Scientometrics (2013) 94:75–93
DOI 10.1007/s11192-012-0820-8

The bibliographic coupling approach to filter the cited and uncited patent citations: a case of electric vehicle technology

Hsi-Yin Yeh · Yi-Shan Sung · Hsiao-Wen Yang · Wan-Chu Tsai · Dar-Zen Chen

Received: 5 January 2012 / Published online: 1 August 2012

Abstract Because some cited references are not relevant to the citing patent and not all the relevant references are cited, the study attempts to use the bibliographic coupling (BC) approach to filter the irrelevant patent citations and supplement the relevant uncited patent citations to construct a patent citation network (PCN). The study selected the field of electric vehicle technology to explore the phenomenon and examined the characteristics of PCNs in terms of the average BC strength and the average citation time lag. Four PCNs were constructed in this study. The aggregated PCN (APCN) excluded the irrelevant patent citations and added the relevant uncited patent citations, which has brought out significant improvement. The APCN became more concentrated and the information which reserved in the APCN was the most current. Additionally, some invisible technology clusters and relationships were also manifested in the APCN.

Keywords Patent citation analysis · Bibliographic coupling · Citation time lag · Electric vehicle technology

Introduction

Patents are one of the important indicators to evaluate the performance of industry research and development (R&D) (Griliches 1990). The World Intellectual Property Organization (WIPO) indicated that 80 % of innovative technologies do not appear in academic publications; however, 90–95 % of inventions can be found in the patents (Liu and Shyu 1997; Chen and Chen 2007). Patents can be viewed as a representative proxy for new technology

H.-Y. Yeh · H.-W. Yang · W.-C. Tsai
Department of Library and Information Science, National Taiwan University, Taipei, Taiwan

Y.-S. Sung
Institute of Industrial Engineering, National Taiwan University, Taipei, Taiwan

D.-Z. Chen (✉)
Department of Mechanical Engineering and Institute of Industrial Engineering, National Taiwan University, Taipei, Taiwan
e-mail: dzchen@ntu.edu.tw

due to a large amount of technological information (Ernst 2003; Lee et al. 2009). In addition, patents can also reflect the industrial changes and trends. If companies get the most information from patents, the time and expenditure on R&D can be significantly reduced (Wartburg et al. 2005).

The general method for early patent data analysis was to count patents and to compare how many patents have been assigned to different entities, e.g. firms, nations and technology fields. However, patent citation analysis was another method of analyzing the development of patents (Yoon and Park 2004; Lee et al. 2009). Patent citation analysis was based on the examination of citation links among different patents, and the citation links between patents and scientific literatures (Narin 1994). It was also regarded as providing crucial information for analyzing technology innovation (Wartburg et al. 2005). Furthermore, patent citation analysis could be used to understand technology diffusion. With patent citation networks (PCN), the relationships among technological patents could be more soundly laid out (Trajtenberg et al. 1997).

Several studies indicated that patent citation analysis was an effective approach to evaluate the importance of patents (Carpenter and Narin 1983; Albert et al. 1991). Lanjouw and Schankeman (2004) compared indicators such as the number of claims, the size of patent family, the number of cited patents and the number of citing patents, and found that the number of citing patents was a better way to explore the importance of patents. Atallah and Rodriguez (2006) integrated direct and indirect patent citations with the weight mechanism to investigate the importance of the patent. Moreover, other studies have conducted patent citation analysis for exploring the relationships among technologies (Ellis et al. 1978; Verspagen 2000; Meng et al. 2000; Lo 2010). Ellis et al. (1978) studied US patents in five domains, evaluated the correlation of citation and plotted a patent citation map to show the highly influential technologies with the citation frequency. Meng et al. (2000) analyzed the patent citations of US patents through the CHI database, and collected US patents issued to Taiwanese assignees between 1980 and 1996 to analyze innovation indicators and industrial technological trends in Taiwan. These prior studies demonstrated that patent citation analysis was a proper way to understand the interrelations of technologies. The structure of the PCN also could be viewed as a good pattern for understanding technological changes and trends (Wartburg et al. 2005; Fontanaa et al. 2009).

Case and Higgins (2000) argued that patent citation presented a review of prior art. The use of patent citation data may be advantageous in terms of validity and reliability. To receive a patent, an inventor must carefully explore the state of the prior art. Foremost, an examiner has to evaluate both the invention and the scope of the claims applied for by performing an in-depth review of existing patents. Meyer (2000, p. 112) concluded that "one needs to have a thorough understanding of patenting practices in order to interpret patent citation data properly." Thus, citing references properly was a critical issue to be concerned about (Michel and Bettels 2001; Breschi and Lissoni 2005).

The relevance of patent citations was mainly based on the assumption that citations could clearly reflect the interrelatedness of patents. However, how a patent cites references is dictated by various motivations (Michel and Bettels 2001; Hall et al. 2005). Michel and Bettels (2001) indicated that a good search report based on the rule of European Patent Organization (EPO) should contain all the technical relevant information within a minimum number of citations. Hall et al. (2005) also noted that the purpose of USPTO patent citations was to identify all the prior developments of technology which might be similar to the demanded invention and might reveal the state of the technology. However, no matter how the patent citation rules are identified, some cited patents still lack relevance to the citing patent. For example, Wartburg et al. (2005) indicated that the cited references of some patents might only focus on one specific technical aspect of a patent but which is not related to the main concepts of the patent itself. This phenomenon showed that there was still much noise in the patent citations. Furthermore, the subjective judgments of patent citation depended on inventors and examiners may introduce some bias (Michel and Bettels 2001). Accordingly, we suspect that irrelevant patent citations exist. Thus, this study attempts to set a threshold to filter out irrelevant patent citations and construct an extracted PCN to clarify these problems.

Prior studies also found that not all relevant references were cited (Smith 1981; Meyer 2000; Chen et al. 2011). Meyer (2000) indicated that both patent and paper citations had much similar characteristics, and found that many references which should have been cited were missed. Chen et al. (2011) also contended that some missing relevant patent citations may exist. They added the missing relevant patent links into the PCN and obtained a more comprehensive view of the relationships between patents. According to the arguments posited by prior research, we can infer that some references which should be cited by the patents may indeed be neglected. Thus, the study also attempts to set a threshold to filter relevant uncited patent citations and construct a supplemented PCN to understand this situation.

