions of 49,873 BC pairs having above average BCS (7.32 for all BC pairs with BCS > 1) and the right one showing 55,954 BC pairs having above average CR (0.043 for all BC pairs with BCS > 1). As illustrated, BCS filters out most pairs having spans above 30, whereas a number of them are still retained by CR.

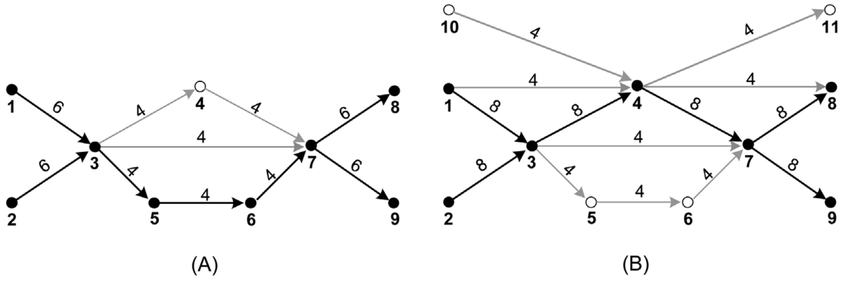

Fig. 9. Fictitious citation networks with arc weights assigned and main paths revealed.

CR is aimed to reduce the age and span impact on assessing BC pairs' relatedness. It is not ideal as the span effect is more satisfactorily resolved whereas the age effect is only improved to a lesser degree. CR, however, is as simple as the conventional measures, both conceptually and computationally. A fixed threshold then may be more safely applied with a reduced possibility of erroneously removing BC pairs having true relatedness.

The above discussion also reveals that BCS and CR seem to complement each in a certain manner. The left and right diagrams of Fig. 8 have 27,369 BC pairs in common, accounting for 55% of the pairs in the left diagram (BCS), and 49% in the right diagram (CR). In other words, for about half of the BC pairs considered to reflect relatedness or un-relatedness by BCS, they are indicated otherwise by CR and vice versa.

Therefore, it should be noted that there are advantages and disadvantages to both BCS and CR. CR may overlook patent pairs of young ages. BCS, on the other hand, prefers more young and short-spanned pairs. Their difference is more clearly revealed in the following section.

4. Comparison using main path analysis

To see how CR differs from conventional BCS within an application setting, this study applies *main path analysis* (MPA) on patent networks (PNs) supplemented with the so-called missing links (ML) (Kuan, Chiu et al., 2018; Kuan, Huang et al., 2018), which are BC pairs without DC but strongly bibliographically coupled.

This study obtains two sets of MLs, one according to a BCS threshold (ML_{BCS}) and one according to a CR threshold (ML_{CR}). The study then constructs three PNs, the traditional patent citation network PN_{DC}, PN_{BCS} which is PN_{DC} supplemented with ML_{BCS}, and PN_{CR} which is PN_{DC} supplemented with ML_{CR}. Each ML within PN_{BCS} and PN_{CR} is simulated as a fictitious DC whose arc is incident from the earlier, lower-number patent into the later, higher-number patent. This study then derives three representative trajectories, or *main paths*, denoted as MP_{DC}, MP_{BCS}, and MP_{CR}, through MPA.

The rationale behind the empirical analysis is that, on one hand, the three networks are epitomized by their MPs so that BCS and CR's effects may be compared efficiently by examining their MPs. On the other hand, MP_{DC} functions as a reference frame against which MP_{BCS} and MP_{CR} are positioned so as to see how MP_{DC} is altered by the supplement of ML_{BCS} and ML_{CR}.

4.1. Main path analysis

Main path analysis (MPA) is a network-based analytic method first used to determine a representative development trajectory for a scientific field using a citation network constructed from the field's research articles (Hummon & Doreian, 1989). This method is well-received and adopted in a wide range of bibliometric and technological management applications. For some examples, MPA has been employed to detect technological changes and knowledge transformation (cf. Martinelli, 2012), to review a field's literature (cf. Calero-Medina & Noyons, 2008; Lu, Hsieh, & Liu, 2016), and to map technological development (cf. Park & Magee, 2017). The popular social network analysis software Pajek (De Nooy, Mrvar, & Batagelj, 2011) has built-in MPA functions.

