The Missing Link Effect: A Case Study Using Patent Main Path Analysis

Chung-Huei Kuan[1] and Dar-Zen Chen[2]

[1] *maxkuan@mail.ntust.edu.tw*
National Taiwan University of Science and Technology, Taipei (R.O.C., Taiwan)

[2] *dzchen@ntu.edu.tw*
National Taiwan University, Taipei (R.O.C., Taiwan)

Abstract
Both direct citation and Bibliographic Coupling (BC) are valid tools in detecting similarity or relatedness among patents. This study is then motivated by the question: what is the difference between a direct citation and a strong BC but without direct citation? We found that the latter, referred to as *missing link* here, seems to be useful in capturing the technology relatedness among patents filed or issued at close times where direct citations are less likely to be established. We then applied the Main Path Analysis to a patent citation network involving 38,652 patents/publications, 188,721 direct citations, and 9,213 missing links, and derived two main paths with and without the missing links added. By comparing these main paths, we found that the missing links, merely accounting for less than 5% of all connections in this case, may identify a series of intermediate and related patents reflecting elaborative evolvement, which may well be overlooked when only direct citations are considered.

Conference Topic
Patent analysis

Introduction
A first patent is cited or referenced by a second, subsequent patent because the applicant of the second patent considers that the first patent provides, addresses, or discloses similar or related technical background, issue, or solution. The first patent may also be cited by the examiner of the second patent as the first patent may challenge the patentability of the second one, in addition to those reasons by the applicant.

Patent citations have long been accepted as having high bibliometric value (cf. Hall, Jaffe, & Trajtenberg, 2005; Jaffe, Fogarty, & Banks, 1997; Trajtenberg, 1990). In many applications, they are assumed to have revealed the similarity or relatedness between patents.

A representative research area based on the above assumption is the knowledge or technology spillover where knowledge flow between similar or related patents associated with different geographic areas or industries is observed. Here we list a small number of studies in this area published after 2013, demonstrating that utilizing patent citations as a proxy for knowledge flow is a widely accepted approach. Murata et al. (2014) confirmed knowledge spillover are significantly localized using cited-citing relationships between U.S. patents granted between 1975 and 1999. Figueiredo, Guimarães, and Woodward (2015) found that knowledge spillover is positively correlated with industry localization and that the localization of an industry may offset the adverse effect of distance, using a set of cited-citing pairs of U.S. patents. Li (2014) studied the effects of distance and subnational/national borders on international and intranational knowledge spillovers through patent citations across some most cited countries and metropolitan statistical areas within U.S, and claimed that border and distance effects increase over time, but fade as patents age. Karvonen and Kässi (2013) collected patent data from main players of the radio-frequency identification (RFID) value chain in EPO database, and used citation data to investigate the emergence of new industry segment. Kim, Lee, and Sohn (2016) quantified the spillover of unmanned aerial vehicle (UAV) technology to various industries by retrieving U.S. patents that cited UAV patents and matching these citing patents to industries. Noailly and Shestalova (2017) used citations of European patents to see what technologies are built on the knowledge developed in renewable energy.

Researchers have been comparing applicant-submitted references (hereinafter, applicant citations) and examiner-located references (hereinafter, examiner citations) for their difference. Criscuolo and Verspagen (2008) found that examiners more often provide citations that may compromise patentability, and that applicants tend to ignore prior art that may endanger their patent applications. Hegde and Sampat (2009) found that examiner citations are better value indicator as they have a much stronger relationship with patent renewal probability than the applicant citations do. Cotropia, Lemley, and Sampat (2013) argued that patent examiners rely almost exclusively on prior art they find themselves, instead of applicant citations, in rejecting patent applications or in limiting patent scopes. In contrast, Thompson (2006) indicated that applicant citations are more relevant as applicants are more familiar with their inventions than the examiners are. Alcacer, Gittelman, and Sampat (2009) suggested that applicant citations express knowledge flows better than examiner citations. Park, Jeong, and Yoon (2017) found that the patents cited by applicants have higher quality than those cited by examiners in four industries.

A patent citation, whether it is from the applicant or the examiner, indeed reflect a connection between the cited and the citing. The difference seems to lie only in the degree of similarity, relatedness, or how much knowledge indeed flows from the cited to the citing. What intrigues us here is not the origin of the citation, but the absence of direct citation between two seemingly related patents.

