Data-Driven Science Policy

Katy Börner
@katycns

Victor H. Yngve Distinguished Professor of Engineering and Information Science
Director, Cyberinfrastructure for Network Science Center
Indiana University
United States

Reference:

Science Center World Congress, Tokyo, Japan

November 16, 2017
Maps of Science & Technology

Using large scale datasets, advanced data mining and visualization techniques, and substantial computing resources.
Maps of Science & Technology
http://scimaps.org

101st Annual Meeting of the Association of American Geographers, Denver, CO.
April 5th - 9th, 2005 (First showing of Places & Spaces)

University of Miami, Miami, FL.
September 4 - December 11, 2014.

Duke University, Durham, NC.
January 12 - April 10, 2015

The David J. Sencer CDC Museum, Atlanta, GA.

100 maps and 12 macroscopes by 215 experts on display at 354 venues in 28 countries.
Map of Scientific Collaborations from 2005-2009

Computed Using Data from Elsevier's Scopus

Stream of Scientific Collaborations Between World Cities - Olivier H. Beauchesne - 2012
Examining the Evolution & Distribution of Patent Classifications

Managing Growing Patent Portfolios
Organizations, businesses, and individuals rely on patents to protect their intellectual property and business models. As market competition increases, protecting innovation and intellectual property rights becomes ever more important.

Managing the staggering number of patents demands new tools and methodologies. Grouping patents by their classifications offers an ideal resolution for better understanding how intellectual borders are established and change over time.

The charts below show the annual number of patents granted from January 1, 1993 to December 31, 2002. In the United States Patent and Trademark Office (USPTO) patent archive, slow and fast growing patent classes; the top 10 fast growing patent subclasses; and two evolving patent portfolios.

The Structure and Evolution of the Patent Space
The United States Patent and Trademark Office assigns each patent to one of more than 450 classes covering broad application domains. For example, class 514 encompasses all patents dealing with "Drug, Bio-Affecting and Body Treating Compositions." Classes are further broken down by subclasses that have hierarchical associations. As one example, class 455 features subclass 99 entitled "with vehicle."

The top 10 fast growing patent classes for 1998–2002 are listed together with the number of patents granted. Most come from the 'Computer and Communications' and the 'Drugs and Medical' area.

Top-10 Subclasses

<table>
<thead>
<tr>
<th>Class</th>
<th>Title</th>
<th># of Patents</th>
</tr>
</thead>
<tbody>
<tr>
<td>514</td>
<td>Drug, Bio-Affecting and Body Treating Compositions</td>
<td>18,778</td>
</tr>
<tr>
<td>438</td>
<td>Semiconductors Device Manufacturing/Process</td>
<td>17,275</td>
</tr>
<tr>
<td>435</td>
<td>Chemistry, Molecular Biology and Microbiology</td>
<td>17,474</td>
</tr>
<tr>
<td>454</td>
<td>Drug, Bio-Affecting and Body Treating Compositions</td>
<td>13,577</td>
</tr>
<tr>
<td>438</td>
<td>Stock, Materials or Miscellaneous Articles</td>
<td>13,314</td>
</tr>
<tr>
<td>577</td>
<td>Active Solid-State Devices (e.g., Transistors, Solid-State Diode)</td>
<td>12,824</td>
</tr>
<tr>
<td>395</td>
<td>Information Processing System Organization</td>
<td>9,955</td>
</tr>
<tr>
<td>359</td>
<td>Optical Systems and Elements</td>
<td>9,151</td>
</tr>
<tr>
<td>365</td>
<td>Static Information Storage and Retrieval</td>
<td>8,392</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>130,910</td>
</tr>
</tbody>
</table>

Patent Portfolio Analysis
A longitudinal analysis of portfolios reveals different patenting strategies. For each year (given in gray above each treemap), a treemap of all new patents granted to the assignees is shown. The number of patents is given below each treemap. The same size and color coding as above was used. In addition, yellow indicates that no patent has been granted in that class in the last 5 years.