Bibliographic coupling (BC), proposed by Kessler (1963), was an appropriate method to evaluate the relevance of patent citation. The BC relationship of two citing patents is constructed by the third cited patent. The more cited patents the two citing patents have, the higher the relevance they have (Kessler 1963). Although Chen et al. (2011) used the BC approach to evaluate the missing links of patent citations and added those missing links to the construction of the PCN, the study further examined the relevance of cited patent citations and supplemented the relevant uncited patent citations to construct an aggregated PCN. Compared to the study of Chen et al. (2011), this study provided a different viewpoint for the discussion of patent citation issues.

This study selected the field of electric vehicle technology to explore this phenomenon. Because of the changing regulatory environment, a shift in the technological environment of automotive manufacturers has emerged. Thus, this study attempts to use PCN to understand the technological pattern in electric vehicle technology. First, the original PCN of electric vehicle technology was plotted. Then, irrelevant cited patent citations were removed and the extracted PCN constructed. Next, relevant uncited patent citations were added and the supplemented PCN constructed. Finally, the extracted PCN and supplemented PCN were integrated into the aggregated PCN. The characteristics of PCN in terms of the average BC strength and the average citation time lag (CTL) were examined. The greater the average BC strength of the PCN was, the more relevant were the patent citations of the PCN; the shorter the average CTL of the PCN was, the more current PCN information was reserved. Based on these characteristics, we can obtain a better understanding about the improvement of these PCNs.

Research methodology

The field of electric vehicle technology

At the beginning of the 21st century, the changing regulatory environment has created a shift in the technological environment of automotive manufacturers from a proven sustaining technology to an unproven disruptive technology, namely electric vehicle technology. In addition, due to concerns about its environmental impact, petroleum-based transportation has also gradually shifted toward electric vehicle technology. In response to these crucial changes, this study attempts to construct a PCN to understand the technological pattern in the electric vehicle technology.

The electric vehicle technology was chosen for the case study. According to the announcement of USPTO posted on its official website, most patents about electric vehicle technology were located in three categories of Current United States Patent Classification (USPC), including 180/65.1 (motor vehicles/electric), 180/65.21 (motor vehicles/hybrid vehicle) and 701/22 (data processing: vehicles, navigation, and relative location/electric vehicle). The study retrieved the patents of these three categories from 01/01/1976 to 12/31/2009 and obtained 1,493 issued patents to analyze.

Evaluation of the irrelevant patent citations

Patent citation refers to references cited from previous patents and has helped both examiners and applicants judge the degree of patent novelty. Bichteler and Parsons (1974) argued that if two patents cite the same references, they may tend to deal with the same subjects. However, evaluating the relevance of patent citations is an important issue. This study argues that the relevance of each patent citation pair could be measured well by using the BC approach.

Given a PCN $P = \{p_1, p_2, \ldots, p_n\}$, for any p_i and p_j where $i! = j$. If P_i is cited by P_j, the pair (P_i, P_j) is called 'BC pairs with citation'. If P_i is not cited by P_j, the pair is called 'BC pair without citation'. P_{F} is defined as a set of BC pairs with citation. The more cited references they share, the more similar their technical backgrounds. P'_{F} is defined as the set of BC pair without citation. BC pair (i, j) is used to express that two patents cited the same patent, and the BC strength refers the number of cited patents. The BC strength of patent i and patent j is denoted as b_{ij}. Huang et al. (2003) noted that the greater the BC strength of the patents, the more relevant the patents may be. The study proposed that the average BC strength of citation pairs would be useful to measure the relevance of PCN.

To further understand how BC strength reflects citation relations, this study divided BC pairs into two groups: "BC pairs with citation" and "BC pairs without citation". If BC strength could reflect the citation relations, the average BC strength with citation should be greater than the average BC strength without citation. The validity of BC relations was denoted as β_{F}. The greater value of β_{F} was, the better BC citation relationship could represent.

$$\beta_{\mathrm{F}} = \frac{\sum_{(i,j)\in P_{\mathrm{F}}} b_{ij}}{|P_{\mathrm{B}} \cap P_{\mathrm{F}}|} - \frac{\sum_{(i,j)\in P'_{\mathrm{F}}} b_{ij}}{|P_{\mathrm{B}} \cap P'_{\mathrm{F}}|} \quad (1)$$

The BC strength of each citation pair may be calculated to present the degree of relevance. However, some studies indicated that not all citation pairs with BC strength greater than zero can be considered relevant citations (Swanson 1971; Jarneving 2007). Swanson (1971) calculated the BC strength of all document citation pairs and identified the relevant citation as the threshold of BC strength which should be greater than the value of seven. This study also argues that it is necessary to set a threshold of BC strength to determine the relevant degree of patent citation.

This study set the average BC strength of BC pairs without citation as the threshold for the evaluation of irrelevant citation pairs. Because the value of BC strength within some citation pairs may be lower than that within BC pairs without citation, the average BC strength of BC pairs without citation could be considered the critical value for measuring the degree of relevance to citation pairs. Thus, if the value of BC strength within citation pairs is not higher than the average BC strength of BC pairs without citation, the citation pairs should be viewed as irrelevant. BM_F was denoted as the threshold value of BC strength.

$$BM_F = \left\lceil \frac{\sum_{(i,j) \in P'_F} b_{ij}}{\left|P_b \cap P'_F\right|} \right\rceil \quad (2)$$

The ratio of extracted citation pairs was further calculated. Extracted citation pairs were denoted as P_E and referred the number of patent citations which the value of BC strength was higher than BM_F. Excluded citation pairs were denoted as P_{EX} and referred the number of patent citations which the value of BC strength was lower than BM_F. α_F was denoted as the ratio of extracted citation pairs.

$$\alpha_F = \frac{|P_E|}{|P_{EX}|} \quad (3)$$

Evaluation of relevant uncited patent citations

Smith (1981) indicated that many references which should be cited may be missed. Due to the similarities between paper citation and patent citation, this study examined relevant uncited patent citations, arguing that the more references both patents cite, the more analogous their technical background, even though patent j does not cite patent i.

