Modelling Science, Technology, and Innovation

Katy Börner

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

ECON 2888R: Economics of Science and Engineering Workshop
The National Bureau of Economic Research (NBER), Cambridge, MA

October 6, 2016

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

Computed Using Data from Elsevier's Scopus

Models of Science, Technology, and Innovation

STI models use qualitative and quantitative data about scholars, papers, patents, grants, jobs, news, etc. to describe and predict the probable structure and/or dynamics of STI itself.

They are developed in economics, science policy, social science, scientometrics and bibliometrics, information science, physics, and other domains.

Computed Using Data from Elsevier's Scopus

Modelling Approaches

- Qualitative and quantitative models
- Deductive, abductive, and inductive models
- Analytic and predictive models
- Universal and domain specific models
- Multi-level and multi-perspective models

Computed Using Data from Elsevier's Scopus

Model Types

- Deterministic models
- Stochastic models
- Epidemic models
- Game-theoretic models
- Network models
- Agent-based models

Computed Using Data from Elsevier's Scopus

References

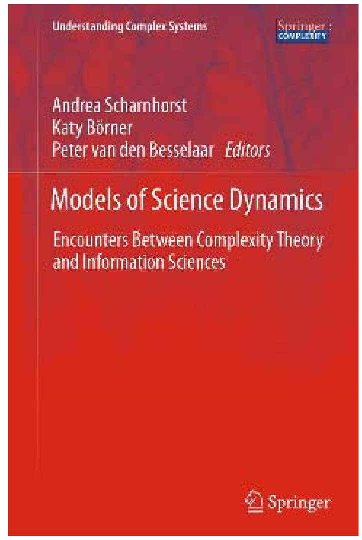
Shiffrin, Richard M. and Börner, Katy (Eds.) (2004). **Mapping Knowledge Domains**. *Proceedings of the National Academy of Sciences of the United States of America*, 101(Suppl_1). http://www.pnas.org/content/vol101/suppl_1/

Börner, Katy, and Andrea Scharnhorst, eds. 2009. **Science of Science: Conceptualizations and Models of Science**. *Journal of Informetrics* 3 (3).

Börner, Katy, Wolfgang Glänzel, Andrea Scharnhorst, and Peter van den Besselaar, eds. 2011. **Modeling science: Studying the structure and dynamics of science**. *Scientometrics* 89 (1): 346-463.

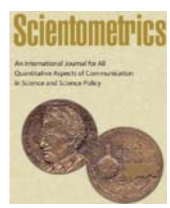
Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2012) **Models of Science Dynamics**. Springer Verlag.

Börner, Katy, Bruce Edmonds, Stasa Milojevic, and Andrea Scharnhorst, eds. 2016. **Simulating the Processes of Science, Technology, and Innovation**. *Scientometrics*.



5

Special Issue of *Scientometrics*:
Simulating the Processes of Science, Technology, and Innovation
Bruce Edmonds, Andrea Scharnhorst, Katy Börner & Staša Milojević (Editors)



- ***Rogier De Langhe:*** Towards the Discovery of Scientific Revolutions in Scientometric Data
- ***Sabine Brunswicker, Sorin Matei, Michael Zentner, Lynn Zentner and Gerhard Klimeck:*** Creating Impact in the Digital Space: Digital Practice Dependency in Scientific Developer Communities
- ***Johan Bollen et al.:*** An Efficient System to Fund Science: From Proposal Review to Peer-to-Peer Distributions
- ***Petra Ahrweiler:*** Agent-based Simulation for Science, Technology and Innovation Policy
- ***David Chavalarias:*** What's Wrong With Science? Modeling Collective Discovery Processes With the Nobel Game
- ***Jeff Alstott, Giorgio Triulzi, Bowen Yan and Jianxi Luo:*** Mapping Technology Space by Normalizing Patent Technology Networks

Government, academic, and industry leaders discussed challenges and opportunities associated with using big data, visual analytics, and computational models in STI decision-making.

