Envisioning Communication and Collaboration

Katy Börner
Victor H. Yngve Professor of Information Science
Director, Cyberinfrastructure for Network Science Center
School of Informatics and Computing, Indiana University, USA

Rice University, Houston, TX
November 11, 2014

Language Communities of Twitter - Eric Fischer - 2012

Humanexus
Watch the official trailer »

http://cns.iu.edu/humanexus
READINGS

Papers


Books

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>Type 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>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 USA, all of science.</td>
</tr>
<tr>
<td>Temporal Analysis (When?)</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>Geospatial Analysis (Where?)</td>
<td>Career trajectory of one individual</td>
<td>Mapping a states intellectual landscape</td>
<td>PNAS publications</td>
</tr>
<tr>
<td>Topical Analysis (What?)</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>Network Analysis (With Whom?)</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.

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 homogeneous 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).
### Type of Analysis vs. Level of Analysis

<table>
<thead>
<tr>
<th>Analysis Type</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>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 USA, all of science</td>
</tr>
<tr>
<td>Temporal Analysis (When?)</td>
<td>Funding portfolio of one individual</td>
<td>Dynamic bursts in PNAS</td>
<td>113 Years of Physics Research</td>
</tr>
<tr>
<td>Geospatial Analysis (Where?)</td>
<td>Career trajectory of one individual</td>
<td>Geospatial interconnections</td>
<td>PNAS publications</td>
</tr>
<tr>
<td>Topical Analysis (What?)</td>
<td>Base knowledge from one individual</td>
<td>Knowledge flows in Chemistry</td>
<td>VxOrd/Topic NIH funding</td>
</tr>
<tr>
<td>Network Analysis (With Whom?)</td>
<td>NSF-CO-PI network of one</td>
<td>Co-author network</td>
<td>NIH’s core competency</td>
</tr>
</tbody>
</table>

### 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.

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


Assume
Total funding budget in year y is \( t_y \)
Number of qualified scientists is \( n \)

Each year,
the funding agency deposits a fixed amount into each account, equal to the total funding budget divided by the total number of scientists: \( t_y/n \).
Each scientist must distribute a fixed fraction, e.g., 50%, of received funding to other scientists (no self-funding, COIs respected).

Result
Scientists collectively assess each others’ merit based on different criteria; they “fund-rank” scientists; highly ranked scientists have to distribute more money.

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

**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.

---

**NEWSFOCUS**

Making Every Scientist a Research Funder

When it comes to using peer review to distribute research dollars, John Bollen finds radical simplicity.

Over the years, many scientists have suggested that the current system could be improved by changing the composition of the review panels, breaking the interactions 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 *EMBO Reports*, retains 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 supernetwork crowdfunding, scientists would fill out a form once a year listing their favored researchers, and a preselected 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 reviews, and there’s a high level of frustation,” 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 John 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 striking 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 health 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 on citation data for 37 million papers over 20 years, the Indiana researchers conducted a simulation premised on the idea that scientists would allocate 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 distributions” at NSF and the National Institutes of Health for the past decade—at a fraction of the overhead required by the current system.

March 7, 2014

Science 7 February 2014: Vol. 343 no. 6171 p. 598
DOI: 10.1126/science.343.6171.598
http://www.sciencemag.org/content/343/6171/598.full?sid=4f40a7f0-6ba2-4ad8-a181-7ab394fe2178
Visualizing STI Model Results

Example: Places & Spaces: Mapping Science Exhibit

Mapping Science Exhibit on display at MEDIA X, Stanford University

Illuminated Diagram Display on display at the Smithsonian in DC.
http://scimaps.org/exhibit_info/#ID
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

The Information Visualization MOOC
ivmooc.cns.iu.edu

Students from more than 100 countries
350+ faculty members
#ivmooc

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

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

Data

READ

ANALYZE

VISUALIZE

DEPLOY

Validation

Interpretation

Visually encode data

Overlay data

Select visualiz. type
Clients

Information Visualization MOOC

List of Clients

1. Project Title: Isis: 100 Years
   Client Name: Jay Malone
   Project goal: Scientific or practical value: A visual representation of Isis' contributors and locales over the past 100 years. Isis is the journal of the History of Science Society. This representation will provide a dynamic picture of how scholarship in the history of science has shifted over the past century. Relevant publications, websites, etc: http://press.chicago.edu/ujp/journal/journal.html
   Conditions under which students can publish results and/or add project results to their resume: Client would like to approve results.

2. Project Title: e-Xploration
   Client Name: Lily
   Project goal: Scientific or practical value: e-Xploration is an agent-based model for the ethnographic observation and the registry, analysis, and interpretation of social practices in virtual communities for intervention in the development of collaboration and cooperation. This project will analyze the interactions between subjects and objects in a platform collaborative community called OxyCIB, a project based on e-Xploration (e-cisc.net). Information on datasets to be used: I can provide a data base in .graph format for the students. The file .graphmi contains the interactions between subjects and objects in a platform collaborative community called OxyCIB. In the level of practice, it is not necessary that students know agent-based models for using the database. But, in another level, for example: the collaborative level for the OxyCIB development, it is necessary to have basic knowledge in AMIS or MAS and another competences like PHP and MySQL. Relevant publications, websites, etc: http://www.e-cisc.net/typ
   Conditions under which students can publish results and/or add project results to their resume: If any person or institution use my dataset or another info about eXploration (e-cisc.net, oxycib.net), I need to approve the results and appear as co-author.

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


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

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