The “missing” citation

In addition to the utilization of direct citations outlined above, another approach in detecting or measuring the similarity or relatedness between two documents or patents is Bibliographic Coupling (BC). Two documents or patents are bibliographically coupled if they both cite at least one common reference. The Bibliographic Coupling Strength (BCS) represents the number of common references. Kessler (1963) proposed the concept Bibliographic Coupling and used it to measure subject similarity. The concept then quickly drew significant interest from researchers. Various applications have been designed and we listed a small number of publications in using BC to detect or measure the similarity or relatedness of patents as follows. Wartburg, Teichert, and Rost (2005) proposed a multi-stage patent citation analysis for the measurement of inventive progress using BC. Park et al. (2015) used BC to measure technological similarity in the fuel cell membrane electrode assembly technology so as to locate potential R&D collaboration partners. Huang and Chang (2014, 2015) applied BC with sliding window to track the generation, growth, decline, and disappearance of research fronts.

If both direct citation and BC reflect the similarity or relatedness between two patents, it is interesting to notice that sometimes there is no direct citation between two patents that are strongly bibliographically coupled.

Taking two U.S. utility patents, US8,622,222 and US8,623,202 as an example, they are both related to membrane bioreactor technologies, were both filed by the same company, one in January 2011 and the other October 2012, but with different inventors. They were both granted within January 2014 (therefore, their patent numbers are very close) after being examined by different examiners. The two patents do not have direction citation between them, but their similarity is clearly reflected by their exceptionally high BCS 1,039 (US8,622,222 has total 1,109 references and US8,623,202 has 1,108 references).

Taking another two more distant U.S. utility patents, US8,585,882 and US9,586,842, as an example, they are both related to water treatment, but filed by different companies, one in December 2008 and the other December 2015. They are granted in November 2013 and March 2017, respectively, by different examiners. The two patents do not have direction citation, but their similarity is also clearly reflected by their high BCS 465 (US8,585,882 has 558 references and US9,586,842 has 622 references).

The term *missing link* is often used loosely to describe a newly found evidence indicating a possible evolution path between a predecessor and a successor such as a new piece of fossil manifests a possible ancestor to the human lineage. We borrowed the term here to denote this latent connection between patents suggested by their strong BC.

Some researchers had already noticed the occurrence of missing links and made use of them. Chen et al. (2011) used missing links to construct more comprehensive technological clusters using light emitting diode (LED) as a case study. Yeh et al. (2013) took one step further by filtering out direct citations between patents with weak BC, in addition to incorporating missing links into a patent citation network.

Why do missing links occur? One possible explanation for the lack of direct citation is that applicant citations or examiner citations are by no means exhaustive, and it is surely possible that the applicant or examiner of one patent fails to learn or discover the presence of the other.

On the other hand, Chen et al. (2011) calculated the time lags for missing links and found that, for a majority of them (65.52%) in their case study, one has application date before the other's issued date (i.e., their application processes are overlapped in time). The authors then argued that the later patents' applicants at the times of filing could not be aware of and cite the earlier ones since information about the earlier ones was not available to the public then.

We are then motivated by the question: what is the difference between a direct citation and a strong BC but without direction citation? Prompted by Chen et al. (2011), the answer seems to lie in the timing characteristics of missing links, and an empirical study was conducted for this purpose.

Empirical Data

We chose the technical area of carbon dioxide (CO_2) capture, storage, recovery, delivery, and regeneration as our empirical case. We collected the related utility patents and their references (including U.S. patents and publications) all issued or published after 1976 and before 2017/03/28 from USPTO database. These patents/publications have at least one of some specific keywords in at least one of the relevant fields (i.e., title, abstract, specification, and claims) and also have at least one of some specific Cooperative Patent Classification (CPC) symbol prefixes. There are then total 38,652 patents and publications with 188,721 pairs having direct citations and 1,609,549 bibliographically coupled pairs.

The 1,609,549 BC pairs have a significantly skewed BCS distribution with a mean 2.74 (μ), a standard deviation 15.66 (σ), and a maximum 1,123. Among them, 1,477,783 pairs, accounting for 92% of all BC pairs, have BCS not greater than 3 (the closest integer to the mean BCS). Since missing links do not actually exist, we decided to be conservative and considered only BC pairs having BCS greater than 34 (=μ+2σ). There are only 13,013 such pairs, accounting for only 0.81% of all BC pairs. We then further removed 3,800 of them that already have direct citations. Finally we were left with as few as 9,213 (=13,013-3800) BC pairs having a range of the BCS between 35 and 1,123. The BCS distribution of these BC pairs is shown in Figure 1. A great majority of the BC pairs (8,674 or 94%) have BCS less than or equal to 300.

