

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

European Food Safety Authority (EFSA)
Scientific Conference: Shaping the Future of Food Safety, Together
Milan, Italy
October 14-16, 2015

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

Computed Using Data from Elsevier's Scopus

Map of Scientific Collaborations from 2005-2009



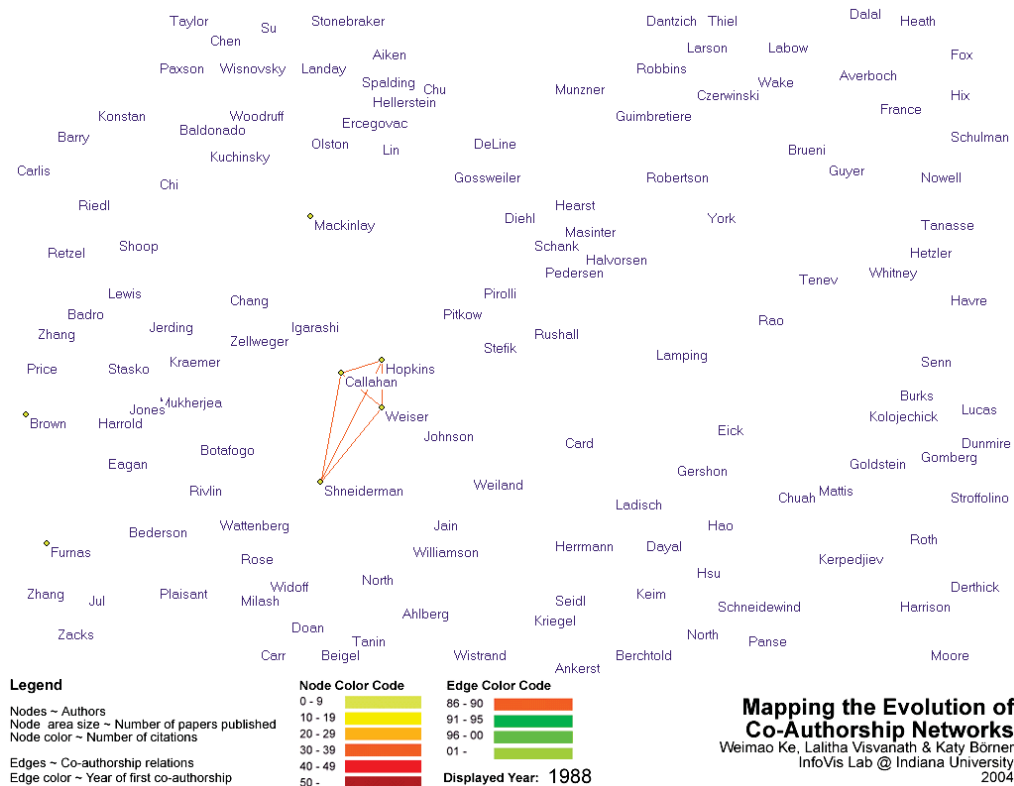
Olivier H. Beauchesne, 2011. Map of Scientific Collaborations from 2005-2009.

Computed Using Data from Elsevier's Scopus

Analyzing and Visualizing S&T

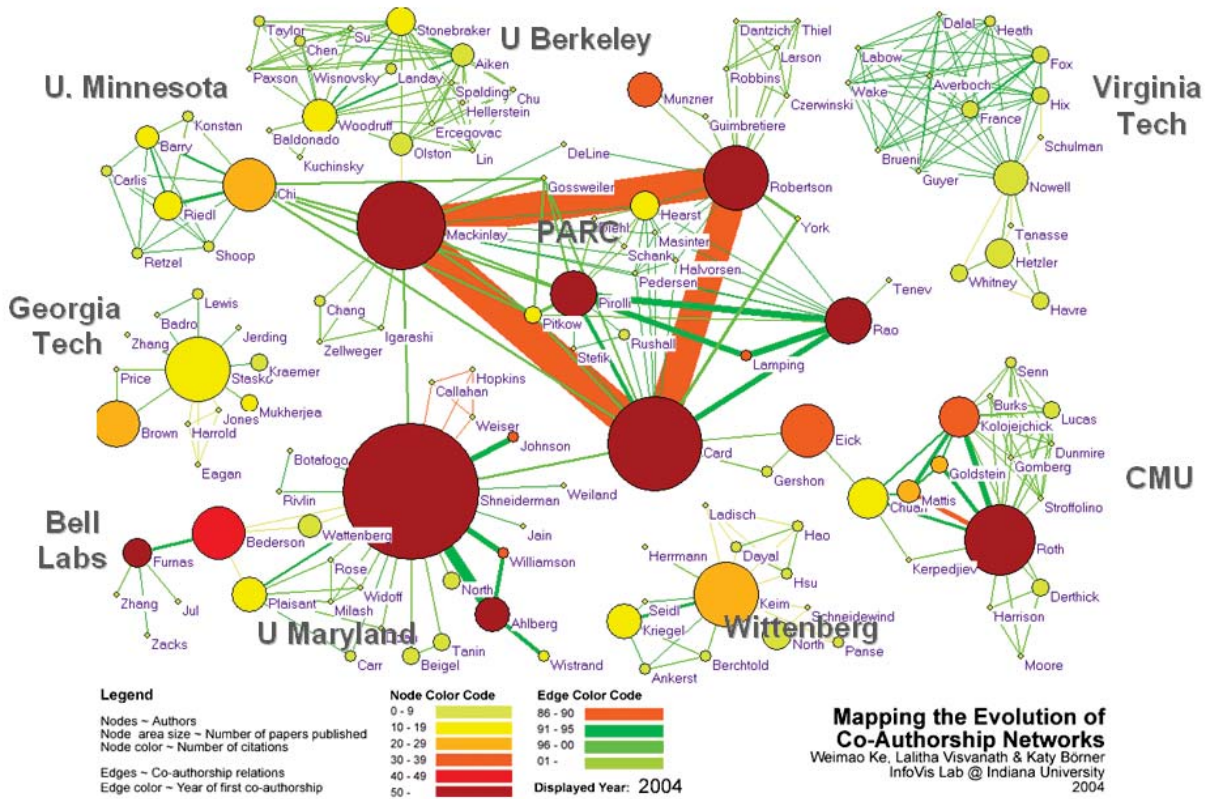
Mapping the Evolution of Co-Authorship Networks

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.



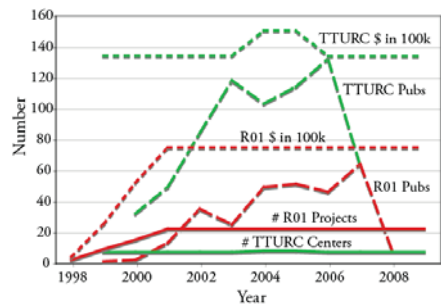
Mapping Transdisciplinary Tobacco Use Research Centers Publications

Compare R01 investigator-based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.

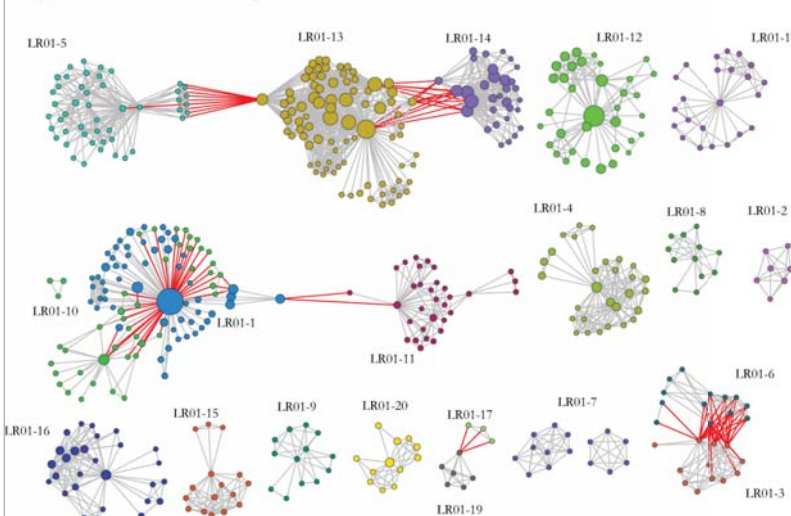
Stipelman, Hall, Zoss, Okamoto, Stokols, Börner, 2014.

