Data Visualizations: Drawing Actionable Insights from Science and Technology Data

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

Center For Complex Systems and Enterprises Seminar
Stevens Institute of Technology, Hudson, Hoboken NJ
September 1, 2015

Olivier H. Beauchesne, 2011. Map of Scientific Collaborations from 2005-2009.
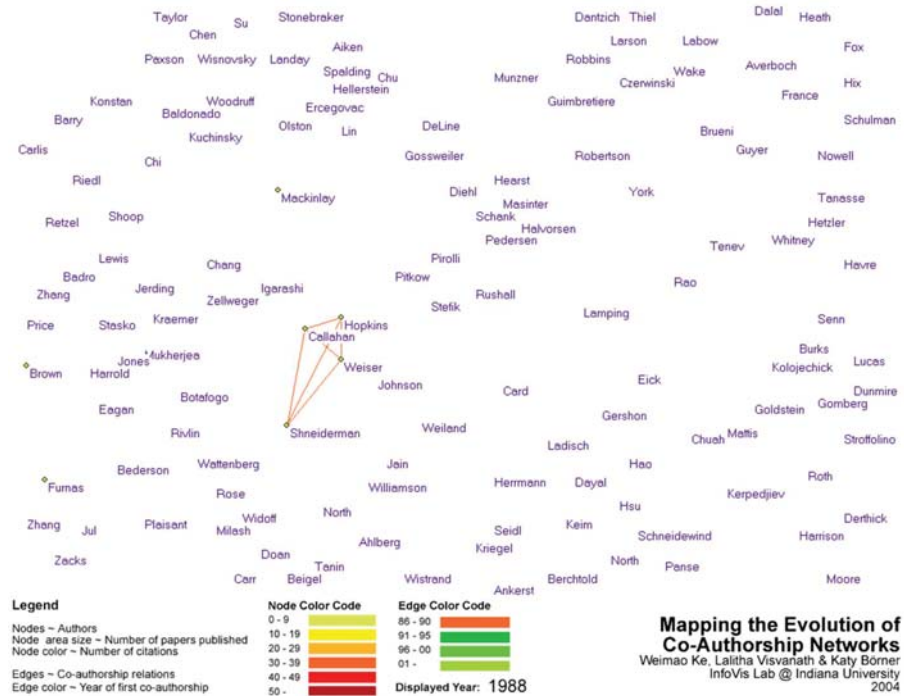
Computed Using Data from Elsevier's Scopus

Olivier H. Beauchesne, 2011. Map of Scientific Collaborations from 2005-2009.

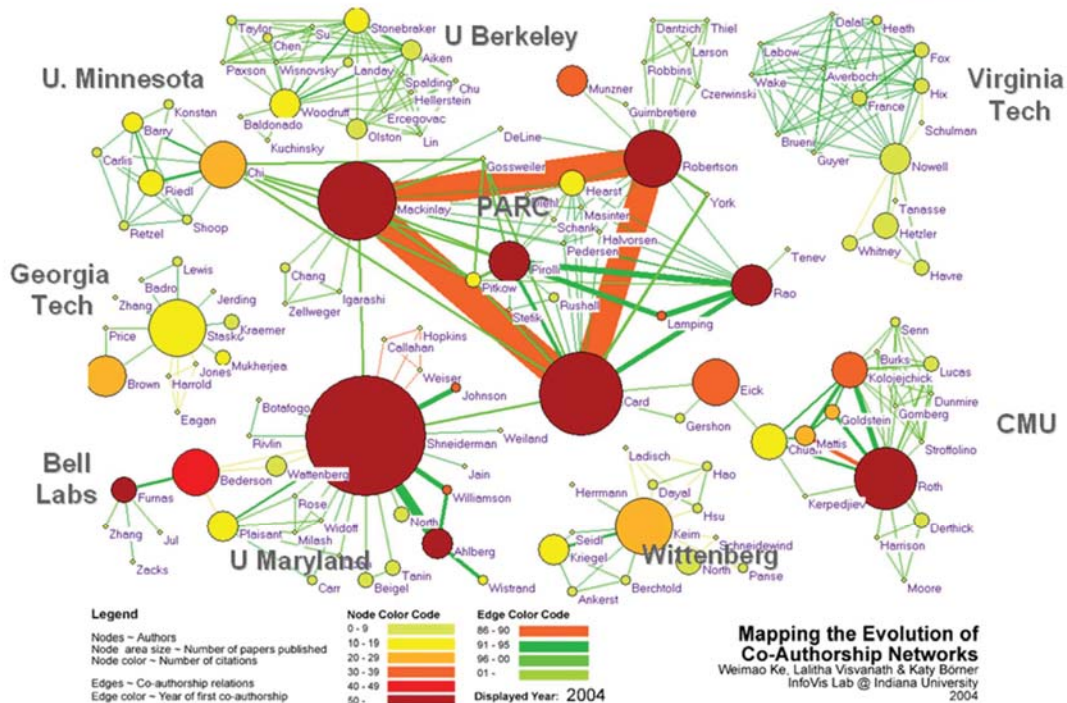# Mapping the Evolution of Co-Authorship Networks

Ke, Visvanath & Börner. 2004. Won 1st prize at the IEEE InfoVis Contest.

Mapping the Evolution of Co-Authorship Networks

*Ke, Visvanath & Börner. 2004. Won 1st prize at the IEEE InfoVis Contest.*

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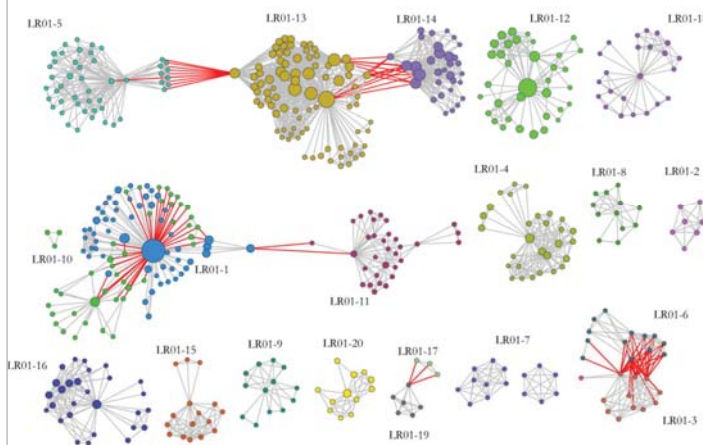
Stipelman, Hall, Zoss, Okamoto, Stokols, Börner, 2014.
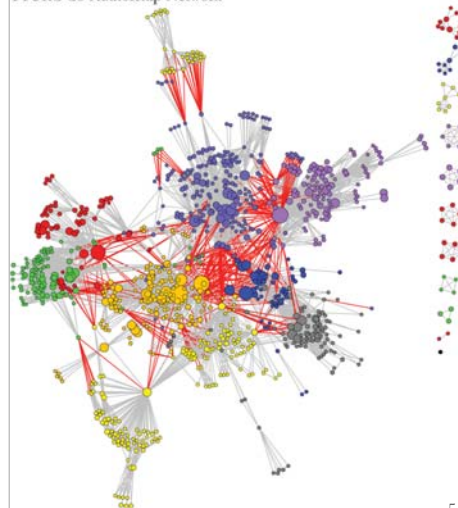
Supported by NIH/NCI Contract HHSN261200800812