Conference slides, recordings, and report are available via http://modsti.cns.iu.edu/report

Modelling Advantage

Models are widely used in the construction of scientific theories as they help

- Make assumptions explicit
- Describe the structure and dynamics of systems
- Communicate and explain systems
- Suggest possible interventions
- Identify new questions

Computed Using Data from Elsevier's Scopus

Modelling Challenges

Comprise among others:

- Model utility and usability
- Model credibility and validation
- Model extendibility and reproducibility
- Model sharing and retrieval

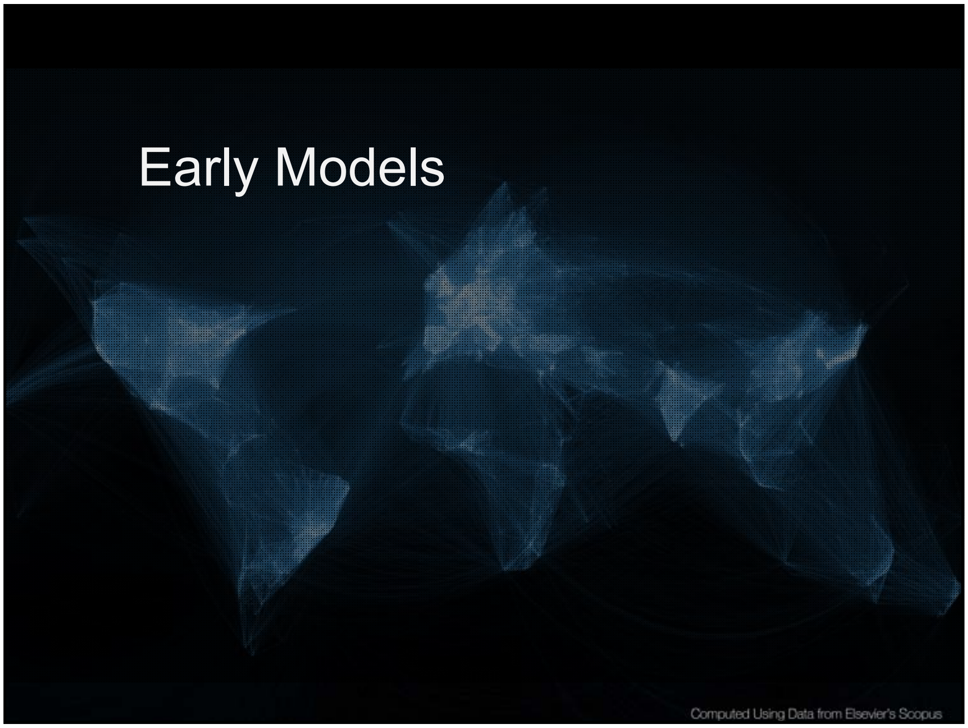
Computed Using Data from Elsevier's Scopus

Modelling Opportunities: Data-Driven Decision Making

Now available:

- high-quality, high coverage, interlinked data
- cost-effective storage and computation
- validated, scalable algorithms
- visualization and animations capabilities

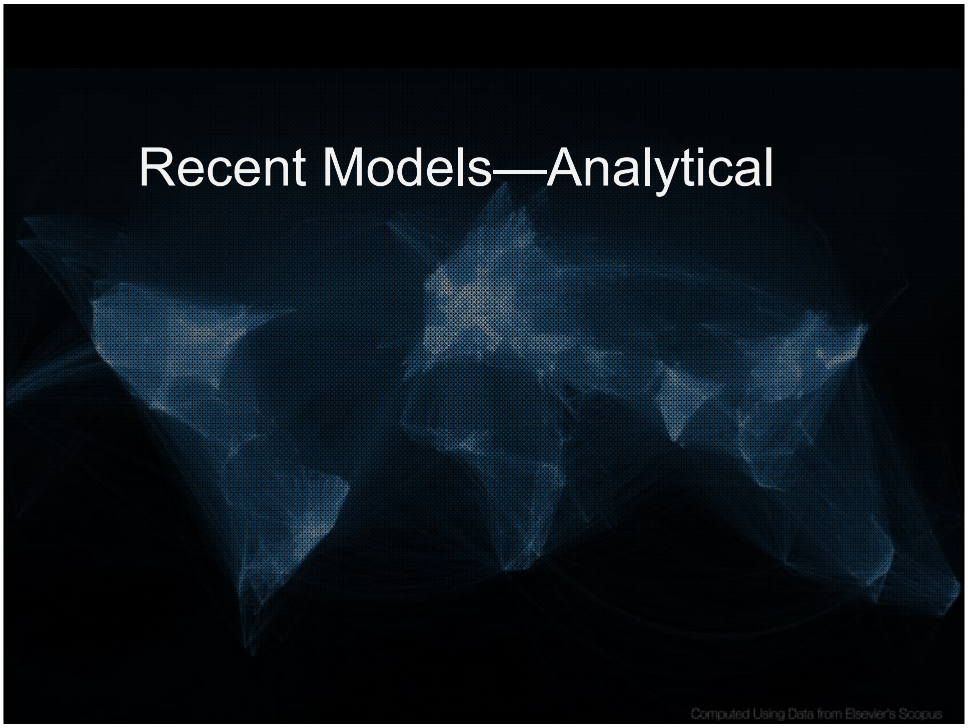
Computed Using Data from Elsevier's Scopus