Supported by NIH/NCI Contract HHSN261200800812

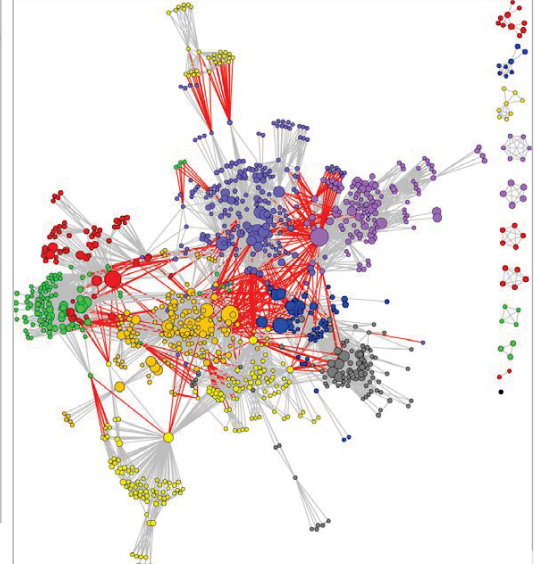
R01 & TTURC Project Information



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.

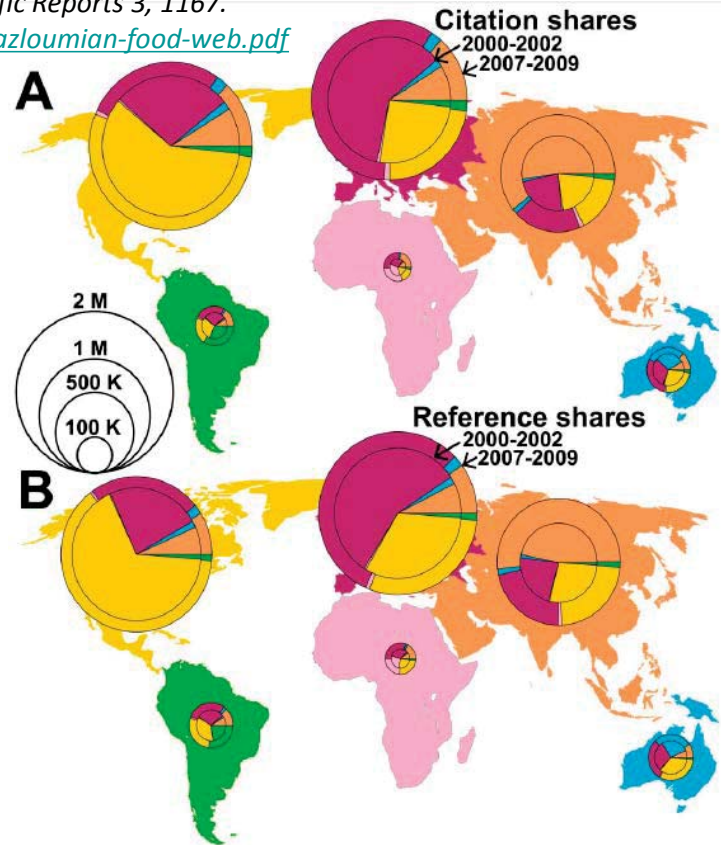
<http://cns.iu.edu/docs/publications/2013-mazloumian-food-web.pdf>

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.



7

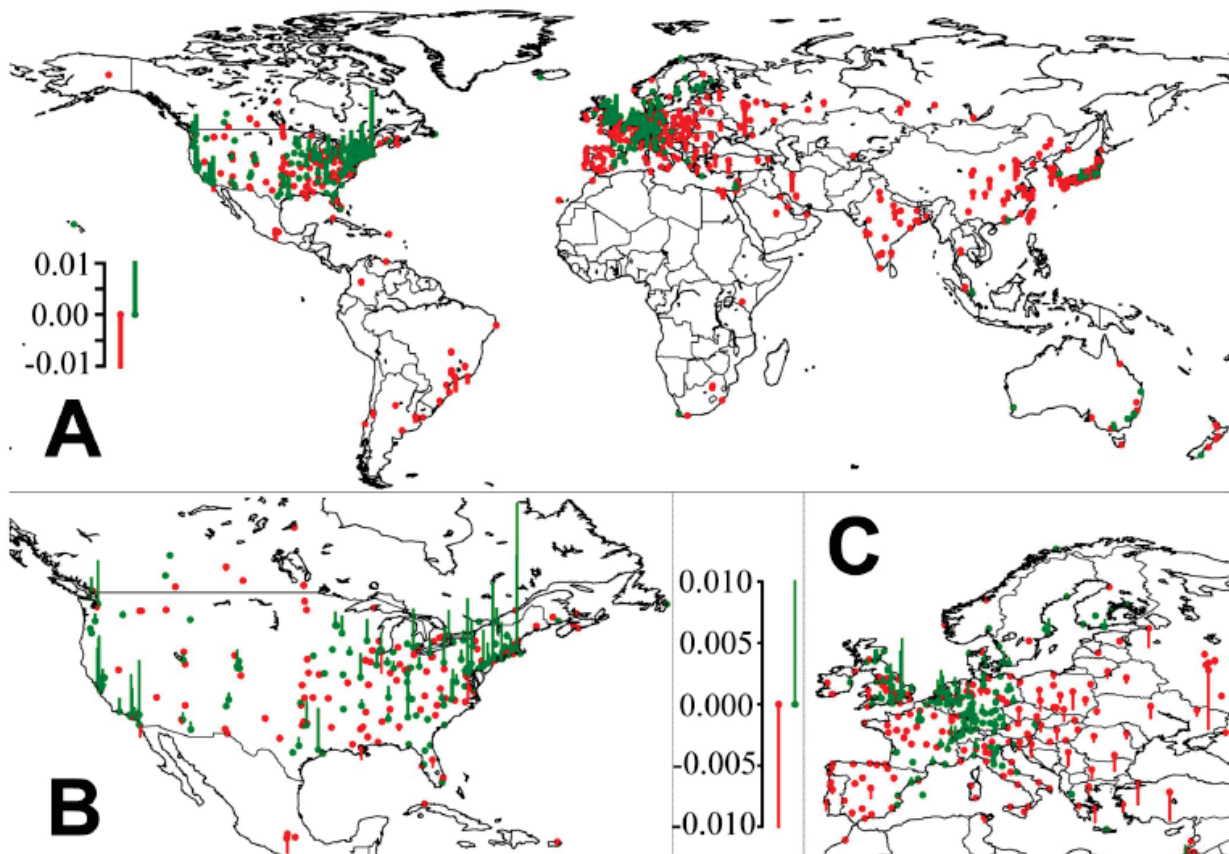


Figure 2 | World map of the greatest knowledge sources and sinks, based on our scientific fitness index. Green bars indicate that the number of citations received is over-proportional, red that the number of citations received is lower than expected (according to a homogeneous distribution of citations 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).

8

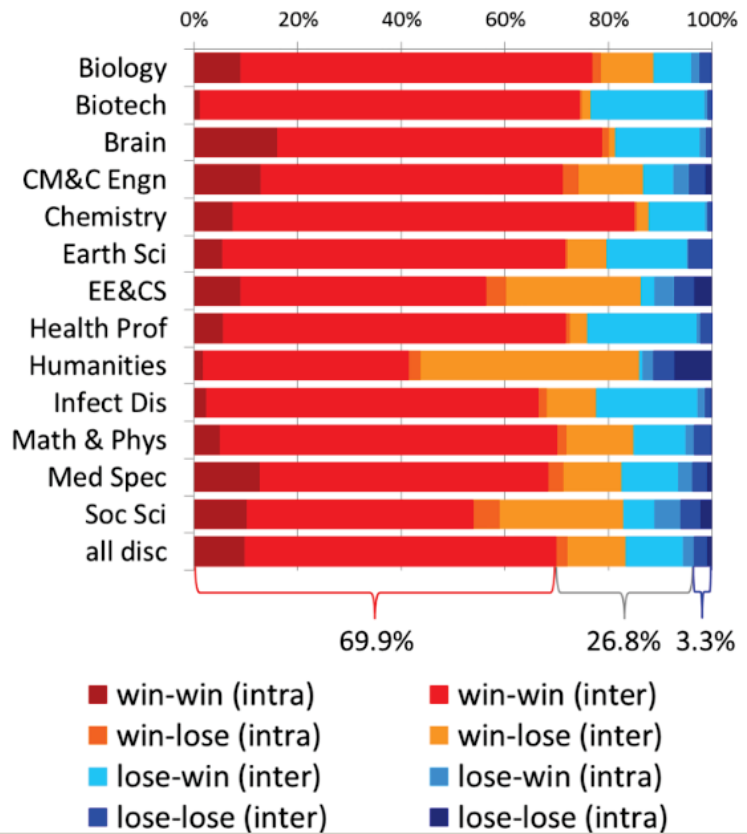
Long-Distance Interdisciplinarity Leads to Higher Scientific Impact

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.

