Science Maps: How to Analyze, Map, and Make Sense of Science

Dr. 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

National Research Council, Ottawa, Canada
10:00am, July 7th, 2008

Science Maps in Action

Computational Scientometrics:
Studying Science by Scientific Means

- Places & Spaces: Mapping Science exhibit, see also http://scimaps.org.

Mapping Science Exhibit – 10 Iterations in 10 years

The Power of Maps (2005)


The Power of Forecasts (2007)

Science Maps for Economic Decision Makers (2008)

Science Maps for Science Policy Makers (2009)

Science Maps for Scholars (2010)

Science Maps as Visual Interfaces to Digital Libraries (2011)

Science Maps for Kids (2012)

Science Forecasts (2013)

How to Lie with Science Maps (2014)

Illuminated Diagram Display


Mapping, Illuminating, and Interacting with Science.

SIGGRAPH 2007, San Diego, C.A.
Advantages for Funding Agencies

- Supports monitoring of (long-term) money flow and research developments, evaluation of funding strategies for different programs, decisions on project durations, funding patterns.
- Staff resources can be used for scientific program development, to identify areas for future development, and the stimulation of new research areas.

Advantages for Researchers

- Easy access to research results, relevant funding programs and their success rates, potential collaborators, competitors, related projects/publications (research push).
- More time for research and teaching.

Advantages for Industry

- Fast and easy access to major results, experts, etc.
- Can influence the direction of research by entering information on needed technologies (industry-pull).

Advantages for Publishers

- Unique interface to their data.
- Publicly funded development of databases and their interlinkage.

For Society

- Dramatically improved access to scientific knowledge and expertise.

“Science of Science” Opportunities
Cyberinfrastructures in Support of Computational Scientometrics

- **Scholarly Database at Indiana University** of 18 million scholarly records, [https://sdb.slis.indiana.edu](https://sdb.slis.indiana.edu)

- **OSGi/CIShell based Plug-and-Play CIs**
  - Information Visualization Cyberinfrastructure, [http://iv.slis.indiana.edu](http://iv.slis.indiana.edu)
  - Network Workbench Tool and Community Wiki, [http://nwb.slis.indiana.edu](http://nwb.slis.indiana.edu)
  - Epidemics Cyberinfrastructure (EPIC), soon at [http://epic.slis.indiana.edu](http://epic.slis.indiana.edu)

Designed and build in support of Knowledge Management

Katy Börner, Science Maps: How to Analyze, Map, and Make Sense of Science

Scholarly Database at Indiana University

See also


Contact Nianli Ma at nianma@indiana.edu
Scholarly Database: Web Interface

Search across publications, patents, grants.
Download records and/or (evolving) co-author, paper-citation networks.

Register for free access at https://sdb.slis.indiana.edu.

Challenges
Interlink $ Input & Publication/Patent Citation Output

Need to interlink
- Grants and papers/patents.
- Grants/papers/patents and their PIs/authors/inventors, etc.

Use resulting networks to
- Count #papers, #citations, etc.
- Determine strength of co-PI/author/inventor relations, etc.
Semantic Association Networks


Improved Representation of Scholarly Knowledge

Entity and link types:

Entity Types
- Authors
- Records

Authors

Link Types
- associated
- cites
- co-authors_with

Attributes:
- Records often have a publication date, a publication type (e.g., journal paper, book, patents, grant, etc.), topics (e.g., keywords or classifications assigned by authors and/or publishers).
- Authors have an address with information on affiliation and geo-location.

Derived attributes:
- Because authors and records are associated, the geo-location(s) and affiliation(s) of an author can be attributed to the authors’ papers.
- Similarly, the publication date, publication type and topic(s) can be associated with a paper’s author(s).
Improved Representation of Scholarly Knowledge makes possible

Statistics:
- Number of papers, grants, co-authorships, citation (over time) per author.
- Bursts of activity (#citations, #$, #patents, #collaborators, etc.).
- Changes of topics and geo-locations for authors and their institutions over time.

Visualizations:
- Geospatial and topical distribution of funding input & research output.
- Structure and evolution of research topics.
- Evolving research areas (e.g., based on young yet highly cited papers).
- Diffusion of information, people, $s over geospatial and topic space.