R01 & TTURC Project Information
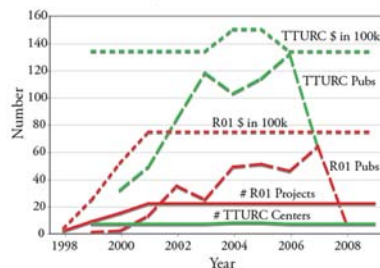
Longitudinal R01 Co-Authorship Network

TTURC Co-Authorship Network

The Global 'Scientific Food Web'

Mazloumian, Amin, Dirk Helbing, Sergi Lozano, Robert Light, and Katy Börner. 2013. "Global Multi-Level Analysis of the 'Scientific Food Web'". Scientific Reports 3, 1167.
http://cns.iu.edu/docs/publications/2013-mazloumian-food-web.pdf

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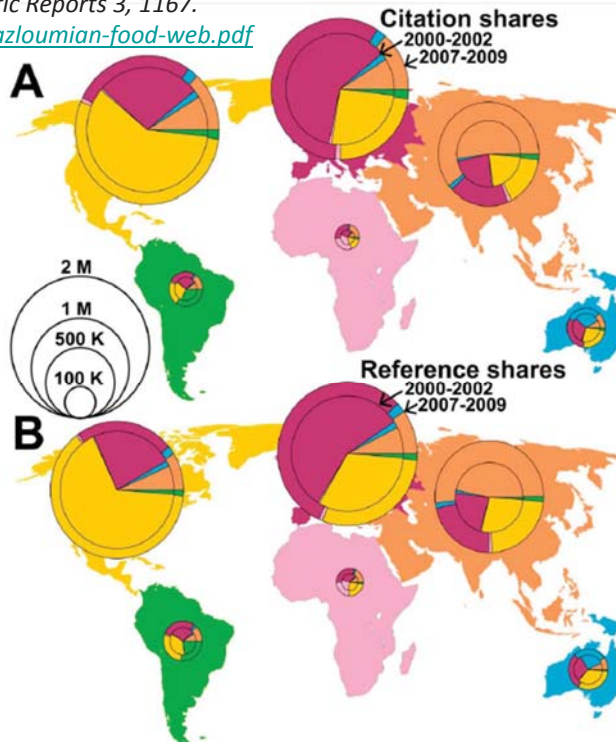
Figure 2 | World map of the greatest knowledge sources and sinks, based on our scientific fitness index. Green bars indicate that the number of citations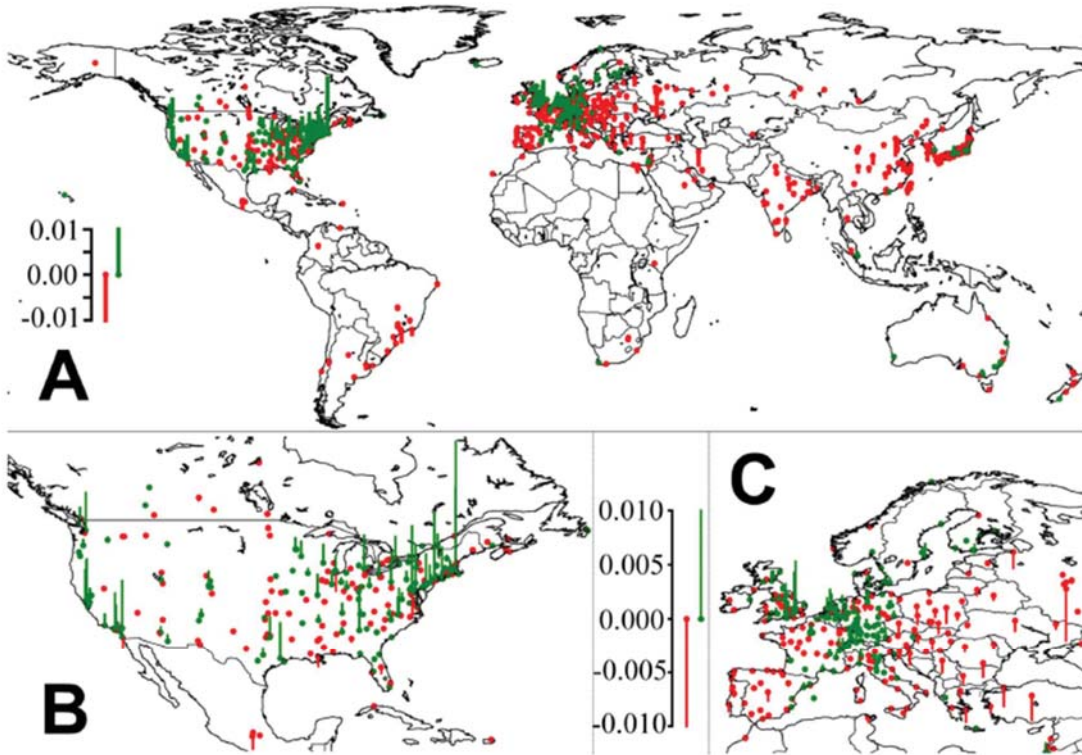 received is over-proportional, red that the number of citations received is lower than expected (according to a homogeneous distribution of citations over all cities that have published more than 500 papers). It can be seen that most scientific activity occurs in the temperate zone. Moreover, areas of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

Long-Distance Interdisciplinarity Leads to Higher Scientific Impact

Larivière, Vincent, Stefanie Haustein, and Katy Börner. 2015. PLOS ONE DOI: 10.1371.