It should be noted that PN_{BCS} and PN_{CR} are no longer citation networks as their supplemented ML_{BCS} and ML_{CR} are "fake" citations and, therefore, MP_{BCS} and MP_{CR} shouldn't be interpreted from the traditional perspective of knowledge flow. This analysis here is mainly for comparison's sake. However, it is not without merit. Our previous work (Kuan, Chiu et al., 2018; Kuan, Huang et al., 2018) found that ML_{BCS} is able to detect concurrent patents embodying cotemporaneous technological development. From the previous discussion, we now understand that ML_{BCS}, following the conventional BCS measure, inevitably includes those short-spanned and more recent pairs, thereby leading to the detection of concurrent patents.

MPA involves two major steps. Each arc of the network is first assigned a weight related to its traversal count within the network. Then, a series of connected arcs is determined as the main path. There are different weight assignment and path determination algorithms. This study chooses to use the *search path count* (SPC) algorithm (Batagelj, 2003) and the *global search* method (Liu & Lu, 2012). Fig. 9 provides two fictitious citation networks (A) and (B) whose arc weights assigned using SPC are shown along the arrows, and main paths determined by the global search method include the dark arcs.

Table 2
Summary statistics for Comparisons I and II.

PN_{DC}	*Comparison I (>μ)*		*Comparison II (>μ + 2σ)*	
	PN_{BCS}	PN_{CR}	PN_{BCS}	PN_{CR}
$\|DC\| / \|MP_{DC}\|$	$\|ML_{BCS}\| / \|MP_{BCS}\|$	$\|ML_{CR}\| / \|MP_{CR}\|$	$\|ML_{BCS}\| / \|MP_{BCS}\|$	$\|ML_{CR}\| / \|MP_{CR}\|$
154,505 / 27	49,873 / 130	55,954 / 127	7750 / 53	17,889 / 61

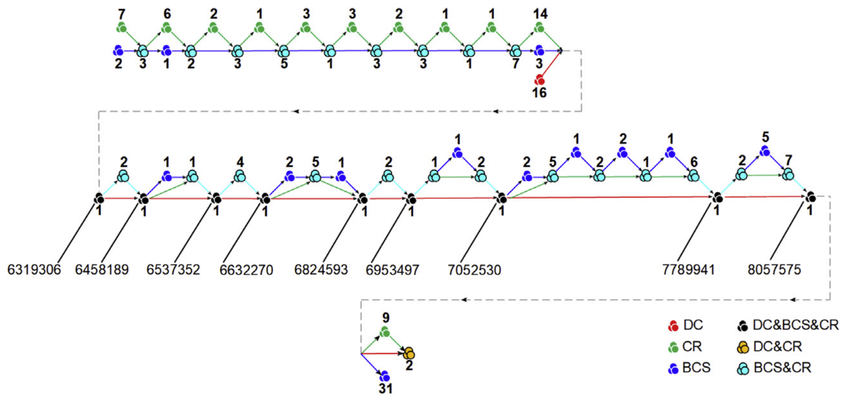

Fig. 10. Combined MP_{DC}, MP_{CR}, and MP_{BCS} from Comparison I. (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

SPC and other algorithms available from Pajek all decide an arc's weight based on its structural connectivity within the network (Hummon & Doreian, 1989), i.e., how many source and/or preceding nodes reaching the arc and how many sink and/or succeeding nodes reached through the arc. For example, an arc's SPC weight is the traversal count of the arc from all source nodes to all sink nodes. Therefore, the arc 3→4 of network (A) has a weight 4, as it is traversed four times from the two source nodes (1, 2) to the two sink nodes (8, 9). Similarly, the arc 3→4 of network (B) is traversed from two source nodes (1, 2) twice to reach sink node 11, four times to reach the sink node 8, and twice to the sink node 9, and therefore has a weight 8.

The global search method selects one or more paths from source to sink nodes having the highest total weight. Therefore, there are four main paths in the network (B), all through the internal nodes (3, 4, 7) and all with the combined weight 32 (=8 + 8+8 + 8). The global search method favors longer paths because they tend to have higher combined weights, as can be seen from network (A). A shorter path must have some significantly weighted links so that it may outweigh the longer path. For example, by having an additional sink node 11 to reach through the arc 3→4 and an additional source node 10 to reach the arc 4→7, the shorter segment 3→4→7 outweighs the longer segment 3→5→6→7, and becomes part of the main paths.

