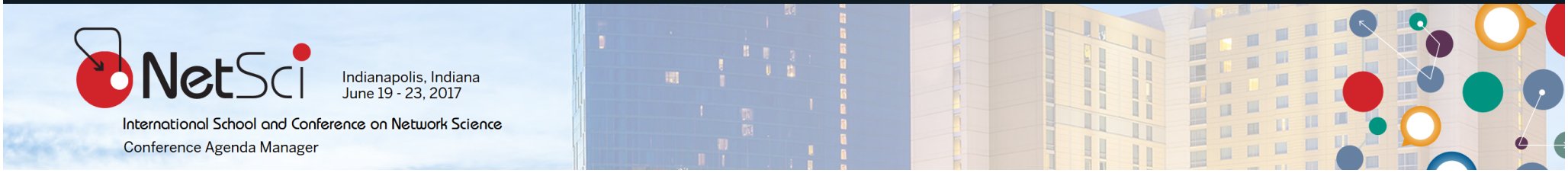
Visualizing Networks of Knowledge

Katy Börner @katycns

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

Knowledge Networks in Science and Technology
NetSci Conferences, Indianapolis, IN
June 20, 2017



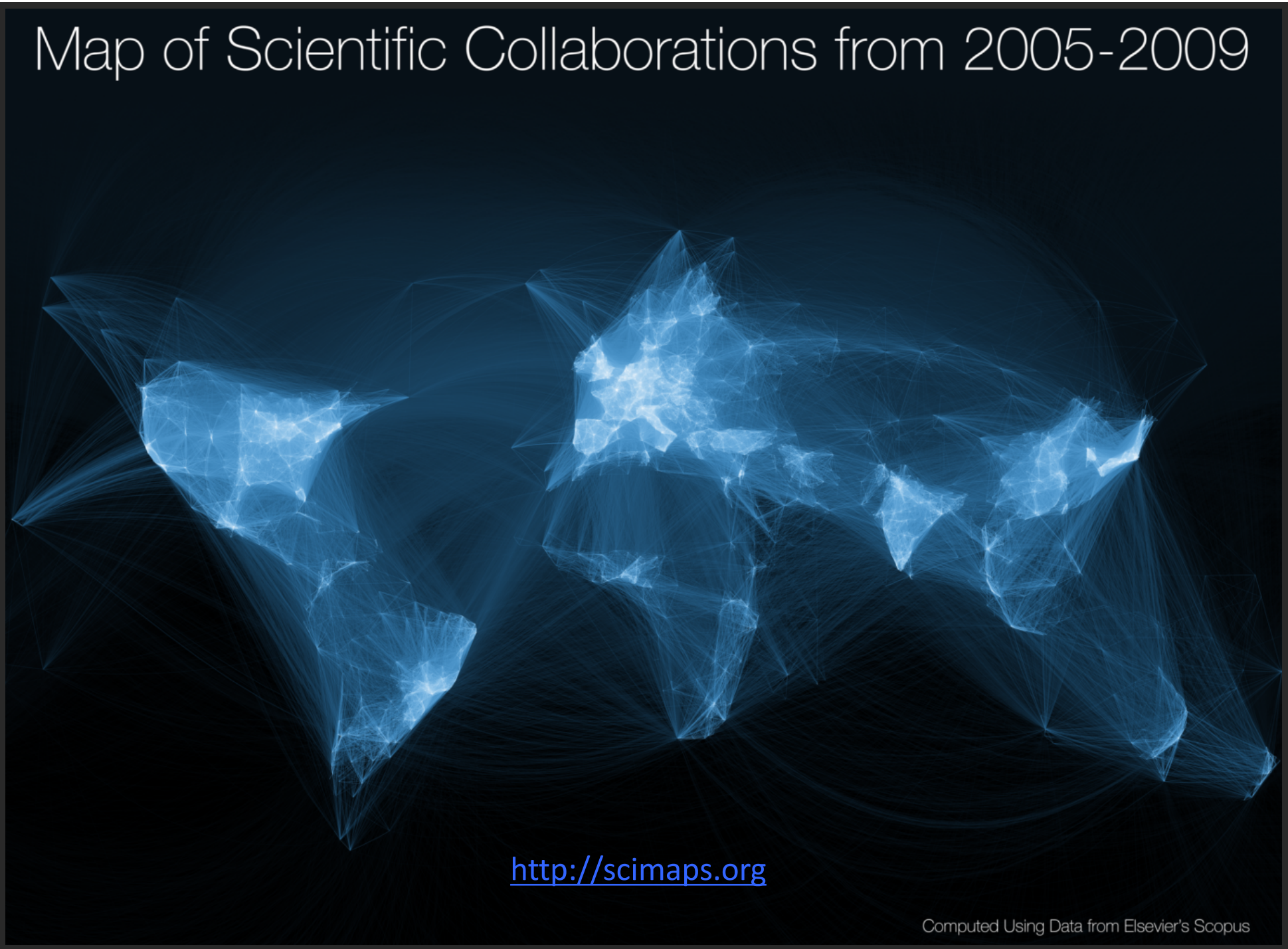
Examples

PLACES & SPACES

MAPPING SCIENCE

scimaps.org

NetSci
Indianapolis, Indiana
June 19 - 23, 2017
International School and Conference on Network Science
Conference Agenda Manager