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)

citing papers

Material Science (387; 91.9%)
Chemical, Mechanical, & Civil Engineering

Surface Science (387; 80.3%)
Math & Physics

Applied Optics (350; 81.4%)
Electrical Engineering & Computer Science

Statistics (349; 73.6%)
Social Sciences

Molecular Medicine (367; 72.8%)
Medical Specialties

Pharmaceutical Design (401; 84.6%)
Chemistry

Magnetic Resonance Imagery (385; 83.9%)
Health Professionals

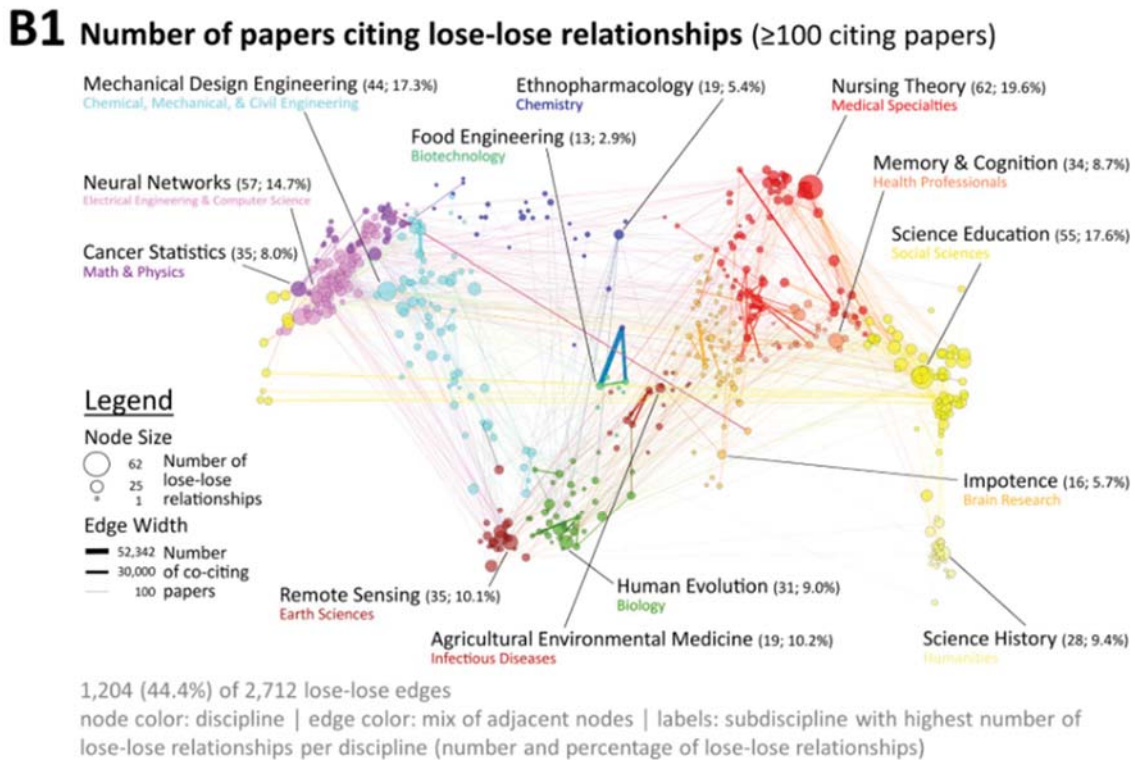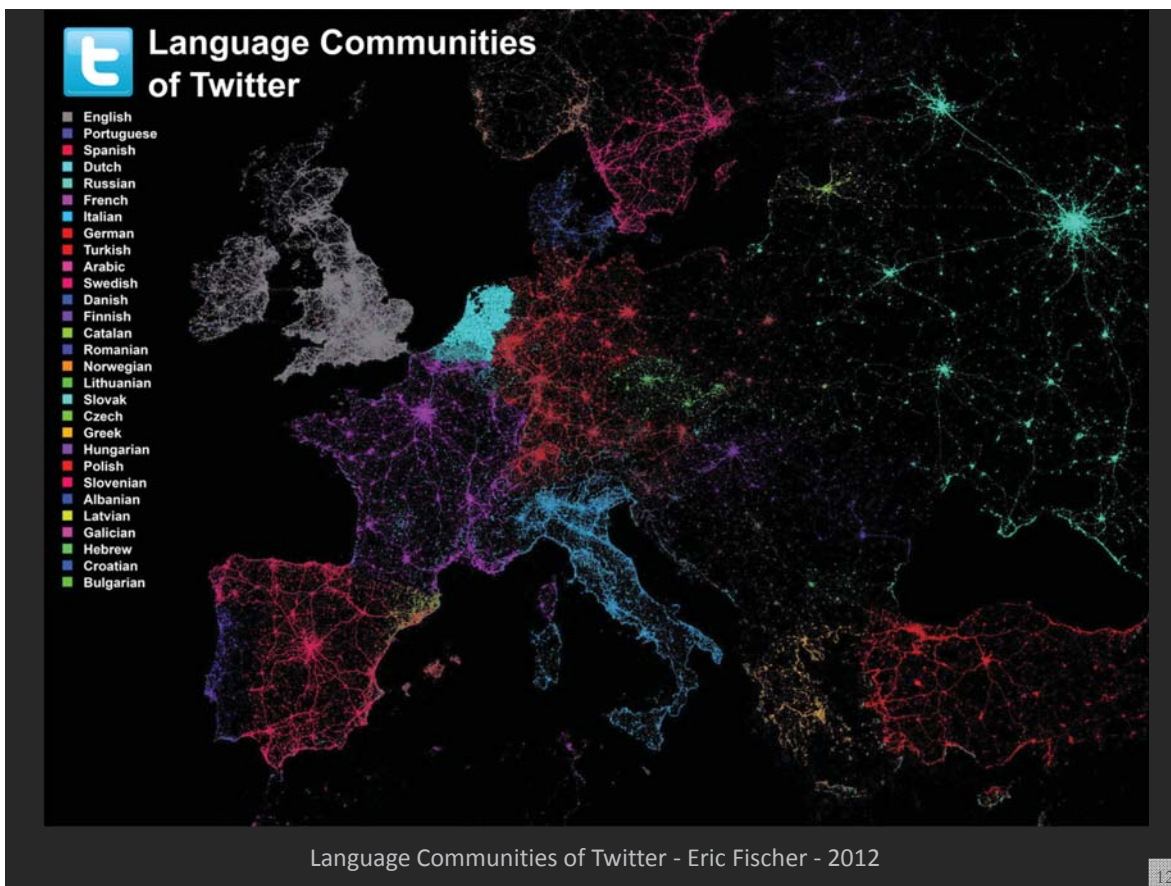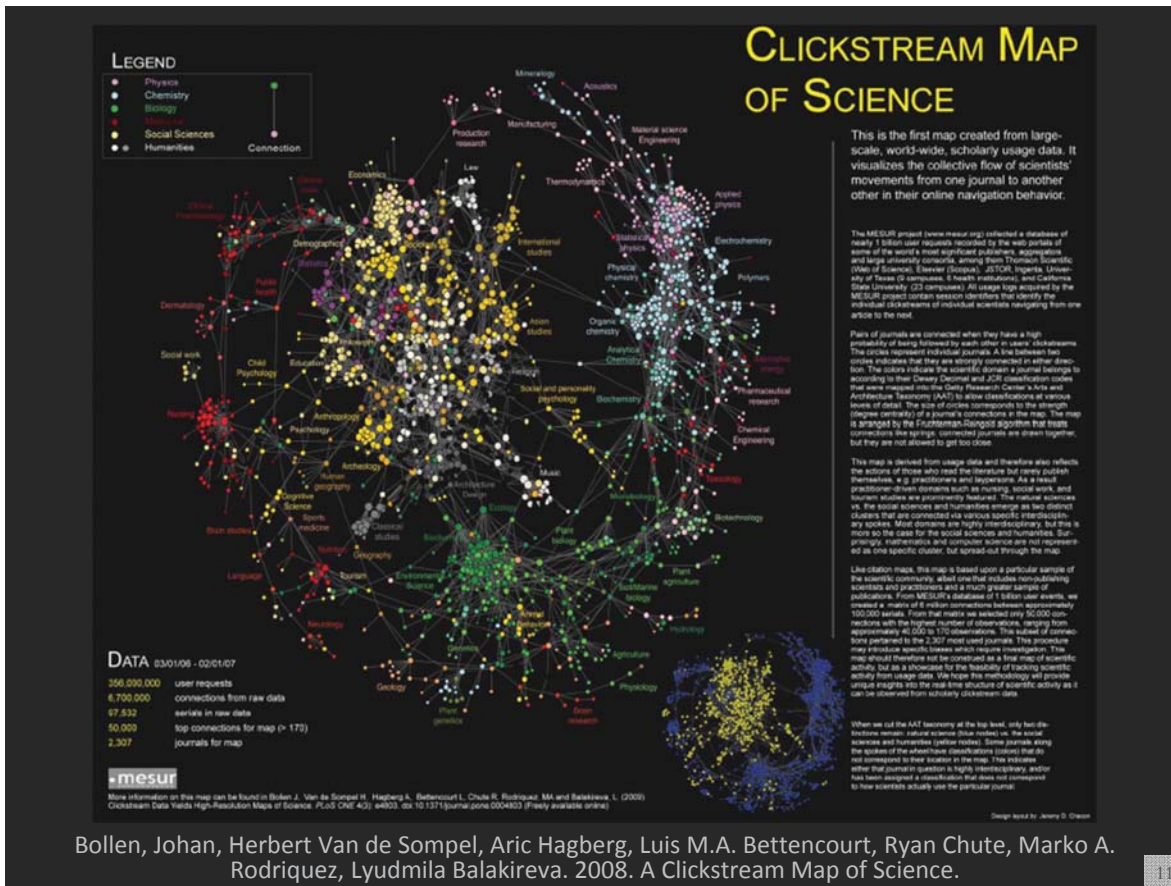Pharmacology Science (390; 84.8%)
Brain Research