The relevant degree of BC pairs without citation was still measured by BC strength. The study argued that BC strength in BC pairs without citation should be higher than the minimum value of BC strength in citation pairs. The average BC strength of citation pairs was set as the threshold of relevant uncited citation pairs. When the value of BC strength in BC pairs without citation passed the threshold, these BC pairs without citation were referred to as supplementary citation pairs. S_F was denoted as the threshold of relevant uncited citation pairs.

$$S_F = \left\lceil \frac{\sum_{(i,j) \in P_F} b_{ij}}{|P_b \cap P_F|} \right\rceil \quad (4)$$

The ratio of supplementary citation pairs to original patent citation pairs was calculated. Supplementary citation pairs were denoted as P_S and referred to the number of supplementary citation pairs with BC strength greater than S_F. P_F was defined as the number of original citation pairs. γ_F was denoted as the supplementary ratio of citation pairs which represented the degree of improvement after supplementary patent citations were added.

$$\gamma_F = \frac{|P_S|}{|P_F|} \quad (5)$$

Citation time lag

Citation behavior is related to the time gap between application or grant date of citing patent and that of the cited patents (Hall et al. 2001). Nagaoka (2004) proposed that patents are most valuable within 5 years of their issue dates. Jaffe and Trajtenberg (1999) suggested that the better inventors know current patents, the more advantages they would have in mastering changes. Thus, the time gap between citing patent and cited patents was an important indicator for evaluating the degree of currency in patent citations. CTL refers to the time that patents take to be cited. The shorter the average CTL of PCN, the more current the information in the PCN. Moreover, CTL could also be used to understand phenomena with added relevant uncited patent citations. A shorter citation time gap between two relevant patents might cause the possibility of links without citations between them. If relevant patent *i* was issued later than the application date of patent *j*, patent *i* was highly likely to be missed by patent *j*.

The CTL between paten *i* and patent *j*, denoted as CTL_{ij}, is defined as the time gap between the issued date of patent *i* and the application date of patent *j*. The formula is $CTL_{ij} = APD_j - ISD_i$ where APD_j is referred to as the application date of patent *j*, and ISD_i is referred to as the issue date of patent *i*. This study divided citation pairs into three zones based on their CTL. In the first zone, the CTL with a value below zero indicated that citation time gap was the smallest and the information was the most current. In the second zone, CTL was higher than zero but below the average CTL of PCN. In the third zone, CTL was higher than the average CTL of PCN, indicating that the citation time gap was the largest and the information was the least current. Three CTL zones were used to discuss the relationships between BC strength and CTL.

Results and discussion

Original patent citation network

From 1976 to 2009, the retrieved citing and cited relationship among the issued patents in the electric vehicle technology field is as follows. There are a total of 1,493 issued patents, which include 1,015 (81 %) patents with citing, 770 (62 %) patents that have been cited, and 245 (16.4 %) patents with non-citing and non-cited. Within these issued patents the total citing number is 3,275, while the average citing number is 4.25 and the average cited number is 3.23. Furthermore, excluding the non-citing and non-cited patents, there are 1,248 (84 %) patents connected by the citation relationship, and the average links between each pair of patents is approximately 2.6, as shown in Table 1.

The social network analysis software UCINET, developed by Borgatti and Cross (2003), was used to visualize the PCN. In the UCINET the clustering method proposed by Johnson (1967) is based on the distances (similarities) between these clusters. In this research, the average-link method is chosen for computing the distance (similarities), which is considered to be the average distance from all members of one cluster to all members of another cluster. UCINET provided an automatic analyzed technique to assign

Table 1 Number of patents in the OPCN

	Number	%
Cited patent	770	62
Citing patent	1,015	81
Cited and citing patent (remained)	1,248	100
Total citation pairs	3,275	
Average cited count	4.25	
Average citing count	3.23	

individual actors to unique groups. According to the results of UCINET provided, the proper clusters were identified in the networks.

Figure 1 shows the results of the original PCN (OPCN), where the dots represent the patents and the arrow represents the citations from the cited towards the citing patents. In Fig. 1 UCINET obtained three clusters in the OPCN and circles were used to mark each cluster. By referring to their contents, these clusters were roughly categorized by three technological characteristics: electric automobile, electric wheelchair and mobility scooter, and children vehicle. The cluster of electric automobile had the largest number of patents, while children vehicle cluster had the least. Most citations belonged to one cluster. Although a few citations were intermediary between two clusters, which were marked as disordered citations, the results showed that these citations were probably not crucial in this field.

The degree of relevance in the original patent citation network

There are 9,443 distinct BC pairs with 13, 788 units of BC. Over 92 % of the BC pairs have no existing citation. The BC strength of 690 citation pairs was greater than the value of one, which accounted for 21.07 % of original citations (see Table 2). The BC strength of 25.8 % of citation pairs was greater than three. The BC strength of 85.54 % of non-citation pairs was greater than one, as shown in Table 3. These results indicate that many without citation pairs ought not to be ignored. In addition, the average BC strength of citation pairs (2.33) was greater than that of without citation pairs (1.39). The citation validity (0.94) was sufficient.

The average CTL of OPCN was 213 weeks. Figure 2 presented the CTL and BC strength of the OPCN. It showed that citation pairs with strong BC strength mainly

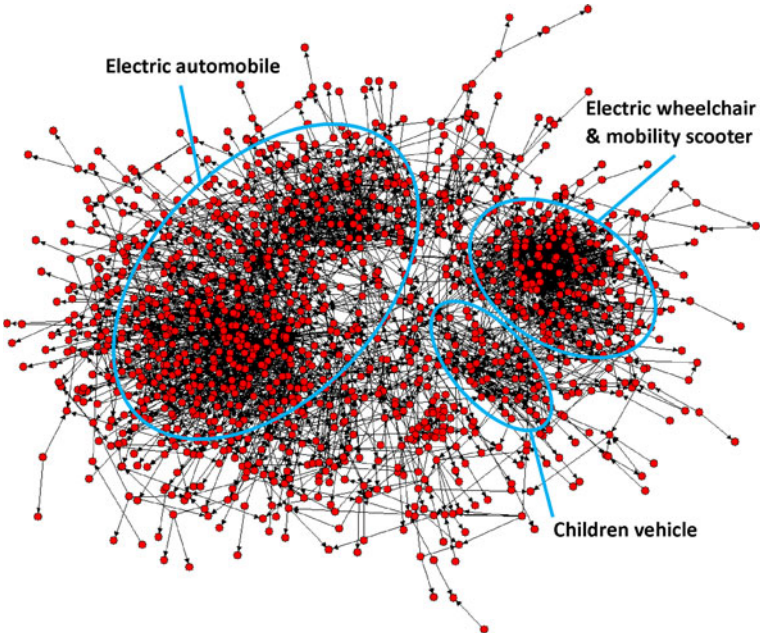

Fig. 1 The OPCN of electric vehicle technology

Table 2 Citation states of BC pairs in the OPCN

	BC pairs	
	With citation	Without citation
Number	690	8,753
%	7.31	92.69
Total	9,443	