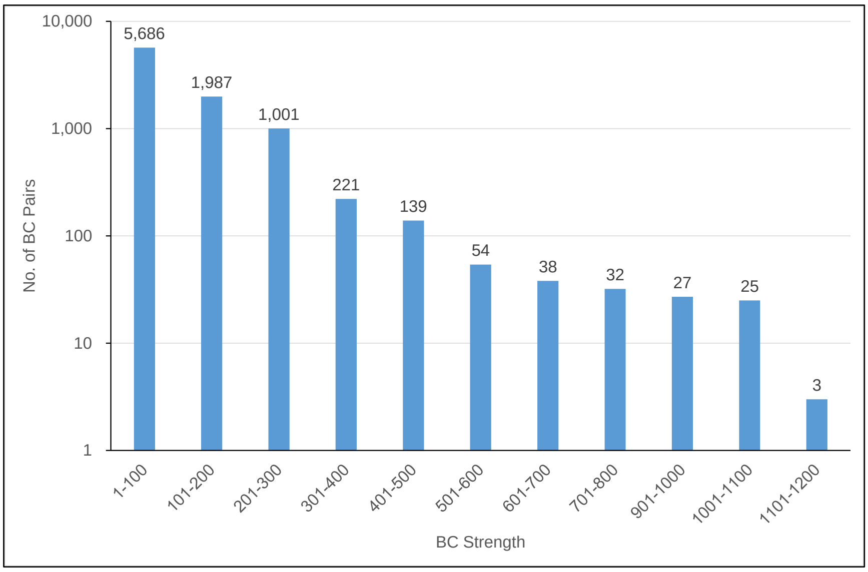

Figure 1. The BCS distribution of the 9,213 BC pairs (vertical axis is in log scale).

Figure 2 shows the distributions of time spans between the filing dates (in lighter bars) and between the issued dates (in darker bars) for the 9,213 BC pairs. As illustrated, 58% of the 9,213 BC pairs have filing date differences fall within 2 years, and 65% have issued date differences fall within 2 years. Then, as the time span between patents' filing and issued dates increases, it seems that they are less likely to form strong BC pairs as manifested by the BC pairs' decreasing trends to the right shown in Figure 2.

This seems to confirm the observation of Chen et al. (2011) that, if two patents are filed or issued within a short time span, their applicants or examiners are more difficult to be aware of each other, as both patents may remain undisclosed before filing or remain pending in the patent office before issuance.

Overall, the above observation seems to suggest that the missing links would be most useful in capturing those technology relatedness among patents where they are filed or issued at close times and where direct citations are less likely, in not impossible, to be established.

To further investigate whether our speculation is reasonable, we conducted the so-called *Main Path Analysis* for our empirical case.

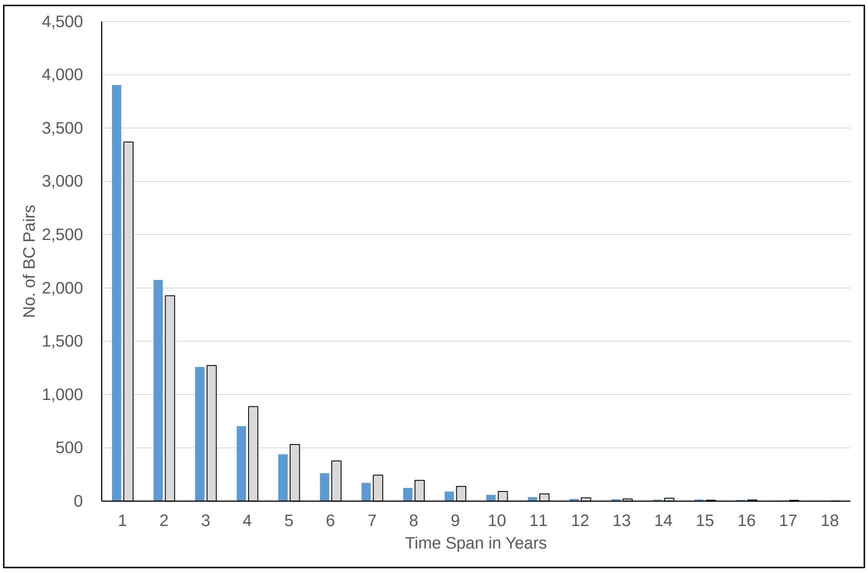

Figure 2. The distributions of time spans of the 9,213 BC pairs.

Main Path Analysis

Main Path Analysis (MPA) was first taught by Hummon and Doreian (1989) to discover the major development trajectory of a scientific field through identifying the most significant chain of citations in a citation network. This method generally involves the following steps: (a) constructing a citation network from relevant documents of the scientific field; (b) calculating a weight for each link of the citation network related to the link's traversal counts; and (c) searching for a series of connected links across the network according to their weight. This series of links is then referred to as the *main path* of the scientific field, and considered as having embodied a development trajectory of the field.