Peptides (381; 79.5%)
Infectious Diseases

Protein Science (369; 74.5%)
Biotechnology

Mineralogy (322; 75.6%)
Earth Sciences

Molecular Ecology (409; 81.2%)
Biology

Science History (204; 68.7%)
Humanities

Legend

Node Size
- 409 — Number of
- 350 — win-win
- 1 — relationships

Edge Width
- 562,384 — Number
- 350,000 — of co-citing
- 10,000 — 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)

9

B1 Number of papers citing lose-lose relationships (≥100 citing papers)

Mechanical Design Engineering (44; 17.3%)
Chemical, Mechanical, & Civil Engineering

Neural Networks (57; 14.7%)
Electrical Engineering & Computer Science

Cancer Statistics (35; 8.0%)
Math & Physics

Ethnopharmacology (19; 5.4%)
Chemistry

Food Engineering (13; 2.9%)
Biotechnology

Nursing Theory (62; 19.6%)
Medical Specialties

Memory & Cognition (34; 8.7%)
Health Professionals

Science Education (55; 17.6%)
Social Sciences

Impotence (16; 5.7%)
Brain Research

Legend

Node Size
- 62 — Number of
- 25 — lose-lose
- 1 — relationships

Edge Width
- 52,342 — Number
- 30,000 — of co-citing
- 100 — papers

Remote Sensing (35; 10.1%)
Earth Sciences

Human Evolution (31; 9.0%)
Biology

Agricultural Environmental Medicine (19; 10.2%)
Infectious Diseases

Science History (28; 9.4%)
Humanities

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)