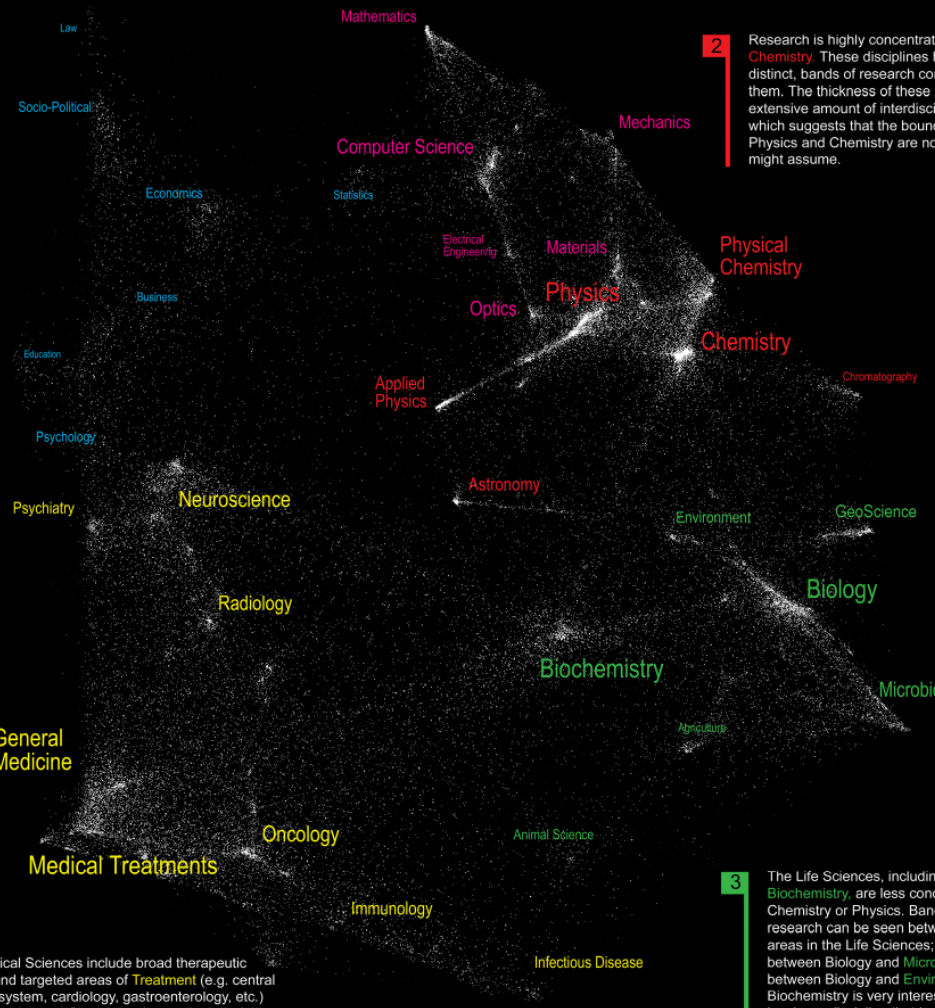


VII.6 *Map of Scientific Collaborations from 2005-2009* - Olivier H. Beauchesne - 2011

The Structure of Science

5 The Social Sciences are the smallest and most diffuse of all the sciences. Psychology serves as the link between Medical Sciences (Psychiatry) and the Social Sciences. Statistics serves as the link with Computer Science and Mathematics.

1 Mathematics is our starting point, the purest of all sciences. It lies at the outer edge of the map. Computer Science, Electrical Engineering, and Optics are applied sciences that draw upon knowledge in Mathematics and Physics. These three disciplines provide a good example of a linear progression from one pure science (Mathematics) to another (Physics) through multiple disciplines. Although applied, these disciplines are highly concentrated with distinct bands of research communities that link them. Bands indicate interdisciplinary research.



2 Research is highly concentrated in Physics and Chemistry. These disciplines have few, but very distinct, bands of research communities that link them. The thickness of these bands indicates an extensive amount of interdisciplinary research, which suggests that the boundaries between Physics and Chemistry are not as distinct as one might assume.

4 The Medical Sciences include broad therapeutic studies and targeted areas of Treatment (e.g. central nervous system, cardiology, gastroenterology, etc.) Unlike Physics and Chemistry, the medical disciplines are more spread out, suggesting a more multi-disciplinary approach to research. The transition into Life Sciences (via Animal Science and Biochemistry) is gradual.