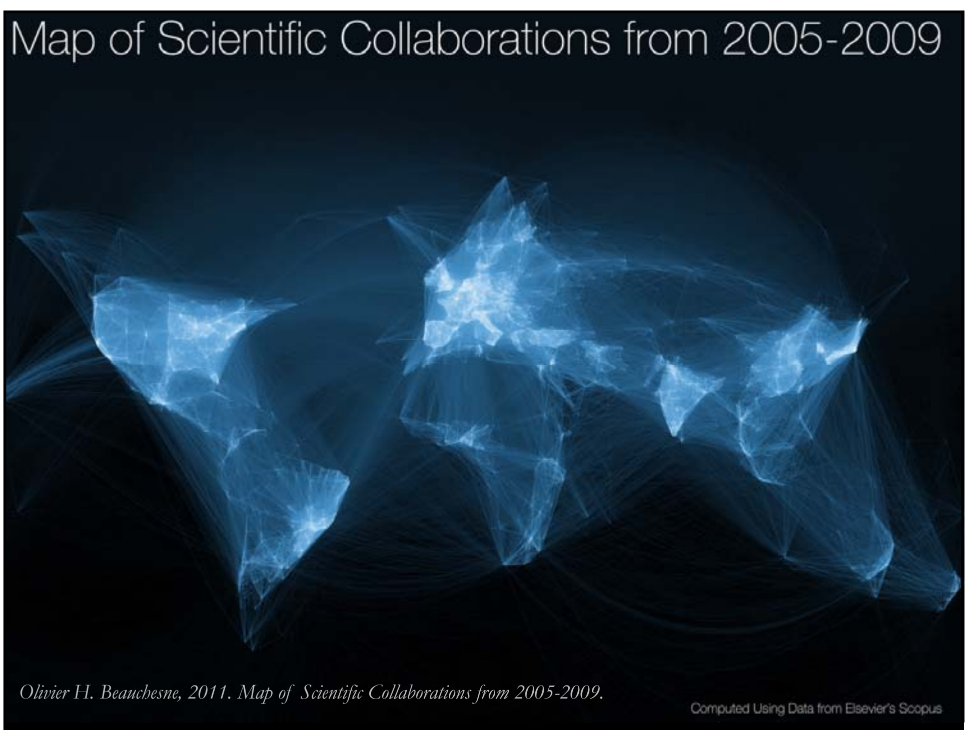


Monetary National Income Analogue Computer (MONIAC) developed by New Zealand economist and inventor Bill Phillips and displayed at the London School of Economics in **1949**. Demonstrates Keynesian and classical economic principles.

"Separate water tanks represent households, business, government, and the exporting and importing sectors of the economy. Colored water pumped around the system measures income, spending and GDP. The system is programmable and capable of solving nine simultaneous equations in response to any change of the parameters."

https://en.wikipedia.org/wiki/MONIAC_Computer


Recent Models—Analytical
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

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

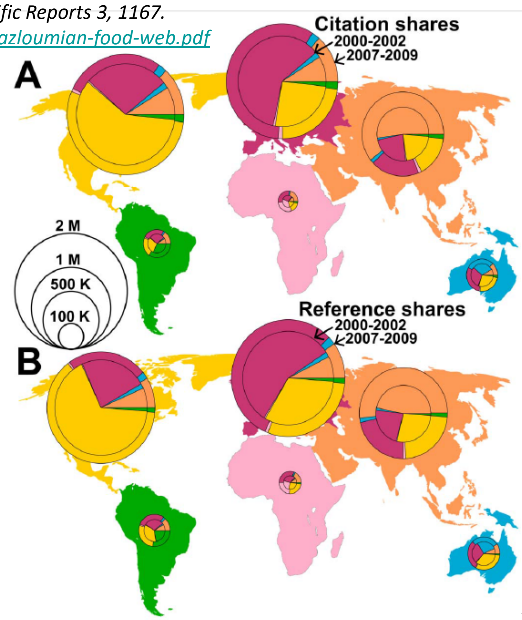
1

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.