Since its inception, researchers have been applying this method in, for example, detecting technological changes, knowledge transformation (Lucio-Arias & Leydesdorff, 2008; Martinelli, 2012; Mina et al., 2007), and reviewing literatures (Bhupatiraju et al., 2012; Calero-Medina & Noyons, 2008; Colicchia & Strozzi, 2012; Harris et al., 2011; Liu et al., 2013; Lu, Hsieh, & Liu, 2016). The same method is also applied to patent citation networks in mapping technological evolvements (Fontana, Nuvolari, & Verspagen, 2009; Park & Magee, 2017; Verspagen, 2007). MPA capability is built in the well-known network analysis application Pajek (Batagelj & Mrvar, 1998).

The rationale of MPA is to consider each link of the citation network as a path of knowledge flow from the cited to the citing. Then, an algorithm is applied to determine a weight for each link related to the number of times the link is traversed. For example, Figure 3 is a fictitious citation network where the link weights are determined using the algorithm Search Path Link Count (SPLC) (Hummon & Doreian, 1989). The weight of the link 5→6 (i.e., from node 5 to node 6) is 15, because it counts the number of traversals of the link 5→6 from all preceding nodes (1 to 5) to the sink nodes (7, 8, 9). Similarly, the weight of the link 6→7 is 6 because there are 6 preceding nodes (1 to 6) and each traverses the link 6→7 once to reach the sink node 7.

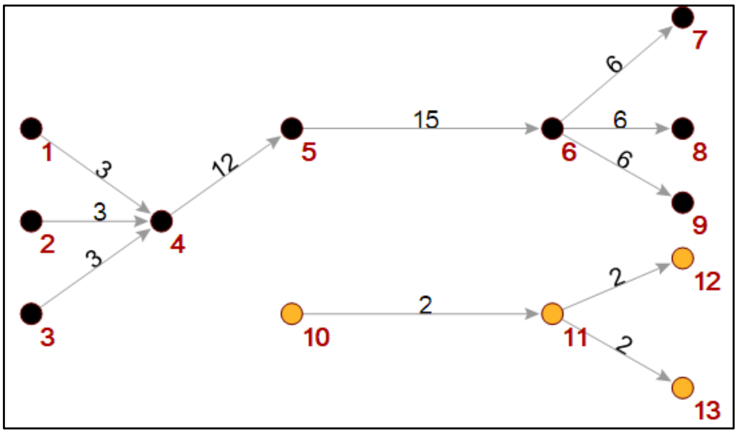

Figure 3. A fictitious citation network with link weights assigned using SPLC.

There are other algorithms for weight calculation such as SPC (search path count) (Batagelj & Mrvar, 1998), SPNP (Search Path Node Pair) (Hummon & Doreian, 1989), etc. These algorithms all determine the weight solely based on the topological location of the link, and a link would have a greater weight if it can be reached from more nodes or more nodes may be reached through it. The link weight based on the traversal count therefore functions as a proxy to the total amount of knowledge flow through the link.

There are also different methods in determining the main path once the link weights are set. The main path of Figure 3 is denoted by black nodes using global search method (Liu and Lu, 2012) which selects a path with the greatest total weight from source to sink nodes. The local search method starts from the source or sink nodes and selects the link(s) from/to the source/sink nodes with the greatest weight(s) and works forward/backward for the next search until a sink/source node is reached (Hummon & Doreian, 1989). The key-route method (Liu and Lu, 2012) determines one or more main paths by locating the link(s) with the greatest weight first and tracing backward and forward until a source or sink node is reached.

By treating missing links as real citations and applying MPA, we are able to see whether the missing links would lead to any difference in the resulted main path. We then may gain some insight into the effect of missing links if some major difference does occur.

Main Paths

A patent citation network (PCN) was constructed out of our empirical case and involved 38,652 nodes and 188,721 arcs. We then adjusted the original patent citation network (OPCN) by incorporating 9,213 missing links as real citations. Each missing link always originates from the lower numbered patent to the higher numbered patent. As there are 188,721 links in the OPCN, these missing links account for only 4.6% (=9,213/ (188,721+9,213)) of the links of the adjusted citation network (APCN).

In determining the main path, we chose to use the SPLC algorithm in calculating the link weights as prior researches had reported that algorithms SPC, SPLC, SPNP, etc. all derive similar results (cf. Batagelj, 2003; Verspagen, 2007). Further, we chose to use the global search method (Liu & Lu, 2012) in developing the main path. The key-route method is a recent variant to the MPA, but it will usually produce more than one main path. To simplify the comparison, we therefore chose the global search method as it will usually produce a single main path having the globally greatest total weight.