Table 3 BC strength of citation pairs in the OPCN

BC strength	BC pairs with citation ($P_B \cap P_F$)		BC pairs without citation ($P_B \cap P'_F$)	
	Number	%	Number	%
1	392	56.81	7,487	85.54
2	120	17.39	712	8.13
≥3	178	25.80	554	6.33
Total	690		8,753	
Average BC strength	2.33		1.39	
β_F	0.94			

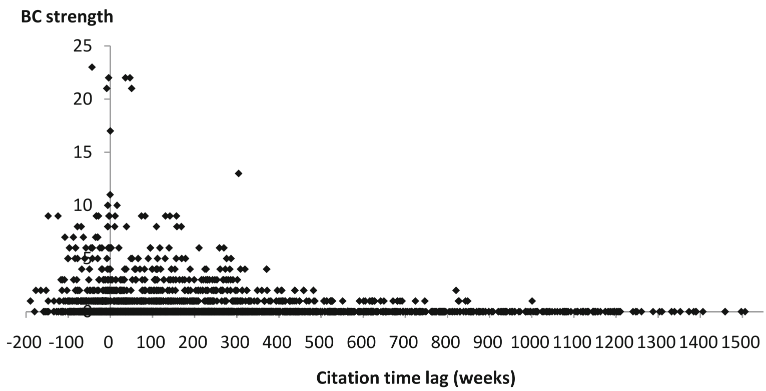

Fig. 2 Relationships between CTL and BC strength in the OPCN

concentrated on CTLs below 213 weeks. The results showed that the citation pairs with higher BC strength had shorter CTLs. This study further divided the original patent citations into three zones based on their CTL, including CTL below zero, CTL greater than zero and less than 213 weeks, and CTL greater than 213 weeks; the results are shown in Table 4. There were 665 (20 %) citation pairs in the first zone, and the average BC strength was the greatest (0.90). The average BC strength in the third zone was the lowest (0.26). This indicates that a negative relationship between BC strength and CTL occurred.

Extracted patent citation network

According to formula (2) and the results of Table 3, the BC threshold of extracted citation pairs was set to 2. Thus, the citation pairs with BC strength greater than two were marked

Table 4 Citation and BC strength of three CTL zones in the OPCN

Type of citation	Total			CTL ≤ 0		0 < CTL < 213 weeks		CTL > 213 weeks	
	Patent	Citation	Avg BC	Citation	Avg BC	Citation	Avg BC	Citation	Avg BC
Original citation	1,248	3,275	0.49	665	0.90	1,403	0.50	1,207	0.26

Table 5 Number of patents in the EPCN

Basic data	Excluded citation		Extracted citation	
	Number	% to PCN	Number	% to PCN
Cited patents	741	96	97	12
Citing patents	1,015	100	109	11
Cited or citing patents	1,248	100	166	13
Total citation pairs	2,977		298	
A_{BF}	9.10 %			

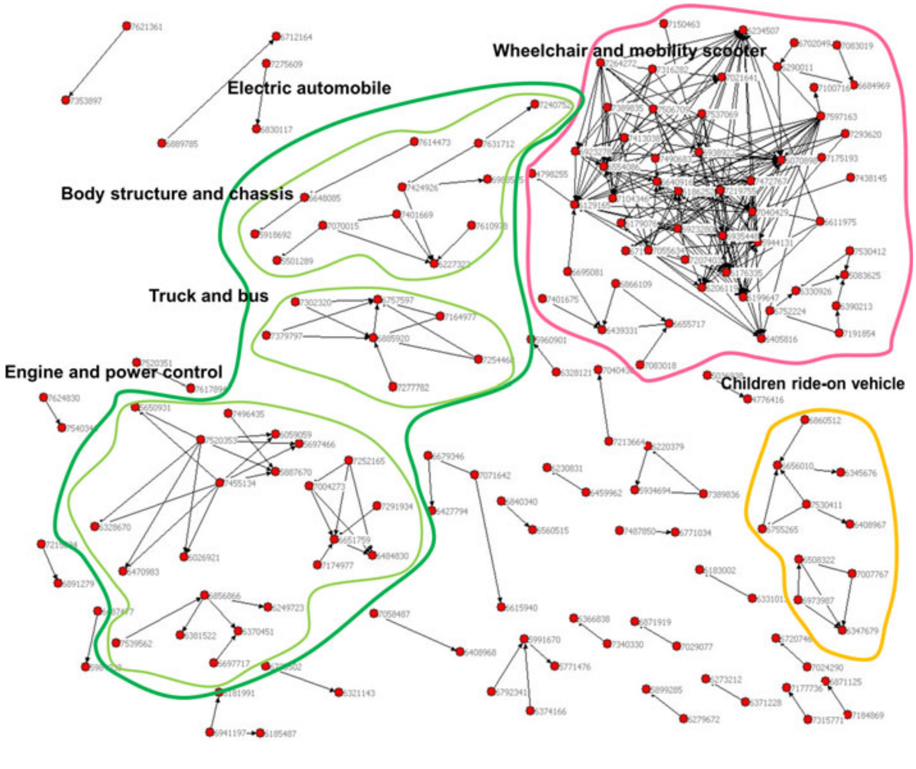

Fig. 3 The EPCN of electric vehicle technology

as extracted citation pairs, and the others were marked as excluded citation pairs. Only 298 (9 %) citation pairs were extracted and 2,977 (91 %) citation pairs were excluded, with results shown in Table 5. 1,248 patents were involved in these 2,977 excluded citations,

