

MIT Open Access Articles

Tracing Technological Development Trajectories: A Genetic Knowledge Persistence-Based Main Path Approach

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation: Park, Hyunseok and Magee, Christopher L. "Tracing Technological Development Trajectories: A Genetic Knowledge Persistence-Based Main Path Approach." Edited by Zhong-Ke Gao. PLOS ONE 12, no. 1 (January 2017): e0170895 © 2017 Park, Magee

As Published: <http://dx.doi.org/10.1371/journal.pone.0170895>

Publisher: Public Library of Science

Persistent URL: <http://hdl.handle.net/1721.1/110026>

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Terms of use: Creative Commons Attribution 4.0 International License



RESEARCH ARTICLE

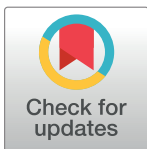
Tracing Technological Development Trajectories: A Genetic Knowledge Persistence-Based Main Path Approach

Hyunseok Park^{1,2}*, Christopher L. Magee^{2,3}

1 Department of Information System, Hanyang University, Seoul, Republic of Korea, **2** Institute of Data, Systems, and Society, Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts, United States of America, **3** SUTD-MIT International Design Center, Massachusetts Institute of Technology (MIT), Cambridge, Massachusetts, United States of America

* These authors contributed equally to this work.

* hp@hanyang.ac.kr



Abstract

The aim of this paper is to propose a new method to identify main paths in a technological domain using patent citations. Previous approaches for using main path analysis have greatly improved our understanding of actual technological trajectories but nonetheless have some limitations. They have high potential to miss some dominant patents from the identified main paths; nonetheless, the high network complexity of their main paths makes qualitative tracing of trajectories problematic. The proposed method searches *backward* and *forward* paths from the high-persistence patents which are identified based on a standard genetic knowledge persistence algorithm. We tested the new method by applying it to the desalination and the solar photovoltaic domains and compared the results to output from the same domains using a prior method. The empirical results show that the proposed method can dramatically reduce network complexity without missing any dominantly important patents. The main paths identified by our approach for two test cases are almost 10x less complex than the main paths identified by the existing approach. The proposed approach identifies all dominantly important patents on the main paths, but the main paths identified by the existing approach miss about 20% of dominantly important patents.

OPEN ACCESS

Citation: Park H, Magee CL (2017) Tracing Technological Development Trajectories: A Genetic Knowledge Persistence-Based Main Path Approach. PLoS ONE 12(1): e0170895. doi:10.1371/journal.pone.0170895

Editor: Zhong-Ke Gao, Tianjin University, CHINA

Received: October 18, 2016

Accepted: January 12, 2017

Published: January 30, 2017

Copyright: © 2017 Park, Magee. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are available from the United States Patent and Trademark Office (USPTO) or any patent database, including PatSnap (www.patsnap.com).

Funding: This work was supported by Hanyang University (Grant number: HY-2016, <http://www.hanyang.ac.kr>) and Singapore University of Technology and Design (Grant number: 6921538, <http://www.sutd.edu.sg>).

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Technological progress is a major factor enabling economic growth [1, 2]. Better understanding of technological change and innovation is essential for informing policy to enable sustainable economic and social growth. An important qualitative aspect of technological change is changes in underlying knowledge bases. Dosi's seminal work [3] delineated the concepts of technological paradigm and trajectories: these have been widely used as a foundation for innovation studies [4–10]. Tracing radical change and incremental development processes through technological trajectories provides essential insights into the evolutionary process and

regularities in a technological domain [10, 11]. This paper contributes to the existing objective methodology for doing such studies.

In the past decade, studies on innovation have exploited patent citation networks to identify and visualize the technological trajectories from empirical data [10]. A patent citation as a reference to prior art for legal purposes represents a proportion of inventive knowledge in the citing patent in originated from or already disclosed by the cited patent [12] and thus a patent citation can be considered to be a knowledge flow and sequential evolutionary path [13–16].

To reduce the complexity of a citation-based knowledge network in order to identify the most significant trajectories, Hummon and Doreian [17] first introduced a main path algorithm and much research has applied main path analysis to technologies to investigate the patterns of technological changes [10, 11, 18–21]. However, previous main path approaches have some limitations. First, most of them identify only one single main path. Second, they cannot show combinatorial relationships between sub-fields in a technological domain (we adopt the definition of a technological domain by Magee, et al., [22]: The set of artifacts that fulfill a specific generic function utilizing a particular, recognizable body of knowledge). Third, traversal counts based forward searching from the starting nodes can ignore other important patents and knowledge flows (details on the limitations are described in section 2).

Therefore, the aim of this research is to propose a new main path approach to overcome the identified limitations. For this, we adopt the genetic knowledge persistence measurement (GKPM), suggested by [23], and differently from others, identify the main paths by *forward* and *backward* tracing. A GKPM quantifies how much knowledge of an invention is retained in and contributes to recent inventions based on structural and topological positions of patents in a citation network and identify high-persistence patents, whose inventive knowledge dominantly persists and contributes to recent inventions in a technological domain. Since the proposed method searches backward and forward paths from the most genetically dominant patents, it can generate multiple interconnected main paths without missing any dominantly important knowledge in the domain (we label the new method genetic backward-forward path analysis and use the acronym GBFP throughout this paper).

To test GBFP and compare it to the existing methods, we conducted empirical analyses for two technological domains, Desalination and Solar photovoltaics (PV). The results show that GBFP identifies multiple main paths for each case with an easily recognizable number of nodes and links, including all high-persistence patents, whereas the existing approach does not identify some dominantly important patents on the identified main paths and also yields high network complexity that makes qualitatively tracing the main paths problematic. Therefore, the proposed method provides clear advantages on reducing network complexity of the identified main paths without missing any developmental paths in a technological domain, which increase the quality and efficiency of qualitative tracing of technological trajectories.

The rest of this paper is structured as follows: Section 2 reviews the literature on the main path analysis, Section 3 describes GBFP, Section 4 presents the empirical analysis and discussion of the results, and finally conclusions are drawn in Section 5.

Main Path Analysis

Citation information indicates knowledge diffusing from the cited to citing documents, so knowledge flows in a citation network are used to trace evolutionary trajectories of technological or scientific knowledge. Although conventional network analysis techniques, such as betweenness and closeness centrality, can be used to analyze the structure or topology of a network, the acyclic character of citations (a patent never cites later patents) makes it difficult to employ the conventional techniques for citation network analysis. Recent research in complex

networks has greatly improved the understanding of structures and dynamics of complex networks [24–31]. However, such general complex network approaches are not as useful for innovation studies particularly because of knowledge inheritance, which is usually not considered in generalized complex network research. This is the primary reason that main path analysis has been widely used to reduce the network complexity and identify the most important knowledge trajectory in a citation networks that characterize invention and innovation trajectories. In order to identify the main sequences of citations in a large citation network, Hummon and Doreian [17] suggested three indices, SPLC (search path link count), SPNP (search path node pair), and NPPC (node pair projection count), which calculate the ‘link connectivity’ based on traversal counts in search paths through the network and assign weight to each link: Batagelj [32] later proposed one more similar index, SPC (search path count): he tested the performance of the search path indices (SPC, SPLC and NPPC) and concluded that these indices provide almost the same performance. The basic logic behind these indices is that a link (and node) included in many search paths in a citation network plays a critical role in the knowledge diffusion, so a sequence of high-weighted links (or nodes) constructs a main path.

In the initial research, main path analysis was applied to academic publications: Hummon and Doreian [17] applied it to DNA development, Hummon and Carley [33] applied it to the social network analysis field, Carley, Hummon [34] analyzed scientific influence in the Conflict resolution field by using main path analysis. More recent research has applied the main path analysis to investigate developmental trajectories of technological fields using patent citation information: Verspagen [10] investigated the technological trajectories of Fuel cell technology, Fontana, Nuvolari [35] traced evolution of Local Area Network (LAN) technology, Mina, Ramlogan [18] analyzed growth and transformation of Coronary artery disease treatment technology, Martinelli [11] applied a main path analysis to trace the Telecommunications switching industry, Epicoco [36] examined the long-term evolution of the Semiconductor miniaturization trajectory, Ho, Lin [21] explored the knowledge diffusion of Membrane electrode assembly technology, and Huenteler, Schmidt [20] analyzed the long-term pattern of innovation and technological life-cycles in the Wind power and Solar PV fields.