15

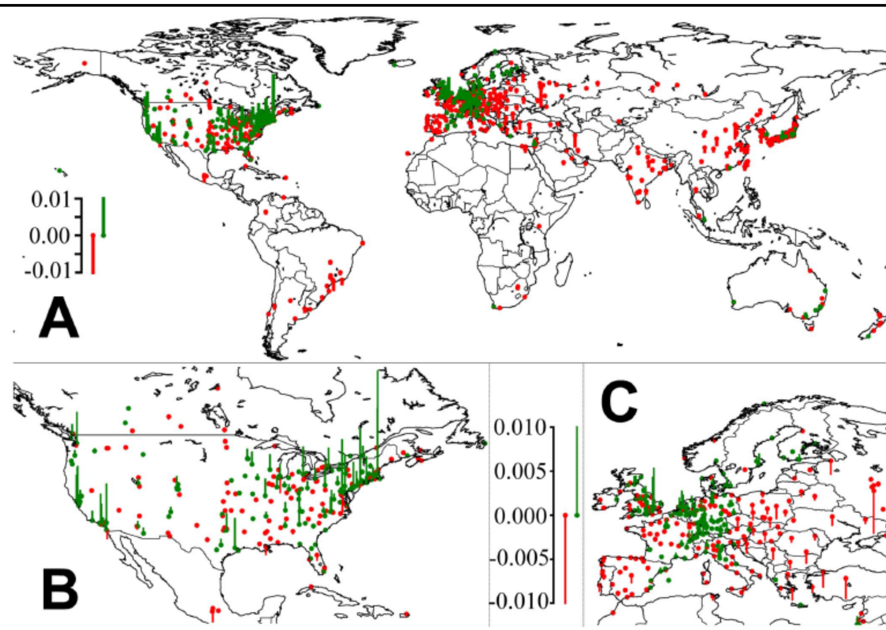


Figure 2 | World map of the greatest knowledge sources and sinks, based on our scientific fitness index. Green bars indicate that the number of citations received is over-proportional, red that the number of citations received is lower than expected (according to a homogeneous distribution of citations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

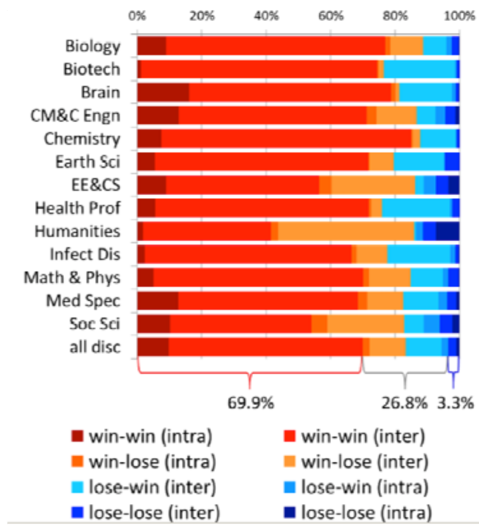
16

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.

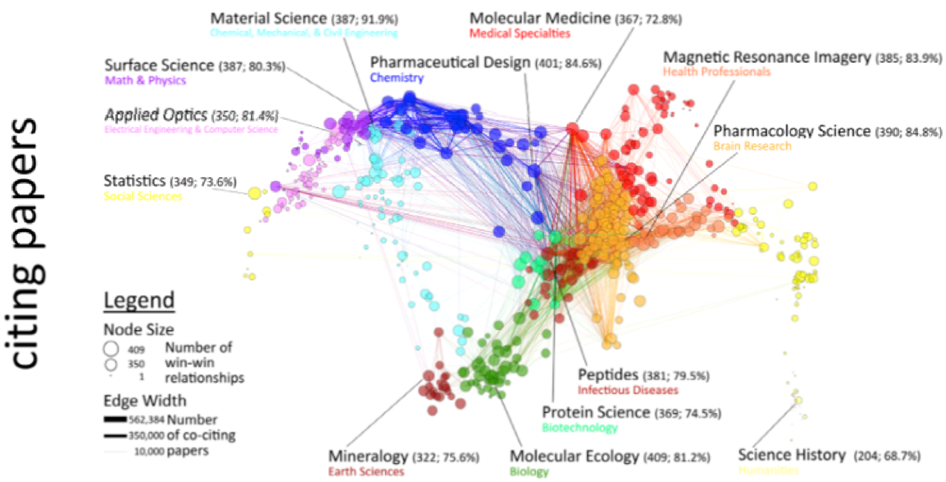
Results: majority (69.9%) of co-cited interdisciplinary pairs are "win-win" relationships, i.e., papers that cite them have higher citation impact and there are as few as 3.3% "lose-lose" relationships. UCSD map of science is used to compute "distance."



