Patents as instruments for exploring innovation dynamics: Different perspectives on "photovoltaic cells"
Leydesdorff, L.A.; Alkemade, F.; Heimeriks, G.

Published in:
"Context Counts: Pathways to Master Big and Little Data": proceedings of the science and technology indicators conference 2014 Leiden: 3-5 September 2014 in Leiden, the Netherlands

Citation for published version (APA):
Proceedings of the science and technology indicators conference 2014 Leiden

“Context Counts: Pathways to Master Big and Little Data”

3 - 5 September 2014 in Leiden, the Netherlands

Electronic version - Edited by Ed Noyons

Published by Universiteit Leiden - CWTS 2014
Introduction
Patents are framed in different contexts: in addition to being among the outputs of the production system of knowledge, patents can also serve as input to the economic process of innovation. Furthermore, intellectual property in patents is legally regulated, for example, in national patent offices (e.g., Granstrand, 1999). Thus, different selection environments are relevant to patenting: the context of technological knowledge production, the economic context, and the legal framework of the state. Patents reflect these different contexts in terms of attributes: names and addresses of inventors and assignees provide information about the locations of inventions, patent classifications and claims within the patents can be used to map technological developments, citations provide measures of impact and value, etc. (Porter & Cunningham, 2005). Can patent analysis and patent maps provide us with an analytical lens for studying the complex dynamics of technological innovations? (e.g., Jaffe & Trajtenberg, 2002; Balconi et al., 2004; Feldman & Audretsch, 1999; Mowery et al., 2001).

In this study, we argue that a further development of methodologies is required more than of theories (Griliches, 1984) when one understands technologies as complex adaptive systems. The diffusion of a new technology in different dimensions can be simultaneous, but also delayed or changing direction. Thus, one is challenged to combine the different perspectives heuristically and yet analytically. We explore comprehensive base maps in different dimensions (cognitive, geographical, etc.) that (i) can be overlaid with information about specifically selected samples, and (ii) show the evolution of the technologies over time. Whereas several teams have generated patent maps and overlays for patent classes (Kay et al., in press; Schoen et al., 2012), our main objective is to make these overlays dynamic and interactive so that one can use them as versatile instruments across samples gathered for different purposes.

Data
Recently, USPTO and EPO introduced a new system of so-called Cooperative Patent Classifications (CPC) that unlike patent classifications such as International Patent Classifications IPC, and its American or European equivalents, is also indexed with a focus on emerging technologies using specific tags in the new Y-class (Scheu et al., 2006; Veefkind et al., 2012). Whereas the previous classification systems have grown historically with the institutions, and combine patents that cover product and process innovations at different scales, the classification in terms of CPC provides the opportunity to take a reflexive turn since technological classes are added under the category “Y” from the perspective of
hindsight. The new classifications have been backtracked into the existing databases for indexing.\(^1\)

A new CPC tag for emerging technologies was developed as Y02: “Climate Change Mitigating Technologies.” This latter tag and its subclasses are now operational in both USPTO and EPO data. The new class follows up on the “Pilot Program for Green Technologies Including Greenhouse Gas Reduction” that USPTO launched in 2009. In the meantime, more than 150,000 patents are tagged with Y02 in USPTO, among which 5,021 US patents with the search string cpc/y02e10/54$ for material technologies in photovoltaic cells (PV). We focus on developing the relevant instruments using the first subclass Y02E10/541 that covers “CuInSe\(_2\) material PV Cells.” In a next study (Heimeriks, Alkemade, & Leydesdorff, 2014), we upscale to comparisons among the nine material technologies, and including PatStat as another database of patent statistics.

Figure 1: 419 Patents granted in USPTO under the CPC tag Y02E10/541 for “CuInSe\(_2\) material PV cells”, 1975-2010; September 5, 2013

CuInSe\(_2\) was first synthesized in 1953 (Hahn et al., 1953), and proposed as a photovoltaic material in 1974 (Shafarman & Stolt, 2003: 567f.). Thin-film technology for solar cells is still considered as the commercially most promising candidate for generating energy at competitive prices although monocrystalline silicon cell technology currently dominates the market at a cost of about 0.5 Euro/Watt-peak.\(^2\) Although this technology has only a small share of the market, it continues to attract most of the funding for R&\(D\) among the material technologies for photovoltaic cells (刘壮 and 卢兰兰, 2011, p. 12).\(^3\) We retrieved 419 granted

\(^1\) The USPTO envisages replacing the US Patent Classification System (USPC) with CPC during a period of transition to 2015; at EPO, however, the European classification ECLA has already been replaced with CPC.

\(^2\) One Watts-peak (Wp) is defined as the maximum power output of a one square meter solar panel at 25 degrees centigrade.

\(^3\) The transcripts of these names in the Latin alphabet are: Liu, Zhuang and Lu, Lanlan, respectively.

365
patents at USPTO with the CPC “Y02E10/541”, and brought these records under the control of a relational database management system. Figure 1 shows the trend.