Apple Computer, Inc.
Apple Computer, Inc.'s portfolio starts in 1980 and increases considerably in size over time. In most years, more than half of Apple Computer's patent filings were placed into four classes, namely 350 Information Processing System Organization, 345 Computer Graphics Processing, Operator Interface Processing, and Selective Visual Display Systems, 395 Image Analysis, and 707 Data Processing: Database and File Management or Data Structures. These four classes are an integral part of Apple Computer, Inc.'s patent portfolio, receiving patents every year.

Jerome Lemelson
The patent portfolio of Jerome Lemelson shows a very different activity pattern. Starting in 1976, he publishes between 6–20 patents each year. However, the predominance of yellow shows that there is little continuity from previous years in regards to the classes into which patents are filed. No class dominates. Instead, more and more new intellectual space is claimed.

IV.5 Examining the Evolution & Distribution of Patent Classifications - Daniel O. Kutz, Katy Borner, and Elisha F. Hardy - 2004
The Product Space

World trade flow data compiled by Feenstra et al. and available at the National Bureau of Economic Research were used to identify the complete co-export matrix of 775 industrial products for 1998-2000. A Minimum Spanning Tree (MST) algorithm was used to reduce the complete co-export matrix to less than 1% of the links. The resulting network, which combines the MST plus all links with a co-export frequency of at least 0.1% was laid out using a force-directed layout algorithm. Node sizes represent the value of traded products in millions of U.S. dollars. Their color corresponds to ten product groups identified using the Leamer classification. Each product class is labeled by an icon. Link color and width indicate the frequency of joint exports.

The network has a core-periphery structure with higher value product classes, e.g., machinery and chemicals, in the core and lower quality classes, e.g., fishing and garments, in the periphery. Products at the core of the network are highly interconnected while products in the periphery are sparsely interlinked.

Each country has a certain product export footprint. Relevant exports by ‘industrialized Countries’, ‘East Asia Pacific’ and ‘Latin America & the Caribbean’ are given on the right.

Traditional growth theory assumes that there is always a more sophisticated product within reach. However, given the core-periphery structure of the product space, the distances between products differ considerably.

Countries that operate in the core have capabilities to develop and manufacture a wide range of products. Yet, countries that mostly operate in the periphery of the product space have much fewer opportunities for diversification. A country’s current footprint and the structure of the product space have a major impact on a country’s future development.
Pulse of the Nation: U.S. Mood Variations Inferred From Twitter

Mood Variations
A number of interesting trends can be observed in the data. First, overall daily variations can be seen from the graph, with the early morning and late evening having the highest levels of tweeting. Second, geographic variations can be observed from the graph, with a slight majority of tweets west, likely due to the consistency of these trends being the result of daily habits.

Weekly Variations
Weekly trends can be observed as well, with weekends much higher than weekdays.

About the Data and Visualization
The posts were calculated using over 300 million tweets (Sep 2009 – Aug 2009) collected by MPI-SBS researchers, represented as density preserving cartograms. The mood of each tweet was inferred using a raw word list (Bradley, M. & Lang, P. J. Affective norms for English words (Anew): Stimuli, instructions, manuals and affective ratings. T. Technical report 2.1, The Center for Research on Psychopathology and University of Oxford). Location data was taken from the U.S. Census Bureau at http://tigerweb.usgs.gov, and the baseline U.S. map was taken from NatureMaps (Common). All location’s were inferred using the Google Maps API and mapped into countries using PostGIS and U.S. county maps from the U.S. National Atlas. Mood colors were created using Color Brewer 2.

About Cartograms
A cartogram is a map in which the mapped variable (in this case, the number of tweets) is scaled for the true land area. Thus, the geometry of the actual map is altered so that the shape of each region is maintained as much as possible, but the area is scaled in order to be proportional to the number of tweets that it contains in the region. This is a density equalizing map. The cartograms in this work were generated using the Carto software by Mark E. J. Newman.