The original main path derived without missing links and the adjusted main path obtained with the missing links supplemented are displayed together in Figure 4 using nodes and arcs of different styles for easier comparison. There are total 67 different nodes (patents) from both the original and adjusted main paths, and they are numbered from 1 to 67 in ascending order of their patent numbers and therefore their issued dates. A complete list of these 67 patents is provided in the Appendix.

There are 19 black nodes representing those patents present on both the original and adjusted main paths. There are 8 white nodes representing those appearing only on the original main path, and the 41 grey nodes are those present only on the adjusted main path. The solid grey arcs are direct citations, and the dashed black arcs are missing links.

The original main path involves 27 patents and is represented by the chain of black and white nodes connected by solid grey arcs (direct citations). The adjusted main path contains total 59 patents and is represented by the chain of black and grey nodes connected either by solid grey arcs (direct citations) or dashed black arcs (missing links).

As can be observed from Figure 4, the adjusted main path has captured two third (18) of the patents from the original main path. It seems fair to claim that both the original and adjusted main paths derive a substantially identical trajectory of knowledge flow in our case study, but the adjusted main path has offered significantly more information, as it provides additional 41 (=59-18) patents and additional 49 links in the adjusted main path. For the 49 newly added links, 6 of them are direct citations, whereas the other 43 are all supplemented missing links.

Based on where the newly included nodes and arcs occur, the adjusted main path may be partitioned into 4 sections, and their end nodes and span years (based on end patents' issued years) are listed in Table 1. These 4 sections allow us to examine more closely the effect of missing links.

In Section 1, the original and adjusted main paths are completely identical. This is because a patent in an early stage of a technology field has fewer references. For example, the end nodes 1 (US3,977,845 issued in 1976) and 8 (US4,813,980 issued in 1989) have only 7 and 15 references. As we considered only BC pairs having BCS greater than 34, no BC pair within this stage would satisfy the threshold.

In Section 2, the end node 19 (US6,221,117 issued in 2001) has 140 references, indicating that there would be BC pairs satisfying the threshold and missing links are supplemented. These missing links help to identify and add new nodes 10, 11, and 18 to the adjusted main path, but these 3 nodes are all connected by direct citations, not missing links. The reason is that the supplemented missing links are few and dispersed, and they are not able to cooperatively thrust any one of them into a status as a major link of knowledge flow. However, they are able to strengthen some existing direct citations (i.e., 8→10→11→14, and 16→18→19) to become part of the adjusted main path.