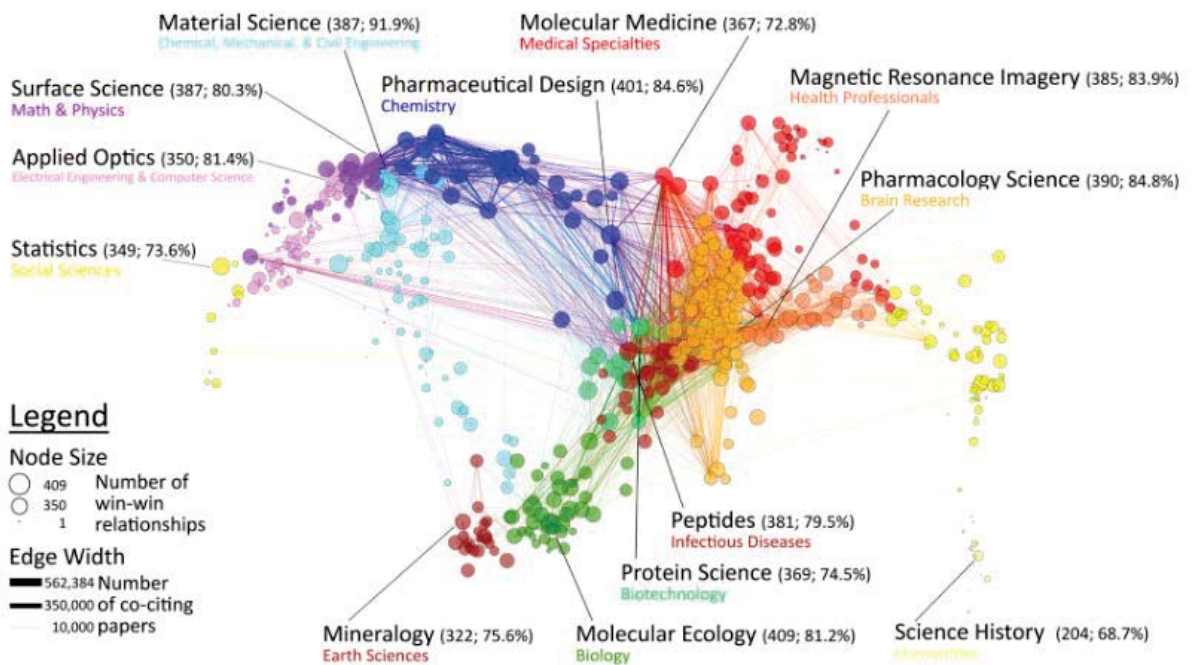
UCSD map of science is used to compute “distance.”



9

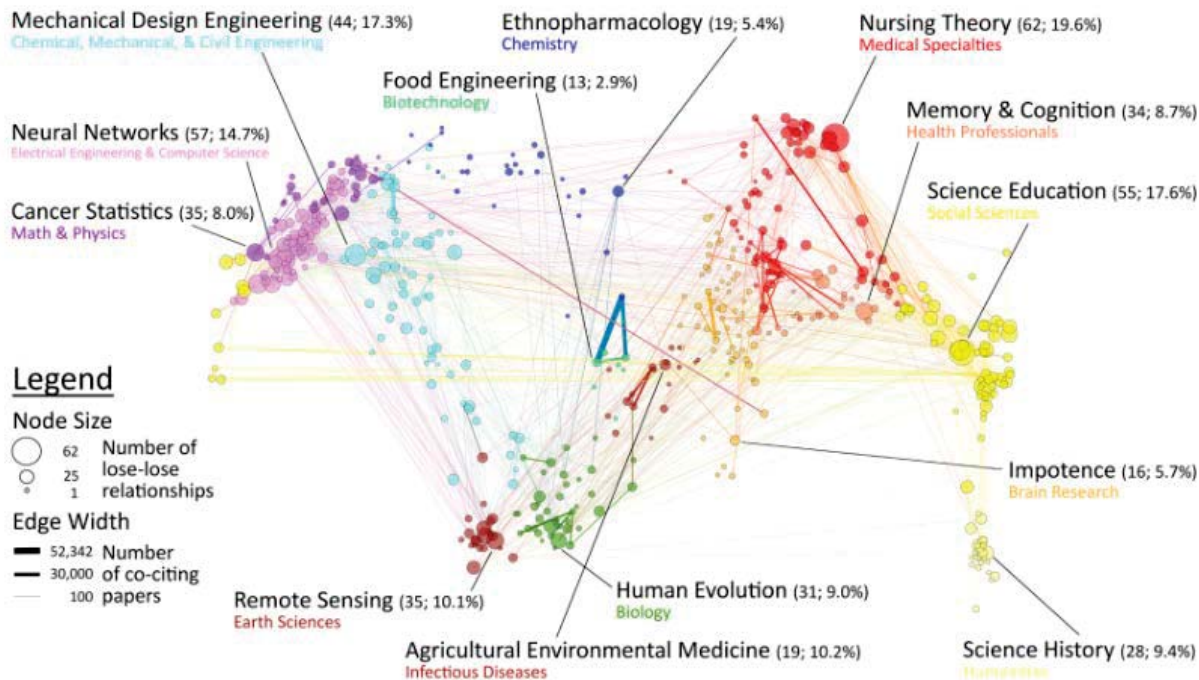
A1 Number of papers citing win-win relationships (≥10,000 citing papers)

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)

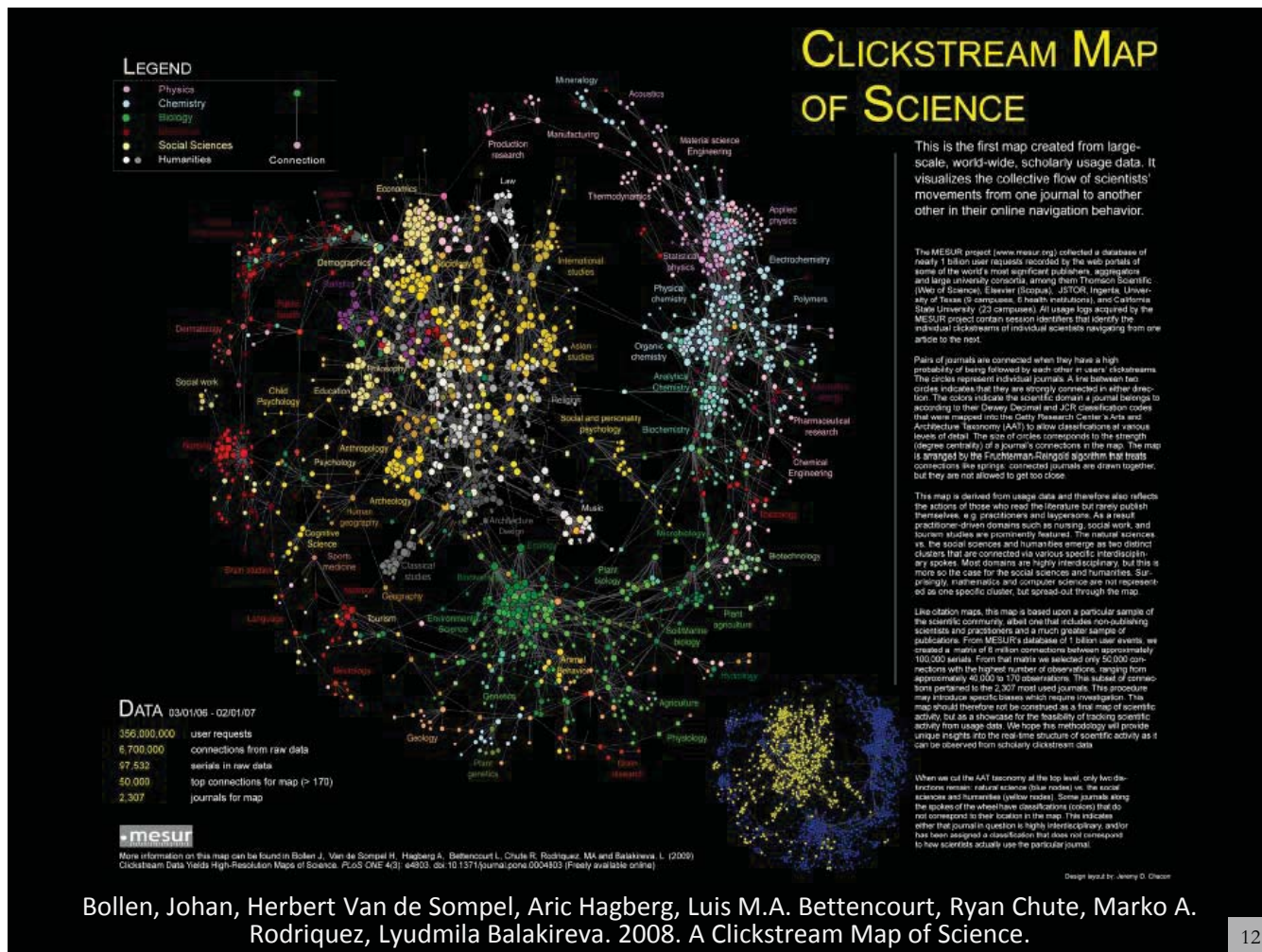
B1 Number of papers citing lose-lose relationships (≥100 citing papers)



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)

11



12

Science Phylomemy

THE RISE AND FALL OF SCIENTIFIC FIELDS

David Chavalarias (dchav@ISIC-CNRS-BES) and Jean-Philippe Cointet (jpc@ISIC-CNRS-BES)

Phylomemes are based on the analysis of the textual content of publications. They describe how the scientific fields evolve and provide a convenient model to investigate science evolution.