3 The Life Sciences, including Biology and Biochemistry, are less concentrated than Chemistry or Physics. Bands of linking research can be seen between the larger areas in the Life Sciences; for instance between Biology and Microbiology, and between Biology and Environmental Science. Biochemistry is very interesting in that it is a large discipline that has visible links to disciplines in many areas of the map, including Biology, Chemistry, Neuroscience, and General Medicine. It is perhaps the most interdisciplinary of the sciences.

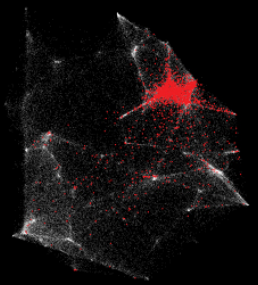
We are all familiar with traditional maps that show the relationships between countries, provinces, states, and cities. Similar relationships exist between the various disciplines and research topics in science. This allows us to map the structure of science.

One of the first maps of science was developed at the Institute for Scientific Information over 30 years ago. It identified 41 areas of science from the citation patterns in 17,000 scientific papers. That early map was intriguing, but it didn't cover enough of science to accurately define its structure.

Things are different today. We have enormous computing power and advanced visualization software that make mapping of the structure of science possible. This galaxy-like map of science (left) was generated at Sandia National Laboratories using an advanced graph layout routine (VxOrd) from the citation patterns in 800,000 scientific papers published in 2002. Each dot in the galaxy represents one of the 96,000 research communities active in science in 2002. A research community is a group of papers (9 on average) that are written on the same research topic in a given year. Over time, communities can be born, continue, split, merge, or die.

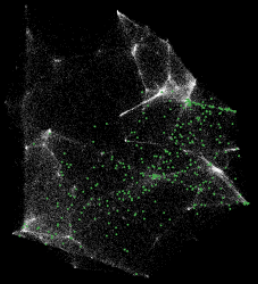
The map of science can be used as a tool for science strategy. This is the terrain in which organizations and institutions locate their scientific capabilities. Additional information about the scientific and economic impact of each research community allows policy makers to decide which areas to explore, exploit, abandon, or ignore.

We also envision the map as an educational tool. For children, the theoretical relationship between areas of science can be replaced with a concrete map showing how math, physics, chemistry, biology and social studies interact. For advanced students, areas of interest can be located and neighboring areas can be explored.



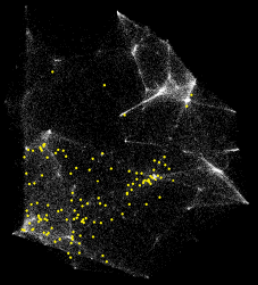
Nanotechnology

Most research communities in nanotechnology are concentrated in Physics, Chemistry, and Materials Science. However, many disciplines in the Life and Medical Sciences also have nanotechnology applications.



Proteomics

Research communities in proteomics are centered in Biochemistry. In addition, there is a heavy focus in the tools section of chemistry, such as Chromatography. The balance of the proteomics communities are widely dispersed among the Life and Medical Sciences.



Pharmacogenomics

Pharmacogenomics is a relatively new field with most of its activity in Medicine. It also has many communities in Biochemistry and two communities in the Social Sciences.

I.10 *The Structure of Science* - Kevin W. Boyack and Richard Klavans - 2005

The EMERGENCE of NANOTECHNOLOGY

MAPPING THE NANO REVOLUTION

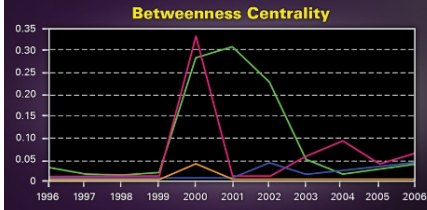
The emergence of nanotechnology has been one of the major scientific-technological revolutions in the last decade and it led to a structural reorganization of major fields of science. Price (1965) showed that fields of science and their development can be mapped using aggregated citations among the journals in the fields and their relevant environments.

The frames to the right show the evolving journal citation network for the years 1998-2003. Distances are proportional to cosine values between the citation patterns of the respective journals. Textual descriptions of key events during the development of *Nanotechnology* are given below each frame. Most notably, leading papers in *Science* and *Nature* catalyzed the breakthrough around 2000.

CHANGING ROLES OF DIFFERENT JOURNALS

The interdisciplinarity of a journal can be measured using betweenness centrality (BC)—journals that occur on many shortest paths between other journals in a network have higher BC value than those that do not. In the maps, sizes of nodes are proportional to the betweenness centrality of the respective journal in the citation network.

From being a specialist journal in applied physics, the journal *Nanotechnology* obtains a high BC value in the years of the transition, ca. 2001. This is preceded by the "intervention" of *Science*. After the transition, the new field of nanotechnology is established, new journals such as *Nano Letters* published by the influential American Chemical Society take the lead, and a new specialty structure with low BC value journals results.

