“© 2017 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.”

Exploring technology evolution pathways to facilitate technology management: a study of Dye-sensitized solar cells (DSSCs)

Ying Huang[a], **Fujin Zhu**[a,d], **Ying Guo** [a], **Alan L. Porter** [b,c], **Yi Zhang**[d], **Donghua Zhu** [a]

[a] School of Management and Economics, Beijing Institute of Technology, Beijing 100081, China

[b] School of Public Policy, Georgia Institute of Technology, Atlanta, GA 30313, USA

[c] Search Technology, Inc., Atlanta, GA 30092, USA

[d] Centre for Quantum Computation and Intelligent Systems, Faculty of Engineering and Information Technology, University of Technology Sydney, Australia

ABSTRACT

Market competition drives attention to the prospects of New and Emerging Science & Technologies (NESTs), which are fast changing and, so far, have relatively limited applications. Technology evolution pathways, as a powerful representation of the development of technology, have caught researchers' interest as a tool to trace historical progression, explore knowledge diffusion, and forecast future NESTs trends. Citation analysis approaches are actively applied to structure a large number of patents, map patent distribution, and capture knowledge transfer and change in technologies or industries. This paper (1) introduces the indicator of connectivity and modularity in the interior citation network to identify the technology development stage; (2) takes family patent information into the process of building a comprehensive patent citation network; (3) extracts technological trajectories by applying integrated approaches of main path analyses, namely global main path analysis and global key-route main analysis, among different technological stages. We illustrate this approach with Dye-sensitized solar cells (DSSCs), as an example of a promising NEST, contributing to the remarkable growth in the renewable energy industry. The results show how our method can trace the main development trajectory of a research field and discern the technology focus to help decision-makers facilitate technology management.

KEYWORDS

Technology Evolution Pathways; Patent Citation Network; Main Path Analysis (MPA); Dye-sensitized solar cells (DSSCs)

1. Introduction

Identifying core and emerging technologies is essential for formulating technology strategies and policies that achieve competitive advantage [1]. The ability to analyze and monitor the history and current stage of a particular technology is a critical asset to gain competitive advantage and to identify promising opportunities [2]. Many researchers have attempted to identify current technology structures and trace technological trends by performing patent analyses [3-7]. Early patent analysis methods mostly compared the numbers of patents assigned to different entities (e.g., nations, companies, inventors, and technological fields, over time [8-10]). Such indicators cannot reflect micro-level technology changes effectively, especially for New and Emerging Science & Technologies (NESTs).

Patents, as important carriers of technological knowledge, often interconnect with other patents; such technological structures and linkages are called networks. Patent citations represent previous knowledge underlying a specific inventive step. A patent citation network consists of groups of related patents in which the largest groups can be defined by patent technology categories [11, 12]. Therefore, some researchers use patent citations to investigate knowledge flows and technology diffusion [13-16]. Patent citations have also been used to evaluate corporate performance and Merger & Acquisition (M&A) candidates [17, 18]. Tracking the number of patent citations to non-patent sources can illuminate the role of science in technological innovation [19-21]. Researchers have considered the role of science in technological innovation by investigating citations between patents and the scientific literature [22, 23]. Moreover, patent citation analyses are applied to confirm rapid growth [24], explain the development trajectory of a technology [25], trace knowledge growth and transformation [26], and identify major technological barriers to mass application [27].

Patent citations, provided by patent applicants and examiners to clarify the reach of intellectual property (IP) rights, are regarded as important information in generating a highly concentrated collection of relevant patents by practitioners [28]. A patent citation indicates a technological relationship between the inventions claimed in the citing and cited patents and provides a reasonable "proxy" for their technological significance, as they generally appear to be highly correlated with other measures of the value of innovations [29].

Patent citation networks can be understood as representing relationships among the pieces of knowledge contained in individual patents to trace "technological trajectories" in a given technology field. One patent citing another implies that the cited patent reflects a piece of previously existing knowledge upon which the citing patent builds [30]. For a vast citation network, some patents usually represent key technologies in the field that play a notable role in the overall progression. These patents are usually located on the "main trajectory" of citation networks. By extracting these patents, one can gain insight into the main technical developments of a certain technology. In addition, for large technical systems, patent citation analysis provides potential clarity. By examining the structure of a patent citation network, one should be able to identify the critical trajectories characterizing a target technology's evolution [31].

Among the approaches of discovering citation trajectories, main path analysis (MPA) is one of the most attractive methods to determine the critical developing paths. The theory of MPA was first introduced by Hummon and Doreian, who called the sequences of links and nodes in the network *search paths*, and calculated a traversal count for each link to quantify the connectivity [32]. In their study, MPA aimed to seek the primary development trend in research fields through identifying the maximum connectivity from a series of studies in the literature. Subsequently, Hummon and his colleagues conducted further studies to test the method. Those studies include applying MPA to the centrality-productivity citation structure [33], to social network analysis [34], as well as to the conflict-resolution field [35]. Afterwards, MPA has been extended using bibliographical citation data and/or patent citation data to explore the history of fuel cell research [25], to map the emergence, growth, and transformation of medical knowledge [26], to highlight the development of a structural backbone in the field of fullerenes and fullerene-like structures of nanotubes [36], to identify the development trajectories of the data envelopment analysis (DEA) literature [13], and to investigate the knowledge diffusion structure for the field of data quality [37]. Recently, Choi and Park suggested a forward citation node pair (FCNP) algorithm to distinguish patent development paths from a large patent citation network by evaluating the weight of citations among patents [2].

Previous patent citation approaches of MPA and analyzing technology evolution offer some fruitful discussions and findings, but there are two notable limitations that need to be addressed. First, most of these analyses ignore the essential role of patent families when constructing citation networks. Consolidating data into patent families not only helps avoid duplicate data retrieval during a search of patents across patent authorities' databases, but also helps show the geographical focus of the patentee and the patentee's evaluation of the value of the patent [38]. We do not find much literature on MPA using patent families. Although some scholars had realized the importance of including patent family information in the analysis of a patent portfolio of a single applicant, their studies were limited to investigating the members of only one patent family [39, 40]. The use of multiple authorities' patent data bundled with the patent family information can significantly improve the coverage and practicability of patent citation analysis [41]. Second, the complex structure of technological changes at different developmental stages should be interpreted in detail. Various citation networks can be generated, even for the same patent, so different periods will show different technological trajectories. Hence, it is necessary to understand the critical technological progress over different periods to better explore the factors determining how technology plays a role in the process of knowledge diffusion.

To overcome these limitations, we focus our analysis of patent data on search results from the title and abstract fields that pertain to a particular technology of interest, and then take a systematic approach to gauge evolutionary pathways. Our approach mainly takes three forms of intelligence into consideration:

1. To identify the Technology Life Circle (TLC) stage by introducing the indicator of connectivity and modularity in the interior citation network;

2. To take family patents into consideration when building a citation network to more fully present the citation relationships among all prominent patent families;

3. To apply integrated main path analyses to construct a set of technology trajectories and then trace the technology evolution pathways.

This paper is organized as follows, Section 2 describes the framework and main methods applied in our study—it details the process to identify patent evolution pathways; Section 3 discusses our search strategy and how the data are acquired; Section 4 presents an application of the suggested approach—evolution pathways for Dye-Sensitized Solar Cells (DSSCs); Finally, in Section 5, we conclude with the summary, discussion, and further research ideas.

2. Methodological Approaches

In this research, we offer a systematic approach to better exploit patent resources concerning NESTs to identify technology evolution pathways. The framework is illustrated as Figure 1. In this process, we use a professional desktop text mining software- VantagePoint (http://www.theVantage-Point.com)—to help identify the field from raw data and show results through a combination of statistics.