The map apposite has been generated by applying the methodology of phylomemy reconstruction to the domain of Future and Emerging Technologies (FET), defined by the FET Open Funding scheme (7th Framework Program of the European Union - EU FP7). We considered all the keyword terms given by authors of projects submitted to FET Open in 2010 (n=2000) in order to delineate the vocabulary associated with FET. These terms have been indexed in the titles and abstracts of a representative sample of worldwide literature, dating from 1900 to 2010 (Thomson Web of Science, >32M publications). Using thematic proximity based on co-occurrence, terms were clustered to identify fields of scientific research.

Each scientific field was then represented by a set of terms, inter-temporal matching between thematic fields results in evolving branches of science that might show several kinds of transition: fields can gain new terms or lose terms, merge with another field, split or even die—if the underlying scientific community loses its thematic cohesion.

Exemplarily shown on the right are all branches in the domain of FET—from Neuroprosthetics Prosthetic Socket on top to Active Circuits, Quantum Metrology on bottom. The evolution of the branches, i.e., changes in the number and composition of their scientific fields—is plotted from left to right. A close-up of the Neuroprosthetics Prosthetic Socket field is given on the far right. Here, time runs top-down and each scientific field has a title and associated keywords that are color and size coded by importance.

The main events in this branch are well recurrent: the emergence of new terms as well as the branching and merging events correspond to important steps in the development of this science such as seminal papers, first clinical trials, etc. Nomadic concepts that migrate from one scientific field to another can be identified. Notice the increase of the branch width when the discipline starts to have commercial applications.

The study of science phylomemes might pave the way towards prediction of science evolution. Indeed, Chavalarias & Cointet (1) used two biomedical datasets (embryology science and network) to demonstrate that fields do not emerge, decline or hybridize at random: the likelihood of observing dynamic events strongly depends on the structural properties of the fields, such as the density index introduced by Callon et al. in 1991 (2). The probability that a scientific field will decline increases drastically as its density decreases—low density fields are more than twice as likely to decline than high-density ones.

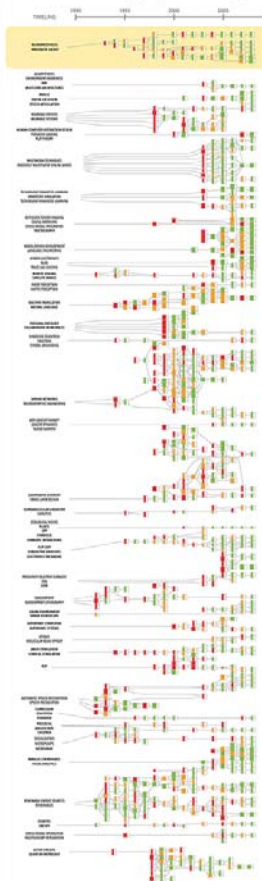


References:

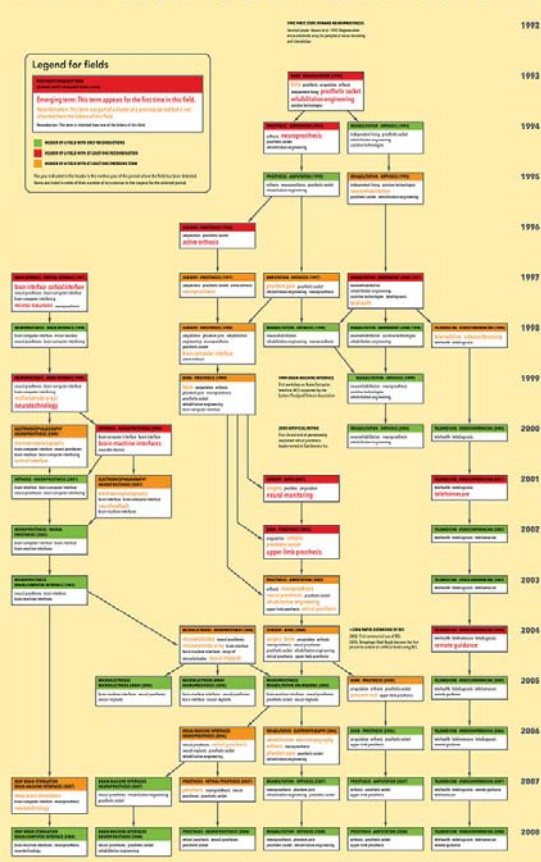
- [1] Chavalarias, David, and Jean-Philippe Cointet. 2013. "Phylomeric Patterns in Science Evolution: The Rise and Fall of Scientific Fields." *PLoS ONE* 8:2.
- [2] Callon, Michel, Jean-Pierre Courtial, and Françoise Laville. 1991. "Co-word Analysis as a Tool for Describing the Network of Interaction between Basic and Technological Research: The Case of Polymer Chemistry." *ScienceMetrics* 22:181-205.

BROWSE THE FULL PHYLOMEMY
fsetphilo.sciencemapping.com

Future and Emerging Technologies



NEUROPROSTHESIS - PROSTHETIC SOCKET

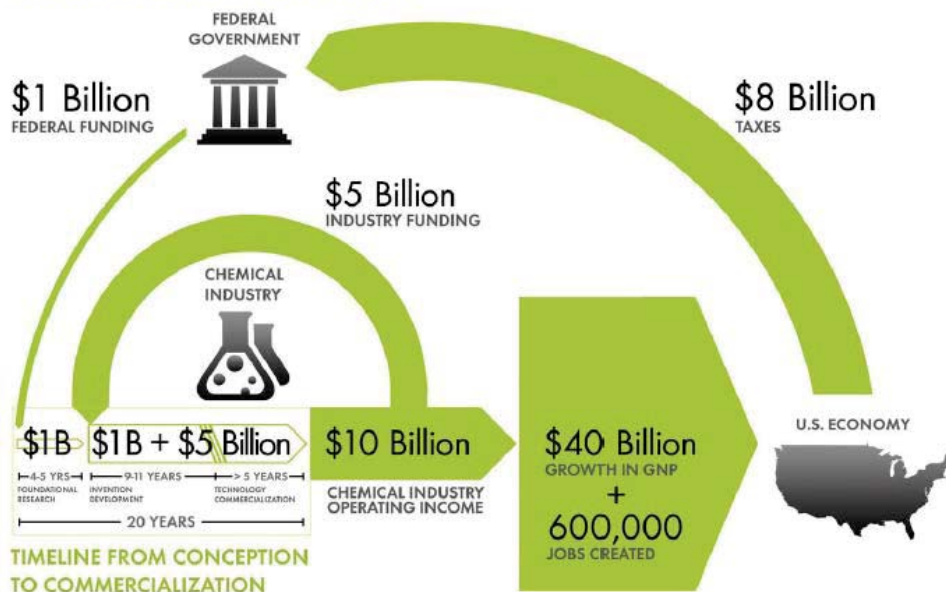


Science Phylomemy - David Chavalarias and Jean-Philippe Cointet - 2013

Chemical Research & Development Powers the U.S. Innovation Engine

Macroeconomic Implications of Public and Private R&D Investments in Chemical Sciences