Scholarly Database: # Records & Years Covered

Datasets available via the Scholarly Database (* future feature)

<table>
<thead>
<tr>
<th>Dataset</th>
<th># Records</th>
<th>Years Covered</th>
<th>Updated</th>
<th>Restricted Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>Medline</td>
<td>13,149,741</td>
<td>1965-2005</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>PhysRev</td>
<td>398,005</td>
<td>1893-2006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PNAS</td>
<td>16,167</td>
<td>1997-2002</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1994-2004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USPTO</td>
<td>3,179,930</td>
<td>1976-2004</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>NSF</td>
<td>174,835</td>
<td>1985-2003</td>
<td>Yes*</td>
<td></td>
</tr>
<tr>
<td>NIH</td>
<td>1,043,804</td>
<td>1972-2002</td>
<td></td>
<td>Yes*</td>
</tr>
<tr>
<td>Total</td>
<td>18,021,560</td>
<td>1893-2006</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

Aim for comprehensive time, geospatial, and topic coverage.
Funding overlay – DOE laboratories
Funding overlay – NIH funded research (extramural)

Funding overlay – NSF funded research

Vitality

High

Low
Mapping the Evolution of Co-Authorship Networks

OSGi/CIKell based Plug-and-Play CIs

See also

Building Market Places not Cathedrals

- ‘Software glue’ has to interlink datasets and algorithms written in different languages using different data formats.
- The smaller the glue or ‘CI Shell’, the more likely it can be maintained.
CIShell is built upon the Open Services Gateway Initiative (OSGi) Framework.

OSGi (http://www.osgi.org) is

- A standardized, component oriented, computing environment for networked services.
- Successfully used in the industry from high-end servers to embedded mobile devices since 7 years.
- Alliance members include IBM (Eclipse), Sun, Intel, Oracle, Motorola, NEC and many others.
- Widely adopted in open source realm, especially since Eclipse 3.0 that uses OSGi R4 for its plugin model.

Advantages of Using OSGi

- Any CIShell algorithm is a service that can be used in any OSGi-framework based system.
- Using OSGi, running CIShells/tools can connected via RPC/RMI supporting peer-to-peer sharing of data, algorithms, and computing power.

Ideally, CIShell becomes a standard for creating OSGi Services for algorithms.
CIShell – Technical Details

CIShell layer cake.

CIShell – Deployment

CIShell applications can be deployed as distributed data and algorithm repositories, stand alone applications, peer-to-peer architectures, and server-client architectures.
The Network Workbench Tool

Investigators: Katy Börner, Albert-Laszlo Barabasi, Santiago Schnell, Alessandro Vespignani & Stanley Wasserman, Eric Wernert

Software Team: Lead: Weixia (Bonnie) Huang
Developers: Bruce Herr, Ben Markines, Santo Fortunato, Cesar Hidalgo, Ramya Sabbineni, Vivek S. Thakre, & Russell Duhon

Goal: Develop a large-scale network analysis, modeling and visualization toolkit for biomedical, social science and physics research.

Amount: $1,120,926 NSF IIS-0513650 award.
Website: http://nwb.slis.indiana.edu
The NWB Advisory Board consists of the following members:

- Ulrik Brandes, University of Konstanz, Germany (Graph Theory)
- Noshier Contractor, Northwestern University (Communication Theory)
- Mark Gerstein, Yale University (Bioinformatics)
- James Hendler, Rensselaer Polytechnic Institute (Semantic Web)
- Jason Leigh, Electronic Visualization Laboratory, University of Illinois at Chicago (Visualization & CI)
- Neo Martinez, Pacific Ecoinformatics and Computational Ecology Lab (Biology)
- Michael Macy, Cornell University (Sociology)
- Stephen North, AT&T (Graph Visualization)
- Tom Snijders, University of Groningen (Social Network Analysis)

The NWB CI Deliverables include:

**Cyberglue:**
- **CIShell** Core programmer team lead by Bonnie Huang

**Tools & Services:**
- **NWB Tool** Lead by Alex Vespignani with input from other PIs
- **SciMaps Service** Lead by Katy Börner
- **Bio Tool** Lead by Laszlo Barabasi & Santiago Schnell

All three are prototypical instantiations of CIShell serving as reference implementations.

**Documentation/Registry/Market Place:**
- **NWB Community Wiki** Lead by Katy Börner
NWB Tool: Interface Elements

Load Data
Select Preferences
List of Data Models
Console
Scheduler
Open Text Files
Visualize Data

NWB Ecology of Data Formats and Converters

Not shown are 15 sample datasets, 45 data preprocessing, analysis, modeling and visualization algorithms, 9 services.