4.2. Main path comparison

Using the same case data, this study conducts two comparisons. For all BC pairs without DC and having BCS > 1, Comparison I respectively obtains ML_{BCS} and ML_{CR} from these BC pairs further having BCS and CR greater than their respective averages (7.32 for BCS and 0.043 for CR). The resulted ML_{BCS} and ML_{CR} are those whose distributions are shown in Fig. 8. Comparison II adopts much stricter thresholds by respectively requiring BCS and CR to be greater than their averages plus two times of their standard deviations (29.40 for BCS and 0.14 for CR). By doing so, this study reduces the possible impact from the large volume of noises, and is able to see how MP_{BCS} and MP_{CR} differ from each other when different amount of ML_{BCS} and ML_{CR} are supplemented. Some relevant statistics are summarized in Table 2. Please note that MPA may actually derive multiple main paths from a patent network. It so happens that a single main path (i.e., MP_{DC}, MP_{BCS}, and MP_{CR}) is respectively produced from the patent networks PN_{DC}, PN_{BCS}, and PN_{CR}. Therefore, there is no need to determine an optimal one from multiple main paths.

As illustrated, in Comparisons I and II, by adding respectively ML_{BCS} and ML_{CR} that are only fractional in size to DCs, the resulted MP_{BCS} and MP_{CR} have significantly greater lengths (~130 patents in Comparison I and ~60 patents in Comparison II) than that of MP_{DC} (27 patents). Please note that, in Comparison II, even though ML_{CR} is more numerous than ML_{BCS} (17,889 vs 7750), the resulted MP_{CR} and MP_{BCS} are of comparable lengths (61 vs. 53). This is because that, once the MLs are supplemented to the DCs, the obtained networks PN_{CR} and PN_{BCS} are not so different in terms of their number of edges, which are 172,394(=154,505 + 17,889) and 162,255(=154,505 + 7750). However, as will be revealed later, MP_{BCS} and MP_{CR}, despite their comparable lengths, are actually rather different due to the different characteristics of BCS and CR.

The MP_{DC}, MP_{BCS}, and MP_{CR} from Comparisons I and II are respectively combined in Figs. 10 and 11 to facilitate comparison. Each node denotes a set of patents whose size is marked besides the node. Some nodes where the MPs intersect are further

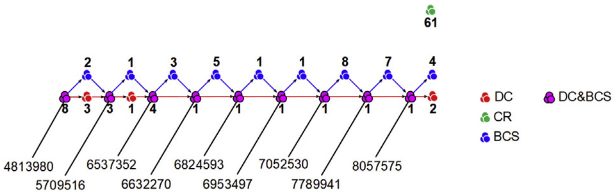

Fig. 11. Combined MP_{DC}, MP_{CR}, and MP_{BCS} from Comparison II. (For interpretation of the references to colour in the text, the reader is referred to the web version of this article.)

Table 3
DC and ML pairs along MP_{CR} of Comparison II.

3941871→4310440	4851106→4882038	5552129→5589147	8529663→8529664
4061724→4310440	4882038→4894213	5589147→5758489	8529664→8545602
4310440→4440871	4894213→4913888	5758489→6272848	8545602→8906138
4440871→4554143	4913888→4917876	6272848→6758036	8906138→8921637
4554143→4683217	4917876→4935216	6758036→6832473	8921637→9017457
4683217→4686092	4935216→4940570	6832473→6976354	9017457→9034078
4686092→4686093	4940570→4952384	6976354→7799314	9034078→9034079
4686093→4735806	4952384→4956164	7799314→7815873	9034079→9067168
4735806→4737353	4956164→4956165	7815873→7875402	9067168→9120049
4737353→4741892	4956165→5019263	7875402→7938886	9120049→9126138
4741892→4744970	5019263→5310714	7938886→7947120	9126138→9162175
4744970→4759919	5310714→5451387	7947120→7959720	9162175→9168485
4759919→4789535	5451387→5482692	7959720→8444750	9168485→9352269
4789535→4793984	5482692→5520895	8444750→8529662	9352269→9358493
4793984→4851106	5520895→5552129	8529662→8529663	9358493→9593778

