Humanexus: Envisioning Communication and Collaboration

Katy Börner
Director, Cyberinfrastructure for Network Science Center
School of Informatics and Computing, Indiana University, USA

Summer School in Cognitive Sciences on
Web Science and The Mind
Université du Québec à Montréal
July 10, 2014

Language Communities of Twitter - Eric Fischer - 2012

READINGS
Papers

Books
Find your way

Terra bytes of data

Find collaborators, friends

Identify trends

Descriptive & Predictive Models

Descriptive Models

Multiple levels: Micro ... Macro

## Different Levels of Abstraction/Analysis

- **Macro/Global Population Level**
- **Meso/Local Group Level**
- **Micro Individual Level**

## Type of Analysis vs. Level of Analysis

<table>
<thead>
<tr>
<th></th>
<th>Micro/Individual (1-100 records)</th>
<th>Meso/Local (101–100,000 records)</th>
<th>Macro/Global (100,000 &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>
Type of Analysis vs. Level of Analysis

<table>
<thead>
<tr>
<th></th>
<th>Micro/Individual (1-100 records)</th>
<th>Meso/Local (101–100,000 records)</th>
<th>Macro/Global (100,000 &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 NSE, all of USA, all of science</td>
</tr>
<tr>
<td><strong>Temporal Analysis (When?)</strong></td>
<td>Funding portfolio of one individual</td>
<td>Temporal bursts in 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>Geospatial analysis</td>
<td>PNAS publications</td>
</tr>
<tr>
<td><strong>Topical Analysis (What?)</strong></td>
<td>Base knowledge from one individual</td>
<td>Topic bursts in 20 years of PNAS</td>
<td>VxOrd/Topic 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 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.

Stipelman, Hall, Zoss, Okamoto, Stokols & Börner, 2014
Supported by NIH/NCI Contract HHSN261200800812
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.

The Global 'Scientific Food Web'


Contributions:
- Comprehensive global analysis of scholarly knowledge production and diffusion on the level of continents, countries, and cities.
- Quantifying knowledge flows between 2000 and 2009, we identify global sources and sinks of knowledge production. Our knowledge flow index reveals, where ideas are born and consumed, thereby defining a global 'scientific food web'.
- While Asia is quickly catching up in terms of publications and citation rates, we find that its dependence on knowledge consumption has further increased.
<table>
<thead>
<tr>
<th>Type of Analysis vs. Level of Analysis</th>
<th>Micro/Individual (1-100 records)</th>
<th>Meso/Local (101–100,000 records)</th>
<th>Macro/Global (100,000 &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 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 state's 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></td>
<td>VxOrd/Topic NIH funding</td>
</tr>
<tr>
<td><strong>Network Analysis (With Whom?)</strong></td>
<td>NSF Co-PI network of one</td>
<td>Co-author network</td>
<td>NIH’s core competencies</td>
</tr>
</tbody>
</table>

Figure 2: World map of the greatest knowledge sources and sinks, based on our scientific fitness index. Green bars indicate that the number of citations received is over-proportional, red that the number of citations received is lower than expected (according to a homogenous distribution of citations over all cities that have published more than 500 papers). It can be seen that most scientific activity occurs in the temperate zone. Moreover, areas of high fitness tend to be areas that are performing economically well (but the opposite does not hold).
Predictive Models (Why?)

Example: Collective allocation of science funding as an alternative to peer review

From funding agencies to scientific agency: Collective allocation of science funding as an alternative to peer review

Existing (left) and proposed (right) funding systems. Reviewers in blue; investigators in red.

In the proposed system, all scientists are both investigators and reviewers: every scientist receives a fixed amount of funding from the government and discretionary distributions from other scientists, but each is required in turn to redistribute some fraction of the total they received to other investigators.
From funding agencies to scientific agency: Collective allocation of science funding as an alternative to peer review

Current Model is Expensive:
If four professors work four weeks full-time on a proposal submission, labor costs are about $30k [1]. With typical funding rates below 20%, about five submission-review cycles might be needed resulting in a total expected labor cost of $150k. The average NSF grant is $128k per year. U.S. universities charge about 50% overhead (ca. $42k), leaving about $86k.
In other words, the four professors lose $150k-$86k= $64k of paid research time by obtaining a grant to perform the proposed research.

To add: Time spent by researchers to review proposals. In 2012 alone, NSF convened more than 17,000 scientists to review 53,556 proposals.

From funding agencies to scientific agency: Collective allocation of science funding as an alternative to peer review

Example:
Total funding budget per year is 2012 NSF budget
Given the number of NSF funded scientists, each receives a $100,000 basic grant.
Fraction is set to 50%

In 2013, scientist S receives a basic grant of $100,000 plus $200,000 from her peers, i.e., a total of $300,000.
In 2013, S can spend 50% of that total sum, $150,000, on her own research program, but must donate 50% to other scientists for their 2014 budget.

Rather than submitting and reviewing project proposals, S donates directly to other scientists by logging into a centralized website and entering the names of the scientists to donate to and how much each should receive.

Model Run and Validation:
It uses citations as a proxy for how each scientist might distribute funds in the proposed system.