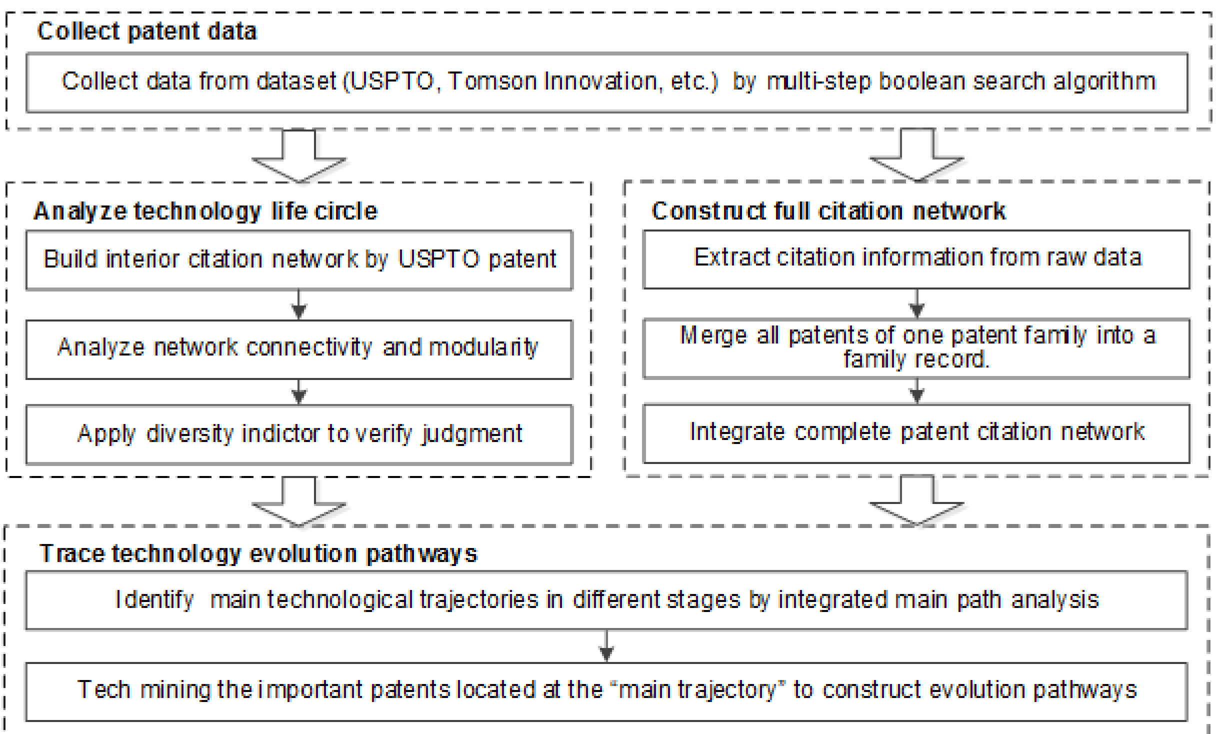

Figure 1. Framework to Explore Technology Evolution

2.1 Technology Life Cycle

Technology often presents different development tracks; therefore, it is necessary to consider the technology life cycle when creating a distinct R&D strategy plan. The technology life cycle comprises a pattern of dynamic characteristics pertaining to technology, in which its innovative and economic outcomes change over time [42].

The concept of the technology life cycle dates back to the 1960s. In 1981, Ford and Ryan (1981) proposed a conceptual standard that allows the base level of technological development to the application level of different technologies be revealed. Also, the consulting firm of Arthur D. Little (1981) developed a technological life cycle model that represents the evolution of the technologies with a system that is similar to the one used to reveal the life cycle of an industry, but it utilizes, on the vertical axis, some qualitative measure of technological changes [43]. Among all the technological life cycle models, the S-curve connects the investment in technology to observe technological performance, either over time or in terms of cumulative R&D expenditures. It is generally accepted that technology life cycle includes two dimensions: (1) competitive impact and (2) integration in products or processes, which can be divided into four stages with different characteristics—emerging, growth, maturity, and decline [44]. Studying technology life cycles is widely used to measure patent activity indices, especially patent applications [45]. Nowadays, more and more researchers tend to introduce multiple indicators to measure the technology life cycle [46-48].

Though such statistical indicators offer a convenient way to make a quick sense of the technological stage, they ignore the technology nature of internal knowledge flow and knowledge overflow. In other words, such traditional indicator-based methods cannot explain the dynamic mechanism of technology evolution and fail to determine inner representation. In this paper, we hold the view that the process of technology evolution can be interpreted through the evolution of patent citation behavior. To some extent, it is similar to the progress of urbanization. The node could be treated as an individual, and the edge could be designated as the community links. At the beginning, the city is a small village, sparsely populated, where hardly anyone has a relationship with others. Thereafter, many people move to this place, and some close community relationships could be observed. Then, more and more people come in and there is more of a community connection, even though a large number of individuals stay isolated; Finally, the urban population keeps increasing to arrive at a relatively stable level, but strong relationships among individuals form a series of communities that can be merged into some larger components. The schematic diagram is presented as Figure 2.

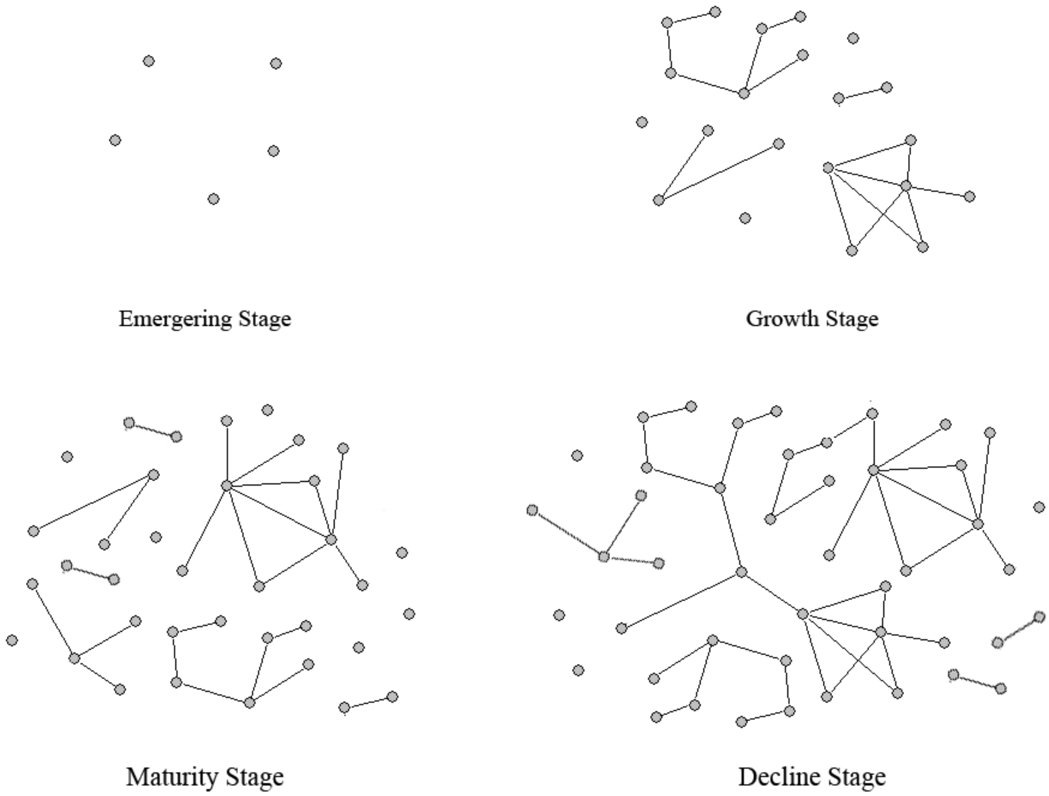

Figure 2. Schematic Diagram of Community Evolution over Different Life Cycle Periods

In order to describe these temporal processes, we introduce the growth rate of connected edges and the growth rate of weakly connected components to observe the technology evolution during the development period. Here, connected edge is defined as citation linkage, and the weakly connected component is deemed as the linkage community, among which all nodes are connected. In general, a certain new technology first appears; the rate of activity increases slowly during the emerging stage, and there are various nodes in the technology field. At the growth stage, the technology develops very fast to form some technology focus, among which the nodes are closely linked. In the technology maturity stage, new patents are filed, typically isolated as independent communities, but the number of such nodes grows rapidly. At the decline stage, technology integration becomes a trend or a pattern; citation linkages also become more frequent and some small components merge into larger technology communities. A summary of the technology life cycle's characteristics in relation to the growth rate of connected edges and weakly connected components is provided in Table 1.