marked with the largest patent number in the set and, as U.S. patents are serially numbered, they serve as a time frame. The nodes and arcs are color-coded as follows. MP_{DC}, MP_{CR}, and MP_{BCS} are in the three primary colors, red, green, and blue, respectively. But, for nodes and arcs where MP_{CR} and MP_{BCS} coincide, they are in cyan; those where MP_{DC} and MP_{CR} coincide are in yellow; and those where MP_{DC} and MP_{BCS} coincide are in magenta. Black nodes are where all three MPs intersect. Therefore, in Figs. 10 and 11, MP_{DC} is the series of red arcs, MP_{CR} green and cyan arcs, and MP_{BCS} blue and cyan arcs.

Let's look at Fig. 10 first. The nodes (patents) in Fig. 10 may be separated into three groups: those before 6,319,306, those from 6,319,306 to 8,057,575, and those after 8,057,575.

The first group of patents, arranged along the top of Fig. 10, includes those issued within the earliest part of the cases time frame. There are fewer DC pairs within this stage and, therefore, the supplemented ML_{BCS} and ML_{CR} lead to completely different MP_{BCS} and MP_{CR}.

CR is particularly strong within this early stage for aged and short-spanned BC pairs, as revealed in Fig. 7. Therefore, not only MP_{CR} in this stage is lengthier than MP_{BCS}, but patents of MP_{CR} (green nodes) actually "fill the gaps" between successive patents of MP_{BCS} (cyan nodes).

For the second group of patents in the middle of Fig. 10, there are more DC pairs due to the patent and reference expansions. Then, patents of MP_{BCS} and MP_{CR} (cyan and blue nodes) fill the gaps between successive patents of MP_{DC} (black nodes). CR's more uniform treatment to short- and long-spanned pairs make MP_{CR} more selective and, contrary to the first group, it is the patents from the MP_{BCS} (blue nodes) fill the gaps between successive patents of MP_{CR} (black and cyan nodes).

Patents in the third group are the latest ones in the time frame. Patents from MP_{CR} (green nodes) continue their gap filling role to patents of MP_{DC} (black and yellow nodes). BCS, however, is particular strong for young and short-spanned pairs in this late stage. MP_{BCS}, therefore, develops totally differently from MP_{DC} and $MP_{CR.}$

Comparing CR and BCS in Fig. 10, CR's favor towards aged patents makes ML_{CR} play a gap-filling role towards ML_{BCS} in the early part of the time frame whereas, in the later part, due to CR's less biased treatment towards short- and long-spanned BC pairs, it is the ML_{BCS} that functions as the gap-filler to ML_{CR}.

In Comparison II, due to the stricter threshold used, ML_{BCS} is significantly reduced, as shown in Table 2. Therefore, in Fig. 11, ML_{BCS} fails to develop differently in the early part of the time frame, but maintains its gap-filling role to MP_{DC}. Then, in the latest part of the time frame where BCS is particularly strong, MP_{BCS} also develops a different path, a phenomenon similar to what is shown in Fig. 10.

The most surprising result from Comparison II and Fig. 11 is that MP_{CR} is totally different from not just MP_{DC} but even MP_{BCS}. However, according to the arcs connecting the 61 MP_{CR} nodes listed in Table 3, 15 of them in grey background are DCs and 19 in dark background are also from ML_{BCS} (i.e., identified by BCS). Only those in white background are exclusively from ML_{CR} (i.e., identified by CR).

Fig. 11 and Table 3 jointly reveal the following: (1) even though a smaller ML_{CR} is supplemented, the structural connectivity of PN_{CR} becomes so different from that of the PN_{DC} and PN_{BCS} that the resulted MP_{CR} actually identifies a set of patents entirely distinct from those on MP_{DC} and MP_{BCS}; (2) CR's favor towards aged BC pairs leads to the inclusion in the MP_{CR} of quite a few earlier and smaller numbered patents linked by ML_{CR} (those in white background); and (3) the younger BC pairs in dark background, even though also included in ML_{BCS} but overlooked by MP_{BCS}, are now elevated into MP_{CR} by the more balanced consideration of CR towards both short- and long-spanned BC pairs.

