RESEARCH HORIZONS 2016

Katy Börner

Modeling Science, Technology, and Innovation

COLLABORATORS

Staša Milojević, Johan Bollen, David Crandall, Damion Junk, Ying Ding (SOIC, IU), James Evans (U Chicago), Susan Fitzpatrick (President, James S. McDonnell Foundation), Richard B. Freeman (Harvard U), Jerome Glenn (The Millennium Project), Caroline Wagner (OSU), Bruce Edmonds (UK), and Andrea Scharnhorst (KNAW, NL). Professional support by Daniel O'Donnell.

INDIANA UNIVERSITY
SCHOOL OF INFORMATICS AND COMPUTING

Models of Science, Technology, and Innovation (STI)

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 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

Model Types

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

Computed Using Data from Elsevier's Scopus

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.
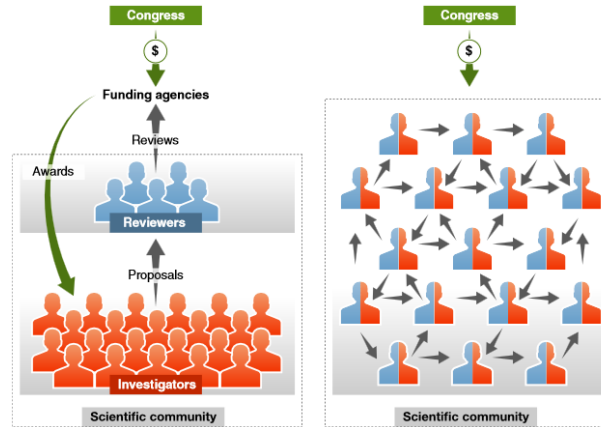
Existing (left) and proposed (right) funding systems. Reviewers in blue; investigators in red.

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.

5

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.
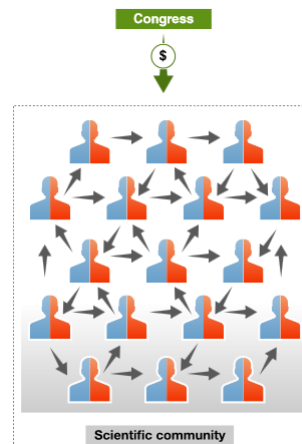

6

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.
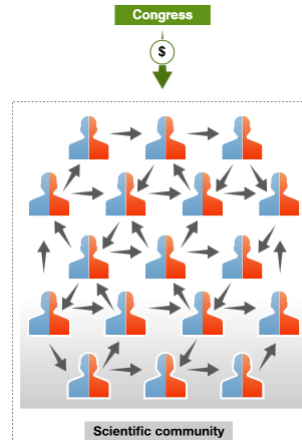
7

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.
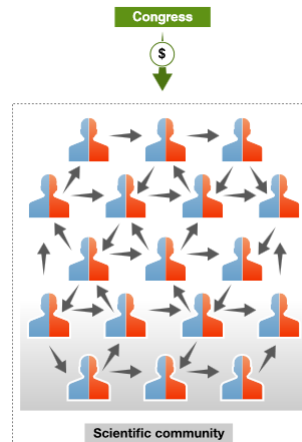
8

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.

In other words, under some conditions, the total cost of application and administration might significantly reduce the monetary value of a grant.

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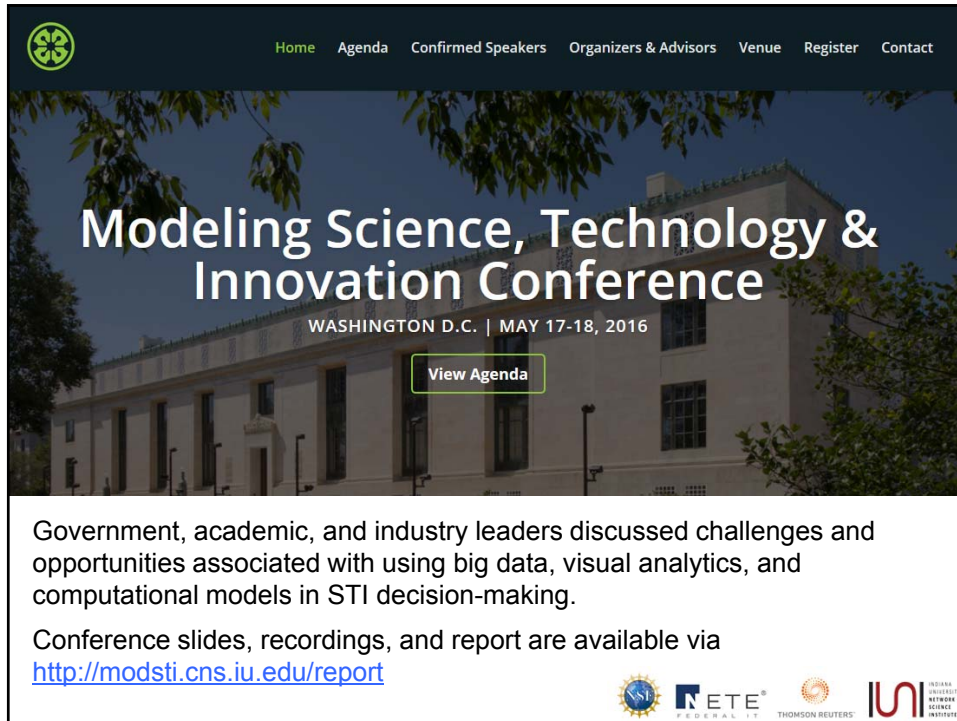

9
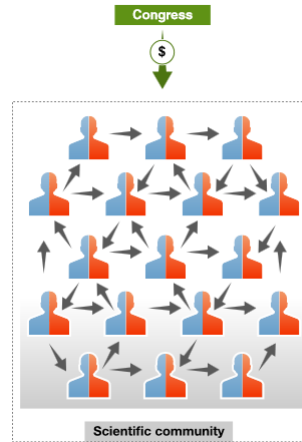
Modelling Challenges

Need to bridge the gap between model development and usage.
(See conference report for several other challenges)

Modeling Science, Technology & Innovation Conference

WASHINGTON D.C. | MAY 17-18, 2016

View Agenda

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 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

Special Issue of *Scientometrics*:
Simulating the Processes of Science, Technology, and Innovation
Bruce Edmonds, Andrea Scharnhorst, Katy Börner & Staša Milojević (Editors)
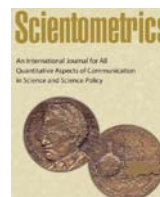
Rogier De Langhe	Towards the discovery of scientific revolutions in scientometric data
Sabine Brunswicker, Sorin Matei, Michael Zentner, Lynn Zentner and Gerhard Klimeck	Creating Impact in the Digital Space: Digital Practice Dependency in Scientific Developer Communities
Johan Bollen et al.	An efficient 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 of collective discovery processes with the Nobel Game
Jeff Alstott, Giorgio Triulzi, Bowen Yan and Jianxi Luo	Mapping Technology Space by Normalizing Patent Technology Networks

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."

Science Forecast
S1:E1, 2015