

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/322788160>

Data-Mining the Foundational Patents of Photovoltaic Materials: An Application of Patent Citation Spectroscopy

Article in Journal of Scientometric Research · January 2018

DOI: 10.5530/jsires.7.2.13

CITATIONS

3

READS

122

2 authors, including:



Loet Leydesdorff

University of Amsterdam

775 PUBLICATIONS 54,101 CITATIONS

SEE PROFILE

Data-mining the Foundational Patents of Photovoltaic Materials:

An application of Patent Citation Spectroscopy

Jordan A. Comins*^a and Loet Leydesdorff^b

Forthcoming In:
Global Innovation Index 2018: ENERGIZING THE WORLD WITH INNOVATION

^a Social and Behavioral Sciences Department, The MITRE Corporation, McLean, VA, United States;
jcomins@gmail.com ; * corresponding author

^b Amsterdam School of Communication Research (ASCoR), University of Amsterdam, PO Box 15793, 1001 NG Amsterdam,
The Netherlands; loet@leydesdorff.net.

1. Introduction

The United States Patent and Trademark Office (USPTO) under the Department of Commerce plays a vital role in American innovation and the global economy by registering and extending legal protections over inventions. In exchange for detailed public disclosure of a technical invention, the patent assignee, the legal entity to which intellectual property rights are assigned, is entitled to a monopoly over the patent's claims. As such, patents provide a unique window on knowledge-based economies (Jaffe & Trajtenberg, 2002) and tend to serve as an indicator of industrial activity, rather than the output of academia (Shelton & Leydesdorff, 2012).

Beyond their critical role in industry, patents are indicators of inventions and thus can be expected to carry information about technological progress (O'Donoghue et al., 1998; Jaffe, Trajtenberg, & Henderson, 1993; Harhoff et al., 1999; Artz et al., 2010; Graevenitz et al., 2013; Comins, 2015). To understand and track technological progress, however, subject matter experts hitherto must review patents and patent applications and maintain an awareness of the most technologically important patents. This is a time-consuming practice which presents several obstacles including difficulty with reliability and replication, and dependence on the availability of experts (Cockburn et al., 2002). As a consequence, patent offices have invested in developing novel automated approaches for identifying landmark patents in technology areas (Jensen and Murray, 2005; Konski and Spielthener, 2009).

Recently, we developed an algorithmic method for Patent Citation Spectroscopy (PCS). PCS enables the user to identify landmark patents via a web-application (url: <http://www.leydesdorff.net/comins/pcs/index.html>). We demonstrated its effectiveness in identifying seminal patents in areas of biotechnology (Comins, Carmack, & Leydesdorff, *under review*). In this previous study, it was demonstrated that the pioneering patent for the invention of RNA molecules that can selectively inhibit gene expression, otherwise known as RNA interference or RNAi, could be successfully data-mined using PCS. More specifically, this data-mining procedure correctly identified the seminal RNAi patent by Andrew Fire and colleagues

(Fire et al., 1998; 2003) by applying the PCS methodology to granted US patents retrieved via both keyword searches and more objective patent search parameters (e.g. CPC classification, described below).

In this chapter, we extend our introduction of PCS by conducting an analysis of the underlying seminal patents for a myriad of material technologies of photovoltaic cells. To do so, we leverage the taxonomy of the patent classification system. While there are numerous patent classification systems, among the most widely-used in patent studies are the United States Patent Classification (USPC) system, which comprises more than 160,000 classes and subclasses of patent functions (USPTO, 2008), and the International Patent Classification (IPC) system, a hierarchical system managed by the World Intellectual Property Organization (WIPO) consisting of more than 70,000 classifications of technical fields (WIPO, 2014). In 2013, the USPTO and the European Patent Office (EPO) adopted a new classification system for patents that will ultimately replace both the USPC and IPC. Known as the Cooperative Patent Classification (CPC) system of these two large agencies, this new taxonomy of patents is a tree-like hierarchy consisting of 5-levels of depth and more than 250,000 classifications at the level of the leaf node and is currently in use for patents filed through the USPTO as well as the European Patent Office (EPO).