2

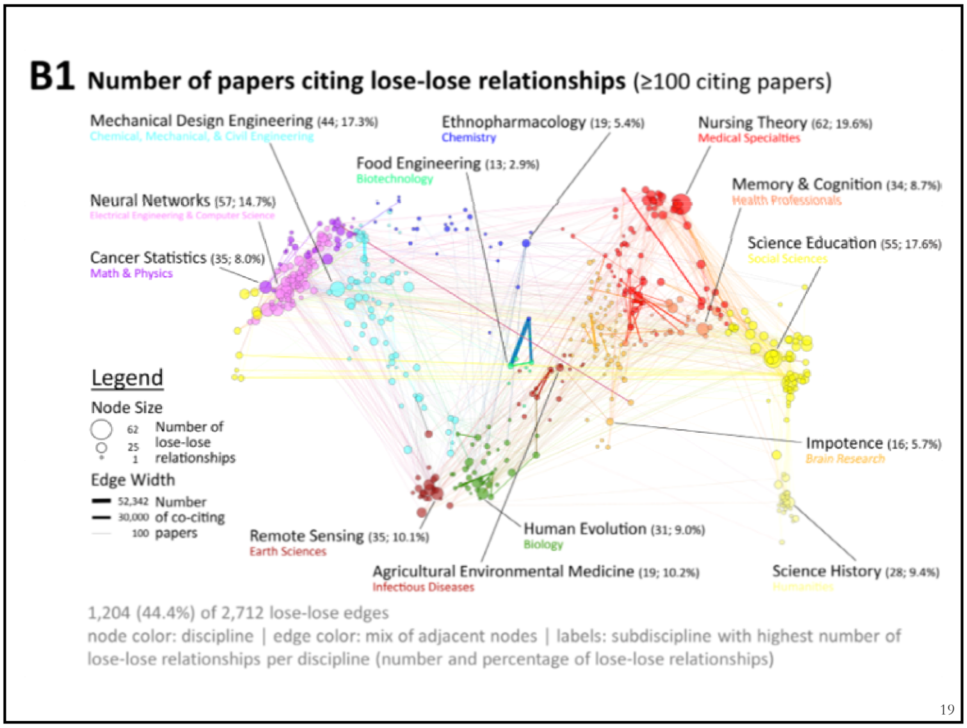
17

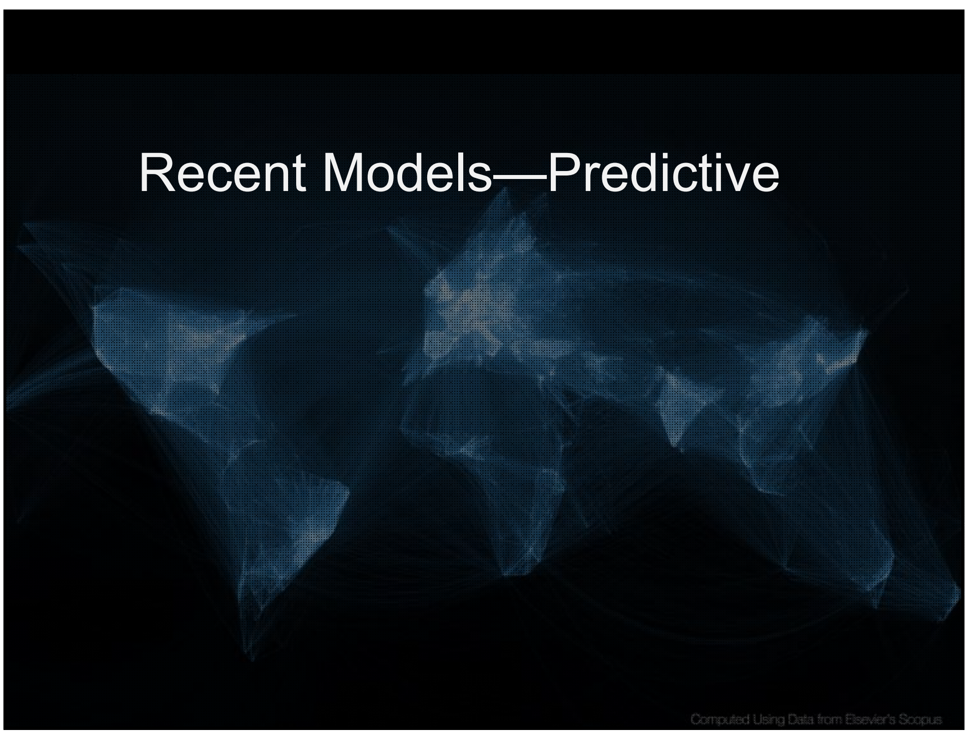
A1 Number of papers citing win-win relationships (≥10,000 citing papers)