INVESTMENT IN CHEMICAL SCIENCE R&D



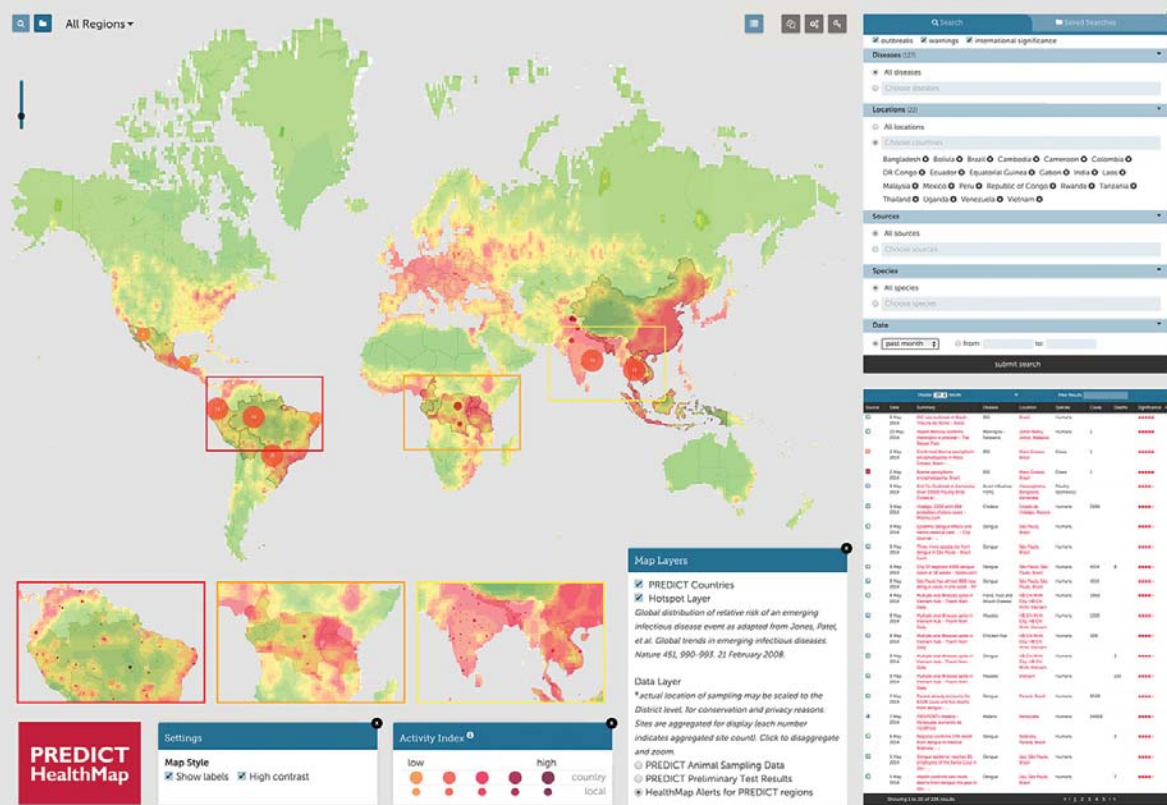
The Council for Chemical Research (CCR)

has provided the U.S. Congress and government policy makers with important results regarding the impact of Federal Research & Development (R&D) investments on U.S. innovation and global competitiveness through its commissioned 5-year two phase study. To take full advantage of typically brief access to policy makers, CCR developed the graphic below as a communication tool that distills the complex data produced by these studies in direct, concise and clear terms.



The design shows that an input of \$1B in federal investment, leveraged by \$5B industry investment, brings new technologies to market and results in \$10B of operating income for the chemical industry, \$40B growth in the Gross National Product (GNP) and further impacts the US economy by generating approximately 600,000 jobs, along with a return of \$8B in taxes. Additional details, also reported in the CCR studies, are depicted in the map to the left. This map clearly shows the two R&D investment cycles; the shorter industry investment at the innovation stage to commercialization cycle; and the longer federal investment cycle which begins in basic research and culminates in national economic and job growth along with the increase tax base that in turn is available for investment in basic research.

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

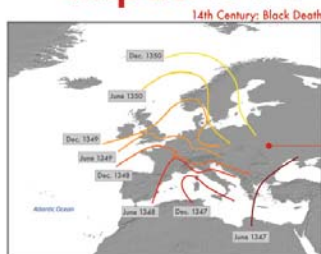


Use the original online tool at healthmap.org/predict



PREDICT: HealthMap - John Brownstein, Damien Joly, William Karesh, Peter Daszak, Nathan Wolfe, Tracey Goldstein, Susan Aman, Clark Freifeld, Sumiko Mekaru, Tammie O'Rourke, Stephen Morse, Christine Kreuder Johnson, Jonna Mazet, and the PREDICT Consortium - 2014

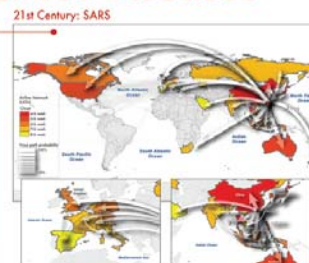
• Impact of Air Travel on Global Spread of Infectious Diseases •



Epidemic spreading pattern changed dramatically after the development of modern transportation systems.

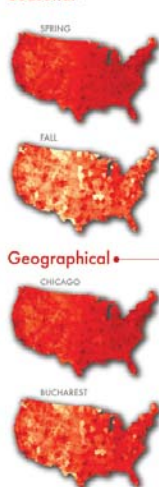
In pre-industrial times disease spread was mainly a spatial diffusion phenomenon. During the spread of Black Death in the 14th century Europe, only few traveling means were available and typical trips were limited to relatively short distances on the time scale of one day. Historical studies confirm that the disease diffused smoothly generating an epidemic front traveling as a continuous wave through the continent at an approximate velocity of 200-400 miles per year.

The SARS outbreak on the other hand was characterized by a patched and heterogeneous spatio-temporal pattern mainly due to the air transportation network identified as the major channel of epidemic diffusion and ability to connect far regions in a short time period. The SARS maps are obtained with a data-driven stochastic computational model aimed at the study of the SARS epidemic pattern and analysis of the accuracy of the model's predictions. Simulation results describe a spatio-temporal evolution of the disease (color coded countries) in agreement with the historical data. Analysis on the robustness of the model's forecasts leads to the emergence and identification of epidemic pathways as the most probable routes of propagation of the disease. Only few preferential channels are selected (arrows; width indicates the probability of propagation along that path) out of the huge number of possible paths the infection could take by following the complex nature of airline connections (light grey; source: IATA).



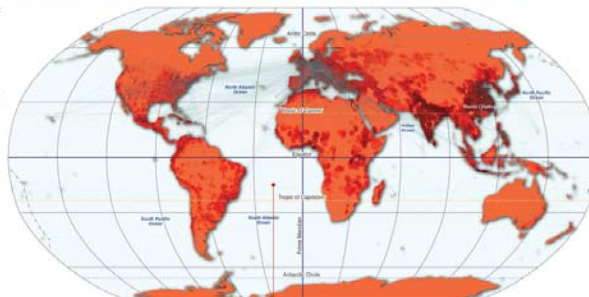
• Forecasts of the Next Pandemic Influenza •

Seasonal

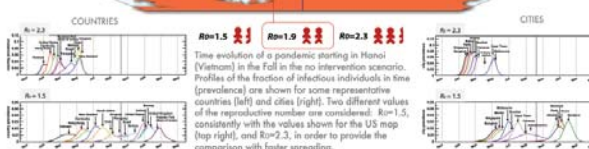


Forecasts are obtained with a stochastic computational model which explicitly incorporates data on worldwide air travel and detailed census data to simulate the global spread of an influenza pandemic. The modeling approach considers infection dynamics (i.e., virus transmission, onset of symptoms, infectiousness, recovery, etc.) among individuals living in urban areas around the world, and assumes that individuals are allowed to travel from one city to another by means of the airline transportation network.

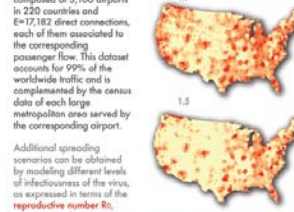
The central map represents the cumulative number of cases in the world after the first year from the start of a pandemic influenza with $R_0=1.9$ originating in Hanoi (Vietnam) in the Spring.



The US maps focus on the situation in the US after one year, and show the effect of changes in the original scenario analyzed. Different color coding is used for the sake of visualization.



Reproductive Number (R_0)



Intervention



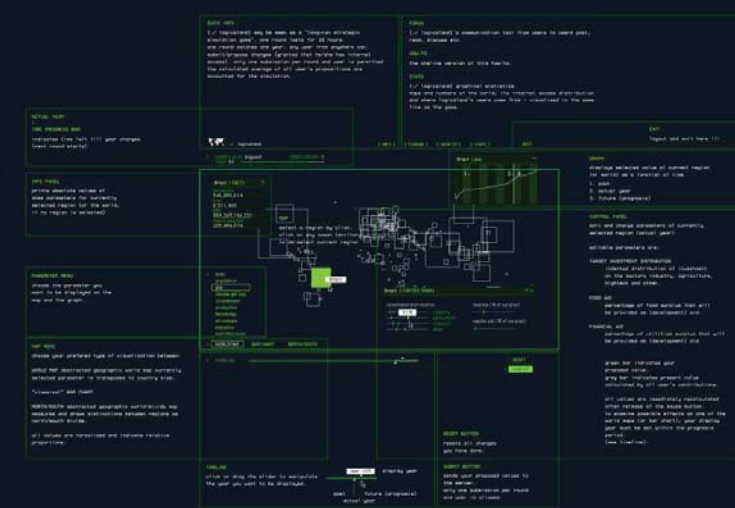
PARTICIPATIVE GLOBAL SIMULATION
WWW.LOGICALAND.NET

LOGICALAND IS A PRIVATE BUREAU FOR VISUALIZING OUR WORLD'S COMPLEX ECONOMIC, POLITICAL AND SOCIAL SYSTEMS.