CPC adds to the previous systems by the introduction of the Y-class of patents representing newly emerging technologies across sectors. The new classes are backtracked into the previous system. Currently, there are nine CPC classifications that describe material photovoltaic technologies (Leydesdorff, Alkemade, Heimeriks, & Hoekstra, 2015):

Y02E 10/541	...	CuInSe ₂ material PV cells
Y02E 10/542	...	Dye sensitized solar cells
Y02E 10/543	...	Solar cells from Group II-VI materials
Y02E 10/544	...	Solar cells from Group III-V materials
Y02E 10/545	...	Microcrystalline silicon PV cells
Y02E 10/546	...	Polycrystalline silicon PV cells
Y02E 10/547	...	Monocrystalline silicon PV cells
Y02E 10/548	...	Amorphous silicon PV cells
Y02E 10/549	...	organic PV cells

We extend our understanding of the performance of PCS by applying the methodology for each of these classifications using the advanced search capability of the online tool. Below we first briefly review the PCS methodology and tool, and then describe our findings pertaining to the landmark patents underlying photovoltaic material technologies.

2. Patent Citation Spectroscopy

PCS is a data mining method that operates over the cited references within sets of patents. The goal is to generate a historical assessment of the most impactful patents within technological areas. The underlying PCS computation is based on a similar data mining methodology developed for use on academic literature, known as Reference Publication Year Spectroscopy (RPYS) technique (Marx, Bornmann, Barth, & Leydesdorff, 2014). This method involves aggregating the cited references across a set of retrieved documents and organizing these cited references by their publication year. For each cited reference year, the total number of references is calculated. Next, data is de-trended by taking the absolute deviation of the number of cited references for a given year from the 5-year median. As specifically applied to patents, this is represented by the equation:

$$f(t) = C_t - med(C_{t-2}, C_{t-1}, C_t, C_{t+1}, C_{t+2}), \quad (1)$$

where C represents the total sum of citations to patents granted in year t and med represents

the median. These steps do not deviate from RPYS in calculation (though RPYS was never applied to patents). However, this de-trending function only considers the aggregated cited reference activity over time. This creates a challenge in identifying seminal works because interesting outliers resulting from the de-trending equation could result from either a large surge in the influence of a single document (i.e., what we might consider a seminal work) or based on several slightly influential documents occurring in the same year. As such, PCS includes an additional normalization calculation to disentangle outliers based on the outstanding performance of a single document as compared to a group of documents:

$$PCS(t) = f(t) \cdot \frac{\text{Count of References to Most Referenced Patent in Year } t}{C_t} \quad (2)$$

This step multiplies the results from equation (1) based on the percentage of all references from that year attributable to the most referenced patent.

3. Applying Patent Citation Spectroscopy to Material Photovoltaic Technologies

At present, PCS can be applied to granted US patents using a web-application produced by Comins et al. (*under review*; <http://www.leydesdorff.net/comins/pcs/index.html>). The web-application leverages the application programming interface (API) to the public data platform PatentsView, which is supported by the USPTO Chief Economist. Users can search for patents using either keyword phrases (e.g., “photovoltaic cells”) or more advanced searches. These advanced searches follow the conventions described by the data-provider ([PatentsView](#)) documentation. Among other things, advanced search queries enable users to apply PCS to patents based on their Cooperative Patent Classification.

Using the PCS web-application, we conducted a search for the seminal patents of the nine CPC subclasses pertaining to photovoltaic solar cells. Here, we walk through the analytic routine for a single case (CPC subclass Y02E 10/541: CuInSe₂ material PV cells). In this case,

an advanced search was conducted in the PCS-application using the following query: ADVANCED={"cpc_subgroup_id":"Y02E10V541"}. This search retrieved metadata on 962 granted US patents and analyzed a total of 3,502 unique patent references. The application yields a visualization of the PCS algorithm output as well as the method's most likely seminal patent (see Figure 1). In the case of CPC subclass Y02E 10/541, the resulting seminal patent is US4335266: "Methods for forming thin-film heterojunction solar cells from I-III-IV₂" by Reid Mickelsen and Wen Chen.