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)

18

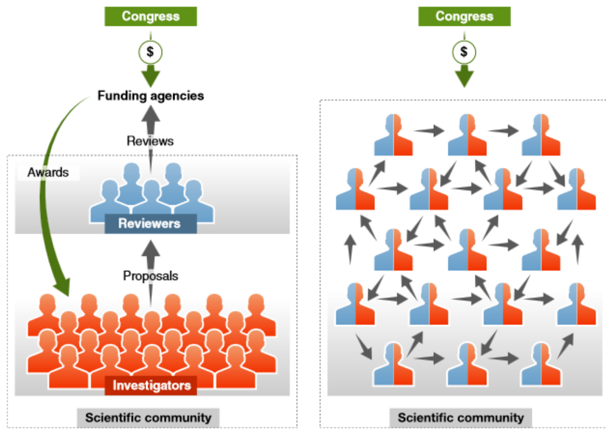

B1 Number of papers citing lose-lose relationships (≥100 citing papers)
Mechanical Design Engineering (44; 17.3%)
Chemical, Mechanical, & Civil Engineering
Ethnopharmacology (19; 5.4%)
Chemistry
Nursing Theory (62; 19.6%)
Medical Specialties
Food Engineering (13; 2.9%)
Biotechnology
Memory & Cognition (34; 8.7%)
Health Professionals
Neural Networks (57; 14.7%)
Electrical Engineering & Computer Science
Science Education (55; 17.6%)
Social Sciences
Cancer Statistics (35; 8.0%)
Math & Physics
Legend
Node Size
62 25 1 Number of lose-lose relationships
Edge Width
52,342 30,000 100 Number of co-citing papers
Impotence (16; 5.7%)
Brain Research
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)
19


Recent Models—Predictive
Computed Using Data from Elsevier's Scopus

3

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.

21

Assume

Total funding budget in year y is t_y

Number of qualified scientists is n

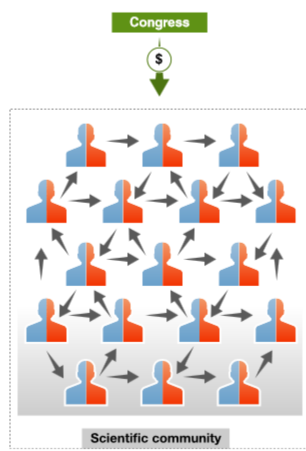
Each year,

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.



22

Example:

Total funding budget in year is 2012 NSF budget

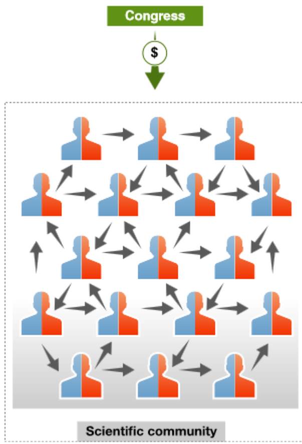
Given the number of NSF funded scientists, each receives a $100,000 basic grant.

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.



23

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.