5. Conclusion

This study notices a temporal pattern between bibliographically coupled patents, indicating that ages and spans of BC pairs may strongly affect their coupling strength measurement. Then, this study reasons that this temporal pattern is inevitable for fields revealing both patent and reference expansions. This study, therefore, proposes a new measure, combined relevance (CR), to reduce such impact.

CR is not ideal as observed in the previous section, and this study does not claim that CR is superior or that conventional BCS measures should be replaced by CR. When BC pairs are collected from patents issued within a time window, as their ages and spans are confined simultaneously in the same period of time, conventional measures are still viable means. But, to observe long-term knowledge dissemination or to trace comprehensive development trajectory, CR may be an alternative.

This study points out that patent and reference expansions are the two factors contributing to the temporal pattern. The patent expansion phenomenon should be applicable to non-U.S. patents as well. However, there is a lack of evidence that non-U.S. patents would also have expanding references. Nonetheless, CR's division of relatedness into factors respectively reflecting the relevance of shared references to the patents seems to be a reasonable design, and should still be applicable to non-U.S. patents as well.

This study may be improved in a number of ways. Firstly, CR itself has room for improvement. For one example, the concept of present value may be borrowed and the denominator $\left|REF_L\right|$ in $\frac{|REF_E \cap REF_L|}{|REF_L|}$ may be discounted to the date t_E to compensate the reference expansion. Secondly, we only observe the difference between the MPs of Figs. 10 and 11 so as to compare the effects of conventional BCS and CR but, for brevity's sake, we do not delve into the details of verifying that MP_{CR} is indeed a more accurate trajectory for a field. This may be conducted in the future.

The study considers that its contribution to the bibliometric community is to point out that conventional BCS measures may be biased against both aged and long-spanned BC pairs, and an improved measure as simple as CR may achieve significant improvement and utterly different analytic result. CR offers a hint that an ideal measure to resolve both age and span issues of BC pairs is possible and probably does not require a sophisticated algorithm.

Even though patents are the main subject of this work, the findings may be applicable to research publications of a field as well, if they reveal similar document and reference expansions. Investigating the effect of CR on BC pairs of research articles would be another interesting endeavor in the future.

Author contributions

Chung-Huei Kuan: Conceived and designed the analysis, Collected the data, Contributed data or analysis tools, Performed the analysis, Wrote the paper.

Dar-Zen Chen: Conceived and designed the analysis, Collected the data, Contributed data or analysis tools, Performed the analysis, Wrote the paper.

Mu-Hsuan Huang: Conceived and designed the analysis, Contributed data or analysis tools, Wrote the paper.

Acknowledgments

The authors express their appreciation to the anonymous reviewers for their careful reading of the manuscript and their helpful comments and suggestions.

This work was financially supported by the Center for Research in Econometric Theory and Applications (Grant no. 108L900204) from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan, and by the Ministry of Science and Technology (MOST), Taiwan, under Grant Nos. MOST 108-3017-F-002-003- and MOST 108-2221-E-011-053-.

References

Angue, K., Ayerbe, C., & Mitkova, L. (2014). A method using two dimensions of the patent classification for measuring the technological proximity: An application in identifying a potential R&D partner in biotechnology. *The Journal of Technology Transfer*, *39*(5), 716–747.

Arts, S., Cassiman, B., & Gomez, J. C. (2018). Text matching to measure patent similarity. *Strategic Management Journal*, *39*(1), 62–84.

Batagelj, V. (2003). *Efficient algorithms for citation network analysis* arXiv preprint cs/0309023.

Bicknell, G. M. (2008). To disclose or not to disclose: Duty of candor obligations of the United States and foreign patent offices. *Chicago-Kent Law Review*, *83*, 425.

Calero-Medina, C., & Noyons, E. C. M. (2008). Combining mapping and citation network analysis for a better understanding of the scientific development: The case of the absorptive capacity field. *Journal of Informetrics*, *2*(4), 272–279.

Chang, S. B. (2012). Using patent analysis to establish technological position: Two different strategic approaches. *Technological Forecasting and Social Change*, *79*(1), 3–15.