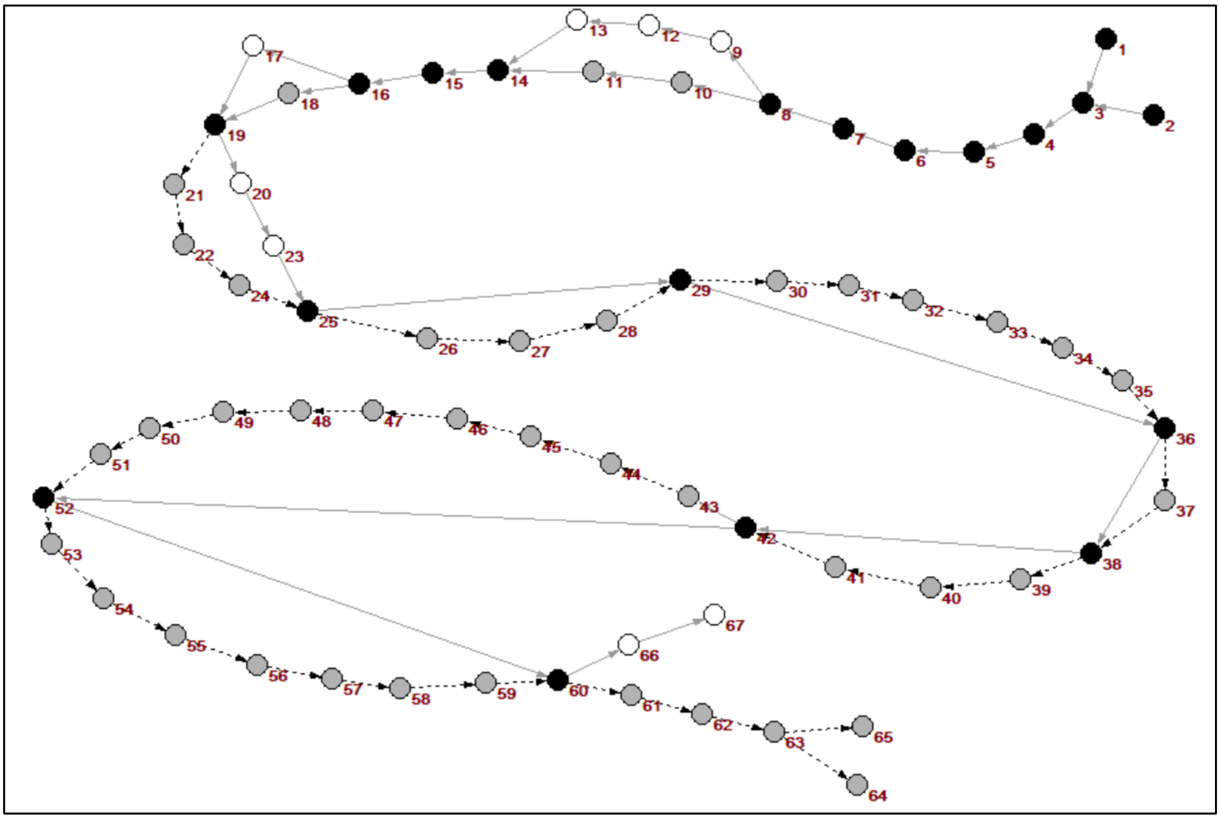

Figure 4. The combined manifestation of the original and adjusted main paths.

Table 1. The 4 sections of the adjusted main path.

Section	*End nodes*	*Span years*	*BCS*		
			Min.	*Max.*	*Mean*
1	1→8	1976-1989			
2	8→19	1989-2001			
3	19→60	2001-2011	36	256	99.2
4	60→67	2011-2014	41	240	159.8

In Section 3, the end node 60 (US8,057,575 issued in 2011) has 276 references, implying that there are more missing links incorporated into the APCN in this stage than in the previous stage. In contrast to what is observed from Section 2, these more numerous missing links are able to cascade together into a number of different path segments. We will talk more about Section 3 later.

Section 4 provides a different scenario as the original and adjusted main paths develop separately. We speculate that the development of the original and adjusted main paths is still in progress. It is possible that the original and adjusted main paths may merge at a future patent, just like what happens, for example, at node 25 that merges the original path segment 19→20→23→25 and the new path segment 19→21→22→24→25.

Table 1 also provides the minimum, maximum, and average BCS for the missing links in Sections 3 and 4, respectively. We can see that their BCS is not particularly strong, and actually none of the strongest 5% shown in Figure 1 has emerged as part of the adjusted main path. This observation does not come as a surprise since, for a missing link to appear on the adjusted main path, it is determined by its topological location, not by its BCS.

Based on the above observation, we can see that there are some key patents denoted by nodes such as 19, 25, 29, 36, 38, 42, 52, and 60 of Section 3 providing a backbone of the knowledge flow as they are present on both the original and adjusted main paths. However, the adjusted main path suggests totally different path segments between these key patents involving completely different sets of patents connected almost entirely by the missing links. This seems to suggest that missing links may indeed deliver more topological significant relatedness than what is captured by the replaced direct citations.

To see that this phenomenon is not coincident, we conducted a number of experiments by manually multiplying the weights of all missing links by 0.5, 0.1, and 0.01 after all link weights are determined and before applying the global search method to determine the main path. In other words, we artificially and purposely reduced the weights of all missing links to one half, one tenth, and one hundredth of their original values. We found that the resulted main paths are exactly the same as the original one. This implies that the weights (i.e., traversal counts) of the 43 missing links on the adjusted main path are so strong that downplaying their importance to a great degree still wouldn't produce any difference. Therefore, the 43 missing links indeed capture some topologically significant connections among patents.

In addition, please pay special attentions to nodes 25, 29, 36, 38, 42, 52, and 60 in Section 3. These patents are issued at close years 2003, 2003, 2004, 2005, 2006, 2010, and 2011, respectively (except the patents denoted by node 42 and 52 where they are issued at years 2006 and 2010). In the OPCN, they are sequentially cascaded by direction citations but, in the APCN, each of these direct citations is replaced by a series of missing links connecting a number of patents not on the original main path.

Taking nodes 52 and 60 as example, the arc in the original main path reflects the knowledge flow between the patents denoted by the two nodes. But with the help of the missing links, patents denoted by nodes 53 to 59 are manifested in the adjusted main path. As we already know, these are related patents occurring at close times (i.e., all issued between 2010 and 2011) and their relatedness may be overlooked by direct citations. Therefore, it seems that this series of intermediate and related patents may give us a more detailed picture regarding, for the original two patents denoted by nodes 52 and 60, how one evolves to the other.

As mentioned earlier, the main idea behind MPA is to use citations as proxies for knowledge flow. But missing links are not real citations, they are only simulations of real citations in the APCN. Therefore the adjusted main path derived from the APCN should not be considered as a better or more accurate path. The adjusted main path in this study is mainly used as a tool for observing the effect of missing links.