5 Supported data formats

5 Output formats for diverse visualization algorithms

8 Intermediate data formats

Supported by 35 data converters.
Education – Learning Modules, NWB User and Developer Workshops

Download from http://nwb.slis.indiana.edu/

Summary
Network Workbench: A Large-Scale Network Analysis, Modeling and Visualization Toolkit for Biomedical, Social Science and Physics Research. This project will design, evaluate, and develop a unique distributed, shared resources environment for large-scale network analysis, modeling, and visualization, named Network Workbench (NWB). The envisioned data-compute-environment will provide...

News & Updates
- 22/08: NWB Tool 0.3.0 Release
- 13/06: NWB Tool 0.2.0 v1 Release
- 12/06: NWB Tool Update (v1.0 supported file format)
- 12/06: NWB at Linked Ed (Poster)
- 12/06: NWB Tool Update (new two-file format)
- 12/06: NWB Tool 0.2.0 Beta Release
- 12/06: NWB Tool Update (new two-file format)
- 12/06: NWB Basic Tutorial Getting Started

Download Latest Release
Note: save the download as .jar
Select Your Operating System
Windows XP
Download

Get Involved
- Sign up for NWB mailing lists
- NWB Tracking System
660 people visited this site

- 930 Visits
- 660 Absolute/Unique Visitors
- 2,989 Pageviews
- 3.21 Average Pageviews
- 00:03:19 Time on Site

Technical Profile

<table>
<thead>
<tr>
<th>Browser</th>
<th>Visits</th>
<th>% visits</th>
</tr>
</thead>
<tbody>
<tr>
<td>Others</td>
<td>650</td>
<td>40.30%</td>
</tr>
<tr>
<td>Internet Explorer</td>
<td>343</td>
<td>36.09%</td>
</tr>
<tr>
<td>Others</td>
<td>87</td>
<td>1.23%</td>
</tr>
<tr>
<td>Mozilla</td>
<td>26</td>
<td>2.80%</td>
</tr>
<tr>
<td>Others</td>
<td>23</td>
<td>2.47%</td>
</tr>
</tbody>
</table>

Connection Speed

<table>
<thead>
<tr>
<th>Speed</th>
<th>Visits</th>
<th>% visits</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSL</td>
<td>36</td>
<td>59.36%</td>
</tr>
<tr>
<td>Unknown</td>
<td>225</td>
<td>34.19%</td>
</tr>
<tr>
<td>Cable</td>
<td>193</td>
<td>17.53%</td>
</tr>
<tr>
<td>DSL</td>
<td>22</td>
<td>2.37%</td>
</tr>
</tbody>
</table>

Visitor Segmentation

- Visitors Profile: languages, network locations, user defined
- Browser Profile: browsers, operating systems, browser and operating system, screen colors, screen resolutions, Java support, Flash
- Map Overlay: Geolocation visualization

Map Overlay

- 930 visits came from 57 countries/territories
Growing a Community of Network Science Researchers

Users come from Social Science, Physics, Biology, Information Science, Telecommunications, Internet Research, Economics, Science Policy, etc.

It takes 9 months to give birth to a human baby and 21 years to raise it.
It takes 3-5 years to build a CI and ??? years to build a vibrant, self-sustaining community.

Quickly identify and serve continuously changing needs of evolving community.