Although previous approaches to main path analysis have been used for many studies, methodological limitations also have been recently discussed [37, 38]. Here, we define the required characteristics of a main path analysis for analyzing ‘technological’ domains by considering the properties of a technological domain and the theoretical perspectives in innovation. First, given that a fundamental purpose of using main path analysis is to minimize the number of patents needed to realistically represent a specific technological domain, the identified main paths have to successfully contain the technologically significant patents [35]. Moreover, since technological discontinuities are usually perceived to be the significant technologies in a domain [23], omission of them on the main paths can make the identified main paths misleading or unreliable trajectories. For example, in the Light Emission Diode (LED) domain, a main path analysis for the domain would be judged unrealistic if the identified main paths do not contain the blue LED related patents. Second, a technological domain usually consists of several specific technological knowledge or sub-knowledge fields, so it is reasonable that developmental trajectories for a technological domain would contain multiple trajectories. For specific scientific fields or methods, such as Green chemistry [39], social network analysis [34] or the Hirsch Index [37], a singular main path might realistically trace the developmental trajectory in that the fields only focus on a narrow knowledge outcome. However, if the technological domain has many sub-technologies, for the example of flexible displays, with wiring material, flexible display substrate material, thin film transistor array, semiconductor material, and so on, is represented by a singular path, this would not be useful to investigate the

technological changes in the domain. Third, one major driver for technological development is the recombination of existing knowledge [40–45], therefore a main path analysis should be able to show the combinatorial relationships between the different knowledge streams which is somewhat equivalent to identifying the discontinuities as discussed above.

However, existing approaches are somewhat insufficient to meet the requirements. Most main path approaches produce only a singular main path [17, 32, 37, 38] and as noted above are inappropriate to analyze some technological domains. Verspagen [10] suggested a modified main path which overcomes the limitation of a singular main path and produces multiple main paths so this is the existing approach considered the baseline method in this research.

Method

In this section, genetic backward-forward path analysis (GBFP) is described in detail and its overall procedure is shown in Fig 1.

Collection of patent set

The unit of analysis of this research is a technological domain; the technological domain is defined as ‘the set of artifacts that fulfill a specific generic function utilizing a particular, recognizable body of knowledge’ [22]. Collection of the right data is a fundamentally important step in that this can seriously affect the result. Therefore, we adopted a highly reliable patent search technique, the classification overlap method (COM) developed by [46, 47], to collect highly relevant patents for a pre-defined technological domain, and downloaded the patents from www.patsnap.com.

Construction of patent citation network

The knowledge network of a technological domain is generated based on patent citations with the basic assumption being that a patent citation represents a knowledge flow from cited patent to citing patent. This paper only considers knowledge flows within the technological domain, so only those patent citations occurring within the technological domain are considered (Fig 2). The cited-citing patent pairs are extracted from patent backward citation information.

Measuring knowledge persistence

GBFP identifies main paths by backward and forward searching from the patents dominantly important in a technological domain, i.e. high persistence patents (HPPs): Martinelli and Nomaler [23] show that patents having high persistence value are technologically important inventions or, often, technological discontinuities in the focal technological domain. Searching from the HPPs can guarantee the inclusion of all significant knowledge on the identified main paths. Both backward and forward searching can identify not only other potential main paths that might be missed by only a forward searching, but can also identify convergence structures in technological trajectories. The methodological difference of GBFP from previous approaches originates from this step. Even though the baseline approach [10] also utilizes genetic knowledge persistence, it identifies main paths by forward searching from the *start-points* and so can miss other main paths that contain dominantly significant patents.

In order to identify the HPPs, knowledge persistence of each patent is measured using the GKPM. GKPM, developed by Martinelli and Nomaler [23], can objectively quantify the persistent knowledge of a patent by a backward mapping of the patent from all connected *end-points*. The main concept of knowledge persistence is that a new invention is created by the

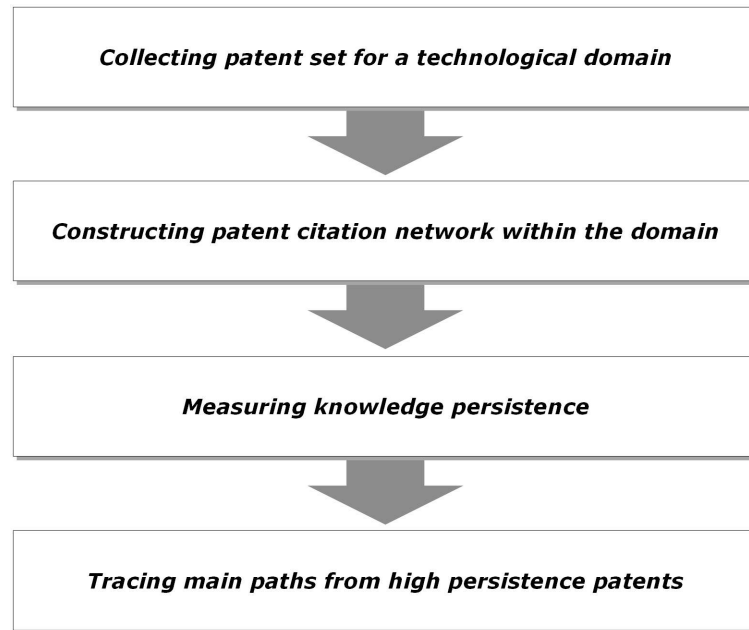


Fig 1. The three steps in the proposed method.

doi:10.1371/journal.pone.0170895.g001

recombination of the existing pieces of knowledge and so, similar to Mendelian genetic inheritance, a proportion of knowledge in a patent is incorporated in its descendant patents. Therefore, in the patent system, cited and citing patents can be interpreted as ancestors and descendants from the genetic inheritance perspective.

The procedure of GKPM is as follows (see Fig 3). First, the overall lineage structure of the technological domain is constructed by assigning each patent to a layer. The endpoints are identified and each patent is assigned to a layer by working backward: the startpoints are assigned as the first layer and then layer numbers for other patents including endpoints are determined. The number of layers of the domain is determined by the longest sequences of citation links from endpoints to startpoints.

Second, based on the topological structure of the layer-based citation network, GKPM measures how much knowledge of a patent is inherited by recently invented patents, i.e. endpoints. Specifically, the proportion of the inherited knowledge of a patent to the next-generation descendant patent is calculated by 1/the number of backward citations of the next descendant patent. Therefore, knowledge persistence of a patent in the network can be calculated by the following equation:

$$KP_A = \sum_{i=1}^n \sum_{j=1}^{m_i} \prod_{k=1}^{l_j-1} \frac{1}{BWDCit(P_{ijk})}, \tag{1}$$

- where KP_A is knowledge persistence value of patent A (P_A),
- n is the number of patents in the last layer, which are (indirectly) connected to P_A ,
- m_i is all possible backward paths from P_i to P_A ,
- l_j is the number of patents on the j -th backward path from P_i to P_A ,