Table 1. Characteristics of the Technology Life Cycle Periods

Life cycle period	Description
Emerging Stage	Growth rate of connected edges (↑), Growth rate of weakly connected components (↑)

Growth Stage	Growth rate of connected edges (↑↑), Growth rate of weakly connected components (↑)
Maturity Stage	Growth rate of connected edges (↑), Growth rate of weakly connected components (↑↑)
Decline Stage	Growth rate of connected edges (↑↑), Growth rate of weakly connected components (↓)

Furthermore, this paper introduces technology specialization to further verify our initial judgment of the technology life cycle of a target technology. We adapt the calculation of technology specialization from the Rao–Stirling diversity, which combines two of the three aspects of interdisciplinarity distinguished by Rafols and Meyer—variety and disparity [49]. Technology specialization (S) is defined as follows, where $COS(IP_i, IP_j)$ is a similarity measure between two classes *i* and *j*—the categories, in this case, are IPC classes at the four-digit level—and IP_i and IP_j is the proportion of elements assigned to each class *i* or *j*. In this study, we use the cosine values between the vectors of the 630 IPC classes at the four-digit level [50].

$$\mathrm{S} = \sum_{ij} IP_i * IP_j * COS(IP_i, IP_j)$$

2.2 Citation Network Construction

As we mentioned above, patent citation analysis approaches are actively applied to structure a large number of patents, to profile the patent landscape, and to capture knowledge transfer and change in technologies or industries. Previous researchers often analyze patent data of a single authority because of the availability of the data and the simplicity of the analysis. A patent family is the collection of patents in different countries referring to the same technical topic, so it can be a useful information source because duplicate data retrieval can be avoided during a search of patents across patent authorities [51]. Citation behavior is different among patent authorities and between parent and child patents, so global technology trends cannot be understood only with the analysis of patent data issued by a single authority. For the sake of statistics, in this paper, an important step is to merge patent documents of a family into a single family record. Family patents are usually identified by the claim of priority or disclosure, and here, the certain patent family is marked by the earliest published patent in this family. Meanwhile, all cited patents of a family's members are merged to form the cited patents of the family record. In regard to patent families, we mainly consider three types of linkage between patents, shown as Figure 3.

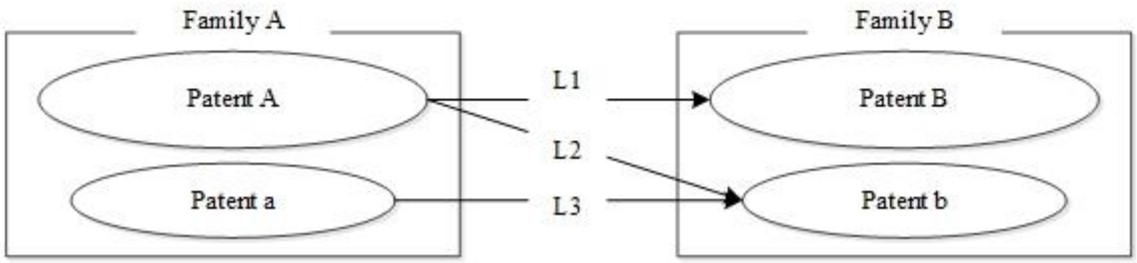

Figure 3. Three Types of Citations Between Patents in a Patent Family [41]

Note: Patent A or Patent B stand for a parent patent, or a patent of a certain authority; Patent a or Patent b represent a child patent of A and B respectively, or the equivalent patent in other patent authorities.

In Figure 3, patent B can cite patent A by direct citation L1. Patent A can also make reference to citation L2, which is a citation to another member of the family to which Patent B belongs. Citation L3 is another citation pattern between patent families including patents A and B. Patent A can be related to patent B with a citation among child patents a and b. Most previous citation analysis research focused on L1 citations, but in this paper, we address all three types of citations to capture the comprehensive structure within and between inventions.

A general directed network (also called a Bayesian Network) consists of vertices and arcs that link two vertices (nodes). A citation network is a standard directed network that can also be represented as a citation matrix. Its columns and rows stand for the nodes, and each value in the matrix is defined as the strength of citation between two nodes. While conducting MPA for a given field of technology based on the patent citation network, we are more concerned about citations between patents within the technology field. These effective citations are extracted from the merged family records.

A patent citation network can be represented as a patent citation matrix. Nodes stand for the individual family records, and arcs between two nodes are citations. The patent (actually meaning family) citation matrix is defined with these nodes and arcs as follows:

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{n2} & \cdots & p_{nn} \end{bmatrix}$$

Where

$$p_{ij} = \begin{cases} 1, & if\ patent\ j\ cites\ patent\ i; \\ 0, & otherwise. \end{cases}$$

The patent citation matrix P is an $n \times n$ square matrix, which represents an acyclic citation network with four types of vertices corresponding to four types of patent family records, respectively: (1) isolated records—the patent families that cite no other patents, and are not cited; (2) source records—the patent families that cite no patents, but are cited; (3) sink records—the patent families that cite others, but are not cited by others; and (4) intermediate records—the patent families that cite other patent families, and are also cited by other patents.

To simplify the process of analyzing the technological trajectories and identifying core patents, we disregard patents neither citing nor cited by at least one patent in the maximum connected component – namely, the “orphan” patents that are digressional from the mainstream of those technological domains are eliminated.

2.3 Main Path Analysis

The main path is a path from a source vertex to a sink vertex with the highest traversal weights on its arcs [52]. MPA is applied to explore the most significant paths in a citation network and is commonly used to trace the developmental trajectory of a research field by using bibliographical citation data and/or patent citation data. In general, there are two factors that should be considered when conducting MPA: (1) calculating the node weight and (2) choosing the path.

How to measure the weights of each citation link from a set of starting vertices to the ending vertex is an important step in MPA. Several indices have been proposed, and the most widespread algorithms are Node Pair Projection Count (NPCC), Search Path Link Count (SPLC), and Search Path Nodes Pair (SPNP), which were proposed by Hummon and Doreian [32]. In 2003, Batagelj proposed a new traversal count, namely the Search Path Count (SPC), concluding that SPC performs a bit better than SPLC and SPNP because of its properties, even though these indices always obtain almost the same results. However, subtle differences exist among them. In this paper, we do not elaborate on the pros and cons of applying each of the traversal counts, but follow the recommendations and apply SPC throughout to count the weight of each citation link.

Based on the previous phases, technology evolution pathways can be finally constructed by identifying the important patents located on the main trajectory at different stages. After getting the SPC weight of each node, we need to choose an algorithm to figure out the main path. Most of the traditionally proposed algorithms represent a “local” approach, which repeatedly chooses the link with the largest traversal count emanating from the current starting node. Such local algorithms highlight significance at a particular point in time and track the most significant citation link at every possible splitting point, whereas the global algorithm emphasizes the overall importance and delivers the path with the largest overall traversal count [53]. In other words, in contrast to the local main path that highlights significance in local progress, the global main path emphasizes the overall importance in knowledge flow [54] . The approaches of main path analysis can be presented as Figure 4.

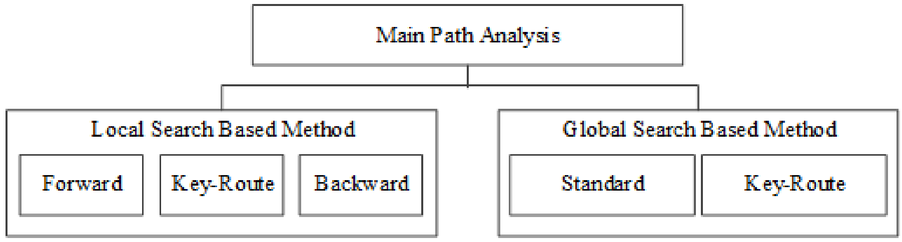

Figure 4. Main Methods in MPA.

Nevertheless, both previous local search based method (Forward and Backward) and the global search based method (standard) may miss the links with the largest traversal count. Liu and colleagues introduced a new method called the *key-route* to enhance MPA by adding an algorithm to search for multiple paths and guarantee inclusion of the top links in these multiple paths. This approach viewed a main path as an extension of a specific key route, and began a search from both ends of that key route [13, 54].

In this paper, we apply the global MPA method for different periods to obtain the path that has the most significant overall traversal count in different technological stages. Furthermore, we use the multiple global key-route method to track the technology flow and evolution from a holistic perspective. The procedure of the multiple global key-route method is as follows: (1) select the link that has the highest traversal count as the key-route; (2) utilize standard search to trace the nodes that have the largest overall traversal counts. The multiple global key-route method not only provides multiple paths, from which we can find the knowledge diffusion trajectory comprehensively, but also contains almost all the important connections and makes the results much more comprehensive.