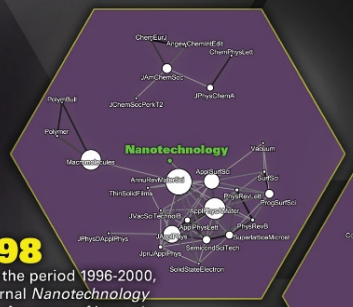


An animated sequence of this evolution is at: http://www.leydesdorff.net/journals/nanotech.

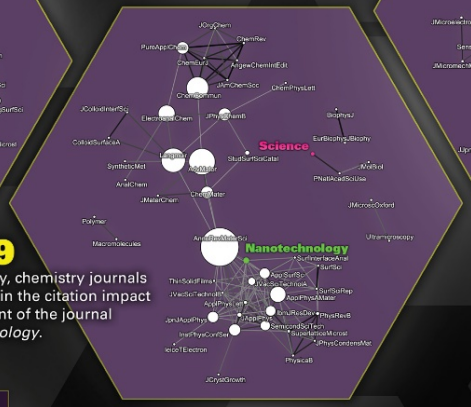
References

Leydesdorff, L. and T. Schank. 2008. Dynamic Animations of Journal Maps: Indicators of Structural Change and Interdisciplinary Developments. *Journal of the American Society for Information Science and Technology,* 59(11), 1810-1818.

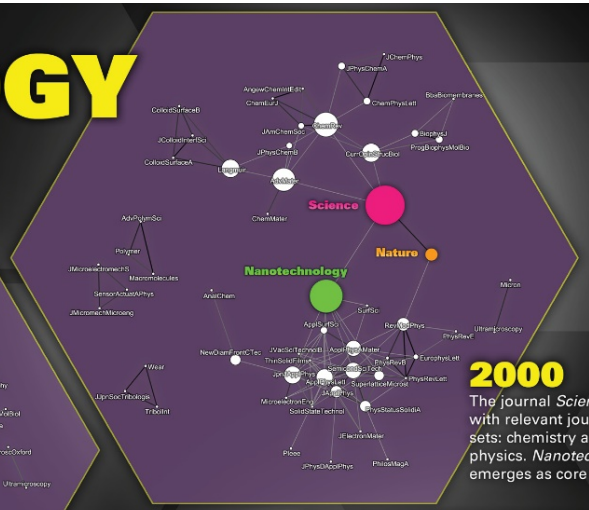
Price, Derek J. de Solla (1965). Networks of scientific papers. *Science*, 149, no. 3683, 510- 515.



1998
During the period 1996-2000, the journal *Nanotechnology* is part of a group of journals in applied physics.



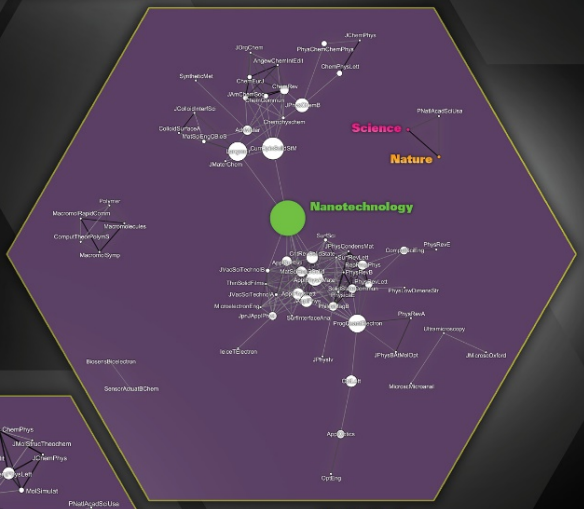
1999
Increasingly, chemistry journals play a role in the citation impact environment of the journal *Nanotechnology*.



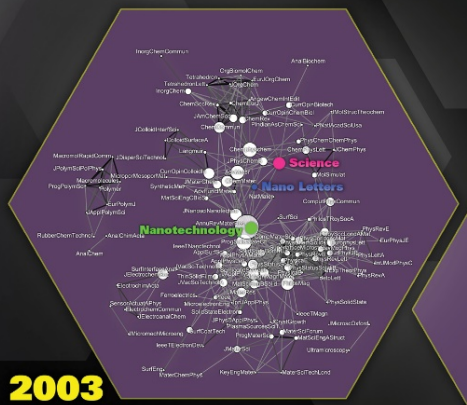
2000
The journal *Science* interfaces with relevant journals in both sets: chemistry and applied physics. *Nanotechnology* emerges as core journal.