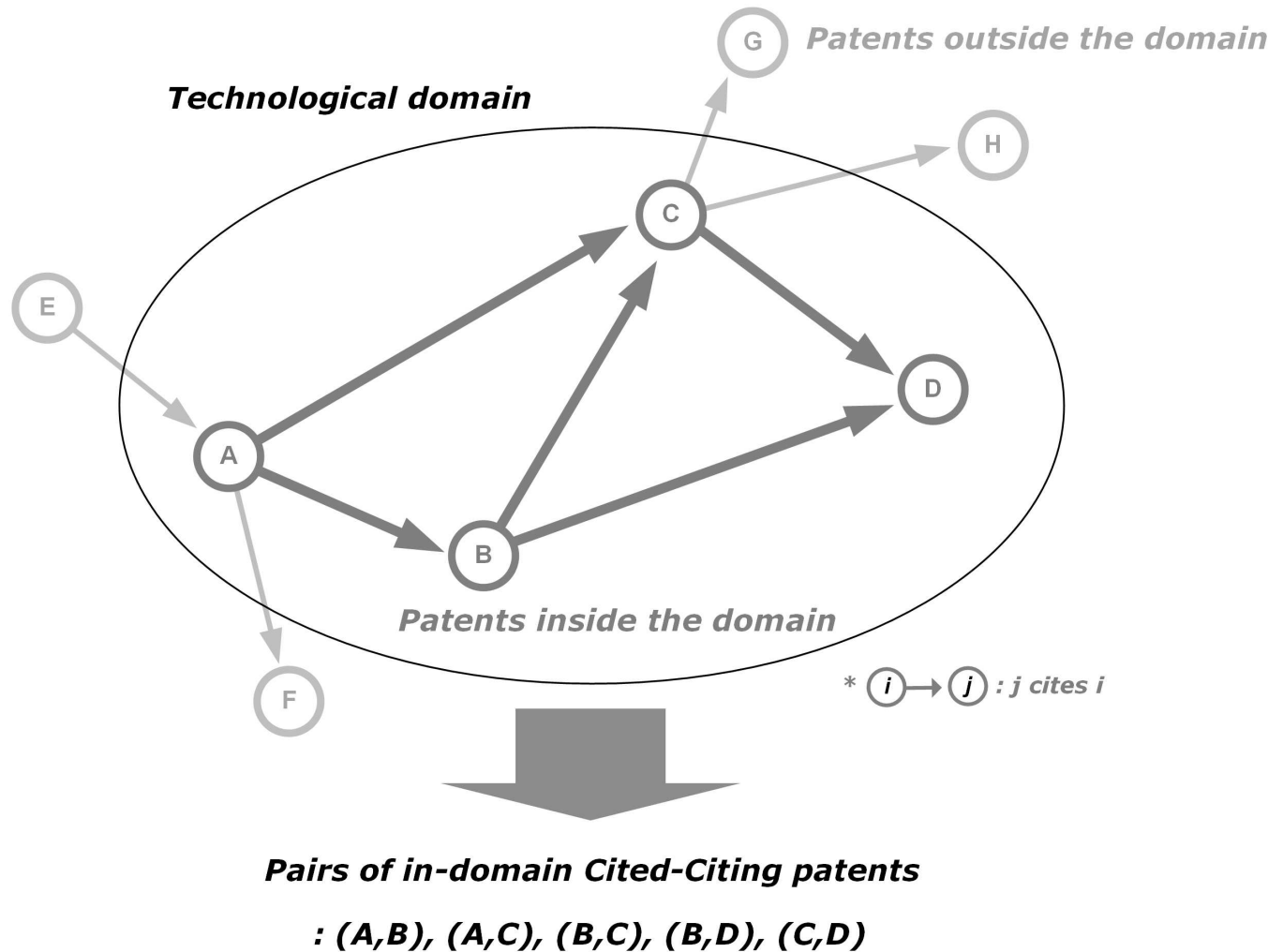


Fig 2. Cited-citing patent pairs. Note: each circle denotes a specific patent; the arrow represents knowledge flow from the cited patent to the citing patent (directed link), e.g. since patent B cites patent A, the arrow from patent A points to patent B; only patents and citations inside the technological domain (A,B,C, and D) are considered as the nodes and links, therefore citations from or to patents outside the technological domain are ignored (E→A, A→F, C→G, and C→H).

doi:10.1371/journal.pone.0170895.g002

- P_{ijk} is the k -th patent on the j -th backward path from P_i to P_A , and
- $BWDCit(P_{ijk})$ is the number of backward citations of P_{ijk} , without considering backward citations by patents in between the first layer and layer $t-1$, when P_A belongs to layer t .

Fig 3 shows a simple patent citation network organized from the layer perspective and gives as an example of use of (Eq 1) for calculation of the knowledge persistence value of patent E. KP_E is calculated by the retained knowledge in the patents J, K and L. J has three backward paths to reach at E (J→G→E, J→E and J→H→E). The number of backward citations for J is 3, for G is 2 (the backward citation to the initial patent A is ignored), and for H is 2. Therefore, 0.167 (= 0.5 × 0.333) of E's knowledge is retained in J through E→G→J path, 0.33 through E→J, and 0.167 (= 0.5 × 0.333) through E→H→J; the total knowledge of E retained in J is 0.667 (= 0.167+0.333+0.167). By the same calculation procedure, 0.5 of E's knowledge is retained in K and 0.75 in L, and so the overall knowledge persistence of E in this simple network is 1.917 (= 0.667+0.5+0.75).

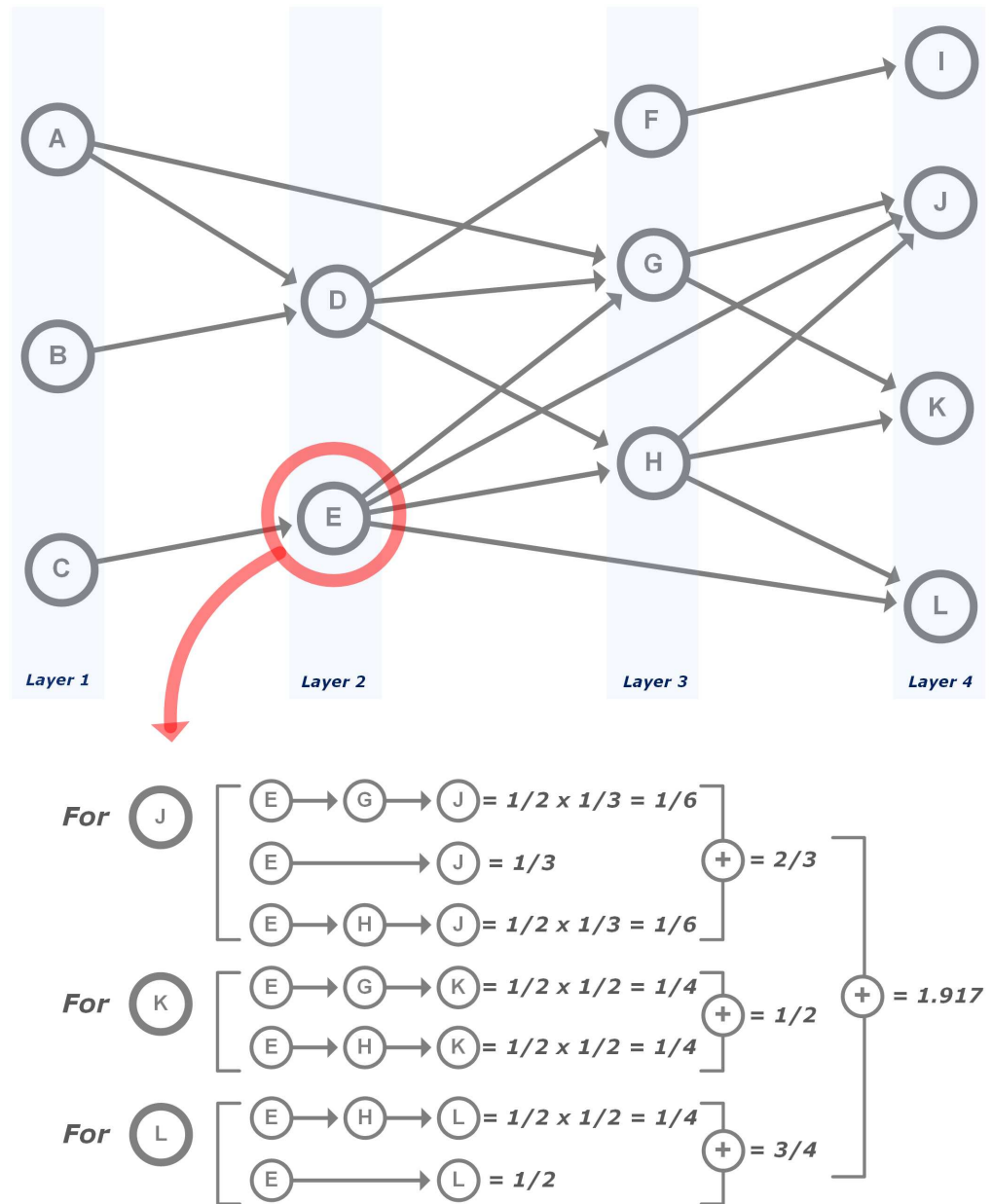


Fig 3. Measurement of knowledge persistence. Note: the layers represent the overall lineage structure of the technological domain. The number of layers in a domain are determined by the longest sequences of citation links from *endpoints* to *startpoints* and the layer number of each patent is determined based on its topological position in the network; inherited knowledge from cited patent to citing patent is measured based on the number of sources, for example, patent D cites two previous patents (patent A and B), so patent D inherits 1/2 of its knowledge from patent A and 1/2 of its knowledge from patent B.

doi:10.1371/journal.pone.0170895.g003

Identification of main paths

GBFP searches paths from the patents having high knowledge persistence value. To select these HPPs, we consider two perspectives: global persistence (GP) and layer (or local) persistence (LP). GP identifies the most important patents in the domain and LP identifies the important patents in each layer: we use LP is to include relatively recent important patents in