10

Bollen, Johan, Herbert Van de Sompel, Aric Hagberg, Luis M.A. Bettencourt, Ryan Chute, Marko A. Rodriquez, Lyudmila Balakireva. 2008. A Clickstream Map of Science.
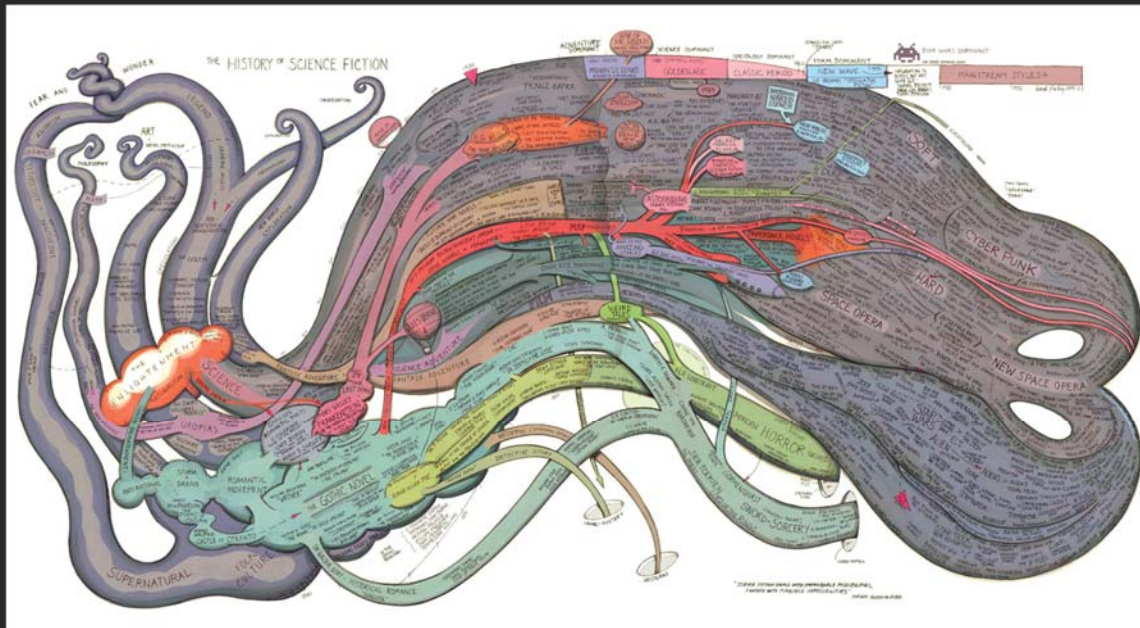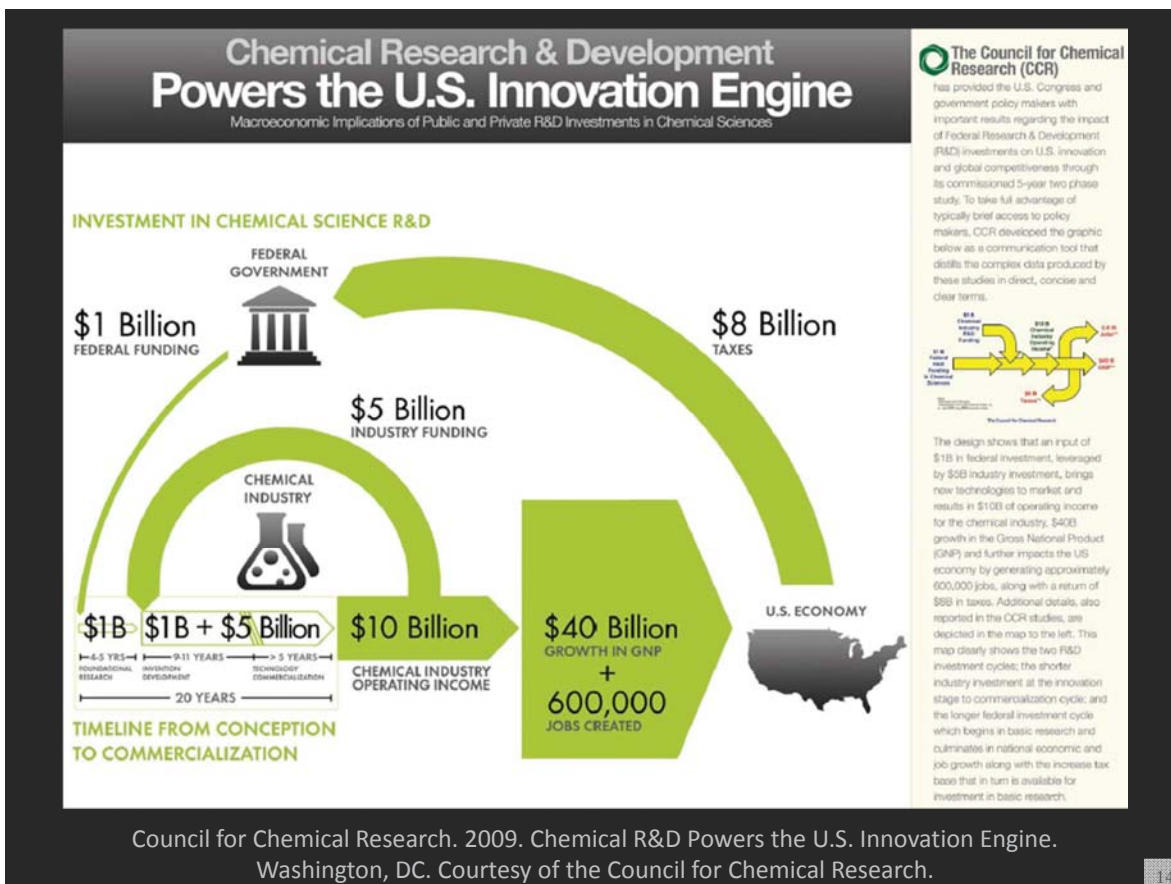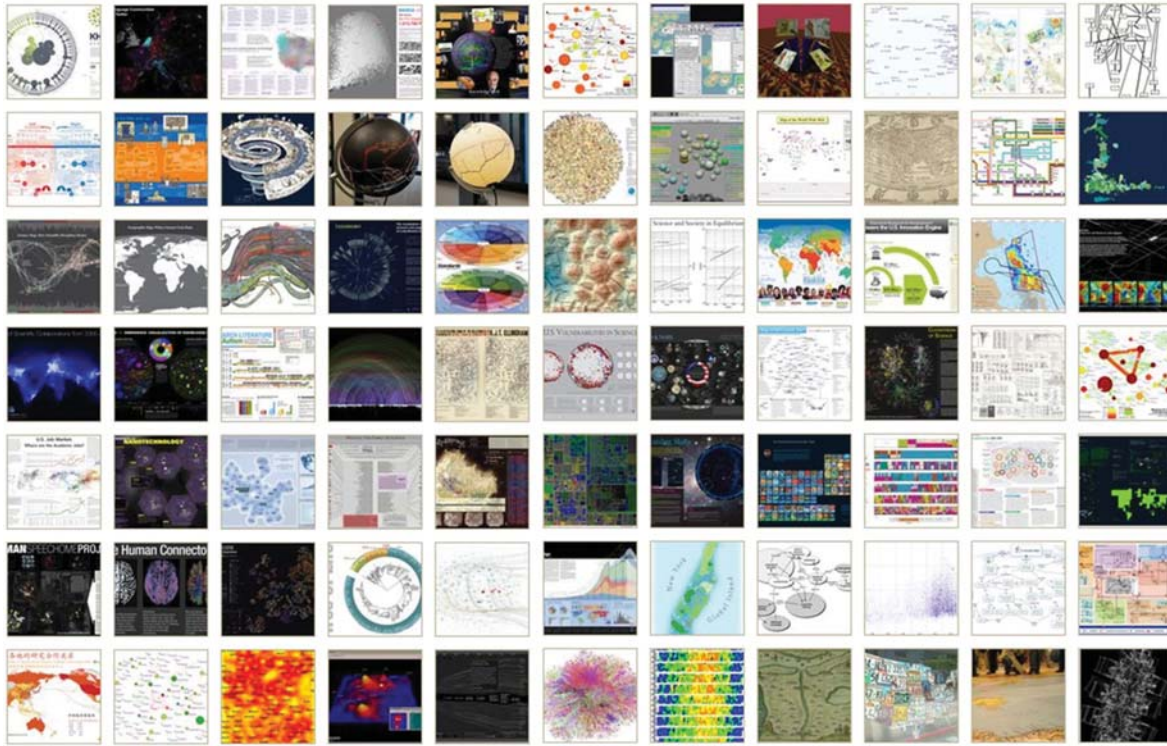
Language Communities of Twitter - Eric Fischer - 2012

