Mapping Our Collective Scholarly Knowledge

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

Symposium Honoring John T. Bruer
Knight Executive Conference Center, Washington University, St. Louis, MO
May 29, 2015


Humanexus


http://cns.iu.edu/humanexus
Mapping the Evolution of Co-Authorship Networks
Ke, Visvanath & Börner. 2004. Won 1st prize at the IEEE InfoVis Contest.

Map of Scientific Collaborations from 2005-2009
Mapping the Evolution of Co-Authorship Networks
Ke, Visvanath & Börner. 2004. Won 1st prize at the IEEE InfoVis Contest.

Mapping the Evolution of Co-Authorship Networks
Supported by NIH/NCI Contract HHSN261200800812

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.

Legend
Node Color Code
-4 to -20
-20 to -50
-50 to -100
-100 to -200

Edge Color Code
-4 to 4
-4 to 10
-10 to 20
-20 to 50
-50 to 100
-100 to 200

Legend
Node Color Code
-4 to -20
-20 to -50
-50 to -100
-100 to -200

Edge Color Code
-4 to 4
-4 to 10
-10 to 20
-20 to 50
-50 to 100
-100 to 200

Legend
Node Color Code
-4 to -20
-20 to -50
-50 to -100
-100 to -200

Edge Color Code
-4 to 4
-4 to 10
-10 to 20
-20 to 50
-50 to 100
-100 to 200

Legend
Node Color Code
-4 to -20
-20 to -50
-50 to -100
-100 to -200

Edge Color Code
-4 to 4
-4 to 10
-10 to 20
-20 to 50
-50 to 100
-100 to 200

Legend
Node Color Code
-4 to -20
-20 to -50
-50 to -100
-100 to -200

Edge Color Code
-4 to 4
-4 to 10
-10 to 20
-20 to 50
-50 to 100
-100 to 200

Legend
Node Color Code
-4 to -20
-20 to -50
-50 to -100
-100 to -200

Edge Color Code
-4 to 4
-4 to 10
-10 to 20
-20 to 50
-50 to 100
-100 to 200
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 300 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).
Long-Distance Interdisciplinarity Leads to Higher Scientific Impact

Data: 9.2 million interdisciplinary research papers published between 2000 and 2012.

Results: majority (69.9%) of co-cited interdisciplinary pairs are “win-win” relationships, i.e., papers that cite them have higher citation impact and there are as few as 3.3% “lose-lose” relationships. UCSD map of science is used to compute “distance.”

A1 Number of papers citing win-win relationships (≥10,000 citing papers)

2,940 (5.19%) of 56,614 win-win edges
node color: discipline | edge color: mix of adjacent nodes | labels: subdiscipline with highest number of win-win relationships per discipline (number and percentage of win-win relationships)
**B1 Number of papers citing lose-lose relationships (≥100 citing papers)**

![Clickstream Map of Science](image)

1,204 (44.4%) of 2,712 lose-lose edges

node color: discipline | edge color: mix of adjacent nodes | labels: subdiscipline with highest number of lose-lose relationships per discipline (number and percentage of lose-lose relationships)

---

Clickstream Map of Science

This is the first map created from large-scale, anonymous, including online data. It visualizes the collective flow of scientific information and provides insights into how scientists interact with each other in their online navigational behavior.

The Clickstream Map of Science is a visual representation of the collective flow of scientific information across different academic disciplines. It is created from large-scale, anonymous clickstream data, capturing the paths and interactions of scientists as they navigate through scholarly resources. The map is designed to illustrate the structure and dynamics of scientific communication, highlighting the connections and overlaps between various fields.

### Data
- **Node Size**: Simulates the number of papers citing a discipline.
- **Edge Width**: Reflects the number of lose-lose relationships between disciplines.

### Legend
- A **Clickstream Map of Science** showing the collective flow of scientific information across different academic disciplines is depicted in the image. The map visualizes the interactions and connections between various fields, providing insights into the patterns of scientific communication and collaboration.

---

Chemical Research & Development
Powers the U.S. Innovation Engine
Macaconomic Implications of Public and Private R&D Investments in Chemical Science

Illuminated Diagram Display on display at the Smithsonian in DC.
http://scimaps.org/exhibit_info/idp/
Call for Macroscopw Tools for the Places & Spaces: Mapping Science Exhibit (2015)

http://scimaps.org/call
Modelling Our Collective Scholarly Knowledge

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 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 EMBR 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 expert-directed crowdsourcing, scientists would fill out a form once a year listing their favored researchers, and a predetermine portion of their annual grant money—a total of, say, 50%—would then be transferred to their choices.

“Too many scientists spend too much time on peer review, and there is a high level of frustration,” Bollen explains. “We already know 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 sciences at the University of Pittsburgh in Pennsylvania. “But there are a number of issues he doesn’t address.”

These 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 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 favor 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 for 37 million papers over 20 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—not a fraction of the overhead required by the current system.

-JDM

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
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.

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 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 in 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.
Using 37M articles from TR 1992 to 2010 Web of Science (WoS) database, we extracted 770M citations. From the same WoS data, we also determined 4,195,734 unique author names and we took the 867,872 names who had authored at least one paper per year in any five years of the period 2000–2010.
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.
Model Efficiency:
Using data from the Taulbee Survey of Salaries Computer Science (http://cra.org/resources/taulbee) and the National Science Foundation (NSF) the following calculation is illuminating:

If four professors work four weeks full-time on a proposal submission, labor costs are about $30k. 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 research.

That is, U.S. universities should forbid professors to apply for grants—if they can afford to forgo the indirect dollars.

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

References
Tasks

LEVELS

MICRO: Individual Level
about 1-5,000 records

Meso: Local Level
about 1,000-500,000 records

MACRO: Global Level
more than 500,000 records

TYPES

Statistical Analysis

Visualizing Time Series

Geospatial Analysis

Visualizing Social Networks

Temporal Analysis

Understanding Temporal Trends

WHERE:

Cell phone usage over time page 52

WHERE:

Proactive space planning: patterns of commuting page 129

WHAT:

Predictive modeling of agent contact

WHAT:

Identifying patterns in social networking page 127

WITH WHOM:

Social media analysis

WITH WHOM:

Visualizing social dynamics in collaborative work

See page 5

See pages 6-7

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/presentations](http://cns.iu.edu/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)