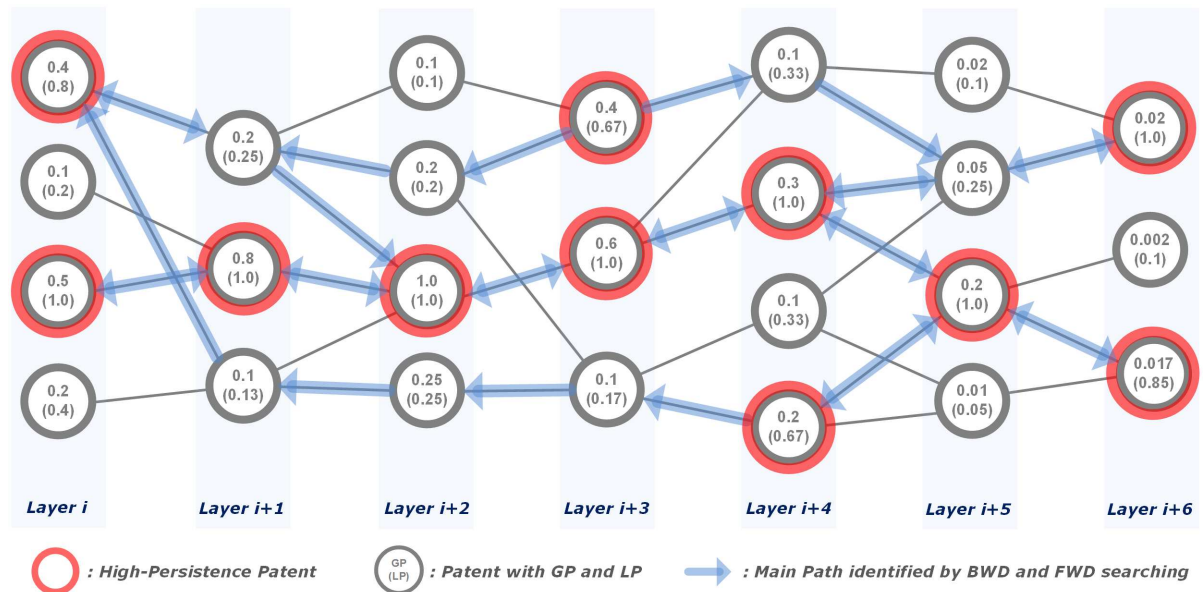


Fig 4. Searching backward and forward paths. Note: every HPP has both left and right arrows for backward and forward searches; HPP (0.2 GP and 1.0 LP) on layer $i+5$ is directly connected with two HPPs on layer $i+4$ and both links are chosen as main paths; if a HPP is not directly connected with other HPPs, e.g. the HPP (0.4 GP and 0.8 LP) on layer i , a patent which is not HPP but having the highest GP among the directly connected patents, e.g. the patent (0.2 GP and 0.25 LP) on layer $i+1$, is chosen and further searching is continued from that patent using the same algorithm.

doi:10.1371/journal.pone.0170895.g004

the main paths whose overall persistence has not yet emerged. To simplify the process, we normalized the GP and LP of each patent by dividing by the maximum persistence value in the domain for GP and dividing by the maximum persistence value in each layer for LP. The cutoff value for HPPs can be set according to the desired complexity of the main paths; based on our heuristic test, $GP > (0.3 \sim 0.5)$ or $LP > (0.7 \sim 0.9)$ are usually most appropriate for further analysis. The lower value for the GP cutoff relative to the LP cutoff value is necessary to maximize retention of dominantly important patents in the main path network. The testing also showed a clear tradeoff between retention of important patents (favored by low cutoff values) and network complexity (favored by high cutoff values). For the rest of this paper, we report results for the cutoff values of $GP = 0.3$ and $LP = 0.8$ but recommend that the range suggested be investigated for completeness of analysis.

After the identification of HPPs, main paths are identified by backward and forward searching from each HPP (Fig 4): if there exist five HPPs, five backward and forward searches are performed. The basic mechanism for a backward/forward search is to choose the patent(s) having the highest GP among the directly connected cited/citing patents, therefore, any direct link between two HPPs is always chosen as a component of the main paths. The backward/forward search is finished when it reaches the *startpoint(s)/endpoint(s)* for all HPPs so all main paths are identified.

Empirical Analysis

In this section, we present empirical analyses for two technological domains: Desalination and Solar PV. In order to show the differences of GBFP from previous methods, we compared the main paths from GBFP with the main paths from the baseline approach.

Solar PV

Introduction to Solar PV. Solar energy is the most abundant energy source on earth and photovoltaic devices have been identified as one promising type of clean energy. PV cells directly generate electricity from sunlight radiation and this PV effect was discovered over 150 years ago but has been practically developed from the 1950s [48]. The Solar PV domain can be broadly classified into three sub-components: solar cell, module and panel, and mounting system. The solar cell as the core component is a form of photoelectric cell which generates electricity. The PV module (or panel) is a bundle of solar cells for practical applications, and mounting systems are related to technologies to install and control a PV system. Major bottlenecks in Solar PV are conversion efficiency and costs, so the overall developmental trajectory consists of inventions that, based on the basic PV effect, adopting new materials or develop new engineering designs for alleviating these bottlenecks. Some of the patents involve sunlight concentration and hybrid structure; moreover, multi-junction cells have been developed to realize cheaper manufacturing costs while maintaining useful conversion efficiencies. Although emerging PVs, such as dye-sensitized solar cells or organic solar cells, have recently received attention, they are still at the laboratory research level [49] due to their currently lower performance (W/\$) compared to the for now dominant PV types using crystalline silicon or amorphous silicon.

Data. The patent set for the Solar PV was obtained by COM specifically using the overlap between UPC 136 (Batteries: thermoelectric and photoelectric) and IPC H01L (Semiconductor devices; Electric solid-state devices). The number of patents in the set is 5,203, from 1976-1-1 to 2013-7-1, and the technological relevancy of the patent set is 0.85.

Result. The main paths for Solar PV drawn by the proposed method are shown in Fig 5; the graphs were drawn by using Gephi (www.gephi.org) and Event graph layout plug-in [50], and serial numbers were given to the patents sorted by ascending order of patent numbers. The identified main paths can be broadly separated into three sub-fields: module and panel, solar cell and mounting system (Fig 5). Overall developmental trajectories are increasing conversion efficiency and reliability with lower cost by adopting new materials or new engineering designs. With the GP cutoff at 0.3 and the LP cutoff at 0.8, the main path network consists of 159 patents (58 are HPP) with 192 citations among them. Qualitatively, the main paths are relatively easy to identify in this network but larger numbers of nodes/patents make this more difficult.

Fig 6. shows the main paths determined by the baseline method [10]. The overall network is quite large (upper left of Fig 6 has 1821 patents and 1729 citations among them which is a factor of ~9 bigger than the GBFP-based network. The larger graphs in Fig 6 were drawn (as was Fig 5) using Gephi and show some similarities to the main paths in Fig 5 with both approaches showing similar trajectories which are mainly constructed by HPPs. However, about 24% (14 patents) of the identified HPPs are only included in the main paths obtained using GBFP (in Fig 5 but not in Fig 6). Some other similarities and differences between the results from the two methods are worth noting. First, in the main paths for the Solar PV module and panel, the paths from 424 to 1645 are similar overall but the path from 1645 to 3492 and 4272 in Figs 5 and 6 are somewhat different: HPPs 1730, 1886 and 2080 are included only in the path in Fig 5. Investigating these additional patents included by BFGP but not in the baseline indicates they are important and appropriate to include. In particular, Patent 1886 (US 5409549) was a new design for the solar cell module panel improving long-term reliability and cost. Specifically, the edge portions of the solar cell modules are fixed and protected by the module fasteners, the modules are not mechanically damaged, and so long-term reliability of the modules is improved. In addition, because a separated base for the work is not required, the facility and