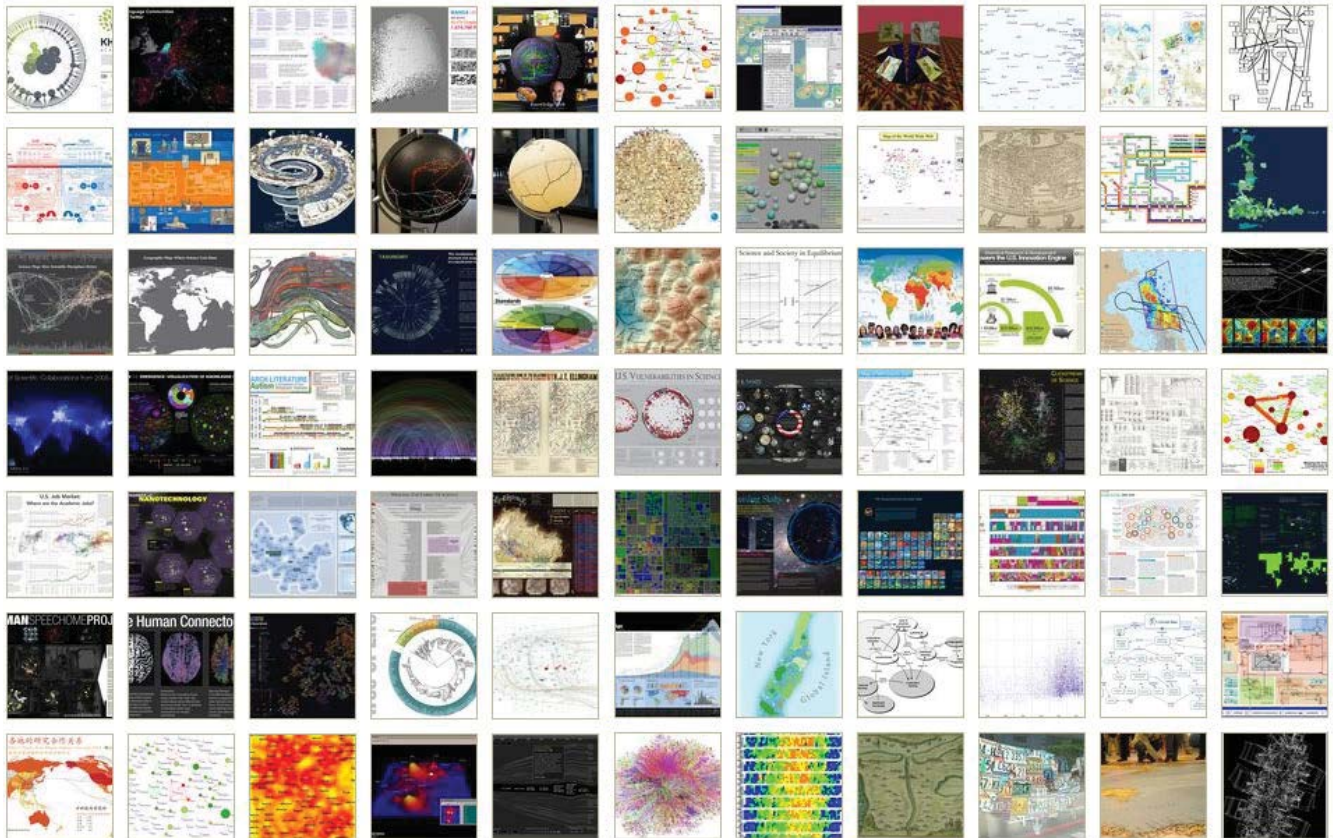
LOGICALAND V0.1 IS THE FIRST ATTEMPT TO REALIZE A PHOTONIC OF A GLOBAL SIMULATION THAT IS TO BE CONTROLLED BY A COMMUNITY OF PARTICIPANTS. IT IS BASED ON A GLOBAL WORLD MODEL DEVELOPED IN THE 1970S THAT HAS BEEN TAKEN OUT OF ITS ORIGINAL CONTEXT AND ADAPTED INTO A PARTICIPATIVE ONLINE GAME. IN ORDER OF PLAY LASTING UP TO 10 HOURS, ECONOMIC AND NATURAL RESOURCES ENDOWMENTS OF 180 STATES ARE SIMULATED IN AN INTERDEPENDENT WORLD OVERVIEW. THE SIMULATION STARTS WITH REAL VALUES FROM THE YEAR 2001, TAKEN FROM THE STATISTICS COMPANION OF THE WORLD FACT BOOK. THE PARAMETERS CHANGES MADE BY PARTICIPANTS SECURE "GAMES" THAT ARE FILLED BY THE SERVER AND FEED BACK INTO THE SIMULATION.

GLOBAL WORLD MODELS CAN BE INTERPRETED AS "COMPLEX PROGRAMS THAT SIMULATE THE WORLD IN VERY SMALL, COMPRESSIVE FORMS. CONCEPTUALLY, THEY ENCOMPASS THE ENTIRE WORLD OR AT LEAST A MAJOR PORTION OF IT. MORE PRACTICALLY, THEY EXPLICITLY LINK CONTENTS IN NUMBERS OF COMMENTS ON ASPECTS OF OUR WORLD SUCH AS ECONOMIC, DEMOGRAPHIC, POLITICAL, AND THE ENVIRONMENT. BECAUSE OF THESE TIGHTLY INTERLINKED GLOBAL MODELS CAN BE AND ARE USED AS TOOLS TO HELP US UNDERSTAND PROCESSES WHOSE EFFECTS CROSS NATIONAL BOUNDARIES AND WHOSE STATE CHANGES PRODUCE MAJOR CONSEQUENCES. (FROM WIKIPEDIA)

LOGICALAND IS A PHOTOGRAPHIC WORLD MODEL AND A PARTICIPATIVE ONLINE GAME. THE MAIN PURPOSE IS TO VISUALIZE THE COMPLEXITY OF OUR WORLD AND TO HELP US UNDERSTAND PROCESSES WHOSE EFFECTS CROSS NATIONAL BOUNDARIES AND WHOSE STATE CHANGES PRODUCE MAJOR CONSEQUENCES. (FROM WIKIPEDIA)



Logicland Participative Global Simulation - Michael Ashauer, Maia Gusberti, Nik Thoenen - 2002





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

21



Places & Spaces Exhibit at the David J. Sencer CDC Museum, Atlanta, GA
January 25-June 17, 2016

22



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

Geographic Map: Where Science Gets Done

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.

Science Map: How Scientific Disciplines Relate

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

Input your search query here.

People & Topics

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.

<http://scimaps.org>

Geographic Map: Where Science Gets Done

Science Map: How Scientific Disciplines Relate

Copyright © 2009 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.

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

Curated by the Cyberinfrastructure for Network Science Center

Search

f

t

✉

[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 Macroscopic 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



Modelling Science

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

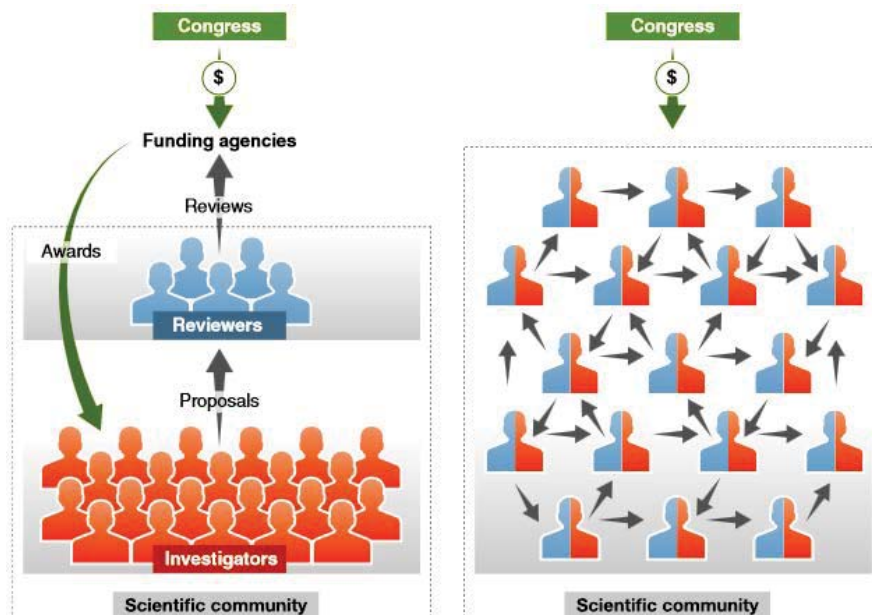
Science 7 February 2014: Vol. 343 no. 6171 p. 598

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

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

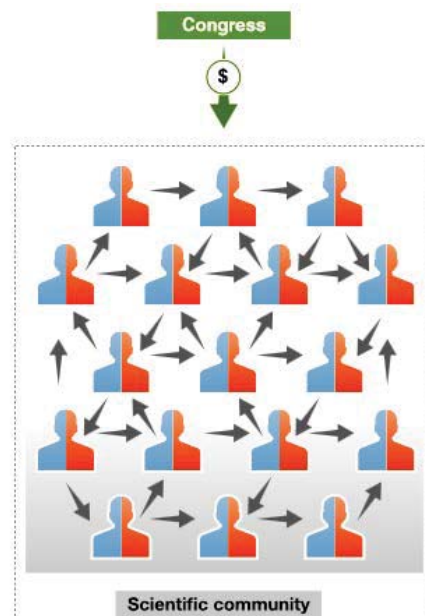
Each year,

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.