Chen, D. Z., Huang, M. H., Hsieh, H. C., & Lin, C. P. (2011). Identifying missing relevant patent citation links by using bibliographic coupling in LED illuminating technology. *Journal of Informetrics*, *5*(3), 400–412.

Chu, Y., Tian, X., & Wang, W. (2018). Corporate innovation along the supply chain. *Management Science*, *65*(6), 2445–2466.

De Nooy, W., Mrvar, A., & Batagelj, V. (2011). . *Exploratory social network analysis with Pajek* (Vol. 27) Cambridge University Press.

Gray, M. L., Hoffman, J. S., Hreha, D. C., Fauth, D. J., Hedges, S. W., Champagne, K. J., et al. (2009). Parametric study of solid amine sorbents for the capture of carbon dioxide. *Energy & Fuels*, *23*(10), 4840–4844.
Giudice, P. L., Musarella, L., Sofo, G., & Ursino, D. (2019). An approach to extracting complex knowledge patterns among concepts belonging to structured, semi-structured and unstructured sources in a data lake. *Information Sciences*, *478*, 606–626.
Glänzel, W., & Czerwon, H. J. (1996). A new methodological approach to bibliographic coupling and its application to the national, regional and institutional level. *Scientometrics*, *37*(2), 195–221.
Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). *The NBER patent citation data file: Lessons, insights and methodological tools (No. w8498)*. National Bureau of Economic Research.
Hummon, N. P., & Doreian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks*, *11*(1), 39–63.
Jaccard, P. (1901). Etude comparative de la distribution florale dans une portion des Alpes et du Jura. *Bulletin de la Société vaudoise des Sciences Naturelles*, *37*, 547–579.
Jaffe, A. B. (1986). *Technological opportunity and spillovers of R&D: Evidence from firms' patents, profits and market value*.
Jarneving, B. (2007a). Bibliographic coupling and its application to research-front and other core documents. *Journal of Informetrics*, *1*(4), 287–307.
Jarneving, B. (2007b). Complete graphs and bibliographic coupling: A test of the applicability of bibliographic coupling for the identification of cognitive cores on the field level. *Journal of Informetrics*, *1*(4), 338–356.
Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *Journal of the Association for Information Science and Technology*, *14*(1), 10–25.
Kuan, C. H., Chiu, C. W., Liu, J. S., Huang, M. H., & Chen, D. Z. (2018). Patent main path analysis with arc weights adjusted by classification similarity. In *23rd international conference on science and technology indicators (STI 2018)*. Leiden, The Netherlands: Centre for Science and Technology Studies (CWTS). September 12-14, 2018.
Kuan, C. H., Huang, M. H., & Chen, D. Z. (2018). Missing links: Timing characteristics and their implications for capturing contemporaneous technological developments. *Journal of Informetrics*, *12*(1), 259–270.
Kuusi, O., & Meyer, M. (2007). Anticipating technological breakthroughs: Using bibliographic coupling to explore the nanotubes paradigm. *Scientometrics*, *70*(3), 759–777.
Leydesdorff, L., Kushnir, D., & Rafols, I. (2014). Interactive overlay maps for US patent (USPTO) data based on International Patent Classification (IPC). *Scientometrics*, *98*(3), 1583–1599.
Liu, J. S., & Lu, L. Y. (2012). An integrated approach for main path analysis: Development of the Hirsch index as an example. *Journal of the American Society for Information Science and Technology*, *63*(3), 528–542.
Lo, S. C. (2007). Patent coupling analysis of primary organizations in genetic engineering research. *Scientometrics*, *74*(1), 143–151.
Lu, L. Y., Hsieh, C. H., & Liu, J. S. (2016). Development trajectory and research themes of foresight. *Technological Forecasting and Social Change*, *112*, 347–356.
Makri, M., Hitt, M. A., & Lane, P. J. (2010). Complementary technologies, knowledge relatedness, and invention outcomes in high technology mergers and acquisitions. *Strategic Management Journal*, *31*(6), 602–628.
Martinelli, A. (2012). An emerging paradigm or just another trajectory? Understanding the nature of technological changes using engineering heuristics in the telecommunications switching industry. *Research Policy*, *41*(2), 414–429.
Moehrle, M. G., & Gerken, J. M. (2012). Measuring textual patent similarity on the basis of combined concepts: Design decisions and their consequences. *Scientometrics*, *91*(3), 805–826.