Ward Shelley . 2011. History of Science Fiction.

Council for Chemical Research. 2009. Chemical R&D Powers the U.S. Innovation Engine.
Washington, DC. Courtesy of the Council for Chemical Research.

Kristi Holmes @kristiholmes · Apr 30
Excited for @cnscenter Places&Spaces at @galterlibrary! @katycns @NUCATSInstitute #unpackingcrates #viz

Places & Spaces at Northwestern University
May 14 - September 23, 2015

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

Geographic Map: Where Science Gets Done

North & West Asia
Europe
North America
Central America
South & East Asia
Africa
Oceania
South America
Oceania
Australia
Antarctica

Science Map: How Scientific Disciplines Relate

Math & Physics
Chemistry
Health Professionals
Aeronautical, Chemical, Mechanical & Civil Engineering
Medical Specialties
Social Sciences
Electrical Engineering & Computer Science
Biotechnology
Brain Research
Infectious Diseases
Earth Sciences
Biology
Humanities

Copyright © 2008 The Regents of the University of California

About
This Illuminated Diagram display adds the flexibility of an interactive program to the incredibly high data density of a print. This technique is generally useful when there is too much pertinent data to be displayed on a screen but the data is relatively stable. The computer can direct the eye to what's important by using projectors or screens as smart spotlights, animating the research impact of individuals, giving a "grand tour" of science, or highlighting query results (as when you touch the lectern or use the keyboard) with an overlay of moving light.

Top Five Continents
North America - 4,000 records
South & East Asia - 3,589
Australia - 2,431
Africa - 2,208
South America - 1,562

Top Five Scientific Disciplines
Math & Physics - 4,000 records
Health Professionals - 3,589
Social Sciences - 2,431
Aeronautical, Chemical, Mechanical & Civil Engineering - 2,208
Humanities - 1,562

Search
The keyboard supports retrieval and display of papers based on their Medical Subject Headings (MeSH) and MeSH qualifier terms. If multiple terms are entered in a field, they are automatically combined using "OR". So, "breast cancer" matches any record with "breast" or "cancer" in that field. You can put AND between terms to combine with "AND". Thus "breast AND cancer" would only match records that contain both terms. Double quotation can be used to match compound terms, e.g., "breast cancer" retrieves records with the phrase "breast cancer", and not records where "breast" and "cancer" are both present, but the exact phrase.

Input your search query here.

Q W E R T Y U I O P
A S D F G H J K L "
Z X C V B N M ⌫
Space   Go

PLACES & SPACES
http://scimaps.org

People & Topics

Geographic Map: Where Science Gets Done

Science Map: How Scientific Disciplines Relate

Copyright © 2008 The Regents of the University of California

About

This Illuminated Diagram display adds the flexibility of an interactive program to the incredibly high data density of a print. This technique is generally useful when there is too much pertinent data to be displayed on a screen but the data is relatively stable. The computer can direct the eye to what's important by using projectors or screens as smart spotlights, animating the research impact of individuals, giving a "grand tour" of science, or highlighting query results (as when you touch the lectern or use the keyboard) with an overlay of moving light.

Elinor Ostrom - Nobel Prize in Economic Sciences 2009

Born: 7 August 1933, New York, NY, USA

Affiliation at the time of the award: Indiana University, Bloomington, IN, USA, Arizona State University, Tempe, AZ, USA

Prize motivation: "for her analysis of economic governance, especially the commons"

Field: Economic governance

Contribution: Challenged the conventional wisdom by demonstrating how local property can be successfully managed by local commons without any regulation by central authorities or privatization.

Interact

Select any location on the *Geographic Map* location (by brushing your finger over an area on the lectern's touch screen) and topics studied in that area will highlight on the *Science Map*: the brighter a topic glows, the more papers on that topic originated in the selected area. Conversely, touching a scientific area in the *Science Map* illuminates places on the *Geographic Map* where that topic is studied. People and topic buttons support the exploration of publication output by selected Noble laureates and particular lines of research using MEDLINE data from 2000-2009.

| Cancer | Cloning | HIV | Robert G. Edwards | Roger D. Kornberg | Elinor Ostrom |
| Obesity | Quality of Life | Smoking | Stanley B. Prusiner | Ahmed H. Zewail | View All |

Keyword Search

PLACES & SPACES
http://scimaps.org

19

PLACES & SPACES
MAPPING SCIENCE

Curated by the Cyberinfrastructure for Network Science Center

search scimaps.org   Search

About   People   Maps & More   Exhibitions   Hosting   Publications   Store   News   Contact

Hidalgo, César A., Bailey Klinger, Albert-László Barabási, and Ricardo Hausmann. 2007. See also The Product Space map from Phase I of *Places & Spaces*.