**Methods**

Existing routines for overlaying patent data to Google Maps (Leydesdorff & Bornmann, 2012) and a map based on aggregated citations among IPC (Leydesdorff, Kushnir, & Rafols, 2012) were further developed for the purpose of dynamic mapping. The resulting routines are available at [http://www.leydesdorff.net/software/patentmaps/dynamic](http://www.leydesdorff.net/software/patentmaps/dynamic).

**Geographic maps**

As specified in Leydesdorff & Bornmann (2012), the proportion of top-cited patents in a sample of USPTO data can be (z-)tested for each location against the expectation, but only in the case of more than five patents at a city-location. Using colors similar to those of traffic lights, cities with patent portfolios significantly below expectation in terms of citedness are colored dark-red and cities with portfolios significantly above expectation dark-green. Lighter colors (lime-green and red-orange) are used for cities with expected values smaller than five patents (which should not statistically be tested) and for non-significant scores above or below expectation (light-green and orange). The precise values are provided in the descriptors which can be accessed by clicking on the respective nodes. See at [http://www.leydesdorff.net/photovoltaic/cuinse2/cuinse2_inventors.htm](http://www.leydesdorff.net/photovoltaic/cuinse2/cuinse2_inventors.htm) for the aggregated set.

**Classification maps**

We use the base map of aggregated citation relations among IPC in the USPTO data 1975-2011 provided by Leydesdorff, Kushnir, and Rafols (2012). These maps are available at [http://www.leydesdorff.net/ipcmaps](http://www.leydesdorff.net/ipcmaps) for both three and four digits of the current IPC version 8. The initial step for the construction of the time-series is again the construction of the overall map for the aggregated set. Subsequently, the time series are generated by setting filters for consecutive years to this aggregate.

The routine `ipcyr.exe` (available at [http://www.leydesdorff.net/software/patentmaps/dynamic](http://www.leydesdorff.net/software/patentmaps/dynamic)) generates input information for consecutive years in the format of VOSviewer for the mapping ([http://vosviewer.com](http://vosviewer.com)). Two time series of files are generated as input for the mapping for three and four digits of IPC, respectively. The routine additionally writes a file “rao.dbf” which contains Rao-Stirling diversity for both three and four-digit IPC-based maps. Rao-Stirling diversity is defined as follows (Rao, 1982; Stirling, 2007):

\[
\Delta = \sum_{ij} p_i p_j d_{ij}
\]

where \(d_{ij}\) is a disparity measure between two IPC classes \(i\) and \(j\) at the respective level of specificity; \(p_i\) is the proportion of elements assigned to each class \(i\). As the disparity measure, we use \((1 - \text{cosine})\) since the cosine values of the citation relations among the aggregated IPC was used for constructing the base map of three and four digits (Jaffe, 1986).

---

4 The first four digits of CPC will be identical to IPC 8.
Results

Geographical diffusion

One obtains a series of maps in which the node sizes are proportionate to the logarithm of the number of patents. [We use log(n+1) in order to prevent cities with single patents from disappearing because log(1) = 0.] As noted, the node colors correspond to the quality of the patents in terms of their citedness. One can click on each node to find statistical details. (This statistical data is also stored in the file “geo.dbf” that is generated and overwritten in each run.)

Figure 2 provides the Google map for the five-year period 2000-2004. The numbers of patents are often too small for significance testing, but one can see at a glance that the US is dominant with green-coloured nodes in this (USPTO!) set in terms of both numbers and quality. In addition to the US, Japan and Europe have developed their own networks. (One can zoom in on the map at http://www.leydesdorff.net/photovoltaic/cuinse2/index.html.) During this period, international co-inventorship between the three world regions was limited to transatlantic collaborations.

Figure 2: Patent configuration during 2000-2004 for CulnSe2 material in PV Cells (Y02E-10/541) in USPTO data; an interactive version of this map is available at http://www.leydesdorff.net/photovoltaic/cuinse2/index.html.

Inspection of the animation informs us that patenting began in isolated centers in the USA, then spread first within the U.S. and thereafter also to some centers in Europe (e.g., 1983-1987). During the second half of the 1980s, Japanese and also isolated inventors in Europe began to patent in the USA. In 1990-1994, co-inventorship is found only in the local environments of Munich (Germany) and within Colorado. The latter network reflects that the
National Renewable Energy Laboratory (NERL) of the US Department of Defense is based in Golden, Colorado.

In the second half of the 1990s, there is also more co-invention in the USA and Japan, but within national boundaries. The technology increasingly becomes commercially viable during this period. The number of cities in Europe and Japan with USPTO patents increases, and transatlantic collaboration is resumed towards the end of the 1990s. Since 2003—the commercial phase—one sees co-invention between Japan and the USA, and within Europe. In the European context, France plays a role in addition to a recurrent collaboration between Germany and Spain. An address in the UK (Stirling in Scotland) joins the US networks in the final periods (2007-2011, 2008-2012). During 2008-2012, Europe is otherwise no longer represented in USPTO data.

In summary, these are sparse networks. The majority of the inventors do not collaborate beyond local environments; collaborations within nations are more important than international collaborations.