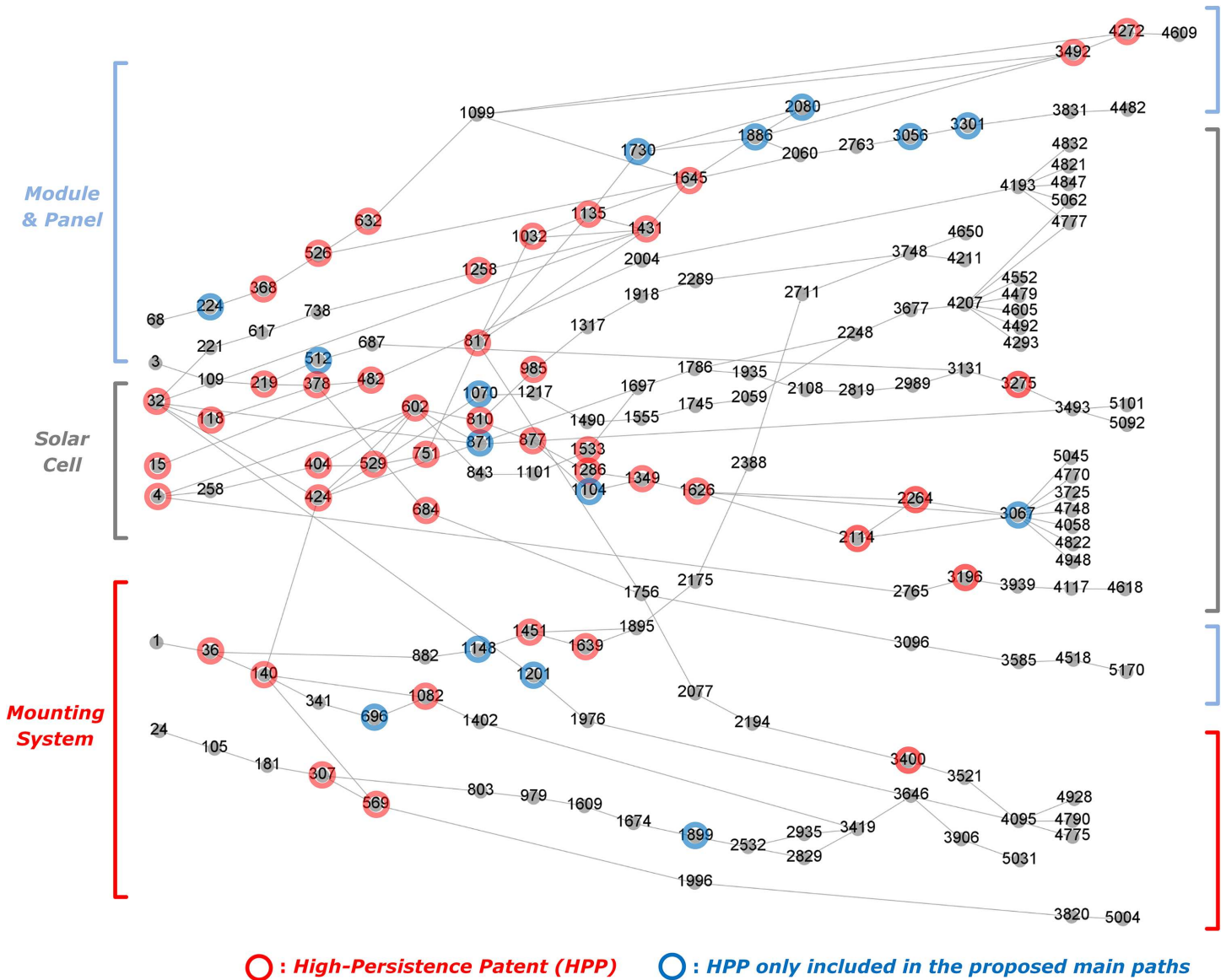


Fig 5. Main paths for Solar PV determined by GBFP. Note: # of nodes and links are 159 and 192; # of high persistence patents (GP>0.3 or LP>0.8) are 58. (see Table A in S2 File for details on HPPs).

doi:10.1371/journal.pone.0170895.g005

safety of the installation is improved, and cost can be reduced. Many other patents also described this invention as one representative type that is compatible with other inventions (US 6119415, US 6182404, and US 7081585).

Patent 2080 (US 5589006) is a solar cell module that can be used with an air heating type passive solar system. Heat from solar energy is usually an inevitable cause of deteriorating conversion efficiency of solar cells. This solar cell module does not require an additional base, which limits the reduction in photoelectric conversion efficiency due to heat. This invention provided an important combination of solar cell modules with a passive solar system by overcoming the efficiency problem introduced by heat.

In addition, in the main paths for the solar cell, the path from 602 to 1349 is about thin film solar cell and patents in the path from both approaches are the same but the main paths in

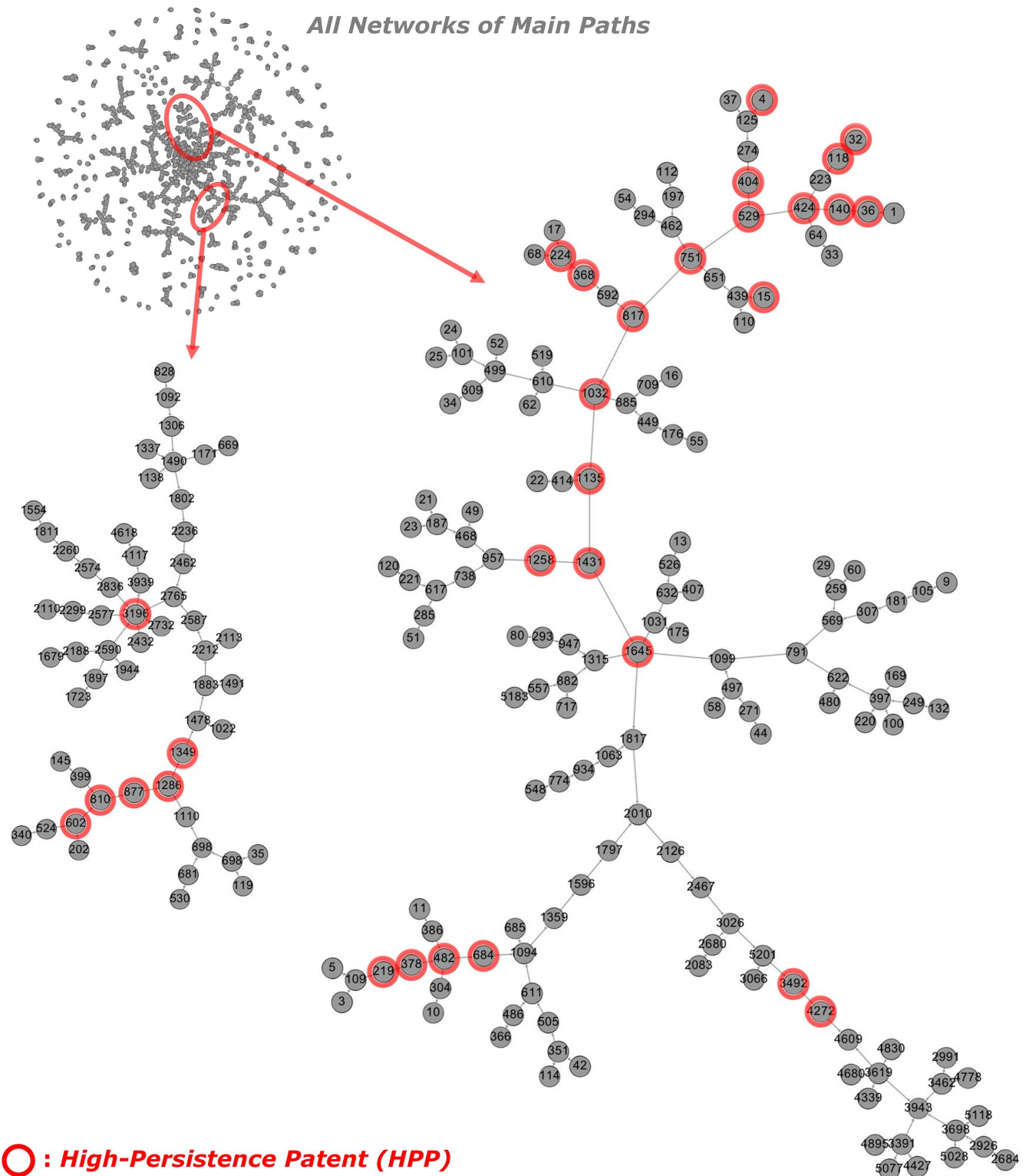


Fig 6. Main paths for Solar PV obtained by the baseline approach. Note: # of nodes and links for all main paths are 1821 and 1729; # of high persistence patents included in all main paths are 44; the left and right graphs are the sub-networks of main paths which contain more than five HPPs.

doi:10.1371/journal.pone.0170895.g006

Fig 5 contain one more patent— 871. Patent 871 (US 4532537) describes a method of fabricating the photo detector comprising the light transmissive electrical contact with the textured surface on the substrate by chemical vapor deposition. This invention is an important application of surface-textured substrates for optical absorption enhancement [51] and many later patents introduced this as one conventional method (US 4664748, US 4880664, US 5078803 and US 5102721).