Call for Macroscope Tools for the *Places & Spaces: Mapping Science* Exhibit (2015)

http://scimaps.org/call

Themes for the upcoming iterations/years are:

- 11th Iteration (2015): Macroscopes for Interacting With Science
- 12th Iteration (2016): Macroscopes for Making Sense of Science
- 13th Iteration (2017): Macroscopes for Forecasting Science
- 14th Iteration (2018): Macroscopes for Economic Decision Makers
- 15th Iteration (2019): Macroscopes for Science Policy Makers
- 16th Iteration (2020): Macroscopes for Scholars
- 17th Iteration (2021): Macroscopes as Visual Interfaces to Digital Libraries

20
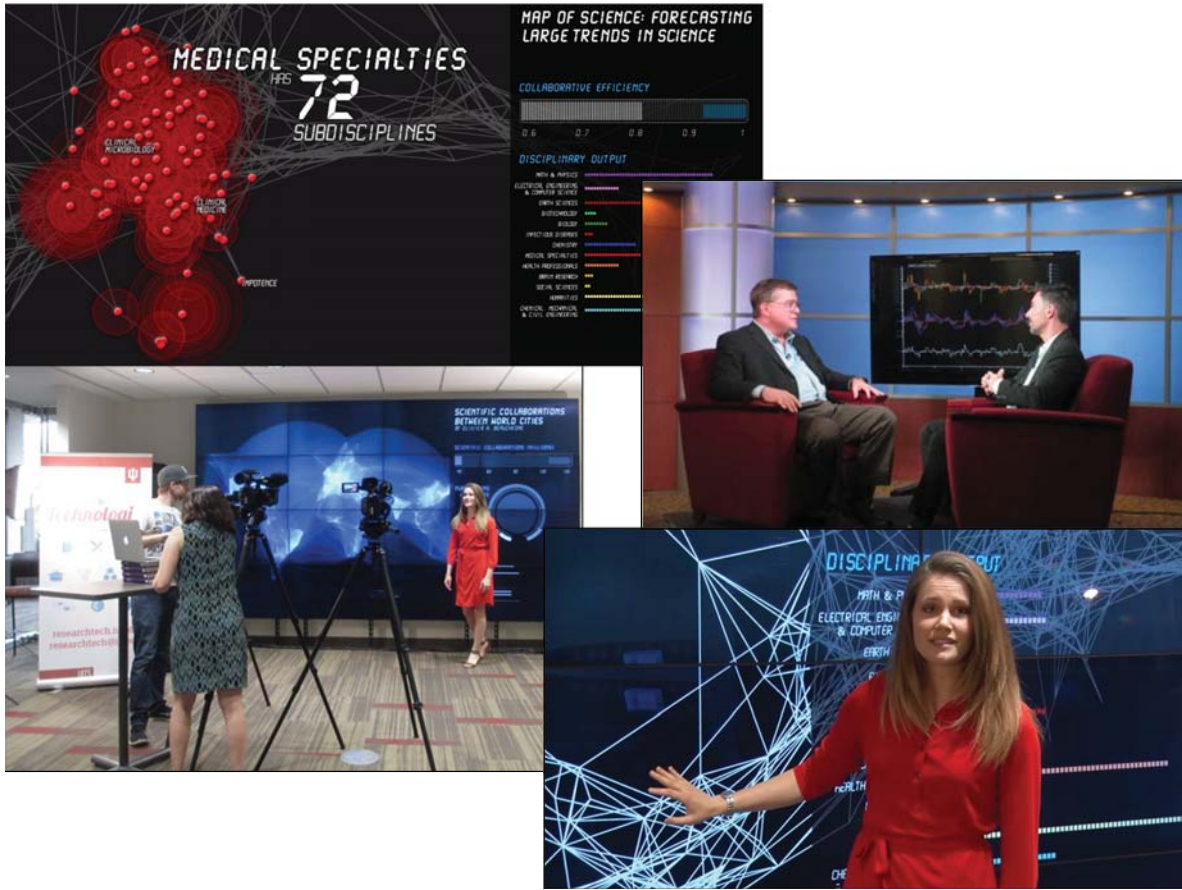
Modelling Our Collective Scholarly Knowledge

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 *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 expert-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 sciences 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 in 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—at a fraction of the overhead required by the current system.
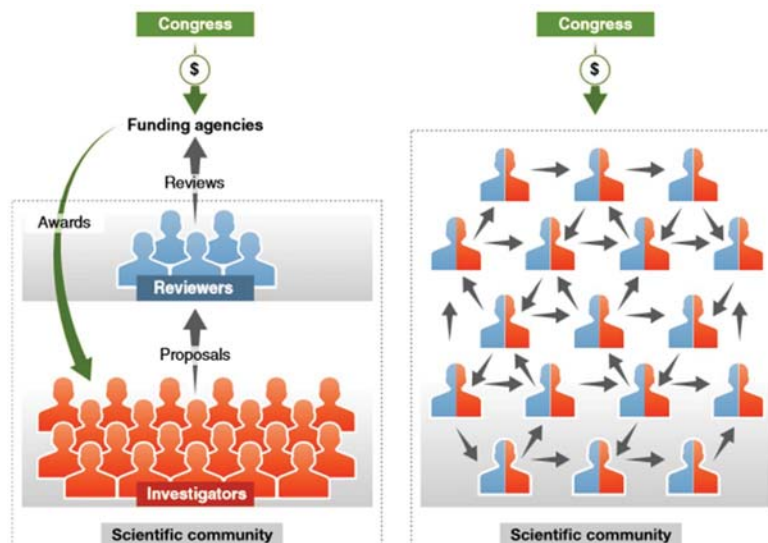
–JDM

February 7, 2014

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

Bollen, Johan, David Crandall, Damion Junk, Ying Ding, and Katy Börner. 2014. EMBO Reports 15 (1): 1-121.

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.
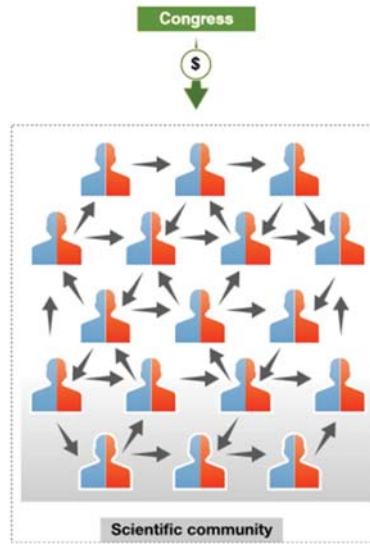
Example:

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.
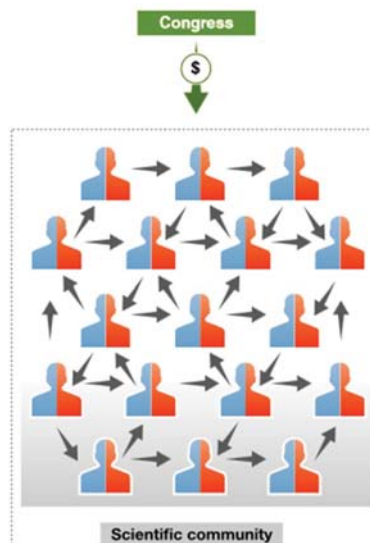
Model Run and Validation:

Model is presented in http://arxiv.org/abs/1304.1067

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.
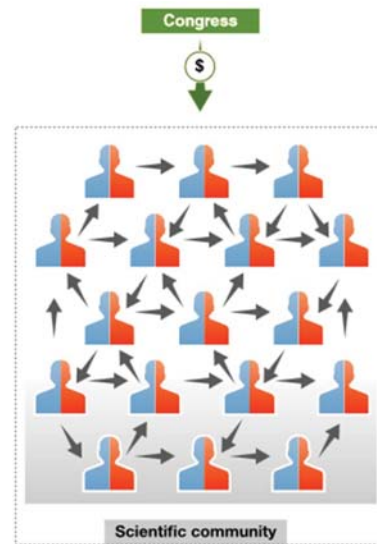
Model Efficiency:

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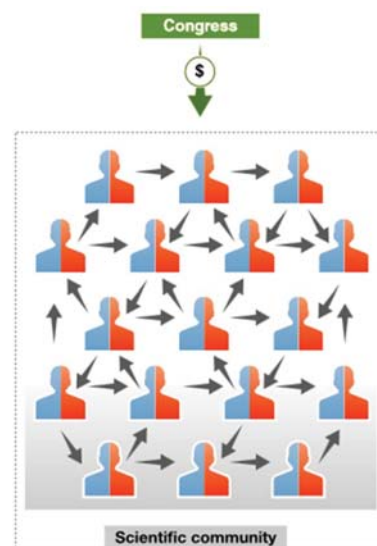
To add: Time spent by researchers to review proposals. In 2012 alone, NSF convened more than 17,000 scientists to review 53,556 proposals.

References

Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003). **Visualizing Knowledge Domains.** In Blaise Cronin (Ed.), *ARIST*, Medford, NJ: Information Today, Volume 37, Chapter 5, pp. 179-255. http://ivl.slis.indiana.edu/km/pub/2003-borner-arist.pdf

Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1). http://www.pnas.org/content/vol101/suppl_1/
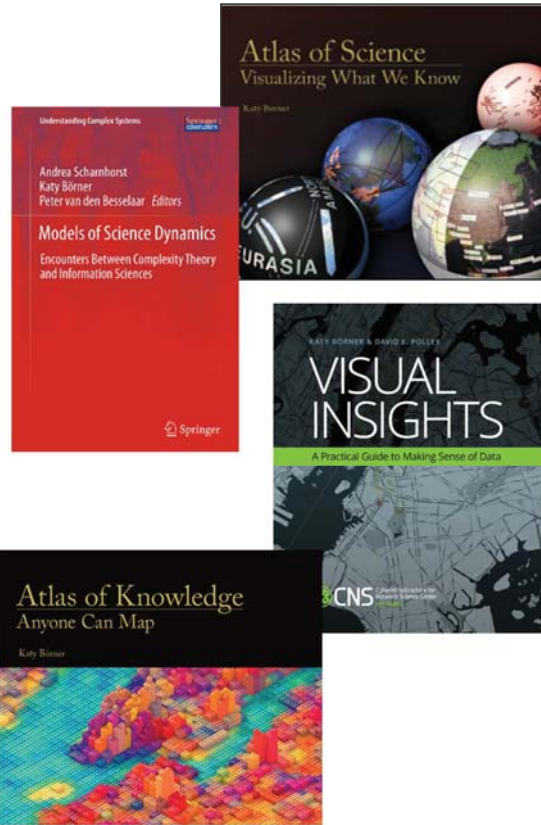
Börner, Katy (2010) **Atlas of Science: Visualizing What We Know**. The MIT Press. http://scimaps.org/atlas

Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2012) **Models of Science Dynamics**. Springer Verlag.

Katy Börner, Michael Conlon, Jon Corson-Rikert, Cornell, Ying Ding (2012) **VIVO: A Semantic Approach to Scholarly Networking and Discovery**. Morgan & Claypool.

Katy Börner and David E Polley (2014) **Visual Insights: A Practical Guide to Making Sense of Data**. The MIT Press.

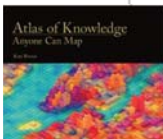Börner, Katy (2015) **Atlas of Knowledge: Anyone Can Map**. The MIT Press. http://scimaps.org/atlas2

Tasks

See page 5

Atlas of Knowledge
Anyone Can Map

See pages 6-7

Insight Need Types page 26	Data Scale Types page 28	Visualization Types page 30	Graphic Symbol Types page 32	Graphic Variable Types page 34	Interaction Types page 26
• categorize/cluster	• nominal	• table	• geometric symbols	• spatial	• overview
• order/rank/sort	• ordinal	• chart	point	position	• zoom
• distributions	• interval	• graph	line	• retinal	• search and locate
(also outliers, gaps)	• ratio	• map	area	form	• filter
• comparisons		• network layout	surface	color	• details-on-demand
• trends			volume	optics	• history
(process and time)			• linguistic symbols	motion	• extract
• geospatial			text		• link and brush
• compositions			numerals		• projection
(also of text)			punctuation marks		• distortion
• correlations/relationships			• pictorial symbols		
			images		
			icons		
			statistical glyphs		

Atlas of Knowledge
Anyone Can Map

See page 24

Graphic Variable Types Versus Graphic Symbol Types

See pages 36-39

Information Visualization MOOC 2015

INDIANA UNIVERSITY · CNS

Overview

This course provides an overview about the state of the art in information visualization. It teaches the process of producing effective visualizations that take the needs of users into account.

The course can be taken for three Indiana University credits as part of the Online Data Science Program, as part of the Information and Library Science M.S. program, and as part of the online Data Science M.S. Program offered by the School of Informatics and Computing. Students seeking enrollment information should contact Rhonda Spencer at 812-855-2018, ilsmain@indiana.edu or datasci@indiana.edu.

Among other topics, the course covers:

- Data analysis algorithms that enable extraction of patterns and trends in data
- Major temporal, geospatial, topical, and network visualization techniques
- Discussions of systems that drive research and development.

IVMOOC: Information Visualization MOOC 2015

Information Visualization MOOC · ivmooc.cns.iu.edu

Register for Course

Already registered? Click here to go to the course.
Forgot your password? Click here to reset it.

Register for free at http://ivmooc.cns.iu.edu. Class restarts January, 2016.

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

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