Usage Statistics for nwb.sils.indiana.edu
### TotalCounter statistics

<table>
<thead>
<tr>
<th>Pages</th>
<th>Percent</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main homepage</td>
<td>8%</td>
<td>1598</td>
</tr>
<tr>
<td>About/FAQ Home Page</td>
<td>6%</td>
<td>1271</td>
</tr>
<tr>
<td>VisualData Office</td>
<td>6%</td>
<td>1159</td>
</tr>
<tr>
<td>VisualData Team/Home</td>
<td>4%</td>
<td>921</td>
</tr>
<tr>
<td>VisualData Pruchnihar-Huang</td>
<td>4%</td>
<td>917</td>
</tr>
<tr>
<td>Main Wiki Tool</td>
<td>4%</td>
<td>577</td>
</tr>
<tr>
<td>Datasets Home Page</td>
<td>4%</td>
<td>797</td>
</tr>
<tr>
<td>VisualData ForceDirected</td>
<td>3%</td>
<td>590</td>
</tr>
<tr>
<td>Tutorials Home Page</td>
<td>2%</td>
<td>409</td>
</tr>
<tr>
<td>Main Blog</td>
<td>2%</td>
<td>400</td>
</tr>
<tr>
<td>Main Publications</td>
<td>2%</td>
<td>304</td>
</tr>
<tr>
<td>Main FAQ</td>
<td>2%</td>
<td>329</td>
</tr>
<tr>
<td>VisualData SpringLayout</td>
<td>1%</td>
<td>120</td>
</tr>
<tr>
<td>VisualData Coauthor</td>
<td>1%</td>
<td>117</td>
</tr>
<tr>
<td>VisualData Citations</td>
<td>1%</td>
<td>115</td>
</tr>
<tr>
<td>VisualData IndexTree</td>
<td>1%</td>
<td>113</td>
</tr>
<tr>
<td>VisualData BetweennessCentralityStakeholder</td>
<td>1%</td>
<td>112</td>
</tr>
<tr>
<td>VisualData HomePage</td>
<td>1%</td>
<td>108</td>
</tr>
<tr>
<td>VisualData Node2vec</td>
<td>1%</td>
<td>106</td>
</tr>
<tr>
<td>CustomFilings Home Page</td>
<td>1%</td>
<td>106</td>
</tr>
<tr>
<td>CustomFilings Analyzing Biological Networks</td>
<td>1%</td>
<td>106</td>
</tr>
</tbody>
</table>

### Users

<table>
<thead>
<tr>
<th>Users</th>
<th>Percent</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>10%</td>
<td>1050</td>
</tr>
<tr>
<td>Guest (not authenticated)</td>
<td>45%</td>
<td>9415</td>
</tr>
<tr>
<td>millican</td>
<td>1%</td>
<td>150</td>
</tr>
<tr>
<td>edenr</td>
<td>1%</td>
<td>147</td>
</tr>
<tr>
<td>book</td>
<td>1%</td>
<td>119</td>
</tr>
<tr>
<td>big</td>
<td>1%</td>
<td>95</td>
</tr>
<tr>
<td>ninnen</td>
<td>0%</td>
<td>69</td>
</tr>
<tr>
<td>cardiff</td>
<td>0%</td>
<td>65</td>
</tr>
<tr>
<td>laty</td>
<td>0%</td>
<td>36</td>
</tr>
<tr>
<td>cesar</td>
<td>0%</td>
<td>32</td>
</tr>
<tr>
<td>taylor</td>
<td>0%</td>
<td>29</td>
</tr>
<tr>
<td>karthip</td>
<td>0%</td>
<td>26</td>
</tr>
<tr>
<td>renamont</td>
<td>0%</td>
<td>24</td>
</tr>
<tr>
<td>holbro</td>
<td>0%</td>
<td>24</td>
</tr>
<tr>
<td>June Young Lee</td>
<td>0%</td>
<td>24</td>
</tr>
<tr>
<td>everyone</td>
<td>0%</td>
<td>24</td>
</tr>
</tbody>
</table>

---

**IU News Room**

**IU News from all eight campuses**

**Sunday, May 4, 2008**

**$1.2 million NIH project will help track and predict epidemics**

**FOR IMMEDIATE RELEASE**

**April 8, 2008**

BLOOMINGTON, Ind. -- The National Institutes of Health has given $1.2 million to Indiana University researchers to build the ultimate international epidemic research tool.
Network Workbench in Action:
Embracing the Diversity of Network Science

See also

Computational Social Science
Studying large scale social networks such as Wikipedia

Vizzards 2007 Entry

Second Sight: An Emergent Mosaic of Wikipedian Activity,
The NewScientist, May 19, 2007
Computational Economics

Does the type of product that a country exports matter for subsequent economic performance?


Computational Proteomics

Computational Epidemics
Forecasting (and preventing the effects of) the next pandemic.


References

- Hidalgo, César A. and C. Rodriguez-Sickert. Persistence, Topology and Sociodemographics of a Mobile Phone Network. 2007. (Submitted to Physica A)
References (Cont.)


More papers are linked from [http://ivl.slis.indiana.edu/publications/](http://ivl.slis.indiana.edu/publications/)

The End. 