Moreover, patent 3301 (US 6784361), a relatively recent invention but not included in the main paths in Fig 6, is about amorphous silicon (a-Si) and CdS/CdTe type thin film solar cells that can provide better efficiency at elevated operation temperatures. Specifically, this includes a front electrode made of a transparent conductive oxide (TCO) and has thick intrinsic layers. Since this invention is one conventional early design of related solar cells (described in US 7846750, US 7875945, US 7888594, US 7964788, US 8203073, US 8012317, US 8022291, US 8076571, US 8133747, US 8236118, US 8334452, US 8338699, and US 8354586), leaving this patent out when tracing technological trajectories of the thin-film solar cell appears unrealistic.

Desalination

Introduction to desalination. The desalination domain consists of artifacts that remove salts and minerals from saline water. The potential for a global water shortage has promoted this technological domain to one of high significance for human welfare. Desalination devices can be broadly categorized as thermal or membrane-based technologies [52]. Thermal desalination is based on water phase changes through distillation or evaporation: Multi effect distillation, multi stage flash and vapor compression distillation are the most representative technologies. Membrane desalination is based on the characteristics of semi-permeable membranes that permit water to pass through it when the pressure of feed water is greater than the osmotic pressure: reverse osmosis (RO) has been widely used for commercial purposes. Although thermal desalination accounts for a significant portion of the entire desalination market, rapid advancement of membrane desalination is apparently leading to it surpassing thermal desalination [52]: most patented inventions, particularly in our patent set, are related to membrane desalination.

Data. The patent set for the Desalination technology is obtained by COM. The specific overlap used was between UPC 210 (Liquid purification or separation) and IPC C02F (Treatment of water, waste water, sewage, or sludge) or B01D (Separation). Since this classification overlap contains patents related to water treatment or purification, we added a keyword search query to isolate only patents relevant to the desalination technology. The number of patents in the set is 3,634, from 1976-1-1 to 2013-7-1, and the technological relevancy of the patent set is 0.87.

Result. The identified main paths for Desalination are shown in Fig 7 for GBFP and in Fig 8 for the baseline method [10]. The networks were again drawn by using Gephi and serial numbers are given to the patents sorted by ascending order of patent numbers. As with Solar PV, the baseline main path network is much larger (1744 patents with 1508 citation links among them) than the network from GBFP (115 patents with 134 citation links among them). Although sections of the baseline network can be isolated as shown in Fig 8, the added complexity makes this main path network less amenable to qualitative analysis than is the GBFP-based network shown in Fig 7.

The identified main paths can be classified into three sub-fields: reverse, or forward, osmosis based membrane process, ion-exchange including water softening and electro dialysis, and preprocess techniques, usually, for membrane processes including dewatering and

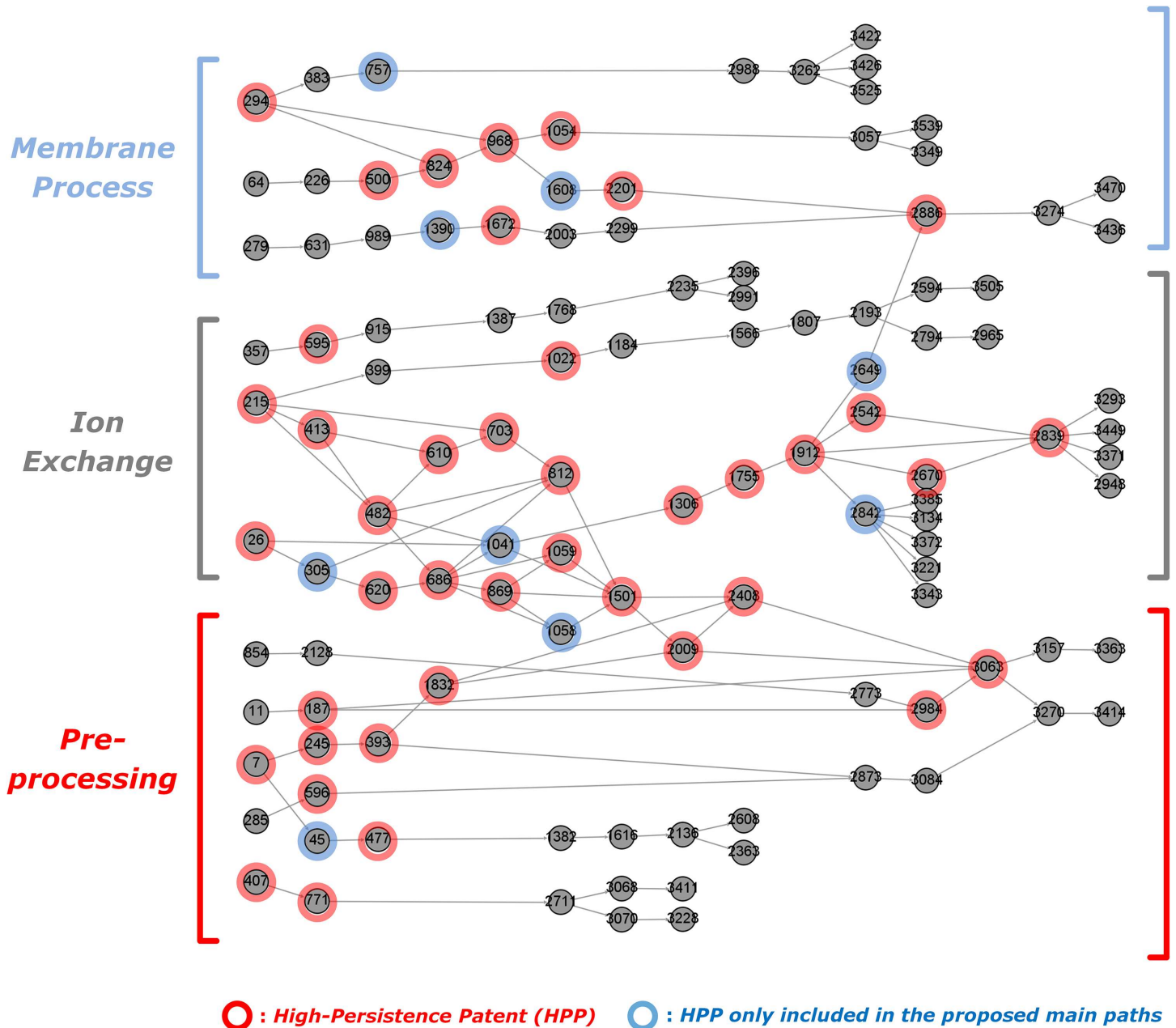


Fig 7. Main paths for Desalination technology drawn by GBFP. Note: # of nodes and links are 115 and 134; # of high persistence patents (GP>0.3 or LP>0.8) are 50. (see Table B in [S2 File](#) for details on HPPs).

doi:10.1371/journal.pone.0170895.g007

precipitation. The major outcomes of the developmental trajectories are increasing desalination efficiency at lower energy consumption, i.e. cost apparently arising from hybrid use of different methods or new materials, e.g. nanofiltration (NF), and new energy sources, e.g. solar energy. Both the proposed and baseline methods' main paths show overall similar trajectories.

However, similar to the PV case, the network obtained from the baseline approach—despite its much larger size—does not contain about 18% of HPPs (9 out of 50) present in the GBFP-based network. As before, qualitative analysis indicates that some of those included by GBFP seem to be significant knowledge inputs in tracing developmental trajectories. First, the patent