31

Example:

Total funding budget in year is 2012 NSF budget

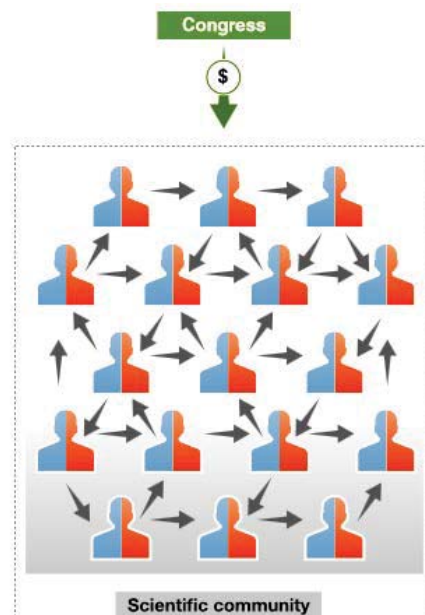
Given the number of NSF funded scientists, each receives a \$100,000 basic grant.

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.



32

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

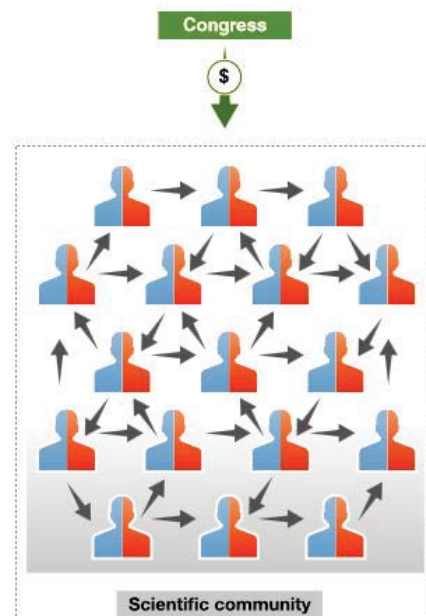
the **867,872 names** who had authored at least one paper per year in any five years of the period 2000–2010.

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.



33

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:

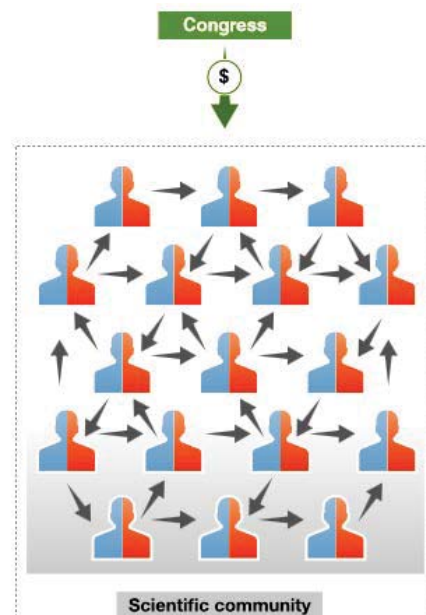
If four professors work four weeks full-time on a proposal submission, labor costs are about \$30k. With typical funding rates below 20%, about five submission-review cycles might be needed resulting in a total expected labor cost of **\$150k**.

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.






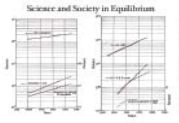



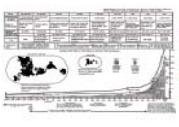

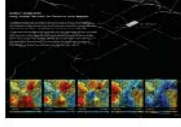
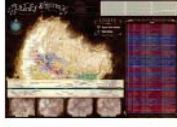









To add: Time spent by researchers to review proposals. In 2012 alone, NSF convened more than 17,000 scientists to review 53,556 proposals.

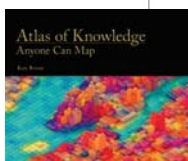
34

Information Visualization Framework & IVMOOC

Tasks

LEVELS

TYPES	MICRO: Individual Level about 1–1,000 records page 6	MESO: Local Level about 1,001–100,000 records page 8	MACRO: Global Level more than 100,000 records page 10
Statistical Analysis page 44 	 Knowledge Cartography page 135	 Productivity of Russian life sciences research teams page 105	 Science and Society in Equilibrium Number of scientists versus population and R&D costs versus GNP. page 103
WHEN: Temporal Analysis page 48 	 Visualizing decision- making processes page 95	 Key events in the development of the video tape recorder page 85	 Increased travel and communication speeds page 83
WHERE: Geospatial Analysis page 52 	 Cell phone usage in Milan, Italy page 109	 Victorian poetry in Europe page 137	 Ecological footprint of countries page 99
WHAT: Topical Analysis page 56 	 Evolving patent holdings of Apple Computer, Inc. and Jerome Lemelson page 89	 Evolving journal networks in nanotechnology page 159	 Product space showing CO-export patterns of countries page 93
WITH WHOM: Network Analysis page 60 	 World Finance Corporation network page 87	 Electronic and new media art networks page 133	 World-wide scholarly collaboration networks page 157



See page 5

Micro: Individual Level

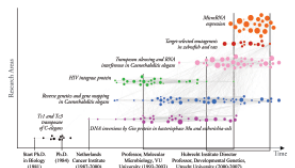
This spread reviews significant findings derived from micro-level studies. Traditionally, personal or product studies start with author, publication, patent, or sales records, from which properties or relations are then extracted. The increasing availability of other important digital data, such as news, stock market, or social media data, offers an opportunity to obtain a richer and more real-time understanding of S&CT developments. In addition to expanding the breadth of data sets studied, that availability also increases the depth of studies by mining the full text for specific tokens, such as chemical compounds, or parsing acknowledgment sections for specific names. Sentiment analysis, also called opinion mining, is now commonly applied to understanding how new ideas and products are perceived and adopted in different markets.

What is your impact if your work is not indexed in the Web of Science database?

Personal Analytics

Insights, governance, and leading agencies are all increased in objective measures that can predict the value of goods, products, or ideas. Items are analyzed with simple numbers, such as words or total number of citations per person, as that enables supporting ranking and the application of feedback. Cohen demands a more holistic understanding of the context in which people or organizations are operating and the many different contributions that individuals make over the course of their careers. The graph below completely shows the trajectory of a whole over time, through backward views, such as starting or completing a PhD and obtaining positions as professor or assistant director, all of

which are included in the horizontal axis. The whole's publications are divided by circles which are size-coded by the number of citations and color-coded and labeled and mostly grouped by research team. Links denote references between publications. The figure was published by Fabrice Huettinger and Thomas Gartner who studied the academic life-cycle of condensed matter physicists and found that scholars commonly work on several lines research until their career, that they likely work in different research areas to parallel, and career changes (e.g., move position or affiliation) have a strong impact on the time and kind of pursued research.

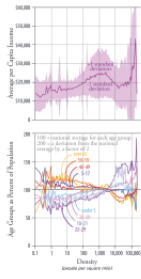


Quantifying Success

People are born to succeed in life, work, and to live. As such, many different approaches exist to quantify the success and reach of individuals based on wealth, tenure, reputation, or other properties. These success stories are often measured in terms of financial wealth, from monthly income to the value of their portfolio. As of March 2014, Bill Gates, co-chair of the Bill & Melinda Gates Foundation, is the world's richest person, with a net worth of \$81.4 billion. Microsoft's success is also reflected in the fact that its stock price has risen from \$28 in 1986 to over \$300 in 2014.