3. Search Strategy and Data Retrieval

Previous citation analysis approaches often chose data from the United States Patent and Trademark Office (USPTO) because data there are rich in citations and are easy to obtain. However, using patents of a single authority may ignore important characteristics of patenting systems and cause a critical oversight of information on patent application status and R&D trends of technology [41]. In this paper, we use the Thomson Innovation patent compilation provided by Thomson Reuters. It offers a comprehensive worldwide patent database system, which covers patents recorded at more than 80 patent authorities, including the USPTO. Thomson Innovation includes Derwent World Patents Index (DWPI) patent data and the Derwent Patent Citation Index (DPCI), by which we can collect data on patent families and citations. While the claim of priority or disclosure as a member of a certain patent family is not mandatory, DWPI defines a family based on both the claim and the investigation of experts. The DWPI bundles patents recorded at 47 worldwide patent authorities as a protection for the same invention as a sort of family [41].

Dye-sensitized solar cells (DSSCs) are an important nanoscience domain contributing to photovoltaic technology development. In contrast to conventional silicon-based solar cells, the demand on purity of materials for DSSCs is lower and forecasted manufacturing costs are approximately halved, making DSSCs an attractive alternative [55]. We have analyzed DSSCs for several years [56-61]. Based on our experience, the seminal paper on DSSCs appeared in 1991 [62]. Therefore, in this paper, we chose the time span from 1991 to 2014. As the quality of the retrieved patent data is important for our analysis and the meaningfulness of the generated results, we selected the terms for Boolean searching with guidance of domain experts. The search strategy was:

TS=((Dye* or Pigment*) and (Sensiti*) and (Solar* or Photovoltaic*) and (Cell* or Batter*))

The total number of patent families obtained from Thomson Innovation was 6,857 (search results updated on February 17, 2016).

4. Results

In this section, we first identify the technology life cycle of DSSCs from 1991 to 2014. In order to better fit the true development level of DSSCs and decrease the influence of commercial activity, we prefer to use raw data from USPTO that provide integral citation information. Thus, we extract 653 USPTO granted patents from our initial dataset to conduct the technology cycle life analysis.

After constructing the citation network by year, we can obtain the information on connected edges and weakly connected components with the help of the interactive visualization and exploration platform, Gephi (http://gephi.org/). We visualized the indicator information and fit these dots in MATLAB (http://www.mathworks.com/), a multi-paradigm numerical computing environment. The results are shown in Figure 5. Before the early 21st century, the granted DSSCs patents were in slow-growth, and most of these patents had rare citations to each other. After 2002, the linkages among DSSCs patents became much more frequent than before, and some technology communities were gradually formed. Upon entering the 2010s, new technology communities grew at a rapid speed; in other words, the new and emerging technology focuses are in the course of forming and developing. After repeated testing, the growth trend of connected edges and weakly connected components can fit three curves in three phases. Based on the analysis, we divide the DSSCs development into three stages: emerging (1991–2001), growth (2002–2011), and maturity (2012–2014).

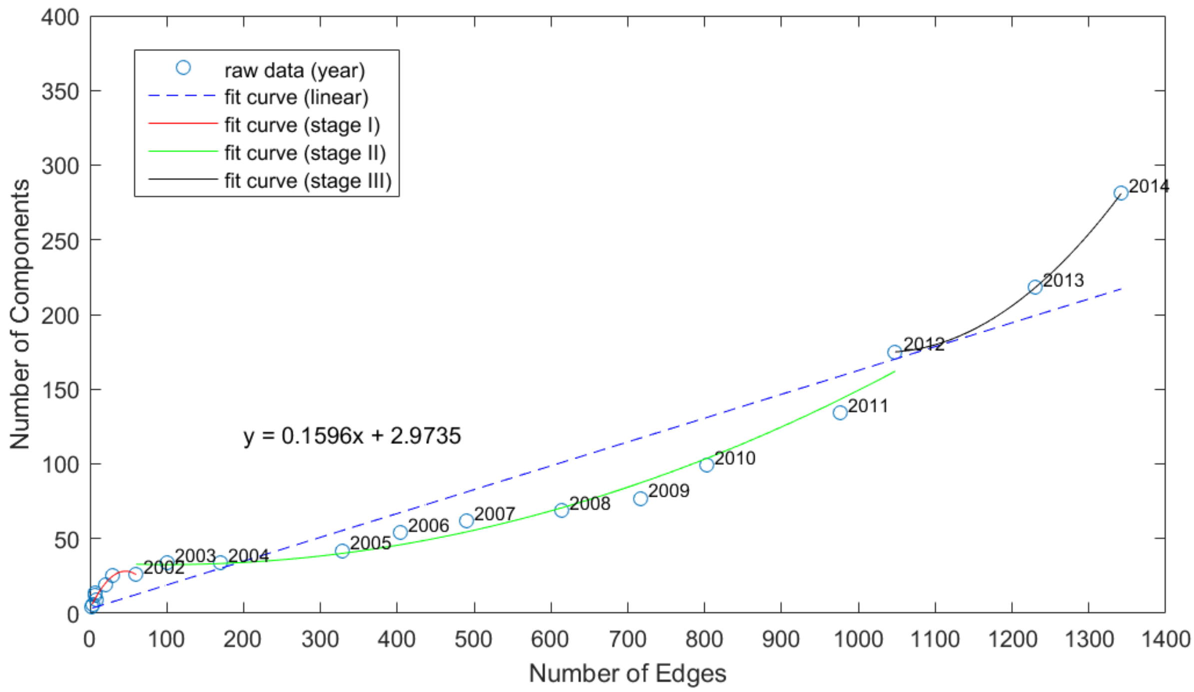

Figure 5. Curve Fitting of Two Indicators in the DSSCs Patent Citation Network

Some scholars reported cyclical patterns in the longitudinal development of the Rao–Stirling diversity in nine material technologies for photovoltaic cells and suggested these cyclical patterns can be used to indicate technological life-cycles [50]. As known to us, in different stages, different technology topics will be targeted, which can be indicated by tracing the changes of International Patent Classifications (IPCs). Therefore, we attempt to introduce the specificity indicator (Specialization score) in patent classifications to indicate the technology life cycle. Figure 6 shows the trends of Specialization scores for 4-digit IPCs of DSSCs patents during the period 1991–2014. This figure suggests that DSSCs technology topics show obvious instability before 2002. After 2003, the specificity indicators show relatively stable increases (except for 2006) and reach their peak in 2012, which tells us that the technology focus became more concentrated on specific sub-topics. Generally, in early phases of a technology, the number of patents is small, and the development of the IPC variety fluctuates. Whereas inventors tend to participate in constructing a research front, assignees can be considered primarily as economic agents who follow another logic than the technology cycle per se [63]. Based on such analyses, the technology stage division basically meets our initial judgment, and the approach we propose offers a reasonable way to set up further MPAs in different time periods. We separately showed our results to two experts who are working for the Georgia Institute of Technology. In their opinion, the results basically fit their understanding of the development of DSSCs.