Table 6 Citation and BC strength of three CTL zones in the EPCN

Type of citation	Total			CTL ≤ 0		0 < CTL < 213 weeks		CTL > 213 weeks	
	Patent	Citation	Avg BC	Citation	Avg BC	Citation	Avg BC	Citation	Avg BC
Extracted citation	166	298 (9 %)	4.08	95 (14 %)	5.12	13 (10 %)	3.95	67 (6 %)	2.89
Children vehicle	10 (6 %)	11	3	4	2.25	6	3.67	1	2
Wheelchair and mobility scooter	55 (33 %)	203 (68 %)	4.53	55	6.13	99	4.33	49	3.08
Electric automobile	41 (25 %)	49	3.27	17	4.35	16	3.00	16	2.37
Structure and chassis	12 (36 %)	11	2.27	4	2.5	4	2.25	3	2
Engine and power control	22 (13 %)	28	3.29	6	5	9	3.33	13	2.46
Truck and bus	7 (5 %)	10	4.3	7	4.86	3	3	0	0
Others	60 (36 %)	35	2.94	19	3.47	15	2.53	1	3

Table 7 Number of patents in the SPCN

Basic data	Supplementary citation		Citation in SPCN	
	Number	% to PCN	Number	% to PCN
Cited patent	124	16	824	107
Citing patent	124	12	1,015	100
Cited or citing patent	160	13	1,248	100
Total citation pair	554		3,829	
γ_F	16.92 %			

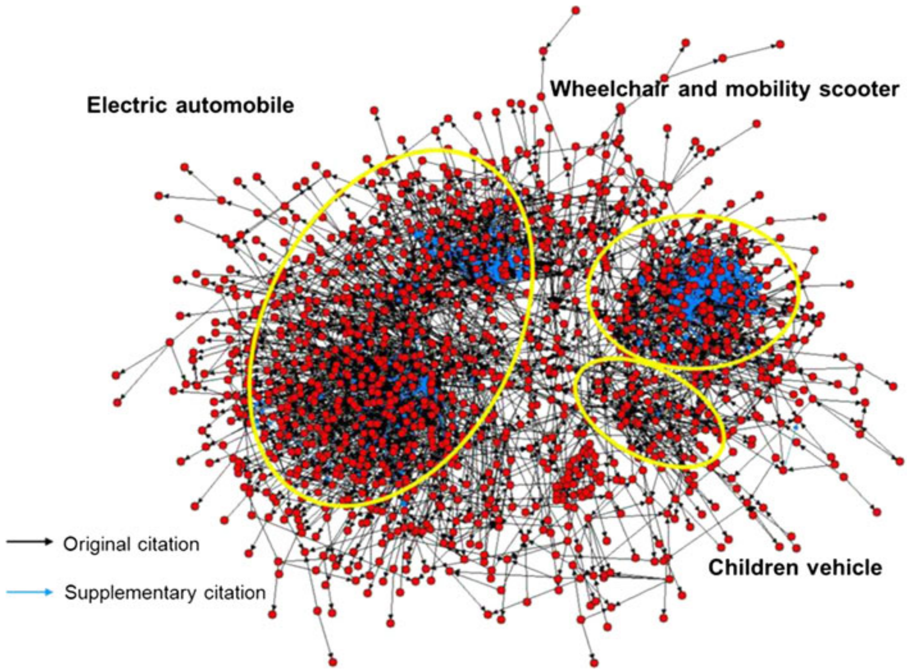

Fig. 4 The SPCN of electric vehicle technology

including 1,015 citing patents and 741 cited patents. In the 298 extracted citation pairs, 166 patents were involved, including 97 citing patents and 109 cited patents. The ratio of extracted citation pairs was 9.10 %.

Extracted PCN (EPCN) was constructed by 298 extracted patent citations as shown in Fig. 3. No patent citation link was intermediary. EPCN were constructed by five clusters. The patents in these clusters could be categorized into three main technological fields, such as electric automobile, wheelchair and mobility scooter, and children ride-on vehicle, which were the same as the OPCN. To further analyze the electric automobile technology cluster, it could be divided into three sub-fields, such as body structure and chassis, engine and power control, and electric truck and bus. These results showed that the technology of “electric automobile” was multiplex and the invisible technology clusters were revealed.

The number of citation pairs and their corresponding average BC strength of these five clusters (Table 6) were analyzed. The results showed that the number of patents in the

Table 8 Citation and BC strength of three CTL zones in the SPCN

Type of citation	Total			CTL $\leq$ 0		0 < CTL < 213 weeks		CTL > 213 weeks	
	Patent	Citation	Avg BC	Citation	Avg BC	Citation	Avg BC	Citation	Avg BC
Supplementary citation	238	554 (17 %)	5.90	400 (60 %)	6.42	135 (10 %)	4.71	19 (2 %)	3.21
Children vehicle	4 (3 %)	2	3.5	2	3.5	0	0	0	0
Wheelchair and mobility scooter	56 (35 %)	328 (59 %)	6.04	214	6.70	100	5.02	14	3.14
Electric automobile	98 (61 %)	223	5.73	183	6.16	35	3.83	5	3.4
Others	2 (1 %)	1	3	1	3	0	0	0	0

cluster of wheelchair and mobility scooter was the greatest. This cluster possessed 68 % of citation pairs and the highest (4.53) average BC strength. It implied that many patents cited the patents in this cluster. Moreover, the results possibly revealed that the technology of wheelchair and mobility scooter was an emerging technology according to the vigorous citation activities (4.53).

The CTL of extracted citation pairs were calculated. The average CTL of EPCN was 110 weeks which reduced 121 weeks from the OPCN. It implied that patents which reserved in the EPCN were the relative current information. The incidence of citation pairs in the first zone was the highest (14 %), and the average BC strength was 5.12. The incidence of the extracted BC citation pairs in the third zone dropped to 6 % and average BC strength reduced to 2.89. The results showed that a negative relationship still existed between BC strength and CTL in the EPCN.

Supplemented patent citation network

According to formula (4) and the results of Table 3, the BC threshold of supplementary citation pairs was set to 3. It means, in the BC pairs without citation, if the BC strength is greater than the value of three, the BC pairs could be added and be granted as supplementary citation pairs. Thus, 554 supplementary citation pairs including 160 patents were involved (see Table 7). The ratio of supplementary citation pairs was 16.92 %. Supplemented PCN (SPCN) was constructed using 554 supplementary citation pairs and 3,275 original citation pairs. SPCN is shown in Fig. 4; the black arcs are the original citations and the blue ones are the supplemented citations. In SPCN, the 1,015 citing patents were same as OPCN, but the number of cited patents had increased to 824. The average cited number and the average citing number increased to 4.65 and 3.77 respectively.