To validate the results of the algorithm, we conduct a search for scholarly articles citing patent US4335266 as the underlying invention of CuInSe₂ material PV cells. In this instance, an article appearing in *Materials Science Forum* states "...in 1980, Boeing Aerospace demonstrated, for the first time, the milestone of 10 % small-area cell efficiency in the form of thin-film solar cells with a CuInSe₂ alloy system, in which they successfully invented how to prepare the p-type absorbers known as so-called 'bilayer' process [Mickelsen and Chen, US4335266]." Such articles provide corroborating evidence as the performance of PCS (cf. Leydesdorff, Alkemade, Heimeriks, & Hoekstra, 2015).

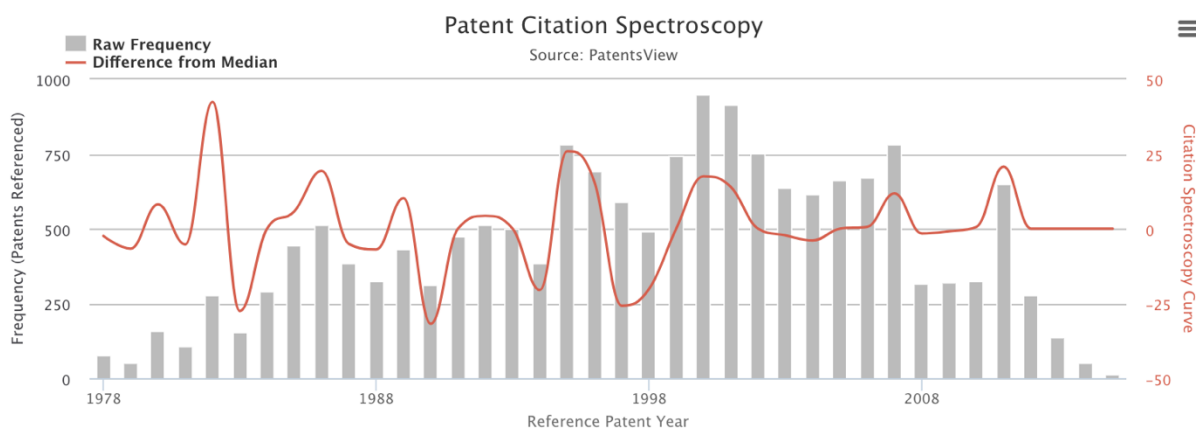


Figure 1. The PCS-derived foundational patent for CPC Subclass Y02E 10/541: CuInSe₂ material PV cells is US4335266: "Methods for forming thin-film heterojunction solar cells from I-III-IV₂" by Reid Mickelsen and Wen Chen.

A summary of results for all nine photovoltaic subclasses examined in our study are detailed in Table 1. Once a seminal patent for a given topic was identified, a traditional search of scholarly literature was performed to identify corroborating evidence of the patent's influence. In the case of 5 of the 9 subgroups analyzed, we found corroborating scholarly evidence for the seminal patents identified by the PCS algorithm.

Patent Topic	CPC Subgroup	PCS Identified Seminal Patent	Corroborating Evidence
CuInSe ₂ material PV cells	Y02E10V541	US4335266	Kushiya et al., 2012
Dye sensitized solar cells	Y02E10V542	US4927721	Longo & Paoli, 2003
Solar cells from Group II-VI materials	Y02E10V543	US5536333	Cheese et al., 2016
Solar cells from Group III-V materials	Y02E10V544	US6252287	Takamoto et al., 2005
Microcrystalline silicon PV cells	Y02E10V545	US5677236	<i>none</i>
Polycrystalline silicon PV cells	Y02E10V546	US5227329	<i>none</i>
Monocrystalline silicon PV cells	Y02E10V547	US5053083	<i>none</i>
Amorphous silicon PV cells	Y02E10V548	US4109271	<i>none</i>
Organic PV cells	Y02E10V549	US4539507	Kiy, 2002

Table 1. Summary results from our application of the PCS algorithm to 9 CPC subclasses related to photovoltaic technology. Table provide corroborating evidence, where available, for the PCS tool.