Northeastern University
College of Computer and Information Science
Center for Complex Network Research
http://www.cos.nyu.edu/home/amjs/twittermood

HARVARD UNIVERSITY

© 2010 Alan Mislove1, Sune Lehmann5, Yong-Yeol Ahn3, Jukka-Pekka Onnela4, and James Niels Rosenquist - 2010

IX.4 Pulse of the Nation - Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and James Niels Rosenquist - 2010
Chemical Research & Development Powers the U.S. Innovation Engine

Macroeconomic Implications of Public and Private R&D Investments in Chemical Sciences

INVESTMENT IN CHEMICAL SCIENCE R&D

FEDERAL GOVERNMENT

$1 Billion
FEDERAL FUNDING

$5 Billion
INDUSTRY FUNDING

$8 Billion
TAXES

$1B
$1B + $5 Billion
$10 Billion
CHEMICAL INDUSTRY OPERATING INCOME

$40 Billion
GROWTH IN GNP

+ 600,000 JOBS CREATED

TIMELINE FROM CONCEPTION TO COMMERCIALIZATION

4-5 YRS
FOUNDATIONAL RESEARCH

9-11 YEARS
INVENTION DEVELOPMENT

> 5 YEARS
TECHNOLOGY COMMERCIALIZATION

20 YEARS

U.S. ECONOMY

The design shows that an input of $1B in federal investment, leveraged by $5B in industry investment, brings new technologies to market and results in $10B of operating income for the chemical industry. $40B of growth in the Gross National Product (GNP) and further impacts the US economy by generating approximately 600,000 jobs, along with a return of $8B in taxes. Additional details, also reported in the CCR studies, are depicted in the map to the left. This map clearly shows the two R&D investment cycles: the shorter industry investment at the innovation stage to commercialization cycle; and the longer federal investment cycle which begins in basic research and culminates in national economic and job growth along with the increase in tax base that in turn are available for investment in basic research.
Check out our Zoom Maps online!

Visit scimaps.org and check out all our maps in stunning detail!
MAPS

vs.

MACROSCOPES
Microscopes & Telescopes vs. MACROSCOPES
Iteration XI (2015): Macrosopes for Interacting with Science
http://scimaps.org/iteration/11
earth ≡

*Earth* – Cameron Beccario
The News Co-occurrence Globe
An interactive visualization of how countries are mentioned together in the world's news media

2.92K COOCCUR%

UNITED KINGDOM cooccurrences in: 2,922%
cooccurrences out: 80%

Mapping Global Society – Kalev Leetaru
Iteration XII (2016): Macroscopes for Making Sense of Science

http://scimaps.org/iteration/12
Smelly Maps – Daniele Quercia, Rossano Schifanella, and Luca Maria Aiello – 2015
This is the Roanoke (Raleigh) megaregion.

Models of Science & Technology

Using large scale datasets, advanced data mining, modeling, and visualization techniques, and substantial computing resources.
Government, academic, and industry leaders discussed challenges and opportunities associated with using big data, visual analytics, and computational models in STI decision-making.

Conference slides, recordings, and report are available via http://modsti.cns.iu.edu/report
Upcoming Colloquia

Unless otherwise indicated, most Sackler colloquia are held at the Arnold and Mabel Beckman Center, in Irvine, California.

*Reproducibility of Research: Issues and Proposed Remedies*

March 8-10, 2017; Washington, D.C.
Organized by David B. Allison, Richard Shiffrin and Victoria Stodden
Registration now open

*Science of Science Communication III*

November 15-16, 2017; Washington, D.C.
Organized by Karen Cook, Baruch Fischhoff, Alan I. Leshner and Dietram A. Scheufele
Registration will open May 2017

*Modelling and Visualizing Science and Technology Developments*

December 4-5, 2017; Irvine, CA
Organized by Katy Börner, William Rouse and H. Eugene Stanley
Registration will open August 2017

http://www.nasonline.org/programs/sackler-colloquia/upcoming-colloquia
References


All papers, maps, tools, talks, press are linked from http://cns.iu.edu
These slides are at http://cns.iu.edu/presentations.html
CNS Facebook: http://www.facebook.com/cnscenter
Mapping Science Exhibit Facebook: http://www.facebook.com/mappingscience