Mining, Mapping, and Accelerating Scholarly Networks

Katy Börner
Cyberinfrastructure for Network Science Center, Director
Information Visualization Laboratory, Director
School of Library and Information Science
Indiana University, Bloomington, IN
katy@indiana.edu

With special thanks to the members at the Cyberinfrastructure for Network Science Center; the Sci2, NWB, and EpiC teams; and the VIVO Collaboration

Shared Horizons: Data, Biomedicine, and the Digital Humanities
Maryland Institute for Technology.

April 11, 2013

Find your way
Find collaborators, friends
Identify trends

Terra bytes of data

Descriptive & Predictive Models
### Type of Analysis vs. Level of Analysis

<table>
<thead>
<tr>
<th></th>
<th>Micro/Individual (1-100 records)</th>
<th>Meso/Local (101–10,000 records)</th>
<th>Macro/Global (10,000 &lt; records)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistical Analysis/Profiling</strong></td>
<td>Individual person and their expertise profiles</td>
<td>Larger labs, centers, universities, research domains, or states</td>
<td>All of NSF, all of USA, all of science.</td>
</tr>
<tr>
<td><strong>Temporal Analysis (When)</strong></td>
<td>Funding portfolio of one individual</td>
<td>Mapping topic bursts in 20-years of PNAS</td>
<td>113 Years of Physics Research</td>
</tr>
<tr>
<td><strong>Geospatial Analysis (Where)</strong></td>
<td>Career trajectory of one individual</td>
<td>Mapping a states intellectual landscape</td>
<td>PNAS publications</td>
</tr>
<tr>
<td><strong>Topical Analysis (What)</strong></td>
<td>Base knowledge from which one grant draws.</td>
<td>Knowledge flows in Chemistry research</td>
<td>VxOrd/Topic maps of NIH funding</td>
</tr>
<tr>
<td><strong>Network Analysis (With Whom?)</strong></td>
<td>NSF Co-PI network of one individual</td>
<td>Co-author network</td>
<td>NIH’s core competency</td>
</tr>
</tbody>
</table>
Mapping Indiana’s Intellectual Space

Identify
- Pockets of innovation
- Pathways from ideas to products
- Interplay of industry and academia

Individual Co-PI Network
Ke & Börner, (2006)
Mapping the Evolution of Co-Authorship Networks

Research question:
• Is science driven by prolific single experts or by high-impact co-authorship teams?

Contributions:
• New approach to allocate citational credit.
• Novel weighted graph representation.
• Visualization of the growth of weighted co-author network.
• Centrality measures to identify author impact.
• Global statistical analysis of paper production and citations in correlation with co-authorship team size over time.
• Local, author-centered entropy measure.

Mapping Transdisciplinary Tobacco Use Research Centers Publications
Compare R01 investigator based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.
Zoss & Börner, forthcoming.
Supported by NIH/NCI Contract HHSN261200800812
Co-word space of the top 50 highly frequent and bursty words used in the top 10% most highly cited PNAS publications in 1982-2001.


Spatio-Temporal Information Production and Consumption of Major U.S. Research Institutions


Research questions:
1. Does space still matter in the Internet age?
2. Does one still have to study and work at major research institutions in order to have access to high quality data and expertise and to produce high quality research?
3. Does the Internet lead to more global citation patterns, i.e., more citation links between papers produced at geographically distant research institutions?

Contributions:
- Answer to Qs 1 + 2 is YES.
- Answer to Qs 3 is NO.
- Novel approach to analyzing the dual role of institutions as information producers and consumers and to study and visualize the diffusion of information among them.
<table>
<thead>
<tr>
<th>Type of Analysis vs. Level of Analysis</th>
<th>Micro/Individual (1-100 records)</th>
<th>Meso/Local (101–10,000 records)</th>
<th>Macro/Global (10,000 &lt; records)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical Analysis/Profiling</td>
<td>Individual person and their expertise profiles</td>
<td>Larger labs, centers, universities, research domains, or states</td>
<td>All of NSF, all of science, all of USA, all of funding</td>
</tr>
<tr>
<td>Temporal Analysis (When)</td>
<td>Funding portfolio of one individual</td>
<td>Mapping topic bursts of PNAS</td>
<td>113 Years of Physical Review</td>
</tr>
<tr>
<td>Geospatial Analysis (Where)</td>
<td>Career trajectory of one individual</td>
<td>Mapping a country's intellectual landscape</td>
<td>PNAS publications</td>
</tr>
<tr>
<td>Topical Analysis (What)</td>
<td>What one grant draws</td>
<td>Knowledge flows in research domains</td>
<td>VxOrd/Topic maps of NIH funding</td>
</tr>
<tr>
<td>Network Analysis (With Whom?)</td>
<td>NSF's network of one individual</td>
<td>NIH's network of core competencies</td>
<td>Cyberinfrastructure for Network Science Center</td>
</tr>
</tbody>
</table>
Mapping Science Exhibit – 10 Iterations in 10 years

http://scimaps.org/

Mapping Science Exhibit at MEDIA X was on May 18, 2009 at Wallenberg Hall, Stanford University,
Places & Spaces Digital Display in North Carolina State's brand new Immersion Theater

Clickstream Map of Science

Eric Fischer. 2012. Language Communities of Twitter.