4. Summary and Conclusions

We used Patent Citation Spectroscopy—originally developed as Reference Publication Year Spectroscopy for studying landmarks and milestones in scientific literature (Comins & Leydesdorff, 2017; Thor, Marx, Leydesdorff, & Bornmann, 2016)—to patent literature classified

into the nine Y-subclasses of CPC that describe material photovoltaic technologies. In five of the nine cases, we found corroborating evidence for the foundational character of the patent indicated by the routine in 5 out of 9 cases.

For this study we extended the routine with the option to use the advanced search queries at PatentsView. On the basis of two normalizations of the longitudinal distribution of the publication years of the patents cited by the retrieved patents, the routine (at <http://www.leydesdorff.net/comins/pcs/index.html>) provides a best guess of the foundational patent for the subject specified in the string. It seems to us that the successful application in five of the nine cases and the previous results in the case of biomedical patents reported by Comins et al. (2017) provide some confidence that this indicator of fundamental patents has potential. However, the normalizations may have to be refined based on further analysis of successful and unsuccessful applications.

References

- Cheese, E., Mapes, M. K., Turo, K. M., & Jones-Albertus, R. (2016, June). US department of energy photovoltaics research evaluation and assessment. In *Photovoltaic Specialists Conference (PVSC), 2016 IEEE 43rd* (pp. 3475-3480). IEEE.
- Comins, J. A., & Leydesdorff, L. (2017). Citation algorithms for identifying research milestones driving biomedical innovation. *Scientometrics*, 110(3), 1495-1504.
- Comins, J. A., Carmack, S. A., & Leydesdorff, L. (under review). Patent Citation Spectroscopy (PCS): Algorithmic retrieval of landmark patents. *arXiv preprint arXiv:1710.03349*.
- Fire, A., Xu, S. Q., Montgomery, M. K., Kostas, S. A., Driver, S. E., & Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391(6669), 806-811.
- Jaffe, A. B., & Trajtenberg, M. (2002). *Patents, Citations, and Innovations: A Window on the Knowledge Economy*. Cambridge, MA/London: MIT Press.
- Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *the Quarterly journal of Economics*, 108(3), 577-598.
- Leydesdorff, L., Alkemade, F., Heimeriks, G., & Hoekstra, R. (2015). Patents as instruments for exploring innovation dynamics: geographic and technological perspectives on “photovoltaic cells”. *Scientometrics*, 102(1), 629-651. doi: 10.1007/s11192-014-1447-8
- Longo, C., & De Paoli, M. A. (2003). Dye-sensitized solar cells: a successful combination of materials. *Journal of the Brazilian Chemical Society*, 14(6), 898-901.
- Kiy, M. (2002). *Charge injection and transport in organic semiconductors* (Doctoral dissertation).

- Kushiya, K., Sugimoto, H., Chiba, Y., Tanaka, Y., & Hakuma, H. (2012). Change of the Characterization Techniques as Progress of CuInSe₂-based Thin-Film PV Technology. In *Materials Science Forum* (Vol. 725, pp. 165-170). Trans Tech Publications.
- Marx, W., Bornmann, L., Barth, A., & Leydesdorff, L. (2014). Detecting the historical roots of research fields by reference publication year spectroscopy (RPYS). *Journal of the Association for Information Science and Technology*, 65(4), 751-764.
- Shelton, R. D., & Leydesdorff, L. (2012). Publish or Patent: Bibliometric evidence for empirical trade-offs in national funding strategies. *Journal of the American Society for Information Science and Technology*, 63(3), 498-511.
- Takamoto, T., Agui, T., Washio, H., Takahashi, N., Nakamura, K., Anzawa, O., ... & Yamaguchi, M. (2005, January). Future development of InGaP/(In) GaAs based multijunction solar cells. In *Photovoltaic Specialists Conference, 2005. Conference Record of the Thirty-first IEEE* (pp. 519-524). IEEE.
- Thor, A., Marx, W., Leydesdorff, L., & Bornmann, L. (2016). Introducing CitedReferencesExplorer : A program for Reference Publication Year Spectroscopy with Cited References Disambiguation. *Journal of Informetrics*, 10(2), 503-515. doi: 10.1016/j.joi.2016.02.005