Nakamura, H., Suzuki, S., Sakata, I., & Kajikawa, Y. (2015). Knowledge combination modeling: The measurement of knowledge similarity between different technological domains. *Technological Forecasting and Social Change*, *94*, 187–201.
Niemann, H., Moehrle, M. G., & Frischkorn, J. (2017). Use of a new patent text-mining and visualization method for identifying patenting patterns over time: Concept, method and test application. *Technological Forecasting and Social Change*, *115*, 210–220.
Ortiz-de-Urbina-Criado, M., Nájera-Sánchez, J. J., & Mora-Valentín, E. M. (2018). A research agenda on open innovation and entrepreneurship: A Co-word analysis. *Administrative Sciences*, *8*(3), 34.
Park, H., & Magee, C. L. (2017). Tracing technological development trajectories: A genetic knowledge persistence-based main path approach. *PLoS One*, *12*(1), e0170895.
Park, I., Jeong, Y., Yoon, B., & Mortara, L. (2015). Exploring potential R&D collaboration partners through patent analysis based on bibliographic coupling and latent semantic analysis. *Technology Analysis & Strategic Management*, *27*(7), 759–781.
Petruzzelli, A. (2011). The impact of technological relatedness, prior ties, and geographical distance on university-industry collaborations: A joint-patent analysis. *Technovation*, *31*, 309–319.
Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for information Science*, *24*(4), 265–269.
Swanson, D. (1971). Some unexplained aspects of the Cranfield tests of indexing performance factors. *The Library Quarterly*, *41*(3), 223–228.
Trajtenberg, M. (1990). A penny for your quotes: Patent citations and the value of innovations. *The Rand Journal of Economics*, 172–187.
Von Wartburg, I., Teichert, T., & Rost, K. (2005). Inventive progress measured by multi-stage patent citation analysis. *Research Policy*, *34*(10), 1591–1607.
Wang, A. P., Hou, C. E., & Hung, S. W. (2018). Exploration of the evolution of nanotechnology from a patent co-classification perspective. *Nanotechnology Reviews*, *7*(3), 233–245.
World Intellectual Property Organization. (2009). *Patent-based technology analysis report—Alternative energy technology* Retrieved from. https://www.wipo.int/edocs/plrdocs/en/landscape_alternative_energy.pdf
Yeh, H. Y., Sung, Y. S., Yang, H. W., Tsai, W. C., & Chen, D. Z. (2013). The bibliographic coupling approach to filter the cited and uncited patent citations: A case of electric vehicle technology. *Scientometrics*, *94*(1), 75–93.
Yoon, B., & Magee, C. L. (2018). Exploring technology opportunities by visualizing patent information based on generative topographic mapping and link prediction. *Technological Forecasting and Social Change*, *132*, 105–117.
Zhang, L., Huang, Y., & Chen, L. (2018). Patent citation inflation: The phenomena, measurement methods, and proposed solutions based on relative citation indicators. In *23rd international conference on science and technology indicators (STI 2018)*. Leiden, The Netherlands: Centre for Science and Technology Studies (CWTS). September 12–14, 2018.

Chung-Huei Kuan is an assistant professor of the Graduate Institute of Patent at National Taiwan University of Science and Technology, Taipei, Taiwan. His research interests include patent bibliometrics, patent information mining and analysis, and practices in patent prosecution, patent specification drafting, and patent/technology transfer and licensing.

Dar-Zen Chen is a professor of the Department of Mechanical Engineering and Institute of Industrial Engineering at National Taiwan University, Taipei, Taiwan. His research interests include intellectual property management, patentometrics, competitive analysis, robotics, automation, kinematics, and mechanism design. He also leads the Intellectual Property Analysis & Innovative Design Laboratory (IAID).

Mu-Hsuan Huang is a chair professor of the Department of Library and Information Science at National Taiwan University, Taipei, Taiwan. Her early research focused on information retrieval and information behavior, and turned to bibliometrics, science and technology policy, intellectual property, and patent information for late years. She is also the project investigator of Performance Ranking of Scientific Papers for World Universities (NTU Ranking).