According to Fig. 4, SPCN was still separated into three clusters. The citation links within each cluster were more concentrated. Few citation links were supplemented between two clusters. This implies that the supplementary citation pairs were in fact highly relevant. For further analysis (see Table 8), 98 patents related to electric automobiles were involved. Although the number of patents involved in the cluster of wheelchair and mobility scooter was not the largest, the 59 % of supplementary citation pairs and average BC strength (6.04) were the highest. This indicates that the supplementary citation links made the cluster of wheelchair and mobility scooter more concentrated.

Table 9 Number of patents in the APCN

Basic data	Extracted citation		Supplementary citation		Citation in APCN	
	Number	% to PCN	Number	% to PCN	Number	% to PCN
Cited patent	97	12	124	16	181	24
Citing patent	109	11	124	12	1,172	17
Cited or citing patent	166	13	160	13	238	19
Average cited count	3.07	72	4.47	105	4.71	111
Average citing count	2.73	85	4.47	138	3.58	111
Total citation pair	298	9	554	17	852	26
% to APCN	34.98		65.02		100	

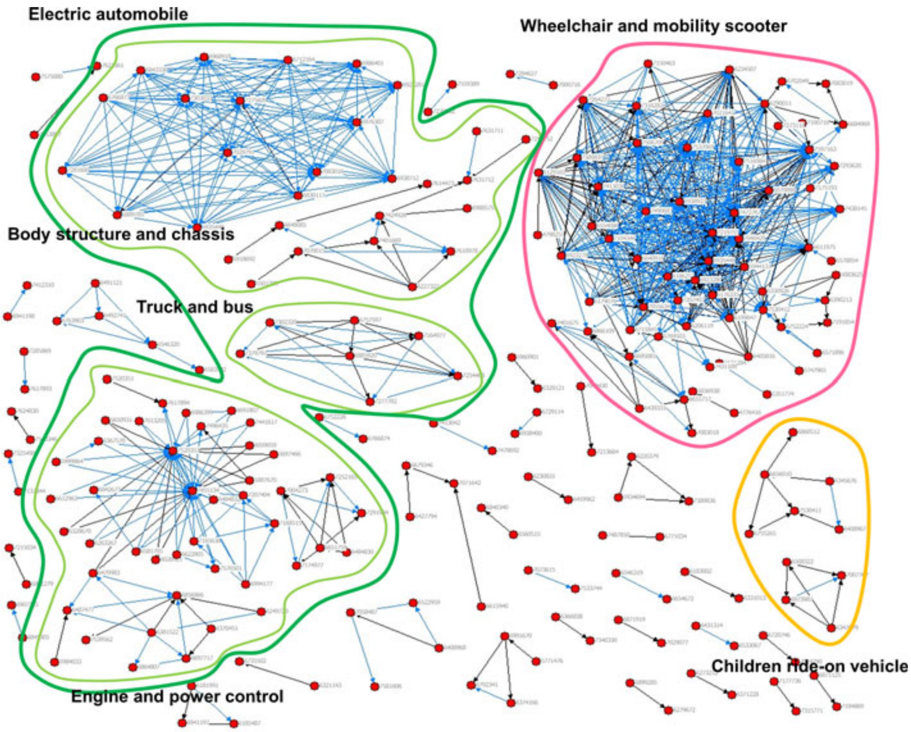

Fig. 5 The APCN of electric vehicle technology

Table 10 Extracted and supplemented citations in the APCN

	Children vehicle	Wheelchair and mobility scooter	Electric automobile				Others	Total
			Structure	Truck	Engine	Total		
Total citation								
Patent	11	62	29	7	44	80	84	238
Citation pair	13	530	138	20	94	252	57	852
Average BC strength	3.08	5.46	6.21	6.4	4.21	5.43	3.11	5.26
Extracted citation								
Patent	10	55	16	7	26	49	52	166
Citation pair	11	203	13	10	30	53	31	298
Average BC strength	3	4.52	3.15	4.3	3.2	3.40	2.7	4.08
Supplementary citation								
Patent	4	54	24	7	32	63	39	160
Citation pair	2	327	125	10	64	199	26	554
Average BC strength	3.5	6.04	6.53	7.9	4.69	5.86	3.44	5.90

Table 11 Citation and BC strength of three CTL zones in the APCN

Type of citation	Total			CTL $\leq$ 0		0 < CTL <213 weeks		CTL > 213 weeks	
	Patent	Citation	Avg BC	Citation	Avg BC	Citation	Avg BC	Citation	Avg BC
Citation in APCN	238	852 (26 %)	5.26	495 (74 %)	6.18	271 (19 %)	4.33	86 (7 %)	2.97
Children vehicle	11 (5 %)	13	3.08	6	2.67	7	3.43	0	0
Wheelchair and mobility scooter	62 (26 %)	530	5.46	268	6.59	199	4.68	63	3.09
Electric automobile	80 (34 %)	252	5.45	182	6.29	49	3.57	21	2.37
Body structure and chassis	29 (12 %)	138	6.21	122	6.60	13	3.54	3	2
Engine and power control	44 (18 %)	94	4.21	43	5.28	33	3.64	18	2.72
Truck and bus	7 (3 %)	20	6.10	17	6.65	3	3	0	0
Others	85 (36 %)	57	3.11	39	3.31	17	2.65	1	3

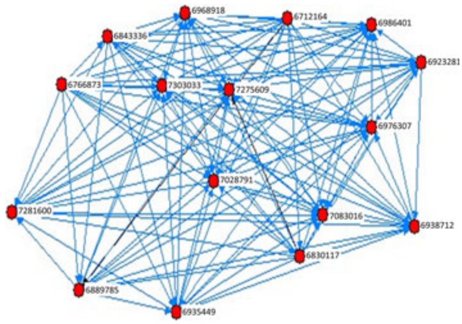

Fig. 6 Patent citations owned by general motor

Table 12 Applications and issued date owned by general motor

Patent	Application date	Issued date
6923281	2002.7.24	2005.8.2
6976307	2002.7.24	2005.12.20
6712164	2002.7.25	2004.3.30
6766873	2002.7.25	2004.7.27
6889785	2002.7.25	2005.5.10
6968918	2002.7.25	2005.11.29
6986401	2002.7.25	2006.1.17
6830117	2002.7.26	2004.12.14
7028791	2002.7.26	2006.4.18
7083016	2002.7.26	2006.8.1
6843336	2002.7.29	2005.1.18
6938712	2002.7.29	2005.9.6
7275609	2002.7.29	2007.10.2
7303033	2002.10.10	2007.12.4
6935449	2003.4.2	2005.8.30
7281600	2005.8.2	2007.10.16