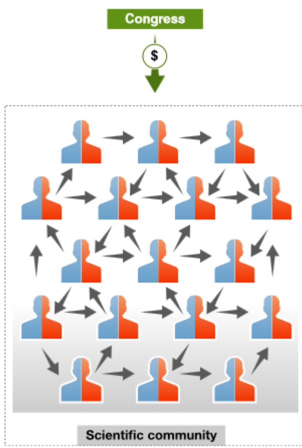
Using 37M articles from TR 1992 to 2010 Web of Science (WoS) database, we extracted **770M citations**. From the same WoS data, we also determined 4,195,734 unique author names and we took the **867,872 names** who had authored at least one paper per year in any five years of the period 2000–2010.

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.



24

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 $35k. With success rates in CS around 20%, about five submission-review cycles might be needed resulting in a total expected labor cost of **$175k**.

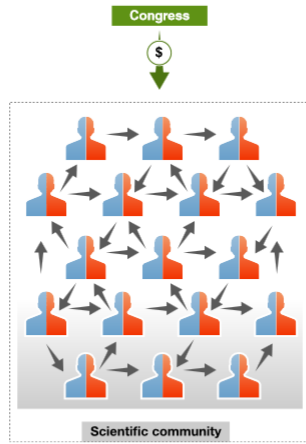
The average NSF grant is **$165k** per year.

U.S. universities charge about 50% overhead (ca. $55k), leaving about **$110k**.

In other words, average success results in a net loss for faculty in terms of paid research time.

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

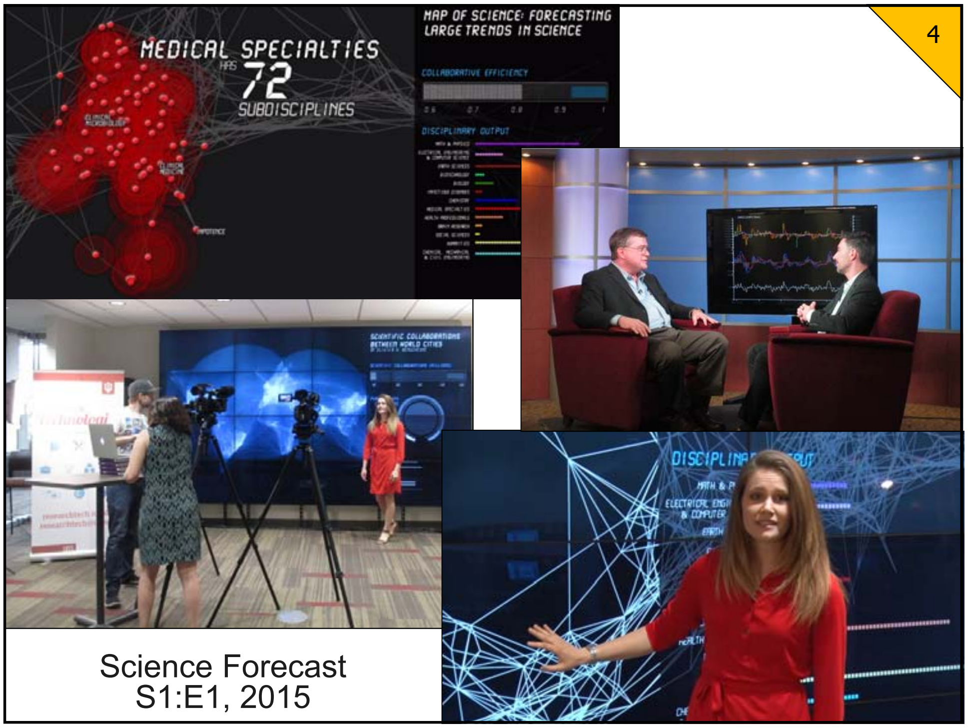
To add: Time spent by researchers to review proposals. In 2015 alone, NSF commissioned more than 231,000 reviews to evaluate 49,600 proposals.



25

Communication of Model Results

4



Science Forecast
S1:E1, 2015

5

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 (2017) http://scimaps.org/call

Background and Goals

The *Places & Spaces: Mapping Science* exhibit was created to in
communicate human activity and scientific progress on a glol
that enable the close inspection of large-scale maps in public
conferences; (2) novel, interactive macroscope tools that let t

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

28

http://scimaps.org/iteration/11

Atlas Trilogy

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

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

Börner, Katy (2018) **Atlas of Forecasts: Predicting and Broadcasting Science, Technology, and Innovation**. The MIT Press.

References/pointers to models that my mom and other key stakeholders should understand and to models that made a true difference are welcome.

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