LEGEND
- *Science*
- *Nature*
- *Nanotechnology*
- *Nano Letters*

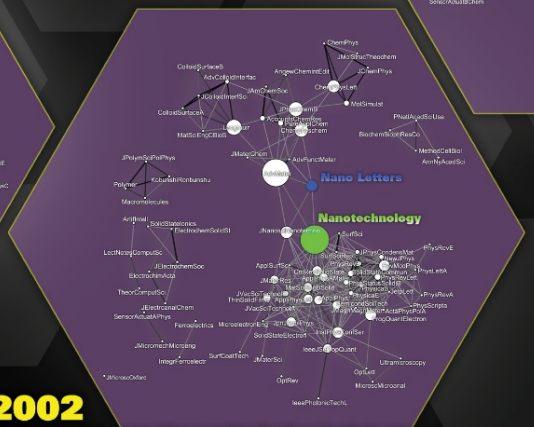
Values
0.8
0.22
0.33



2001
The journal *Nanotechnology* now provides the interface between chemistry and physics. The "intervention" by *Science* is no longer needed.



2003
The journal *Science* is relevant in the citation impact environment, but now functions as one of the specialist journals in nanotechnology. *Nanoscience* further develops as an increasingly integrated network of journals.



2002
Other journals in nanoscience and technology begin to emerge, and the bridging role of the journal *Nanotechnology* gradually subsides. *Nano Letters* and the *Journal of Nanoscience and Nanotechnology* join the new field of nanotechnology.

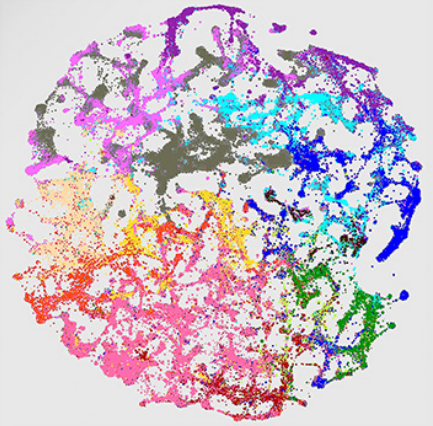
Design by Michael J. Stamper and Katy Börner
Cyberinfrastructure for Network Science Center | Indiana University
cns.iu.edu

IDENTIFYING EMERGING TOPICS IN SCIENCE AND TECHNOLOGY

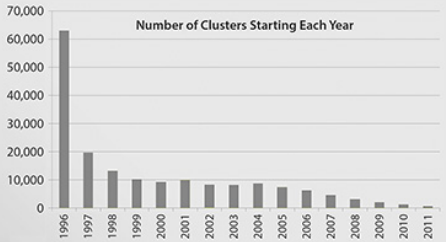
(finding the needles in the haystack)

A novel approach to identifying emerging topics in science and technology has been developed. Two models of science and technology have been created using 16 years (1996-2011) of Scopus (20 million articles) and USPTO (3 million patents) data. These two models—one based on direct citation, and one based on co-citation—are used together to nominate the most emergent topics in S&T at a particular point in time.

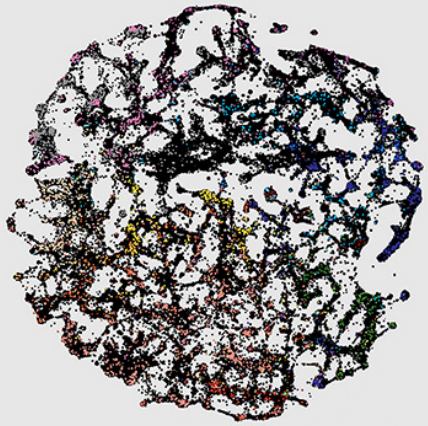
Step 1: Map All of Science and Technology



The map above was created using a combination of direct citation and text mining. First, 20 million articles from Scopus were clustered using direct citation links and the new modularity-based clustering methodology from Waltman & van Eck at Leiden University. Second, 3 million patents from the USPTO were clustered using the same direct citation method. Third, a BM25 text similarity was computed between all clusters, whether articles or patents. Using this text similarity, clusters were positioned relative to each other in the map above. Each of the 149,611 article clusters is colored based on the journals that comprise the cluster. The 27,114 patent clusters are colored gray.