The social structure and connections that one person engages in may be considered another type of wealth. The more connections one has, the higher one's ability to disseminate news or influence others. An individual's position in a social or business network matters. People have a higher value if they are the only ones connecting two communities. Online networking sites such as Facebook or LinkedIn make it easy to map or contact with friends, colleagues, and fans. These links are commonly log user activities. These log files can subsequently be used to produce interactive visualizations, which can be explored as easy links. For example, WilliamsAlpha lets users render their personal Facebook data in different ways. Exemplary visualizations of this data are given in the complete image on the opposite page. The report interpretation shows the name of the Facebook account holder and the name of the report generation. Click below to have personal information. The friend network shows clusters of friends and their interconnections (degree of success, "what" and "when" questions, see page 70). The friend network shows clusters of friends and their interconnections (degree of success, "what" and "when" questions, see page 70). The friend network shows clusters of friends and their interconnections (degree of success, "what" and "when" questions, see page 70).

papers count more than those by authors or papers that are cited or not cited. It is important to distinguish between scholarly impact index calculated by summing up citations, funding, or the number of students taught and actual impact like, for example, patents, health improvements, and new products. The latter is harder to assess and often has a political dimension. These metrics aim to judge the reputation or value of a scholar in a manner that is neither only price or reputation nor based on any specific fields. Ideally, the metrics would present themselves in a transparent, simple, and meaningful way. The metrics would present themselves in a transparent, simple, and meaningful way. The metrics would present themselves in a transparent, simple, and meaningful way.



based approaches that aim to identify a scholar's position in global scholarly networks through international collaboration and diffusion of works, and measures of the scholar's involvement in positive or negative feedback cycles.

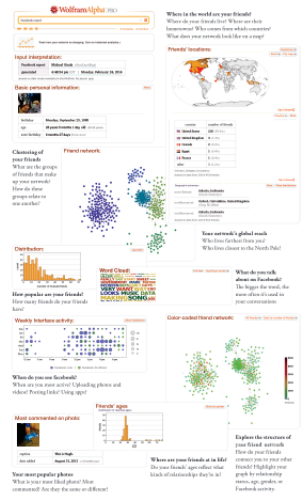
Concentrating Success

Success of any kind depends on one's physical and social environment. For example, rock climbing tends to be done where there is a rock, and where there is an urban environment—where dense populations and high cost of living are counterbalanced by the presence of many jobs, higher salaries, and enough work demands—in a rural area, where a low-density population and low cost of living are usually counterbalanced by fewer jobs, lower salaries, and generally weaker property. Facebook graphs (see page 70) clearly show the relationship between U.S. population density and per capita income. The top graph reveals that there are two distinct people are willing to pay for low-density education from both a local perspective of about 1,000 people per square mile and high-density urban living with more than 100,000 people per square mile, on a horizontal axis, to show that the year are concentrated in the densest areas in cities, whereas in the suburbs middle-class incomes are far more dispersed in terms of living density.

The bottom Facebook graph shows different age groups or percentages of population by density, normalized to national averages. An expected, young adults are more likely to live in dense urban areas, whereas older adults are more likely to move to be located in low-density suburbs. This is an important inflection point at about 3,000 people per square mile that may help define where the "inflection" point and "when" factors.

Academic Products Analytics

Academic products comprise such items as scholarly papers, books, data sets, software articles, and teaching materials. These products are evaluated and analyzed in terms of citations and usage to help measure faculty productivity, impact, and reach. The alternative between the metrics with widely used on the impact and cost of the whole paper—the price scholarly products. These factors may help to produce higher citation counts for a scholar's paper. Among them are author's prestige, based on the existing number of citations of the author's previous work, the prestige of an author's institution, the degree of internationalization of the author's team (especially only those which produce high-quality research will have the maximum to engage in international



collaboration) a journal's prestige, measured mainly by the journal's impact factor; the language of the author and the journal (especially, research from non-English speaking countries tends to be less frequently cited); the number of existing references; an article which has a particularly high number of references and one known to attract higher citation counts; the history of references, as measured by the PageRank; and the topics of the referenced articles, such as the measured by the χ^2 index. Finally, Facebook S&CT research has been done to be more frequently cited than other fields.

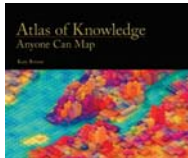
The cost of producing a scholarly paper can range from \$100 to \$1,000. Costs depend primarily on the amount of equipment and material needed to conduct the research. Although some scholarly efforts require specially prepared costs for individual scholars, other research demands expensive instrumentation and big science teamwork.

Commercial Product Analytics

Provides an intended or understood low-costs measure with their products and services. They are data first purchasing, download, service requests, product returns, warranties, customer feedback, and embedded sensors to measure customer reach, engagement, retention, conversion, and future patterns in product usage or capacity, while being all of these factors in transition.

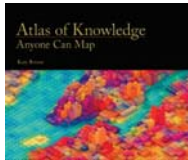
Research

When it comes to analyzing or assessing individual scholars, consumers, or products, high data accuracy and data volume are important. Ideally, we will use data that has been verified. Metrics and expert knowledge are then combined, and multiple indicators are calculated. One needs to clearly distinguish between the descriptive and prescriptive use of metrics. When making a choice, one needs to take into account the phase of that scholar's career and the maintained and research interest of their work. It is important to avoid the use of phrases, such as "journal impact factors," unless one should use micro-level data such as paper citations. The results of a data analysis should be published as a "research presentation." Using the price and cost of a PhD degree analysis together with a life list that details the data sets used, the algorithms and parameters applied, and the decisions made along the way.



See pages 6-7

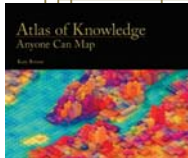
Insight Need Types page 26	Data Scale Types page 28	Visualization Types page 30	Graphic Symbol Types page 32	Graphic Variable Types page 34	Interaction Types page 26
<ul style="list-style-type: none"> • categorize/cluster • order/rank/sort • distributions (also outliers, gaps) • comparisons • trends (process and time) • geospatial • compositions (also of text) • correlations/relationships 	<ul style="list-style-type: none"> • nominal • ordinal • interval • ratio 	<ul style="list-style-type: none"> • table • chart • graph • map • network layout 	<ul style="list-style-type: none"> • geometric symbols <ul style="list-style-type: none"> • point • line • area • surface • volume • linguistic symbols <ul style="list-style-type: none"> • text • numerals • punctuation marks • pictorial symbols <ul style="list-style-type: none"> • images • icons • statistical glyphs 	<ul style="list-style-type: none"> • spatial <ul style="list-style-type: none"> • position • retinal <ul style="list-style-type: none"> • form • color • optics • motion 	<ul style="list-style-type: none"> • overview • zoom • search and locate • filter • details-on-demand • history • extract • link and brush • projection • distortion



See page 24

Graphic Variable Types Versus Graphic Symbol Types

		Point	Line	Geometric Symbols	Surface	Volume	Linguistic Symbols Text, Numbers, Punctuation Marks	Pictorial Symbols Images, Icons, Statistical Graphs
Type	x							
	y							
	z							
Color	hue							
	saturation							
	value							
	opacity							
	blend							
	stroke							
	fill							
Shape	point							
	line							
	area							
	volume							
	text							
	image							
	statistical							
Texture	point							
	line							
	area							
	volume							
	text							
	image							
	statistical							
Motion	point							
	line							
	area							
	volume							
	text							
	image							
	statistical							



See pages 36-39

Information Visualization MOOC 2015



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.



[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 12, 2016.

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/

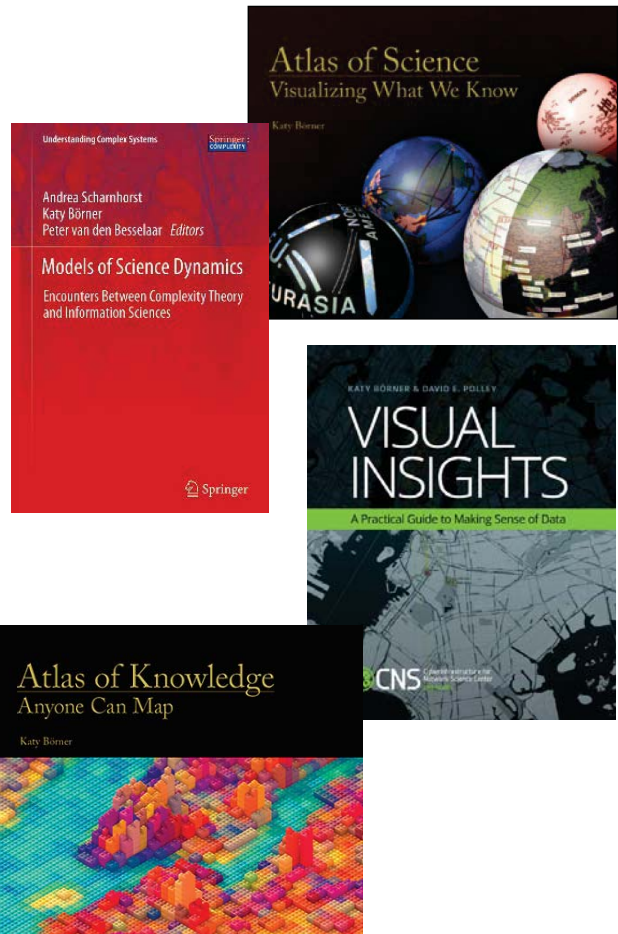
Börner, Katy (2010) **Atlas of Science: Visualizing What We Know**. The MIT Press. <http://scimaps.org/atlas>

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>



41

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>

42