**IPC classes**

How can the map in terms of IPC-classes add to our understanding of these geographical dynamics? Figure 3 shows the IPC-based map (three digits) for the same set of patents as used in Figure 2 (2000-2004). The technology originated during the 1970s in the category of “basic electric elements” and remained there during the next 15 years, but has spread during the 1990s into other domains of technology such as “spraying and atomizing” and machine techniques for making thin films in photovoltaic cells. This diffusion increases further during the 2000s.
Figure 3: Map of USPTO patents in terms of IPC at the three-digit level for the period 2000-2004. A dynamic version of this map is available at http://www.leydesdorff.net/photovoltaic/cuinse2/cuinse2.ppsx.

Figures 2 and 3 can be combined using frames in the html for the splitting of the screens (at http://www.leydesdorff.net/photovoltaic/cuinse2/dualmix.html; not shown here. One can animate this figure as Figure 2, but this animation taught us that dynamic changes in two different (split) screens are difficult to handle for an analyst.

A user needs control over the time steps when focusing on the differences between two dynamics. Therefore, we propose another solution: by clicking on another year, one opens a new window in the browser with the same figures for this different year. A user is then able to compare among years using, for example, different time intervals (such as five or ten years) by going back and forth between windows, at one’s own pace.
Rao-Stirling diversity as a measure of technological change

Figure 4: The development of Rao-Stirling diversity in IPC (three and four digits) among 419 USPTO-patents with CPC Y02E10/541 (“CuInSe₂ material PV cells”) during the period 1975-2012.

Figure 4 shows the development of Rao-Stirling diversity in the IPC-based maps during the entire period. The figure suggests that the technology was developed in three cycles. Two of the valleys, i.e., the period of convergence in the late 1980s and the latest convergent period, correspond with breakthroughs in the efficiency of thin-film solar cells (Green et al., 2013). Combining the maps with split-screens (at http://www.leydesdorff.net/photovoltaic/cuinse2/dualmix.html) for each consecutive year, we suggest specifying these cycles as follows (Shafarman & Stolt, 2003):

1. an early cycle during the 1980s which is almost exclusively American; after initial development of the technology at Bell Laboratories in the ’70s, Boeing further developed the solar cells using these materials;
2. a second cycle during the 1990s that includes transatlantic collaboration and competition with Europe; and
3. a third and current cycle—the commercial phase—in which American-Japanese collaboration, on the one side, and collaboration within Europe, on the other, prevail.

The volume of patents continued to increase more smoothly, but with an increasing (above-exponential) rate during the most recent years (Figure 1). The pronounced articulation of these cycles in terms of Rao-Stirling diversity came as a surprise to us. As the material technology becomes mature, other technologies such as spraying the thin film on carrier materials may become crucial.
Conclusions
The maps of patents in different dimensions are instrumental to understanding the complex
dynamics of innovation by providing different projections of these dynamics. We
distinguished in this study between IPC-based maps that show the technological organization
of the patents in a vector space, the geographic maps as overlays to Google Maps, and the
social networks—not shown here—that can be overlaid to the geographic map, but can also
be studied in themselves using graph-theoretical instruments such as spring-embedded
layouts.

At the theoretical level, we thus aim to address what Griliches (1994) called “the computer
paradox” from a methodological angle: ever more data—nowadays, one would say “big
data”—are stored in ever larger repositories, but the logic of these repositories is institutional,
whereas the logic of innovation is based on the transversal recombination of functions at
interfaces (e.g., supply and demand). The relabeling using the Y-tag in CPC, however,
provides an opportunity to follow delineated technologies within and across databases: recent
agreements of EPO and USPTO with the Chinese, Korean, and Russian patent offices to use
also CPC in the near future show an increased awareness to coordinate the data in a
networked mode.

As innovations relate at structural interfaces—between selection environments—one can
consider them from different perspectives such as market opportunities or technological
novelty. Using a single (theoretical) term such as “diffusion,” is then foreseeably insufficient
without specification of the different systems of reference: diffusion can be defined in terms
of markets/industries, geographies, or also technologies—in terms of branching and
recombination (Arthur, 2009). As we argued, patents provide an (albeit imperfect) lens to this
complex dynamics.

The use of Rao-Stirling diversity in this study (Figure 4) can be considered as a case in point:
the literature pointed us to considering variety versus the loss of variety in shake-out phases
as central to techno-economic developments (Anderson & Tushman, 1990), but the data
allowed us to operationalize this in relation to the new instruments. The extension beyond two
maps to be recombined follows as a progressive research agenda for quantitative innovation
studies (Rotolo et al., in preparation).

Acknowledgements
We are grateful to comments on a previous version of the manuscript by Jan Youtie (Georgia
Tech) and Wilfried van Sark (Utrecht University).

References:


Green, M. A., Emery, K., Hishikawa, Y., Warta, W., & Dunlop, E. D. (2013), Solar cell
efficiency tables (version 41). Progress in Photovoltaics: Research and Applications,
21(1), 1–11; doi: 10.1002/pip.2352