Discussion and Conclusion

Both direct citations and BC are valid tools in detecting similarity or relatedness, or in capturing knowledge flow among patents. Then, we were intrigued by a conflicting scenario that two patents lack direct citation but are strongly bibliographic-coupled.

Based on empirical data, we found that a significant portion of missing links have their patents' filing or issued dates fall within one to two years. In other words, these patents are prosecuted concurrently, and it would be difficult, if not impossible, for their applicants or examiners to become aware of and cite each other. Yet their relations may well be captured by missing links. Therefore, missing links seem to be useful in capturing the technology similarity or relatedness among these concurrent patents while direct citations among them are less likely to be established.

This study then applied MPA to a patent citation network involving 38,652 patents/publications, 188,721 direct citations, and 9,213 missing links. We then derived an original main path without the missing links added, and an adjusted main path with the missing links added to simulate real citations.

We found that some direct citations in the OPCN are replaced by a series of newly found patents connected by missing links in the adjusted main path. We verified that this is not coincident by manually downplaying the weights of the missing links to as low as one hundredth of their original values, and the adjusted main path is not affected at all.

On one hand, this observation suggests that missing links may indeed deliver, from the PCN and main path's point of view, topologically more significant connections between patents.

More importantly, as a single direct citation of the OPCN is elaborated into a chain of newly found patents in the APCN, we found that these newly found patents not only may fill the gap left by direct citations as described above, but also may give us a more detailed picture regarding, for the two patents connected by the direct citation, how one evolves to the other.

This study may be challenged that an absolute BC threshold (34) is adopted in determining the missing links, and one may argue that a normalized or relative BC threshold should be more appropriate. By using an absolute threshold, indeed we may filter out some important BC pairs having fewer references (like what we observed in Section 1 of Figure 4), and may include some not-so-important BC pairs having more references (e.g., those added in Sections 3 and 4 of Figure 4). It is definitely interesting to see if the normalized or relative BC threshold would contribute any difference. We however were not so concerned in this study since MPA itself has an inherent tendency to filter out those irrelevant links.

This study may be extended along several directions. One extension is that, instead of adding the missing links, one can use BC to filter out irrelevant direct citations and see what difference may occur in the resulted main path. Another extension is that, similar to what Persson (2010) suggested, instead of using BC to determine latent connections, the similar and popular approach Co-citation (CC) may be applied to compare the two approaches.

References

Alcacer, J., Gittelman, M., & Sampat, B. (2009). Applicant and examiner citations in US patents: An overview and analysis. *Research Policy, 38*(2), 415-427.

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

Batagelj, V., & Mrvar, A. (1998). Pajek - Program for Large Network Analysis. *Connections*, *21*(2), 47-57.

Bhupatiraju, S., Nomaler, O., Triulzi, G., & Verspagen, B. (2012). Knowledge flows—Analyzing the core literature of innovation, entrepreneurship and science and technology studies. *Research Policy*, *41*(7), 1205-1218.

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.

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.

Colicchia, C., & Strozzi, F. (2012). Supply chain risk management: A new methodology for a systematic literature review. *Supply Chain Management: An International Journal*, *17*(*4*), 403-418.

Criscuolo, P., & Verspagen, B. (2008). Does it matter where patent citations come from? Inventor vs. examiner citations in European patents. *Research Policy*, *37*(10), 1892-1908.

Cotropia, C. A., Lemley, M. A., & Sampat, B. (2013). Do applicant patent citations matter? *Research Policy*, *42*(4), 844-854.

Fontana, R., Nuvolari, A., & Verspagen, 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.

Figueiredo, O., Guimarães, P., & Woodward, D. (2015). Industry localization, distance decay, and knowledge spillovers: Following the patent paper trail. *Journal of Urban Economics*, *89*, 21-31.

Hall, B. H., Jaffe, A., & Trajtenberg, M. (2005). Market value and patent citations. *RAND Journal of economics*, 16-38.

Harris, J.K., Beatty, K.E., Lecy, J.D., Cyr, J.M., & Shapiro, R.M. (2011). Mapping the multidisciplinary field of public health services and systems research. *American Journal of Preventive Medicine*, *41*(1), 105-111.

Huang, M. H., & Chang, C. P. (2014). Detecting research fronts in OLED field using bibliographic coupling with sliding window. *Scientometrics*, *98*(3), 1721-1744.

Huang, M. H., & Chang, C. P. (2015). A comparative study on detecting research fronts in the organic light-emitting diode (OLED) field using bibliographic coupling and co-citation. *Scientometrics*, *102*(3), 2041-2057.

Hegde, D., & Sampat, B. (2009). Examiner citations, applicant citations, and the private value of patents. *Economics Letters*, *105*(3), 287-289.