Step 2: Burn Off the Straw

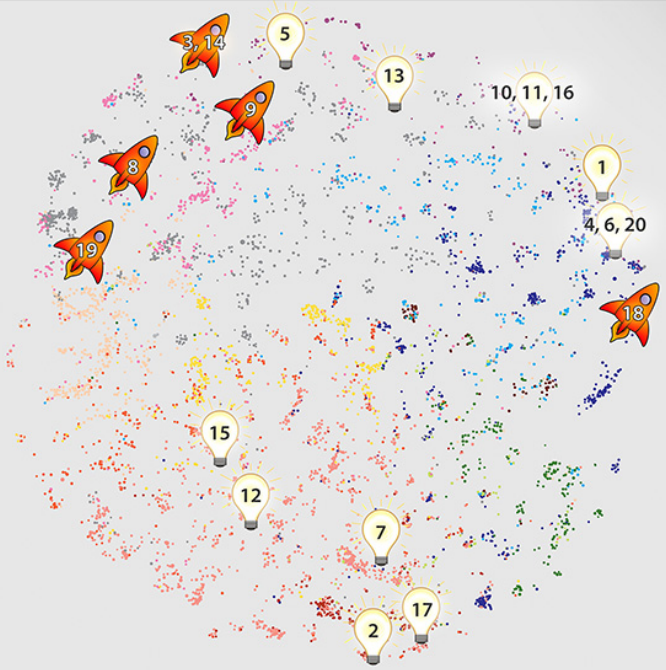


Most of the clusters in the direct citation map are very old (71%), having started a decade or more ago. Another 25% are in their prime, starting between 2002-2007. Emerging topics are, by definition, very new. In the map above, all clusters starting in 2007 or earlier have been blackened, or burned off. The 4% of clusters retaining their original color are those that began in 2008 or later and are still active.

To determine which of these clusters are most emergent, we further filter these clusters to those whose articles are also in new clusters in our co-citation model of science (not shown). Clusters are then ranked using a growth index. The locations of the top-20 emerging clusters for the year 2010 based on the growth index are shown in the map to the right. Emerging clusters can continue to be emergent in additional years if their growth characteristics are strong enough.

Interestingly, some emerging topics are based on scientific discoveries while others spring from technological innovation. Scientific discoveries are those where a new or unexpected finding is made or fundamental knowledge is gained. Technological innovations are those where existing science or technology is used to create new devices or capabilities that serve specific purposes. For the top-20 emerging areas in 2010, roughly one third of them (6) are based on technological innovations. The remainder are based on scientific discoveries.