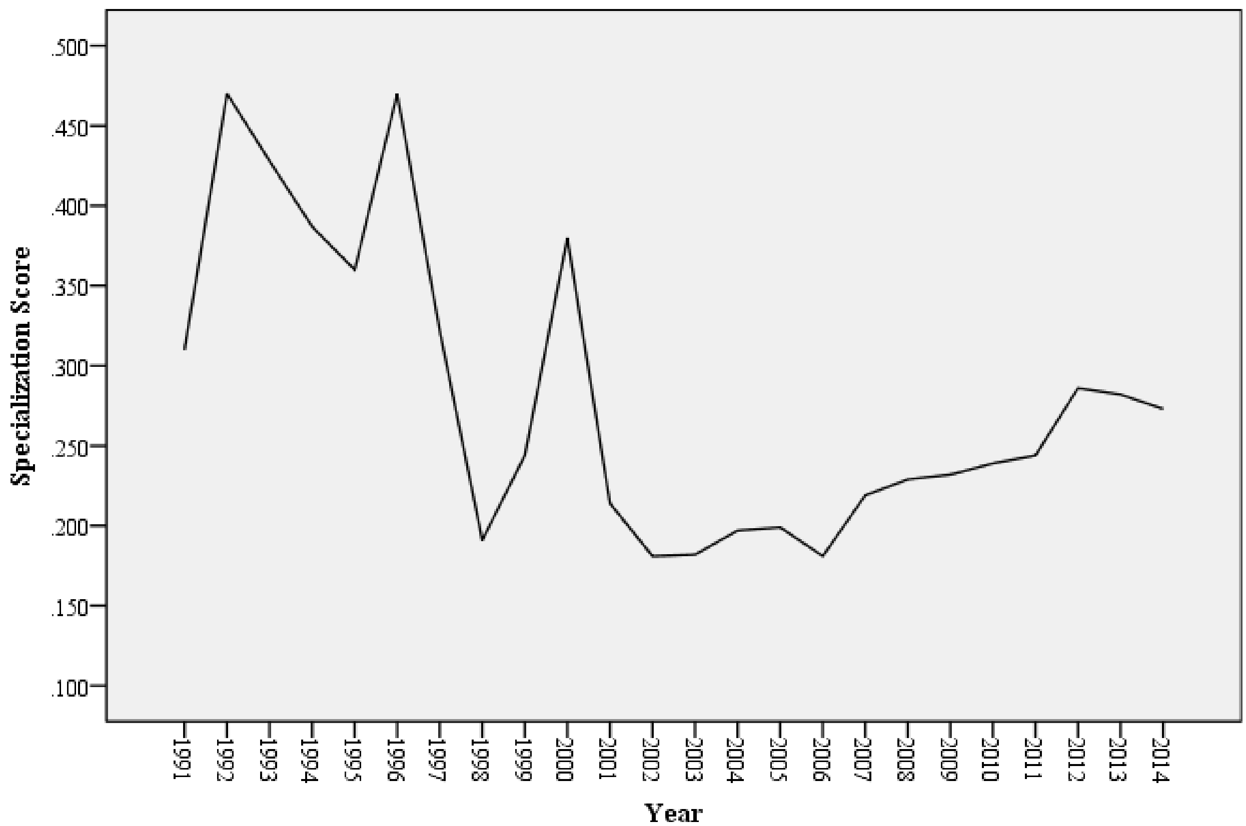

Figure 6. Specialization Scores for DSSCs IPCs by Year

Instead of conducting MPA for every technology life cycle stage, we inspect main paths extracted from the year 1991, which is the beginning of our investigated interval of DSSCs development, to the end of the emerging stage of DSSCs' technology life cycle stage (2001), the end of growth (2011), and the end of our investigated interval (2014). As mentioned earlier, we conduct global MPA in these three periods based on the commonly adopted SPC weighting scheme (for its performance and properties). The extracted main path was drawn with Gephi software and shown in Figure 7. We mark them as $Path_{2001}$, $Path_{2011}$ and $Path_{2014}$ for the convenience of description. The thickness of an arc in a main path indicates the SPC value of the citation it represents; specifically, the thicker arc indicates a higher SPC value and thinner indicates a lower SPC value.

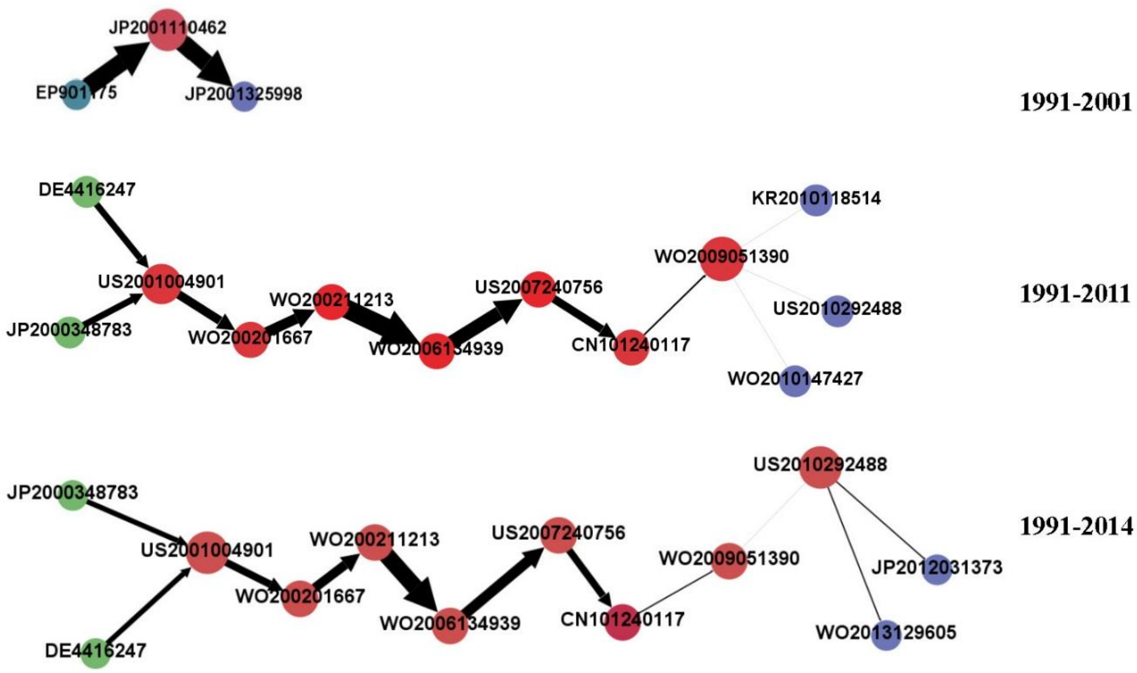

Figure 7. Global Main Paths at Three Periods

As shown in Figure 7, $Path_{2001}$ contains only three patents and two citation pairs. Compared to $Path_{2011}$ and $Path_{2014}$, we find that neither patent nodes nor citation pairs in $Path_{2001}$ remain in $Path_{2011}$ and $Path_{2014}$. One of the reasons is that the amount of patents and citations in the emerging period (1991–2011) is very small (only 131 and 17, respectively). In contrast, there is a significant overlap between $Path_{2011}$ and $Path_{2014}$. Only two patents authorized/applied in 2010 (WO2010147427, KR2010118514) are skipped from $Path_{2011}$ to $Path_{2014}$ and two new patents authorized/applied for in 2013 (WO2013129605, JP2012031373) are appended to $Path_{2014}$. From the above observations, we conclude that a significant overlap exists between global main paths of the overlapping periods.

Table 2. Profiling Information for Patents Located along the DSSCs Technological Trajectory

Patent Number	Publication Year	Technology Focus	Technology life cycle stage	Main IPCs
DE 4416247	1995	Monolithic series-connected dye-sensitized photovoltaic module	G-1; M1	H01L-031/04 (Semiconductor devices adapted as conversion devices)
EP 901175	1999	Photoelectric conversion device for solar cell	E-1	H01L-051/20 (Devices)
JP 2000348783	2000	Pigment sensitization type solar battery manufacture; Space surrounding	G-1; M1	H01M-014/00 (Electrochemical current or voltage generators not provided for in groups)

		semiconductor and counter electrodes		
US 2001004901	2001	Dye sensitizing solar cell having redox electrolyte sealed with vitreous material	G-2; M2	H01L-031/00 (Semiconductor devices)
JP 2001110462	2001	Optoelectric transducer for solar battery	E-2;	H01M-014/00 (Electrochemical current or voltage generators not provided for in groups)
WO 200201667	2002	Photoelectric transducer comprises oxide semi-conductor particles	G-3; M-3	H01M-014/00 (Electrochemical current or voltage generators not provided for in groups)
JP 2001325998	2001	Pigment-sensitizing type solar battery manufacturing method; metallic oxide film	E-3	H01M-014/00 (Electrochemical current or voltage generators not provided for in groups)
WO 200211213	2002	Dye-sensitized photoelectric transducer is composed of fine oxide semiconductor particles	G-4; M-4	H01L-031/00 (Semiconductor devices)
WO 2006134939	2006	Optoelectric transducer is formed using oxide semiconductor microparticles	G-5; M-5	H01M-014/00 (Electrochemical current or voltage generators)
US 2007240756	2007	Fluorenyl-containing compound dyes	G-6; M-6	H01L-051/00 (Solid state devices using organic materials as the active part)
CN 101240117	2008	Pure organic dye	G-7; M-7	C09B-057/00 (Other synthetic dyes of known constitution)
WO 2009051390	2009	Thiophene-based dye	G-8; M-8	C09B-049/00 (Sulfur dyes)
US 2010292488	2010	Organic dye	G-9; M-8	C07D-495/04 (Ortho-condensed systems)
WO 2010147427	2010	New organic dye	G-9;	C09B-049/00 (Sulfur dyes)
KR 2010118514	2010	New organic dye	G-9;	C09B-011/00 (Diaryl- or triarylmethane dyes;)
JP 2012031373	2012	Optical functional material is used for sensitizing dye	M-9	C09B-023/00 (Methine or polymethine dyes)
WO 2013129605	2013	New organic dye compound	M-9	C09B-023/00 (Methine or polymethine dyes)

Note: The technology life cycle stages -- E, G, and M -- stand for the stage of emerging, growth, and maturity, respectively. The number behind the caption means the order in the main path of the corresponding stage.