The average CTL of SPCN was 175 weeks, reduced by 38 weeks from OPCN, and showing that the information in SPCN was more current than OPCN. In-depth observation of the supplementary citation pairs in the three CTL zones indicated that the CTL of 400 supplementary citation pairs was less than zero, and the ratio was up to 60 % in the first zone. This shows that the current citation pairs were certainly revealed by these supplementary citations. The average BC strength of the first zone was 6.42, which was the highest. The ratio of supplementary citation pairs in the third zone reduced to 2 % and the average BC strength was reduced to 3.21. This result still showed that the negative relationship existed between the BC strength and the CTL in the SPCN.

Aggregated patent citation network

According to the improved results of EPCN and SPCN, the study integrated EPCN and SPCN into an aggregated PCN (APCN). 852 citation pairs were established with 298 (34.98 %) extracted citation pairs and 554 (65.02 %) supplementary citation pairs. 19 % patents of the OPCN were reserved. The average cited number and the average citing number were both higher than those in OPCN (see Table 9). This implies that the weak relations in the OPCN were excluded and relatively stronger relations were added in the APCN.

The relationships of patents of APCN were visualized in Fig. 5. The black arcs were the extracted citations and the blue ones were the supplementary citations. Patents in the APCN were still grouped into five clusters similar to EPCN, but the citation links in the APCN were more concentrated. Although the patents in the cluster of engine and power control were still separated into three sub-clusters, the patents in these sub-clusters were aggregated in a more concentrated manner. This was an important finding in which those invisible relationships were manifested.

The information of extracted and supplemented citations of APCN was shown in Tables 9 and 10. 65.02 % of citation pairs in the APCN were supplementary citation pairs. The number of supplementary citation pairs in most technology clusters was greater than that of extracted citation pairs, except for the cluster of “children vehicle”. The number of citation pairs in the cluster of “wheelchair & mobility scooter” was the largest which accounted for more than half of citation pairs in the APCN. The average BC strength of this cluster was 5.46 which higher than the averages of others. Furthermore, body structure and chassis cluster, a sub-cluster of electric automobile technology, comprised 29 patents with 138 citation pairs. The average BC strength in this cluster was 6.21 which even higher than the cluster of wheelchair and mobility scooter. For further analysis, the study found that this cluster had 125 supplementary citations and an average BC strength of 6.53. It implied that many neglected links were found in the cluster of body structure and chassis. The unobvious relationships were manifested in the APCN.

The average CTL of APCN was 56 weeks. Compared to OPCN (213 weeks), EPCN (110 weeks) and SPCN (175 weeks), the average CTL of APCN reduced to 157 weeks (about 3 years). This shows that the information of the APCN was the most current among all PCNs. As presented in Table 11, there were 495 citation pairs in the first zone, which was more than the half of citation pairs in the APCN. The average BC strength of the first zone reached 6.18 was the highest. There were only 86 citation pairs in the third zone and most citation pairs were extracted citations. The average BC strength of the third zone was 2.97. This result demonstrated again that BC strength and CTL have a negative relationship.

Looking more deeply at the citation pairs in the sub-cluster of body structure and chassis, we see 138 citation pairs with an average BC strength of 6.21. As shown in Fig. 5, this cluster was separated into two groups. The first group was more concentrated, as shown in Fig. 6. There were 16 patents in the first group, and all of them were owned by the assignee General Motor. Referring to their application dates and issued dates, as shown in Table 12, we find that most of them were published at roughly the same time, and thus did not have enough time to cite each other. As a result, the APCN was actually helpful to reveal the invisible relationships among patents.

Conclusion

Patent citation analysis is regarded as a crucial way to analyze technological development. When discussing the citation relationship of patent networks, several studies took those citations between patents within the network into account (Chen et al. 2011; Fontanaa et al. 2009). Therefore, the close-world assumption would also be the limitation of the study. This study used the BC approach to filter the irrelevant patent citations and supplement relevant uncited patent citations to construct the PCN. First, we found that few citation links in the OPCN were intermediary. This shows that these citation pairs were probably not crucial. For this reason, the study used BC strength to measure degree of relevance of citation pairs and filtered irrelevant citation pairs. After this, the EPCN was constructed. The average BC strength of EPCN was increased to 4.08 and the average CTL of EPCN was tremendously reduced to 110 weeks. These results showed that the more relevant and shorter CTL of citation pairs were reserved in the EPCN. Next, we measured the BC strength of BC pairs without citation and supplemented the relevant uncited citation pairs. Then the SPCN was constructed. The average BC strength of SPCN was 5.90 and the average CTL of SPCN was 175 weeks. These results showed that the information in SPCN was more current than that in OPCN. Finally, the study integrated the EPCN and SPCN into the APCN. The average BC strength of APCN was 5.26 and the average CTL of APCN was 56 weeks. The results showed that patents in the APCN were more relevant than OPCN and EPCN. The information of APCN was the most current among all PCN.

This study further discussed technology clusters within the PCN. There were three technology clusters in the OPCN, namely electric automobile, wheelchair and mobility scooter, and children vehicle. The electric automobiles cluster had the largest number of patents. The EPCN was constructed from five clusters after removing irrelevant citation pairs. Three clusters were same as OPCN, and the cluster of electric automobile technology was further divided into three sub-clusters: body structure and chassis, engine and power control, and electric truck and bus. The results showed that the technology of electric automobile was multiplex and that unobvious technology clusters were revealed in the EPCN. The SPCN was categorized into three clusters; however, the citation links within each cluster were more concentrated. Finally, we found not only that invisible technology clusters in the APCN were revealed, but that citation links of all the technology clusters were much more concentrated. APCN aimed at retaining more relevant and current patents in the patent citation network. Furthermore, APCN had the benefit to disclose the invisible technology clusters for the field. These findings were the most important in our research.