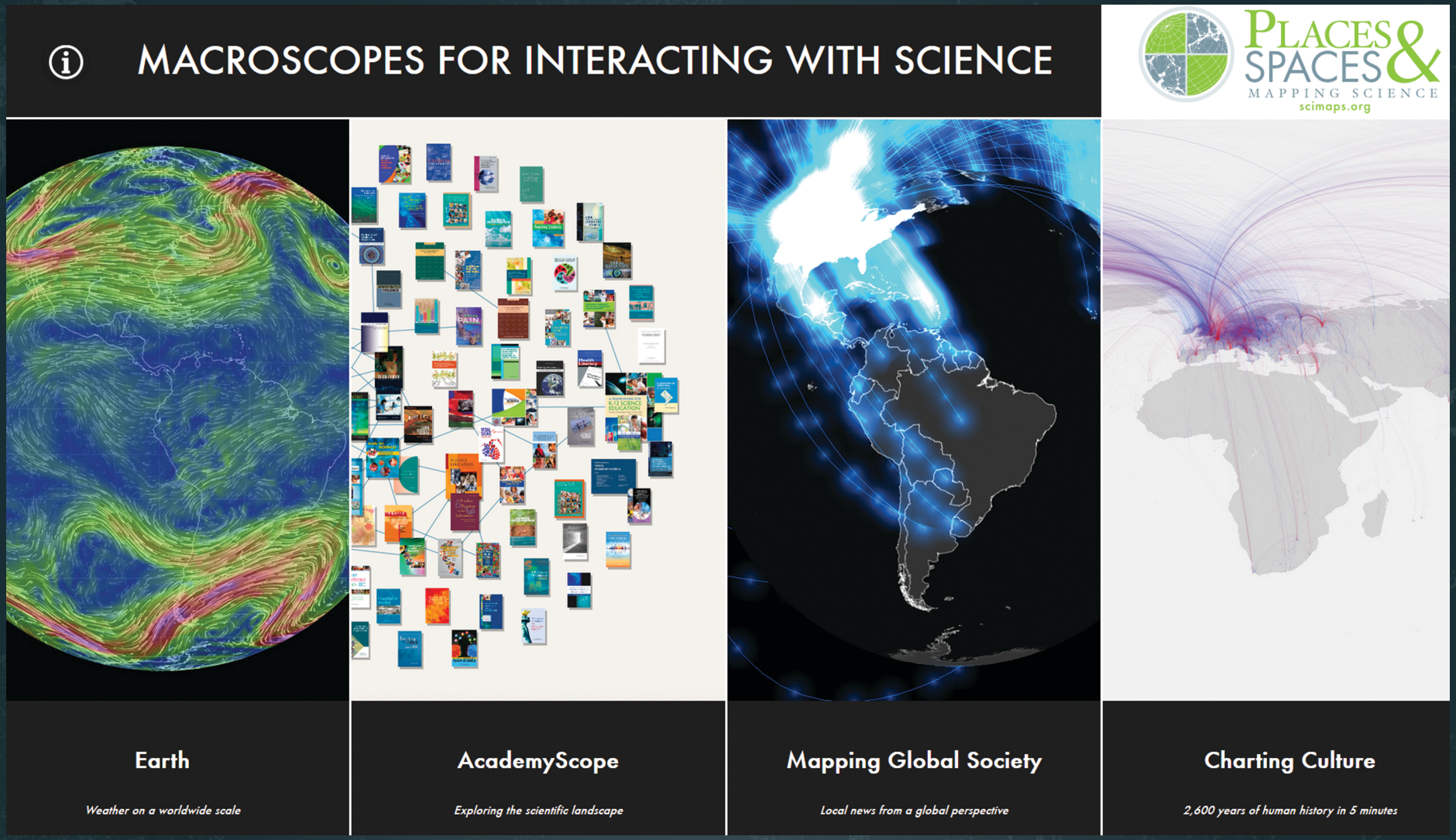
Step 3: Highlight the BRIGHTEST Needles



Top-20 Emerging Topics in 2010

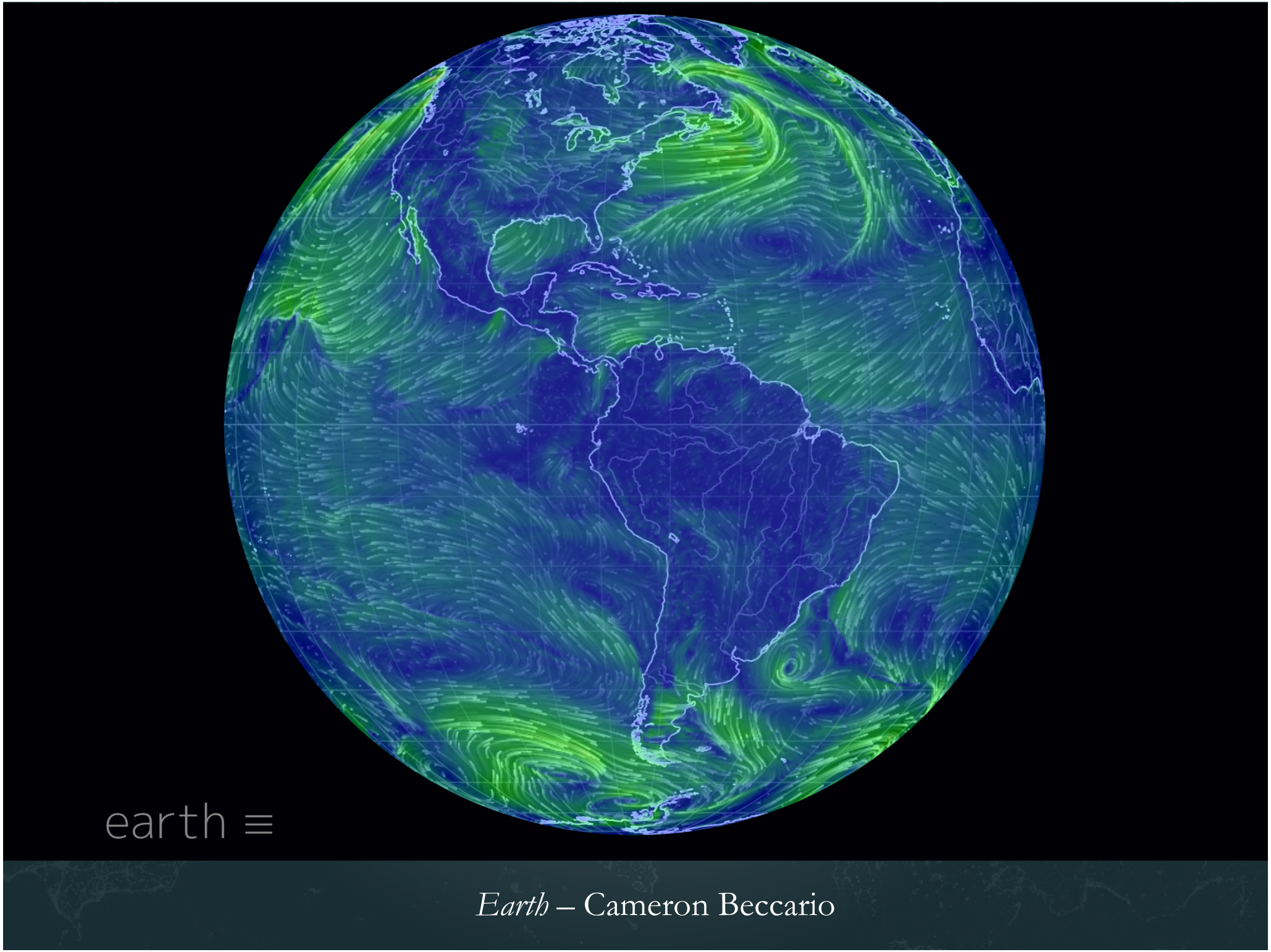
1 – iron-based superconductors
2 – swine flu (H1N1) pandemic
3 – spectrum sensing in cognitive radio
4 – graphene nanosheets and nanocomposites
5 – Horava-Lifshitz quantum gravity
6 – graphene oxide nanosheets
7 – induced pluripotent stem cells
8 – MapReduce framework
9 – signal recovery from compressed sensing
10 – graphene transistors and optical devices
11 – zigzag graphene nanoribbons
12 – cardiovascular events in type 2 diabetes
13 – transformative optics
14 – spectrum allocation in cognitive radio
15 – IDH1 and IDH2 mutations in cancer
16 – epitaxial graphene
17 – H1N1 pandemic and seasonal flu
18 – crystallographic validation
19 – social tagging
20 – mechanical properties of graphene