Table 2 provides profiling information for patents located along the technological trajectory of DSSCs represented by citation main paths $Path_{2001}$, $Path_{2011}$, and $Path_{2014}$. An additional observation is that technological focus indicated by the main IPC numbers of patents drifts along with citation flows in the patent citation network. Take the citation pairs <WO 200211213, WO 2006134939> and <WO 2006134939, US 2007240756> in $Path_{2011}$ and $Path_{2014}$ as an example; the main IPC number of patent "WO 200211213" is *H01L-031/00* (Semiconductor devices) while the main IPC of "WO 2006134939" is *H01M-014/00* (Electrochemical current or voltage generators), and the main IPC of "US 2007240756" is *H01L-051/00* (Solid state devices using organic materials as the active part). Since citations indicate direct relevance between citing and cited patents, it is pretty interesting to find topic drifts in a technological trajectory.

In order to identify other critical R&D directions in the field, we use a global key-route MPA. The key-route approach allows us to examine different levels of details by selecting the number of important links [64]. We explore different sequential numbers from the top 10 to top 50 to seek a reasonable threshold to identify paths exhibiting the greatest weight values in the patent citation network. Finally, we figure out that when the threshold value is set to the top 20 links, a divergent-convergent pattern is clearer than others. We therefore determine to analyze the key-route main paths based on these top 20 links, and the results are shown in Figure 8.

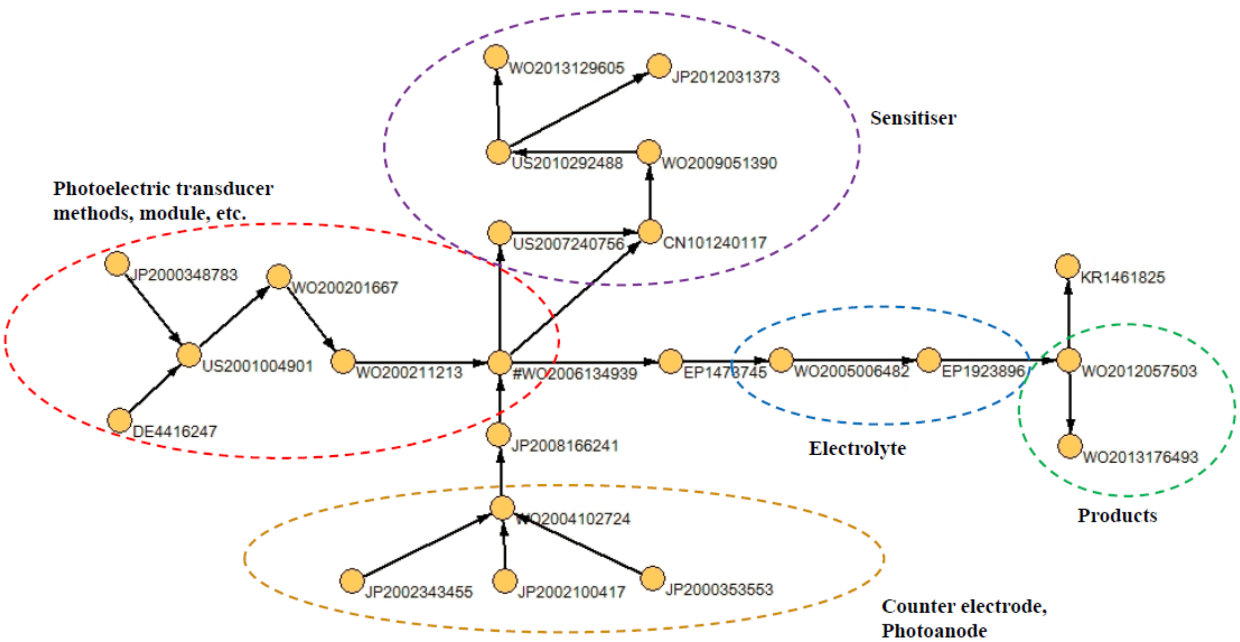

Figure 8. Key-route Main Path of DSSCs from 1991–2014

To better analyze technology focus and evolution, we extract topic information from "Title+Abstract" of the patents located on the key-routes. The key-routes present fruitful technology diffusion and technological topic evolution. Along the main path, most of the DSSCs patents in the early stages primarily discuss the photoelectric transducer and photovoltaic cell module, including DE4416247, JP2000348783, US2001004901, WO200201667 and WO200211213. Thereafter, counter electrode-related and photoanode-related technologies draw wide attention, and key patents in this stage are P2000353553, JP2002100417, JP2002343455, and WO2004102724. WO2006134939, a patent published in 2006, proposed an optoelectric transducer formed using oxide semiconductor microparticles sensitized with a methane type pigment; it plays a vital role in the citation network. Afterwards, one subgroup aims to introduce different improved dyes for DSSCs, especially organics. Such patents include US2007240756, CN101240117, WO2009051390, US2010292488, JP2012031373, and WO2013129605. The other subgroup focuses on developing electrolytes, mainly WO2005006482 and EP1923896. Recently, more and more DSSCs patents are committed to produce environmentally-friendly solar cells to advance the solar energy and energy storage industry.

Are the patents located at the technological trajectory all highly cited by others patents? The answer is *no*. In general, the number of citations a patent receives can be used to show its impact on a field of technology, but an analysis based on citation counts may fail to identify those concepts and principles that could act as "focusing devices" for a sequence of inventive activities [31]. The information shown in Table 3 supports this conclusion. In the process of identifying the technological trajectory, patents cited by the other patents of our targeted data set are highly weighted. Also, those that cite more DSSCs patents are more likely to be selected as nodes located on the main technological trajectory.

Table 3. Network Information of DSSCs Patents Located on the Key-route Main Path

Patent Number	Publication Year	Citation	Times Cited	In-degree	Out-degree	Closeness Centrality	Betweenness Centrality
DE4416247	1995	11	84	0	18	4.179	0.000
JP2000348783	2000	0	21	0	16	4.128	0.000
JP2000353553	2000	0	8	0	3	5.286	0.000
US2001004901	2001	19	0	2	25	3.732	675.667
JP2002100417	2002	0	5	0	2	5.324	0.000
JP2002343455	2002	0	4	0	5	4.936	0.000
WO200201667	2002	30	10	2	7	3.723	585.500
WO200211213	2002	35	22	4	15	3.215	1217.117
EP1473745	2004	31	3	2	13	2.349	1557.583
WO2004102724	2004	13	11	3	5	4.359	721.500
WO2005006482	2005	49	26	2	13	1.805	1096.417
WO2006134939	2006	112	5	14	3	3.865	158.317
US2007240756	2007	28	20	5	19	2.097	907.400
CN101240117	2008	7	21	2	18	1.650	785.500
EP1923896	2008	13	23	4	15	1.211	679.533
JP2008166241	2008	4	6	1	4	3.462	934.500
WO2009051390	2009	20	14	2	4	1.556	127.833
US2010292488	2010	5	16	8	2	1.000	63.000
JP2012031373	2012	8	0	5	0	0.000	0.000
WO2012057503	2012	5	3	4	2	1.000	117.000
WO2013129605	2013	5	0	2	0	0.000	0.000

WO2013176493	2013	9	0	2	0	0.000	0.000
KR1461825	2014	2	0	2	0	0.000	0.000

We examined correlations among some important indicators in social network analysis, shown as Table 4. According to social network theory, the average degree of a node is the number of edges connected to it, and it closely relates to the density of a network, including in-degree (citing other nodes) and out-degree (cited by other nodes). From Table 4, we can figure out citation correlates to in-degree, also with times cited, and out-degree. Betweenness, as a core measure of the centrality of a node in a network, is normally calculated as the fraction of shortest paths between node pairs that pass through the node of interest. Therefore, betweenness, to some extent, is a measure of the influence a node has over the spread of information through the network. According to correlation analyses, betweenness-centrality is closely related to out-degree. In other words, betweenness-centrality, in some sense, correlates with inter-citing activities in an established field or network.