Science Maps in “Expediton Zukunft” science train visiting 62 cities in 7 months 12 coaches, 300 m long
Opening was on April 23rd, 2009 by German Chancellor Merkel
http://www.expedition-zukunft.de

Illuminated Diagram Display
soon on display at the Smithsonian in DC.
http://scimaps.org/exhibit_info/#ID
Ingo Günther’s Worldprocessor globe design now on display at the Giant Geo Cosmos OLED Display at the Museum of Emerging Science and Innovation in Tokyo, Japan.
VIVO International Researcher Network

**VIVO: A Semantic Approach to Creating a National Network of Researchers** ([http://vivoweb.org](http://vivoweb.org))

- Semantic web application and ontology editor originally developed at Cornell U.
- Integrates research and scholarship info from systems of record across institution(s).
- Facilitates research discovery and cross-disciplinary collaboration.
- Simplify reporting tasks, e.g., generate biosketch, department report.

**Funded by $12 million NIH award.**

**Cornell University**: Dean Krafft (Cornell PI), Manolo Bevia, Jim Blake, Nick Cappadona, Brian Caruso, Jon Corson-Rikert, Elly Cramer, Medha Devare, John Fereira, Brian Lowe, Stella Mitchell, Holly Mistlebauer, Anup Sawant, Christopher Westling, Rebecca Younes. **University of Florida**: Mike Conlon (VIVO and UF PI), Cecilia Botero, Kerry Brit, Erin Brooks, Amy Buhler, Elle Bushhouse, Chris Case, Valrie Davis, Nita Ferree, Chris Haines, Rae Jesano, Margeaux Johnson, Sara Kreinert, Yang Li, Paula Markes, Sara Russell Gonzalez, Alexander Rockwell, Nancy Schaefer, Michele R. Tennant, George Hack, Chris Barnes, Narayan Raun, Brenda Stevens, Alicia Turner, Stephen Williams. **Indiana University**: Katy Borner (IU PI), William Barnett, Shanshan Chen, Ying Ding, Russell Duhon, Jon Dunn, Micah Linnemeier, Nian Ma, Robert McDonald, Barbara Ann O’Leary, Mark Price, Yuyin Sun, Alan Walsh, Brian Wheeler, Angela Zoss. **Ponce School of Medicine**: Richard Noel (Ponce PI), Ricardo Espada, Damaris Torres. **The Scripps Research Institute**: Gerald Joyce (Scripps PI), Greg Dunlap, Catherine Dunn, Brant Kelley, Paula King, Angela Murrell, Barbara Noble, Cary Thomas, Michaelene Trimarchi. **Washington University, St. Louis**: Rakesh Nagarajan (WUNTL PI), Kristi L. Holmes, Sunita B. Koul, Leslie D. McIntosh. **Weill Cornell Medical College**: Curtis Cole (Weill PI), Paul Albert, Victor Brodsky, Adam Cheriff, Oscar Cruz, Dan Dickinson, Chris Huang, Itay Klar, Peter Michelini, Grace Migliorisi, John Ruffing, Jason Speckland, Tru Tran, Jesse Turner, Vinay Varughese.
Temporal Analysis *(When)* Temporal visualizations of the number of papers/funding award at the institution, school, department, and people level
**Topical Analysis (What)** Science map overlays will show where a person, department, or university publishes most in the world of science. (in work)

**Network Analysis (With Whom?)** Who is co-authoring, co-investigating, co-inventing with whom? What teams are most productive in what projects?
**Geospatial Analysis (Where)** Where is what science performed by whom? Science is global and needs to be studied globally.

---

**VIVO On-The-Go** Overview, Interactivity, Details on Demand come to commonly used devices and environments.
Online Interactive Maps for Sustainability Research and NIH

http://mapsustain.cns.iu.edu
The geographic map at **state level**.

The geographic map at **city level**.
Search result for “corn”
Icons have same size but represent different #records

Click on one icon to display all records of one type.
Here publications in the state of Florida.
Detailed information on demand via original source site for exploration and study.