All Networks of Main Paths

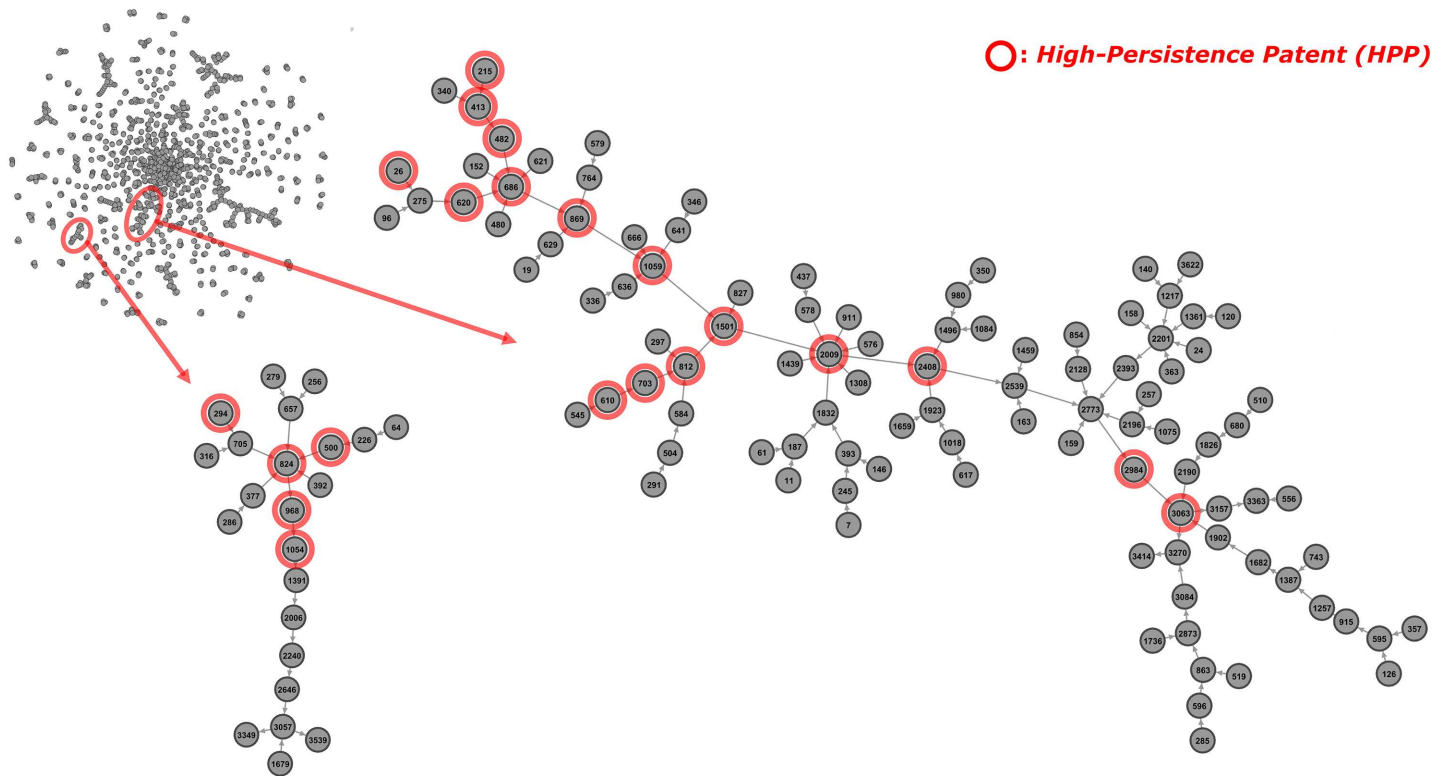


Fig 8. Main paths for Desalination technology obtained by the baseline approach. Note: # of nodes and links are 1774 and 1508; # of high persistence patents included in the all main paths are 41; the left and right graphs are the sub-networks of main paths which contain more than five HPPs.

doi:10.1371/journal.pone.0170895.g008

757 (US 4704324) is an invention describing a method to form a semi-permeable membrane. Specifically, a permselective discriminating layer of the membrane prepared by reaction of onium compounds with nucleophilic compounds can greatly enhance the flux of membranes and thus increase desalination efficiency of RO systems. Since many later patents adopted its inventive knowledge (US 4812238, US 4888116, US 5238747 and US 5310581), patent 757's position on the main paths appears important for further evolution of the trajectory on the flux of membranes (Fig 7). Missing this patent suggests a main path that is probably not a reliable representation of reality.

Second, the patent 1041 (US 4936987) disclosed a water soluble polymer that can prevent the precipitation or crystallization of scale-forming salts of alkaline earth metal cations. Although later inventions pointed out that large dosages of the polymers are required (US 5256302 and US 5393456), this invention is a relatively early technology for scale and corrosion inhibitors which have made substantial contribution to later relevant developments (US 5182028, US 5259974, US 5322636, US 5338477, US 5358642, US 5284590, US 6333005, US 6355214 and US 6646082).

Third, the patent 1608 (US 5458781) is a method to separate the monovalent anion bromide from sea and brackish waters by using a combination of RO and NF membranes. Even though this patent has not been widely cited by the later patents and its major purpose is not exactly for producing potable water from sea or brackish waters, the inventive knowledge of this invention had a significant impact on further developments in combinations of RO and NF membranes; many of the patents that cite the patent 1608 are directly affected by this invention (US 6190556, US 7144511, US 8366924, and US 9205383).

Conclusion

Main path analysis has been widely adopted as a useful tool to empirically trace technological trajectories. Ideally, a main path analysis would be able to reduce network complexity for effective qualitative investigation of the technological trajectories while not eliminating dominantly important knowledge. Our results for the solar PV and the desalination domains clearly show that our proposed new approach (genetic backward-forward main path analysis or GBFP) is a significant step towards this ideal compared to the existing approaches represented by the baseline approach. In the two cases, the GBFP networks are about 10 times *smaller* and also contain about 20% *more* of the dominant genetic knowledge in the domain. Defining less complex networks that nonetheless contain more of the important knowledge makes simultaneous progress along both key dimensions toward a better method. The GBFP does this while adopting the best practice of the baseline approach—the genetic knowledge definition as first defined by Verspagen [10]. This is done through adopting the persistence measurement (GKPM) to first identify the patents having the dominantly significant knowledge bases, i.e. HPPs, in the technological domain. Then, main paths are identified by backward and forward searching from the identified HPPs.

To verify the usefulness of the proposed method, we conducted empirical analyses for the solar PV and desalination technological domains and compared our method with the existing approach. The empirical results show that major technological trajectories on both main paths are quite similar and they are overall represented by HPPs. Most HPPs on the main paths are actually critical for dominantly important knowledge streams for the technological domains. In regard to combinatorial relationships, both approaches appropriately identify the important combinations of knowledge streams (e.g. two converging main paths onto patent 424 in Solar PV). Even though main paths from the baseline method show much more combinatorial relationships, it is apparently due to the baseline method identifying much larger networks, and many of the converging paths seem to be noise, also due to the large network size. Our qualitative analysis also found that some HPPs only included in the main paths obtained by GBFP involve significant domain knowledge and should be contained on realistic main paths. In addition, even though many of previous studies using the existing approach focused on the largest sub-network of main paths to trace technological trajectories, the empirical result shows that the largest, or even second or third largest, sub-network of main paths is not always the most important main path. This means that most of large sub-networks need to be analyzed to uncover the appropriate trajectories. Therefore, given that a major reason to use main path analysis is to reduce the network complexity, the high network complexity of the baseline main paths is not a negligible issue in that the size of the main paths, shown in Figs 6 and 8, is still too large for qualitative analysis.

However, there exist some issues that would be improved in further research. First, we adopt GKPM to identify HPPs in a patent citation network. GKPM considers that a citing patent receives same proportion of knowledge from all cited patents, but the proportions in knowledge inheritance for each citation might be different. Our empirical analysis and Martignelli and Nomaler [23]'s research show that this weighting issue is negligible when the size of a citation network is relatively large, but it might cause a reliability problem when the network size is too small. Therefore, the improvement of this weighting algorithm may increase quality and reliability of the main paths. Second, we use a concept of layer persistence (LP) to identify the patents which are recently invented but have the potential to be dominantly important. Nonetheless, the identified main paths still have a relatively low number of recent HPPs. Therefore, development of other criteria using a text mining technique to identify the recently invented but technologically important patents may provide additional value.

Supporting Information

S1 File. Patent search query for Solar PV and Desalination.
(DOCX)

S2 File. List of high persistence patents in Solar PV (Table A) and Desalination (Table B).
(DOCX)

Acknowledgments

We would like to thank Hanyang University and SUTD/MIT International Design Center for supporting the research.

Author Contributions

Conceptualization: HP CLM.

Formal analysis: HP CLM.

Investigation: HP CLM.

Methodology: HP CLM.

Software: HP.

Visualization: HP.

Writing – original draft: HP CLM.

Writing – review & editing: HP CLM.