As the problem of how to cite proper references is a constant concern in patent citation, this study aimed to provide a different perspective on the issue. Based on the BC approach, three different PCNs were proposed to clarify the situations. Through these PCNs, we obtained a better understanding of how proper patent citation affects the structure of technological networks. If patent citations are proper, the relationships between technologies are more clearly revealed. With regard to practical implications, APCN provided a good approach to find more relevant and more complete patents for a given field. With filtering the irrelevant patents and supplementing the missing patent links, APCN offered an alternative to conduct better patent analysis. Furthermore, it is hoped that these patent citation analysis results can be better applied to future research.

References

Albert, M., Avery, D., Narin, F., & McAllister, P. (1991). Direct validation of citation counts as indicators of industrially important patents. *Research Policy, 20*(3), 251–259.

Atallah, G., & Rodriguez, G. (2006). Indirect patent citations. *Scientometrics, 67*, 437–465.

Bichteler, J., & Parsons, R. G. (1974). Document retrieval by means of an automatic classification algorithm for citations. *Information Storage and Retrieval, 10*(7–8), 267–278.

Borgatti, S. P., & Cross, R. (2003). A relational view of information seeking and learning in social networks. *Management Science, 49*(4), 432–445.

Breschi, S. & Lissoni, F. (2005). Knowledge Networks from Patent Data. *Handbook of Quantitative Science and Technology Research*, pp. 613–643.

Carpenter, M., & Narin, F. (1983). Validation study: patent citations as indicators of science and foreign dependence. *World Patent Information, 5*(3), 180–185.
Case, D. O., & Higgins, G. M. (2000). How we can investigate citation behaviour ? A study of reasons for citing literature in communication. *Journal of the American Society for Information Science, 51*, 635–645.
Chen, A., & Chen, R. (2007). Design patent map: an innovative measure for corporative design strategies. *Engineering Management Journal-Rolla, 19*(3), 14.
Chen, D. Z., Huang, M. H., Hsieh, H. C., & Lin, C. P. (2011). Identifying hidden relevant patent citation links by using bibliographical coupling in LED illuminating technology. *Journal of Informetics, 5*(3), 400–412.
Ellis, P., Hepburn, G., & Oppenheim, C. (1978). Studies on patent citation networks. *Journal of Documentation, 34*, 12–20.
Ernst, H. (2003). Patent information for strategic technology management. *World Patent Information, 25*(3), 233–242.
Fontanaa, R., Nuvolaric, A., & Verspagend, B. (2009). Mapping technological trajectories as patent citation networks: an application to data communication standards. *Economics of Innovation and New Technology, 18*(4), 311–336.
Griliches, Z. (1990). Patent statistics as economic indicators: a survey. *Journal of Economic Literature, 28*, 1661–1707.
Hall, B., Jaffe, A., & Trajtenberg, M. (2001). *The NBER patent citation data file: lessons*. NBER working paper: Insights and Methodological Tools.
Hall, B., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of Economics, 36*(1), 16–38.
Huang, M., Chiang, L., & Chen, D. (2003). Constructing a patent citation map using bibliographic coupling: a study of Taiwan's high-tech companies. *Scientometrics, 58*(3), 489–506.
Jaffe, A., & Trajtenberg, M. (1999). International knowledge flows: evidence from patent citations. *Economics of Innovation and New Technology, 8*(1), 105–136.
Jarneving, B. (2007). Bibliographic coupling and its application to research-front and other core documents. *Journal of Informetrics, 1*(4), 287–307.
Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika, 32*(3), 245–254.
Kessler, M. (1963). Bibliographic coupling between scientific papers. *American Documentation, 14*(1), 10–25.
Lanjouw, J. O., & Schankeman, M. (2004). Patent quality and research productivity: measuring innovation with multiple indicators. *Economic Journal, 114*(495), 441–465.
Lee, H., Kim, C., Cho, H., & Park, Y. (2009). An ANP-based technology network for identification of core technologies: a case of telecommunication technologies. *Expert Systems with Applications, 36*(1), 894–908.
Liu, S., & Shyu, J. (1997). Strategic planning for technology development with patent analysis. *International Journal of Technology Management, 13*(5), 661–680.
Lo, S. C. S. (2010). A comparative study of linkage indexes: co-assignee, reciprocal citation, patent coupling and co-patent. *Journal of Library and Information Studies, 8*(1), 11–27.
Meng, H. C., Hwang, M. J., Chang, D. M., & Kuo, K. H. (2000). Industry innovation indicator: patent citation analysis and patent indicator (I). *Journal of Technology Management, 5*(1), 31–49. (In Chinese).
Meyer, M. (2000). What is special about patent citaions? Differences between scientific and patent citations. *Scientometrics, 51*, 185–201.
Michel, J., & Bettels, B. (2001). Patent citation analysis: a closer look at the basic input data from patent search reports. *Scientometrics, 51*(1), 185–201.
Nagaoka, S. (2004). Assessing the R&D management of firms by patent citation: evidence from the US Patents. In *Proceedings of Nihon Keizai Gakkai Autumn research presentations*.
Narin, F. (1994). Patent bibliometrics. *Scientometrics, 30*(1), 147–155.
Smith, L. (1981). Citation analysis. *Library Trends, 30*(1), 83–106.
Swanson, D. (1971). Some unexplained aspects of the cranfield tests of indexing performance factors. *The Library Quarterly, 41*(3), 223–228.
Trajtenberg, M., Henderson, R., & Jaffe, A. (1997). University versus corporate patents: a window on the basicness of invention. *Economics of Innovation and New Technology, 5*(1), 19–50.
Verspagen, B. (2000). The role of large multinationals in the Dutch technology infrastructure: a patent citation analysis. *Scientometrics, 47*(2), 427–448.
Wartburg, I. V., Teichert, T., & Rost, K. (2005). Inventive progress measured by multi-stage patent citation analysis. *Research Policy, 34*(10), 1591–1607.
Yoon, B., & Park, Y. (2004). A text-mining-based Patent network: analytical tool for high-technology trend. *The Journal of High Technology Management Research, 15*(1), 37–50.