Table 4. Correlations of Some Social Network Indicators

		Citation	Times Cited	In-degree	Out-degree	Closeness Centrality	Betweenness Centrality
Citation	Pearson Correlation	1					
	Sig. (1-tailed)						
Times Cited	Pearson Correlation	0.014	1				
	Sig. (1-tailed)	0.474					
In-degree	Pearson Correlation	.727**	-0.156	1			
	Sig. (1-tailed)	0	0.239				
Out-degree	Pearson Correlation	0.117	.499**	-0.157	1		
	Sig. (1-tailed)	0.297	0.008	0.238			
Closeness Centrality	Pearson Correlation	0.097	0.213	-0.232	0.251	1	
	Sig. (1-tailed)	0.329	0.165	0.143	0.124		
Betweenness Centrality	Pearson Correlation	0.301	0.038	0.013	.547**	0.052	1
	Sig. (1-tailed)	0.081	0.432	0.476	0.003	0.406	

Note: ** indicates correlation is significant at the 0.01 level (1-tailed).

5. Conclusions and Discussion

Tracing technology evolution pathways is essential to track innovation progress, and MPA is one of the effective approaches to identify key technological trajectories within complex patent citation networks. Most previous studies ignore the essential role of patent families when constructing the citation network, which leads to a decrease of coverage and practicability of the analyses. Also, different temporal intervals may affect citation relations, so it is vital to understand the critical technological progress over different periods to better explore knowledge diffusion. Therefore, we introduce technology community evolution and patent diversity indicators to gauge the technology stage over time; and then we merge family patents to build a more comprehensive citation network. Lastly we introduce global MPA and global key-routes MPA to identify a set of main technology trajectories and then trace the technology evolution pathways accordingly.

Based on these analyses we derive several ideas. First, observing technology changes in different stages can help understand the mainstream track and key technologies. Static technical evolution is only well suited for mature technologies or emerging technologies that are in an extended, stable developmental stage. Some patents play a vital role in a certain stage and obtain a remarkable weighting in the patent citation network, but they may lose their advantages in the process of identifying main patents and focus in a longer period. Second, taking patent family into consideration has a remarkably positive effect on constructing the patent citation network and identifying main patents. The performance of centrality, connectivity, and modularity in the citation network of considering family patents are better than the one ignoring family patents. Such good network attributes are beneficial for discovering critical nodes in MPA. Third, patent citation analysis is a useful method to trace technology development. Applying MPA to a citation network simplifies a complicated citation network to a small number of nodes and links. The identified patents located on the main path may prove helpful for decision makers in the field. This method generates technological intelligence, which serves to elucidate technological change processes. We believe it can facilitate the identification of innovation opportunities (i.e., prospective paths to commercialization along with target developmental priorities to attain them).

Furthermore, this study offers some technology management insights for practical applications. On one hand, the main path analysis with the multiple key-route approach is an effective tool to trace technological changes. Based on the assumption that the technological changes are embedded in the governing structure of the knowledge diffusion paths, technology development can be observed by detecting the evolution paradigm and the stories of technological changes speak for themselves. On the other hand, technological emergence, technological convergence, and technological diffusion always occur as a series of evolutionary, variant changes that are gradually combined or fused together to open the industry to successive dominant designs or guideposts. If inventors can publish equivalent patents of an invention in different countries, the multiple authorities' patent data should be analyzed to better grasp the technology development venation. Therefore, patent family information can significantly improve the coverage and practicability of patent citation analysis and also indicate the commercial potential and market distribution in the near future.

Based on current research, future studies could be improved in at least these two fields. First, we applied SPC algorithms to calculate the weight of the vertices in the network. Such algorithms (also including NPCC, SPLC and SPNP) only work on binary citation networks; in other words, all citations between any citing–cited pair are treated the same. The current advances in text mining, especially semantic analysis, can be used to scale the relevancies between any citing-cited pair of patents or publications and then further turn the traditional binary citation networks into weighted networks. Second, in a complex citation network, the method to find main paths seems to achieve the goal of simplifying the citation network and looking for the most significant development path. But in fact, on one hand, the obtained significant route may not be the path with the largest overall impact, even though some vital nodes can be observed; on the other hand, it is also important to identify the potential nodes that will make an enormous impact in the future but are neglected in the process under the evaluation system of citation.

Acknowledgements

This research is mainly undertaken at Georgia Institute of Technology, drawing on support from the United States National Science Foundation (NSF) (Award No.1064146), MOE (Ministry of Education in China) Project of Humanities and Social Sciences (Award No.13YJC630042) and the National High Technology Research and Development Program of China (Grant No.2014AA015105). Besides, we are grateful for the scholarship provided by the China Scholarship Council (CSC Student ID 201406030005). The findings and observations contained in this paper are those of the authors and do not necessarily reflect the views of the supporters.

References

[1] T.S. Cho, H.Y. Shih, Patent citation network analysis of core and emerging technologies in Taiwan: 1997—2008, *Scientometrics*, 89 (2011) 795–811.

[2] C. Choi, Y. Park, Monitoring the organic structure of technology based on the patent development paths, *Technological Forecasting and Social Change*, 76 (2009) 754–768.

[3] J. Yoon, K. Kim, Identifying rapidly evolving technological trends for R&D planning using SAO-based semantic patent networks, *Scientometrics*, 88 (2011) 213–228.

[4] P.L. Chang, C.C. Wu, H.J. Leu, Using patent analyses to monitor the technological trends in an emerging field of technology: a case of carbon nanotube field emission display, *Scientometrics*, 82 (2010) 5–19.

[5] D.Z. Chen, H.W. Chang, M.H. Huang, F.C. Fu, Core technologies and key industries in Taiwan from 1978 to 2002: A perspective from patent analysis, *Scientometrics*, 64 (2005) 31–53.

[6] P. Ganguli, Patents and patent information in 1979 and 2004: A perspective from India, *World Patent Information*, 26 (2004) 61–62.

[7] I.D. Lacasa, H. Grupp, U. Schmoch, Tracing technological change over long periods in Germany in chemicals using patent statistics, *Scientometrics*, 57 (2003) 175–195.

[8] I. von Wartburg, T. Teichert, K. Rost, Inventive progress measured by multi-stage patent citation analysis, *Research Policy*, 34 (2005) 1591–1607.

[9] M. Bengisu, Critical and emerging technologies in Materials, Manufacturing, and Industrial Engineering: A study for priority setting, *Scientometrics*, 58 (2003) 473–487.

[10] D. Harhoff, F.M. Scherer, K. Vopel, Citations, family size, opposition and the value of patent rights, *Research Policy*, 32 (2003) 1343–1363.

[11] J. Shin, Y. Park, Building the national ICT frontier: The case of Korea, *Inf Econ Policy*, 19 (2007) 249–277.

[12] B. Yoon, Y. Park, A text-mining-based patent network: Analytical tool for high-technology trend, *The Journal of High Technology Management Research*, 15 (2004) 37–50.

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

[14] Y. Kajikawa, Y. Takeda, Citation network analysis of organic LEDs, *Technological Forecasting and Social Change*, 76 (2009) 1115–1123.

[15] O. Sorenson, J.W. Rivkin, L. Fleming, Complexity, networks and knowledge flow, *Research Policy*, 35 (2006) 994–1017.

[16] E. Duguet, M. MacGarvie, How well do patent citations measure flows of technology? Evidence from French innovation surveys, *Economics of Innovation and New Technology*, 14 (2005) 375–393.

[17] A. Breitzman, P. Thomas, Using patent citation analysis to target/value M&A candidates, *Research-Technology Management*, 45 (2002) 28–36.

[18] F. Murray, Innovation as co-evolution of scientific and technological networks: exploring tissue engineering, *Research Policy*, 31 (2002) 1389–1403.

[19] L. Fleming, O. Sorenson, Science as a map in technological search, *Strategic Management Journal*, 25 (2004) 909–928.

[20] M.S. Meyer, Patent citation analysis in a novel field of technology: An exploration of nano-science and nano-technology, *Scientometrics*, 51 (2001) 163–183.

