Patent Mapping

Alan L. Porter,[*] Luciano Kay,[**] Nils C. Newman,[***] Jan Youtie[****]

[*] *alan.porter@isye.gatech.edu*
Technology Policy & Assessment Ctr., Georgia Tech, Atlanta GA, 30302-0345 (USA)
[**] luciano.kay@ucsb.edu
Center for Nanotechnology in Society, University of California, Santa Barbara, CA (USA)
[***] *newman@iisco.com*
Intelligent Information Services Co., Norcross GA (USA)
[****] jy5@mail.gatech.edu
Enterprise Innovation Institute, Georgia Tech, Atlanta GA 30302

Introduction

Visual analytics are increasingly being used to help grasp the complexity and evolution of scientific and technological (S&T) activities over time. Aims include understanding concentrations of activity over S&T areas and across organizations. New and diverse analysis and mapping methods, increasing computing power, and new software and network algorithms enable this. Scientometrics is increasingly extending attention from research publications to patent analyses, together aimed at understanding a range of innovation-related phenomena.

Proposed Activities

This workshop presents the new global patent mapping and overlay technique on which several of us have been working (Kay et al., 2012). In this session, we will present this patent mapping technique that disaggregates IPC hierarchies and re-aggregates them based on cross-citation patterns. We compare our approach based on an analysis of European Patent Office patents under the IPC8 regime versus what was previously found in the IPC7 regime.

Differences between the two patent maps and the reasons for these differences will be explored. In addition, measures of technological distance will be presented for a set of patents (case study on Nano-Enabled Drug Delivery -- NEDD) under each of the two regimes to enable further comparisons. The implications for developing patent-mapping based indicators will be explored. We intend to enrich that discussion by demonstrating the maps live, allowing for exploration of suggested alterations real-time (to some degree).

Background

Some other patent maps also use IPC categories (c.f., Hinze et al 1997; Leydesdorff et al., in press). Our approach differs in blending multiple class levels in the visualization. It also gets at technological relationships not well-captured by the IPC classification per se. The approach compensates for these issues by (1) disaggregating IPC categories, and (2) reforming them based on citing-to-cited reference patterns.

Our IPC7 global patent map is based on citing-to-cited relationships among IPCs of European Patent Office (EPO) patents from 2000-2006. This period was chosen because of its stability with respect to IPC 7 categories. The dataset containing IPC relationships, extracted from the *Patstat* database in 2010, represents more than 760,000 patent records in more than 400 IPC categories.

Our methodology involves disaggregating, then folding IPC categories up into the next highest level of aggregation to create 466 relatively similar sized categories (i.e. within two orders of magnitude) for IPC7. The next step involves extracting from *Patstat* the patents cited by the target records. The IPCs of those patents are mapped to the 466 IPC categories. Data processing involves generating a cosine similarity matrix among citing IPC categories and then factor analysis of the IPC categories (following the method used in global science maps by Leydesdorff and Rafols (2009). A factor analysis of the citing-to-cited matrix among IPC categories is then used to consolidate the 466 categories into 35 "macro patent categories" that form the basis for color-coding the maps. The visualizations also require converting IPC codes to succinct text labels. The full map of patents shows all 466 categories in a Kamada-Kawai layout.

The new IPC8 patent map is similarly formulated. We capture some 1,000,000 PATSTAT records for EPO patents from 2007. We increase the threshold for inclusion of IPC categories from 1000 to 1500, to deal with a comparable magnitude of resulting categories. As before, we grab the cited patent records and extract Cited IPC information. We obtain 434 categories that factor analyze into 35 usable macro categories (accounting for some half of the variance) plus a set of categories that do not load well on any factors . We construct both 434-node and 35 node maps, in Pajek and in VOSViewer. We present and contrast the basemaps, noting their limitations, and a set of overlays for NEDD patenting.

Purpose and Intended Audience

The purpose of this workshop is to stimulate the discussion about patent mapping methods, their potential and limitations. Based on the global map/overlay maps method developed by the authors, we discuss findings and issues found in mapping NEDD and graphene patent datasets and address specific questions such as:

- What are the benefits and drawbacks of mapping hybrid patent categories as opposed to staying within patent classification systems (e.g., 3-digit classes, 4-digit subclasses)
- What indicators can be used to compare patent maps?
- How much change takes place in IPC/CPC categories and how does this change affect maps from different classification systems (IPC 7 versus IPC8 versus CPC)?
- What level of change is deemed significant enough to call into question use of these maps?
- What benefits and limitations exist in applying patent maps to different types of technology portfolios?

Researchers, policy makers, and R&D managers can find value in patent maps to assess where and how knowledge migrates and integrates across application areas. Exposure of the approach and variations to the GTM community is vital to obtain feedback on the validity of the representations, alternative presentation modalities, and advice on making the patent overlay mapping accessible to our research community.

Workshop Logistics

1) Presentation and discussion of the basic patent mapping approach;
2) Examination and discussion of case studies (NEDD, Graphene), alternative visualizations and tools;
3) Comparison of maps, indicators, measures of diversity;
4) Discussion of how best to provide easy access for mapping alternative patent data resources; and,

5) Exploration of "next steps," particularly concerning ways to combine science and patent overlay mapping to gain richer perspective on ST&I dynamics.

References & Citations

Hinze, S., Reiss, T., & Schmoch, U. (1997). *Statistical Analysis on the Distance between Fields of Technology*. Innovation Systems and European Integration (ISE), Targeted Socio-Economic Research Program, 4th Framework Program of the European Commission (DGXII), Karlsruhe, Germany.

Kay, L., Newman, N., Youtie, J., Porter, A.L., and Rafols, I. (2012). Patent Overlay Mapping: Visualizing Technological Distance. arXiv:1208.4380 [physics.soc-ph]; and Journal of the American Society for Information Science and Technology (*JASIST)*.

Leydesdorff, L., Kushnir, D., & Rafols, I. (In press). Interactive overlay maps for US patent (USPTO) data based on International Patent Classification (IPC). *Scientometrics*, 1–17.

Leydesdorff, L., & Rafols, I. (2009). A Global Map of Science Based on the ISI Subject Categories. *Journal of the American Society for Information Science and Technology, 60*(2), 348–362.

Technology Review coverage: http://www.technologyreview.com/view/518991/global-patent-map-reveals-the-structure-of-technological-progress/ [Sep. 2, 2013]

InfoPorn, Tech's patently clear connections, *Wired-UK*, February, 2014, 24-25 + notes on p. 145.