References

1. Solow RM. Technical change and the aggregate production function. *The review of Economics and Statistics*. 1957;312–320.
2. Schumpeter J. *The Theory of Economic Development*. New Brunswick: NP Transaction Publishers; 1934.
3. Dosi G. Technological paradigms and technological trajectories: a suggested interpretation of the determinants and directions of technical change. *Research policy*. 1982; 11(3):147–162.
4. Castellacci F. Technological paradigms, regimes and trajectories: Manufacturing and service industries in a new taxonomy of sectoral patterns of innovation. *Research Policy*. 2008; 37(6):978–994.
5. Christensen CM, Rosenbloom RS. Explaining the attacker's advantage: Technological paradigms, organizational dynamics, and the value network. *Research policy*. 1995; 24(2):233–257.
6. Cimoli M, Dosi G. Technological paradigms, patterns of learning and development: an introductory roadmap. *Journal of Evolutionary economics*. 1995; 5(3):243–268.
7. Parayil G. Mapping technological trajectories of the Green Revolution and the Gene Revolution from modernization to globalization. *Research policy*. 2003; 32(6):971–990.
8. Frenken K, Leydesdorff L. Scaling trajectories in civil aircraft (1913–1997). *Research Policy*. 2000; 29(3):331–348.
9. Von Tunzelmann N, Malerba F, Nightingale P, Metcalfe S. Technological paradigms: past, present and future. *Industrial and Corporate Change*. 2008; 17(3):467–484.
10. Verspagen B. Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in Complex Systems*. 2007; 10(01):93–115.
11. Martinelli A. An emerging paradigm or just another trajectory? Understanding the nature of technological changes using engineering heuristics in the telecommunications switching industry. *Research Policy*. 2012; 41(2):414–429.
12. Von Wartburg I, Teichert T, Rost K. Inventive progress measured by multi-stage patent citation analysis. *Research Policy*. 2005; 34(10):1591–1607.

13. Marco AC. The dynamics of patent citations. *Economics Letters*. 2007; 94(2):290–296.
14. Hall BH, Jaffe AB, Trajtenberg M. The NBER patent citation data file: Lessons, insights and methodological tools. National Bureau of Economic Research, 2001.
15. Jaffe AB, Trajtenberg M. *Patents, citations, and innovations: A window on the knowledge economy*: MIT press; 2002.
16. Jaffe AB, Trajtenberg M, Henderson R. Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics*. 1993; 108(3):577–598.
17. Hummon NP, Doreian P. Connectivity in a citation network: The development of DNA theory. *Social networks*. 1989; 11(1):39–63.
18. Mina A, Ramlogan R, Tampubolon G, Metcalfe JS. Mapping evolutionary trajectories: Applications to the growth and transformation of medical knowledge. *Research policy*. 2007; 36(5):789–806.
19. Ho JC, Saw E-C, Lu LY, Liu JS. Technological barriers and research trends in fuel cell technologies: A citation network analysis. *Technological Forecasting and Social Change*. 2014; 82:66–79.
20. Huenteler J, Schmidt TS, Ossenbrink J, Hoffmann VH. Technology life-cycles in the energy sector—Technological characteristics and the role of deployment for innovation. *Technological Forecasting and Social Change*. 2016; 104:102–121.
21. Ho MH-C, Lin VH, Liu JS. Exploring knowledge diffusion among nations: a study of core technologies in fuel cells. *Scientometrics*. 2014; 100(1):149–171.
22. Magee C, Basnet S, Funk J, Benson C. Quantitative empirical trends in technical performance. *Technological Forecasting and Social Change*. 2016; 104:237–246.
23. Martinelli A, Nomaler Ö. Measuring knowledge persistence: a genetic approach to patent citation networks. *Journal of Evolutionary Economics*. 2014; 24(3):623–652.
24. Xia C-y, Wang Z, Sanz J, Meloni S, Moreno Y. Effects of delayed recovery and nonuniform transmission on the spreading of diseases in complex networks. *Physica A: Statistical Mechanics and its Applications*. 2013; 392(7):1577–1585.
25. Xia C, Wang L, Sun S, Wang J. An SIR model with infection delay and propagation vector in complex networks. *Nonlinear Dynamics*. 2012; 69(3):927–934.
26. Chen M-h, Wang L, Wang J, Sun S-w, Xia C-y. Impact of individual response strategy on the spatial public goods game within mobile agents. *Applied Mathematics and Computation*. 2015; 251:192–202.
27. Chen M-h, Wang L, Sun S-w, Wang J, Xia C-y. Evolution of cooperation in the spatial public goods game with adaptive reputation assortment. *Physics Letters A*. 2016; 380(1):40–47.
28. Sun S, Wu Y, Ma Y, Wang L, Gao Z, Xia C. Impact of Degree Heterogeneity on Attack Vulnerability of Interdependent Networks. *Scientific Reports*. 2016; 6:32983. doi: [10.1038/srep32983](https://doi.org/10.1038/srep32983) PMID: [27609483](https://pubmed.ncbi.nlm.nih.gov/27609483/)
29. Gao Z-K, Cai Q, Yang Y-X, Dang W-D, Zhang S-S. Multiscale limited penetrable horizontal visibility graph for analyzing nonlinear time series. *Scientific Reports*. 2016; 6:35622. doi: [10.1038/srep35622](https://doi.org/10.1038/srep35622) PMID: [27759088](https://pubmed.ncbi.nlm.nih.gov/27759088/)
30. Gao Z-K, Fang P-C, Ding M-S, Jin N-D. Multivariate weighted complex network analysis for characterizing nonlinear dynamic behavior in two-phase flow. *Experimental Thermal and Fluid Science*. 2015; 60:157–164.
31. Gao Z-K, Cai Q, Yang Y-X, Dong N, Zhang S-S. Visibility Graph from Adaptive Optimal Kernel Time-Frequency Representation for Classification of Epileptiform EEG. *International Journal of Neural Systems*. 2017; 27(4):1750005.
32. Batagelj V. Efficient algorithms for citation network analysis. arXiv preprint cs/0309023. 2003.
33. Hummon NP, Carley K. Social networks as normal science*. *Social networks*. 1993; 15(1):71–106.
34. Carley KM, Hummon NP, Harty M. Scientific Influence An Analysis of the Main Path Structure in the Journal of Conflict Resolution. *Science Communication*. 1993; 14(4):417–447.
35. Fontana R, Nuvolari A, Verspagen B. Mapping technological trajectories as patent citation networks. An application to data communication standards. *Economics of Innovation and New Technology*. 2009; 18(4):311–336.
36. Epicoco M. Knowledge patterns and sources of leadership: Mapping the semiconductor miniaturization trajectory. *Research Policy*. 2013; 42(1):180–195.
37. Liu JS, Lu LY. 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*. 2012; 63(3):528–542.
38. Yeo W, Kim S, Lee J-M, Kang J. Aggregative and stochastic model of main path identification: a case study on graphene. *Scientometrics*. 2014; 98(1):633–655.

39. Epicoco M, Oltra V, Saint Jean M. Knowledge dynamics and sources of eco-innovation: Mapping the Green Chemistry community. *Technological Forecasting and Social Change*. 2014; 81:388–402.
40. Dahlin KB, Behrens DM. When is an invention really radical?: Defining and measuring technological radicalness. *Research Policy*. 2005; 34(5):717–737.
41. Fleming L. Recombinant uncertainty in technological search. *Management science*. 2001; 47(1):117–132.
42. Nelson RR, Winter SG. *An evolutionary theory of economic change*. Cambridge, MA: Harvard University Press; 1982.
43. Schilling MA, Green E. Recombinant search and breakthrough idea generation: An analysis of high impact papers in the social sciences. *Research Policy*. 2011; 40(10):1321–1331.
44. Youn H, Strumsky D, Bettencourt LM, Lobo J. Invention as a combinatorial process: evidence from US patents. *Journal of The Royal Society Interface*. 2015; 12(106):20150272.
45. Basnet S, Magee CL. Modeling of technological performance trends using design theory. *Design Science*. 2016; 2:1–35.
46. Benson CL, Magee CL. A hybrid keyword and patent class methodology for selecting relevant sets of patents for a technological field. *Scientometrics*. 2013; 96(1):69–82.
47. Benson CL, Magee CL. Technology structural implications from the extension of a patent search method. *Scientometrics*. 2015; 102(3):1965–1985.
48. EPRI. *Solar Photovoltaics: Status, Costs, and Trends*. 2009.
49. Funk JL, Magee CL. Rapid improvements with no commercial production: How do the improvements occur? *Research Policy*. 2015; 44(3):777–788.
50. Spekkink W. Building capacity for sustainable regional industrial systems: an event sequence analysis of developments in the Sloe Area and Canal Zone. *Journal of Cleaner Production*. 2015; 98:133–144.
51. Yablonovitch E, Cody GD. Intensity enhancement in textured optical sheets for solar cells. *IEEE Transactions on Electron Devices*. 1982; 29(2):300–305.
52. Greenlee LF, Lawler DF, Freeman BD, Marrot B, Moulin P. Reverse osmosis desalination: water sources, technology, and today's challenges. *Water research*. 2009; 43(9):2317–2348. doi: [10.1016/j.watres.2009.03.010](https://doi.org/10.1016/j.watres.2009.03.010) PMID: [19371922](https://pubmed.ncbi.nlm.nih.gov/19371922/)