The science map at 13 top-level scientific disciplines level.
The science map at 554 sub-disciplines level.
United States Patent
Ding, et al.

Thermal tolerant amylase from Acidothermus cellulolyticus

Abstract

The invention provides a thermal tolerant (thermostable) amylase, Amy3II, that is a member of the glycoside hydrolase (GH) family. Amy3II was isolated and characterized from Acidothermus cellulolyticus and, like many amylases, the disclosed polypeptide and/or its derivatives may be useful for the conversion of biomass into biofuels and chemicals.

Inventors: Ding; Shi-You (Golden, CO), Adney; William S. (Golden, CO), Vinrant; Todd B. (Golden, CO), Himmel; Michael E. (Lafayette, CO)
Assignee: Midwest Research Institute (Kansas City, MO)

https://app.nihmaps.org
https://app.nihmaps.org
The purpose of this study is to better understand how lifestyle factors and their interaction with genetic factors influence a woman's risk of developing breast cancer. In order to learn more about the causes of breast cancer, we need to compare the lifestyles of people who have breast cancer with those who do not. 600 women are expected to be enrolled.

https://app.nihmaps.org
The Information Visualization MOOC

ivmooc.cns.iu.edu

Students come from 93 countries
300+ faculty members
#ivmooc

Video and paper are at [http://www.scivee.tv/node/27704](http://www.scivee.tv/node/27704)

Sci² Tool – “Open Code for S&T Assessment”

OSGi/CI Shell powered tool with NWB plugins and many new scientometrics and visualizations plugins.

Mapping NEH awards and MEDLINE publications, 1980-2009

**NEH Grants:**
41,258 grants of 47,197 started between 1980 and 2009, encompassing 3.21 billion of the 3.77 billion dollars awarded.

*Geo-coding by zipcode:*
36,512 of 41,258 grants encompassing 3.13 billion of a potential 3.21 billion were geocoded to 3,510 distinct locations.

*Science-coding by topic:*
37,132 of 41,258 grants encompassing 2.09 billion of a potential 3.21 billion were mapped to 42 distinct subdisciplines.

**MEDLINE publications:**
12.95 million papers were published between 1980 and 2009.

*Geo-coding:*
Not possible with the data we have.

*Science-coding by journal:*
11.62 million of 12.95 million papers were science located (89.7%). Science located 5,941 out of 14,561 journals (40.8%) to 415 distinct subdisciplines.
Topical Visualization: UCSD Map of Science
NEH grants, mapped based on categories

- **Chemical, Mechanical, & Civil Engineering**
  - 13,469,200 Mechanical Design Engineering

- **Electrical Engineering & Computer Science**
  - 153,044,896 Library Science, Information Retrieval
  - 724,505 Logic

- **Humanities**
  - 509,399,136 American History
  - 101,708,568 Art History
  - 87,938,056 Asian Studies
  - 15,956,450 Biblical Literature
  - 33,688,248 Classics
  - 347,050 Critical Studies
  - 36,530,032 Cross Disciplinary Study
  - 238,170,928 English Literature
  - 8,471 Ethics
  - 4,126,550 German Studies
  - 38,912,520 Hispanic Studies
  - 2,234,555 Italian Studies
  - 36,181,320 Linguistics
  - 33,062,848 Literary Criticism
  - 56,233,540 Medieval History
  - 65,690,320 Modern Language
  - 98,065,344 Music & Theatre
  - 88,949,420 Philosophy Psychology
  - 42,320,544 Science History
  - 905,530 Semiotics
  - 52,800,752 Social History
  - 100,976,304 Socio-Cultural Anthropology

- **Social Sciences**
  - 2,593,480 Child Development
  - 1,652,100 Communication Research
  - 1,482,022 Economics
  - 58,847,100 Education
  - 53,233,792 Ethnology
  - 79,414,408 Higher Education
  - 71,120 Human Resource Management
  - 15,179,400 International Development
  - 3,008,390 International Economics
  - 10,232,764 Law
  - 4,693,470 Political Geography
  - 28,897,260 Political Science
  - 832,157 Public Administration
  - 21,255,080 Regional Studies
  - 1,967,490 Rural Studies
  - 15,774,390 Sociology
  - 4,333,450 Urban Studies
Topical Visualization: UCSD Map of Science
Medline papers, mapped based on journal names

References


All papers, maps, tools, talks, press are linked from http://cns.iu.edu

CNS Facebook: http://www.facebook.com/cnscenter
Mapping Science Exhibit Facebook: http://www.facebook.com/mappingscience