Dataset: 37M articles from TR 1992 to 2010 Web of Science (WoS) database with 770M citations and 4,195,734 unique author names. The 867,872 names who had authored at least one paper per year in any five years of the period 2000–2010 were used in validation.
For each pair of authors we determined the number of times one had cited the other in each year of our citation data (1992–2010).
NIH and NSF funding records from IU’s Scholarly Database provided 347,364 grant amounts for 109,919 unique scientists for that time period.
Simulation run begins in year 2000, in which every scientist was given a fixed budget of B = $100k. In subsequent years, scientists distribute their funding in proportion to their citations over the prior 5 years.
The model yields funding patterns similar to existing NIH and NSF distributions.
Making Every Scientist a Research Funder

When it comes to using peer review to distribute research dollars, Johan Bollen favors radical simplicity. Over the years, many scientists have suggested that the current system could be improved by changing the composition of the review panels, tweaking the interaction among reviewers, or revising how the proposals are scored. But Bollen, a computer scientist at Indiana University, Bloomington, would simply award all eligible researchers a block grant—and then require them to give some of it away to colleagues they judge most deserving.

That radical step, described in a paper Bollen and four Indiana colleagues recently posted on LANL Reports, returns peer review’s core concept of tapping into the views of the most knowledgeable researchers. But it would eliminate the huge investment in time and money required to submit proposals and assemble panels to judge them.

Bollen’s process would be almost instantaneous: in a version of input-directed crowdsourcing, scientists would fill out a form once a year listing their favored researchers, and a predetermined portion of their annual grant money—a total of say, 50%—would then be transferred to their choices.

“So many scientists spend so much time on peer review, and there’s a high level of frustration,” Bollen explains. “We already know who the best people are. And if you’re doing good work, then you deserve to receive support.”

Others are skeptical. “I’ve known Johan for a long time and have the highest regard for his ability as an out-of-the-box thinker,” says Stephen Griffin, a retired National Science Foundation (NSF) program manager who’s now a visiting professor of information science at the University of Pittsburgh in Pennsylvania. “But there are a number of issues he doesn’t address.”

Those sticking points include the likely mismatch between what researchers need and what their colleagues give them; the absence of any replacement for the overhead payments in today’s grants, which support infrastructure at host institutions; and the dearth of public accountability for the billions of dollars that would flow from public coffers to individuals. “Scientists aren’t really equipped to be a funding agency,” Griffin notes.

Bollen acknowledges that the process would need safeguards to ensure that scientists don’t reward their friends or punish their enemies. But his analysis suggests that the U.S. research landscape would not look all that different if his radical proposal were adopted.

Drawing upon citation data on 37 million papers over 26 years, the Indiana researchers conducted a simulation premised on the idea that scientists would reallocate their federal dollars according to how often they cited their peers. The simulation, he says, yielded a funding pattern “similar in shape to the actual distribution” at NSF and the National Institutes of Health for the past decade—at a fraction of the overhead required by the current system.

—JHN

Visualizing STI Model Results

Example: Places & Spaces: Mapping Science Exhibit
Mapping Science Exhibit on display at MEDIA X, Stanford University

Map of Scientific Collaborations from 2005-2009


Illuminated Diagram Display on display at the Smithsonian in DC. http://scimaps.org/exhibit_info/inf2
Science Maps in “Expedition 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

Places & Spaces Digital Display in North Carolina State’s brand new Immersion Theater
Places & Spaces: Mapping Science Exhibit
http://scimaps.org

Maps are available for sale and the exhibit can be hosted by anyone.

Visualizing STI Model Results

Example: The Information Visualization MOOC
Course Schedule

- **Session 1** – Workflow design and visualization framework
- **Session 2** – “When:” Temporal Data
- **Session 3** – “Where:” Geospatial Data
- **Session 4** – “What:” Topical Data

**Mid-Term**

Students work in teams with clients.
- **Session 5** – “With Whom:” Trees
- **Session 6** – “With Whom:” Networks
- **Session 7** – Dynamic Visualizations and Deployment

**Final Exam**

Final grade is based on Midterm (30%), Final (40%), Client Project (30%).

**Needs-Driven Workflow Design**

- Stakeholders
- Validation Interpretation
- Types and levels of analysis determine data, algorithms & parameters, and deployment
- Data

**DEPLOY**

- Visually encode data
- Overlay data
- Select visualiz. type

**READ**

**ANALYZE**

**VISUALIZE**
Needs-Driven Workflow Design

**Stakeholders**

Validation → Interpretation

**Types and levels of analysis** determine data, algorithms & parameters, and deployment

**Data**

**Read** → **Analyze** → **Visualize**

**DEPLOY**

Visually encode data → Overlay data → Select visualiz. type

Clients

http://ivmooc.cns.iu.edu/clients.html

http://cns.iu.edu/humanexus

References


All papers, maps, tools, talks, press are linked from [http://cns.iu.edu](http://cns.iu.edu)
These slides will soon be at [http://cns.iu.edu/docs/presentations](http://cns.iu.edu/docs/presentations)

CNS Facebook: [http://www.facebook.com/cnscenter](http://www.facebook.com/cnscenter)
Mapping Science Exhibit Facebook: [http://www.facebook.com/mappingscience](http://www.facebook.com/mappingscience)