[21] N. Shibata, Y. Kajikawa, I. Sakata, Extracting the commercialization gap between science and technology—Case study of a solar cell, *Technological Forecasting and Social Change*, 77 (2010) 1147–1155.

[22] R.J.W. Tijssen, Global and domestic utilization of industrial relevant science: patent citation analysis of science-technology interactions and knowledge flows, *Research Policy*, 30 (2001) 35–54.

[23] G.S. McMillan, F. Narin, D.L. Deeds, An analysis of the critical role of public science in innovation: the case of biotechnology, *Research Policy*, 29 (2000) 1–8.

[24] Y. Kajikawa, J. Yoshikawa, Y. Takeda, K. Matsushima, Tracking emerging technologies in energy research: Toward a roadmap for sustainable energy, *Technological Forecasting and Social Change*, 75 (2008) 771–782.

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

[26] A. Mina, R. Ramlogan, G. Tampubolon, J.S. Metcalfe, Mapping evolutionary trajectories: Applications to the growth and transformation of medical knowledge, *Research Policy*, 36 (2007) 789–806.

[27] J.C. Ho, E.C. Saw, L.Y.Y. Lu, J.S. Liu, Technological barriers and research trends in fuel cell technologies: A citation network analysis, *Technological Forecasting and Social Change*, 82 (2014) 66–79.

[28] D. Alberts, C. Yang, D. Fobare-DePonio, K. Koubek, S. Robins, M. Rodgers, E. Simmons, D. DeMarco, Introduction to Patent Searching: Practical experience and requirements for searching the patent space, in: M. Lupu, K. Mayer, J. Tait, A.J. Trippe (Eds.) *Current Challenges in Patent Information Retrieval*, Springer Berlin Heidelberg, 2011, pp. 3–43.

[29] A.B. Jaffe, M. Trajtenberg, Patents, citations, and innovations: A window on the knowledge economy, MIT Press, 2002.

[30] P. Erdi, K. Makovi, Z. Somogyvari, K. Strandburg, J. Tobochnik, P. Volf, L. Zalanyi, Prediction of emerging technologies based on analysis of the US patent citation network, *Scientometrics*, 95 (2013) 225–242.

[31] R. Fontana, A. Nuvolari, B. Verspagen, Mapping technological trajectories as patent citation networks. An application to data communication standards, *Economics of Innovation and New Technology*, 18 (2009) 311–336.

[32] N.P. Hummon, P. Doreian, Connectivity in a citation network: the development of DNA theory, *Social Networks*, 11 (1989) 39–63.

[33] N.P. Hummon, P. Doreian, L.C. Freeman, N.P. Hummon, Analyzing the structure of the centrality-productivity literature created between 1948 and 1979, *Science Communication*, 11 (1990) 459–480.

[34] N.P. Hummon, K. Carley, Social networks as normal science, *Social Networks*, 15 (1993) 71–106.

[35] K.M. Carley, N.P. Hummon, M. Harty, Scientific Influence: An Analysis of the Main Path Structure in the Journal of Conflict Resolution, *Science Communication*, 14 (1993) 417–447.

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

[37] Y. Xiao, L.Y.Y. Lu, J.S. Liu, Z. Zhou, Knowledge diffusion path analysis of data quality literature: A main path analysis, *Journal of Informetrics*, 8 (2014) 594–605.

[38] E.S. Simmons, "Black sheep" in the patent family, *World Patent Information*, 31 (2009) 11–18.

[39] F.P. Su, W.G. Yang, K.K. Lai, A heuristic procedure to identify the most valuable chain of patent priority network, *Technological Forecasting & Social Change*, 78 (2011) 319–331.

[40] S.F. Pei, K.K. Lai, R.R.K. Sharma, K.T. Hsien, Patent priority network: Linking patent portfolio to strategic goals, *Journal of the American Society for Information Science & Technology*, 60 (2009) 2353–2361.

[41] H. Nakamura, S. Suzuki, Y. Kajikawa, M. Osawa, The effect of patent family information in patent citation network analysis: a comparative case study in the drivetrain domain, *Scientometrics*, 104 (2015) 1–16.

[42] J. Park, E. Heo, Patent quality determinants based on technology life cycle with special reference to solar-cell technology field, *Maejo International Journal of Science and Technology*, 7 (2013) 315–328.

[43] A.D. Little, The strategic management of technology, Arthur D. Little, 1981.

[44] D. Miller, P.H. Friesen, A longitudinal-study of the corporate life-cycle, *Management Science*, 30 (1984) 1161–1183.

[45] H. Ernst, The use of patent data for technological forecasting: the diffusion of CNC-technology in the machine tool industry, *Small Business Economics*, 9 (1997) 361–381.

[46] R.J. Watts, A.L. Porter, Innovation forecasting, *Technological Forecasting and Social Change*, 56 (1997) 25–47.

[47] R. Haupt, M. Kloyer, M. Lange, Patent indicators for the technology life cycle development, *Research Policy*, 36 (2007) 387–398.

[48] L.D. Gao, A.L. Porter, J. Wang, S. Fang, X. Zhang, T.T. Ma, W.P. Wang, L. Huang, Technology life cycle analysis method based on patent documents, *Technological Forecasting and Social Change*, 80 (2013) 398–407.

[49] I. Rafols, M. Meyer, Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience, *Scientometrics*, 82 (2010) 263–287.

[50] L. Leydesdorff, F. Alkemade, G. Heimeriks, R. Hoekstra, Patents as instruments for exploring innovation dynamics: geographic and technological perspectives on "photovoltaic cells", *Scientometrics*, 102 (2015) 629–651.

[51] J. Michel, B. Bettels, Patent citation analysis—A closer look at the basic input data from patent search reports, *Scientometrics*, 51 (2001) 185–201.

[52] W. De Nooy, A. Mrvar, V. Batagelj, Exploratory social network analysis with Pajek, Cambridge University Press, 2011.

[53] M. Hsiu-Ching Ho, V.H. Lin, J.S. Liu, Exploring knowledge diffusion among nations: a study of core technologies in fuel cells, *Scientometrics*, 100 (2014) 149–171.

[54] J.S. Liu, L.Y. Lu, 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 (2012) 528–542.

[55] M. Hočevar, M. Berginc, U. Opara Krašovec, M. Topič, Dye-Sensitized Solar Cells, in: M. Aparicio, A. Jitianu, L.C. Klein (Eds.) Sol-Gel Processing for Conventional and Alternative Energy, Springer US, 2012, pp. 147–175.

[56] T.T. Ma, A.L. Porter, Y. Guo, J. Ready, C. Xu, L.D. Gao, A technology opportunities analysis model: applied to dye-sensitised solar cells for China, *Technology Analysis & Strategic Management*, 26 (2014) 87–104.

[57] Y. Guo, L. Huang, A.L. Porter, The research profiling method applied to nano-enhanced, thin-film solar cells, *R & D Management*, 40 (2010) 195–208.

[58] Z.F. Zhao, F.J. Zhang, Y. Huang, C.L. Zhao, F. Guo, An experimental study of the hydraulic free piston engine, *Applied Energy*, 99 (2012) 226–233.

[59] Y. Guo, C. Xu, L. Huang, A. Porter, Empirically informing a technology delivery system model for an emerging technology: illustrated for dye-sensitized solar cells, *R & D Management*, 42 (2012) 133–149.

[60] Y. Zhang, X. Zhou, A.L. Porter, J.M.V. Gomila, How to combine term clumping and technology roadmapping for newly emerging science & technology competitive intelligence: “problem & solution” pattern based semantic TRIZ tool and case study, *Scientometrics*, 101 (2014) 1375–1389.

[61] Y. Zhang, A.L. Porter, Z. Hu, Y. Guo, N.C. Newman, “Term clumping” for technical intelligence: A case study on dye-sensitized solar cells, *Technological Forecasting and Social Change*, 85 (2014) 26–39.

[62] B. Oregan, M. Gratzel, A low-cost,high-efficiency solar cell based on dye-sensitized colloidal TiO2 films, *Nature*, 353 (1991) 737–740.

[63] L. Leydesdorff, Can technology life-cycles be indicated by diversity in patent classifications? The crucial role of variety, *Scientometrics*, 105 (2015) 1441–1451.

[64] Hung, S. C., Liu, J. S., Lu, L. Y. Y., & Tseng, Y. C. (2014). Technological change in lithium iron phosphate battery: the key-route main path analysis, *Scientometrics,* 100(1), 97–120.