Jaffe, A. B., Fogarty, M. S., & Banks, B. A. (1997). Evidence from patents and patent citations on the impact of NASA and other federal labs on commercial innovation (No. w6044). National Bureau of Economic Research.

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *Journal of the Association for Information Science and Technology*, *14*(1), 10-25.

Karvonen, M., & Kässi, T. (2013). Patent citations as a tool for analysing the early stages of convergence. *Technological Forecasting and Social Change*, *80*(6), 1094-1107.

Hegde, D., & Sampat, B. (2009). Examiner citations, applicant citations, and the private value of patents. *Economics Letters*, *105*(3), 287-289.

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

Li, Y. A. (2014). Borders and distance in knowledge spillovers: Dying over time or dying with age?—Evidence from patent citations. *European Economic Review*, *71*, 152-172.

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.

Liu, J.S., Lu, L.Y.Y., Lu, W.M., & Lin, B.J.Y. (2013). Data envelopment analysis 1978-2010: A citation-based literature survey. *OMEGA: The International Journal of Management Science*, *41*(1), 3-15.

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.

Lucio-Arias, D., & Leydesdorff, L. (2008). Main-path analysis and path-dependent transitions in HistCite (TM)-based histograms. *Journal of the American Society for Information Science and Technology*, *59*(12), 1948-1962.

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.

Mina, A., Ramlogan, R., Tampubolon, G., & Metcalfe, J. S. (2007). Mapping evolutionary trajectories: Applications to the growth and transformation of medical knowledge. *Research policy*, *36*(5), 789-806.

Murata, Y., Nakajima, R., Okamoto, R., & Tamura, R. (2014). Localized knowledge spillovers and patent citations: A distance-based approach. *Review of Economics and Statistics*, *96*(5), 967-985.

Noailly, J., & Shestalova, V. (2017). Knowledge spillovers from renewable energy technologies: Lessons from patent citations. *Environmental Innovation and Societal Transitions*, *22*, 1-14.

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.

Park, I., Jeong, Y., & Yoon, B. (2017). Analyzing the value of technology based on the differences of patent citations between applicants and examiners. *Scientometrics*, 1-27.

Persson, O. (2010). Identifying research themes with weighted direct citation links. *Journal of Informetrics*, *4*(3), 415-422.

Thompson, P. (2006). Patent citations and the geography of knowledge spillovers: evidence from inventor- and examiner-added citations. *The Review of Economics and Statistics*, *88*(2), 383-388.

Trajtenberg, M. (1990). A penny for your quotes: patent citations and the value of innovations. *The Rand Journal of Economics*, 172-187.

Verspagen, B. (2007). Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in Complex Systems*, *10*(1), 93-115.

Von Wartburg, I., Teichert, T., & Rost, K. (2005). Inventive progress measured by multi-stage patent citation analysis. *Research Policy*, *34*(10), 1591-1607.

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.

Appendix: List of patents from both the original and adjusted main path

Node	Pat. Num.	Original	Adjusted	Node	Pat. Num.	Original	Adjusted
1	3977845	Y	Y	36	6824593	Y	Y
2	4021210	Y	Y	37	6869707		Y
3	4077779	Y	Y	38	6953497	Y	Y
4	4171207	Y	Y	39	6967063		Y
5	4512780	Y	Y	40	6994927		Y
6	4624841	Y	Y	41	7005113		Y
7	4671893	Y	Y	42	7052530	Y	Y
8	4813980	Y	Y	43	7135048		Y
9	4913709	Y		44	7195663		Y
10	4915711		Y	45	7368194		Y
11	5073356		Y	46	7410531		Y
12	5133785	Y		47	7470293		Y
13	5332424	Y		48	7601302		Y
14	5435836	Y	Y	49	7632322		Y
15	5562754	Y	Y	50	7682718		Y
16	5705916	Y	Y	51	7736596		Y
17	5861137	Y		52	7789941	Y	Y
18	5997594		Y	53	7819955		Y
19	6221117	Y	Y	54	7828864		Y
20	6319306	Y		55	7939051		Y
21	6375906		Y	56	7972420		Y
22	6376113		Y	57	7981172		Y
23	6458189	Y		58	8021446		Y
24	6494937		Y	59	8038748		Y
25	6537352	Y	Y	60	8057575	Y	Y
26	6562111		Y	61	8157900		Y
27	6569227		Y	62	8257466		Y
28	6596057		Y	63	8636828		Y
29	6632270	Y	Y	64	8691463		Y
30	6641625		Y	65	8696772		Y
31	6719831		Y	66	8961627	Y	
32	6719832		Y	67	9187324	Y	
33	6723156		Y				
34	6767389		Y				
35	6783741		Y				