IX.9 *Identifying Emerging Topics in Science and Technology* - Kevin W. Boyack, Richard Klavans, and Henry G. Small - 2013

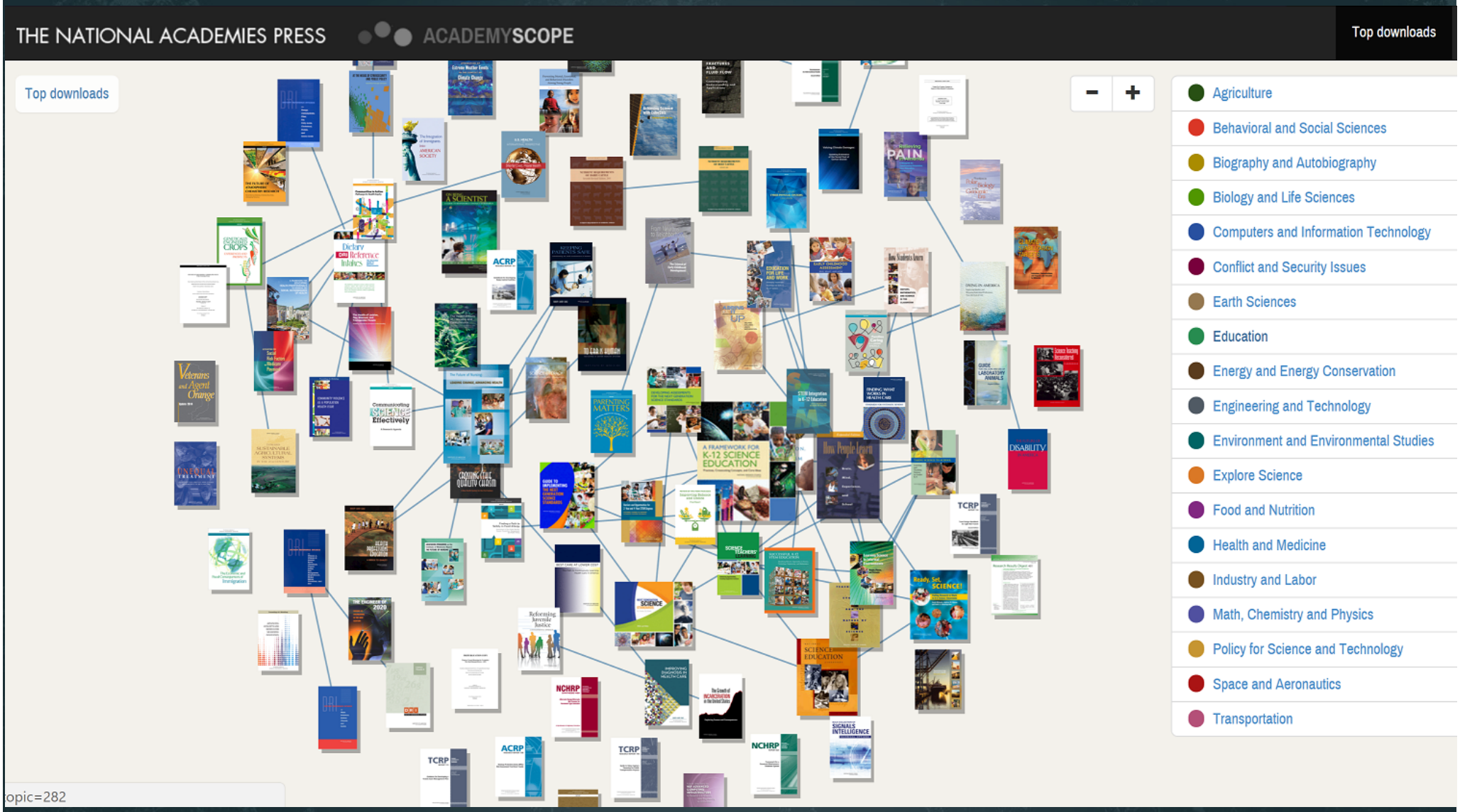


Iteration XI (2015): Macroscopes for Interacting with Science

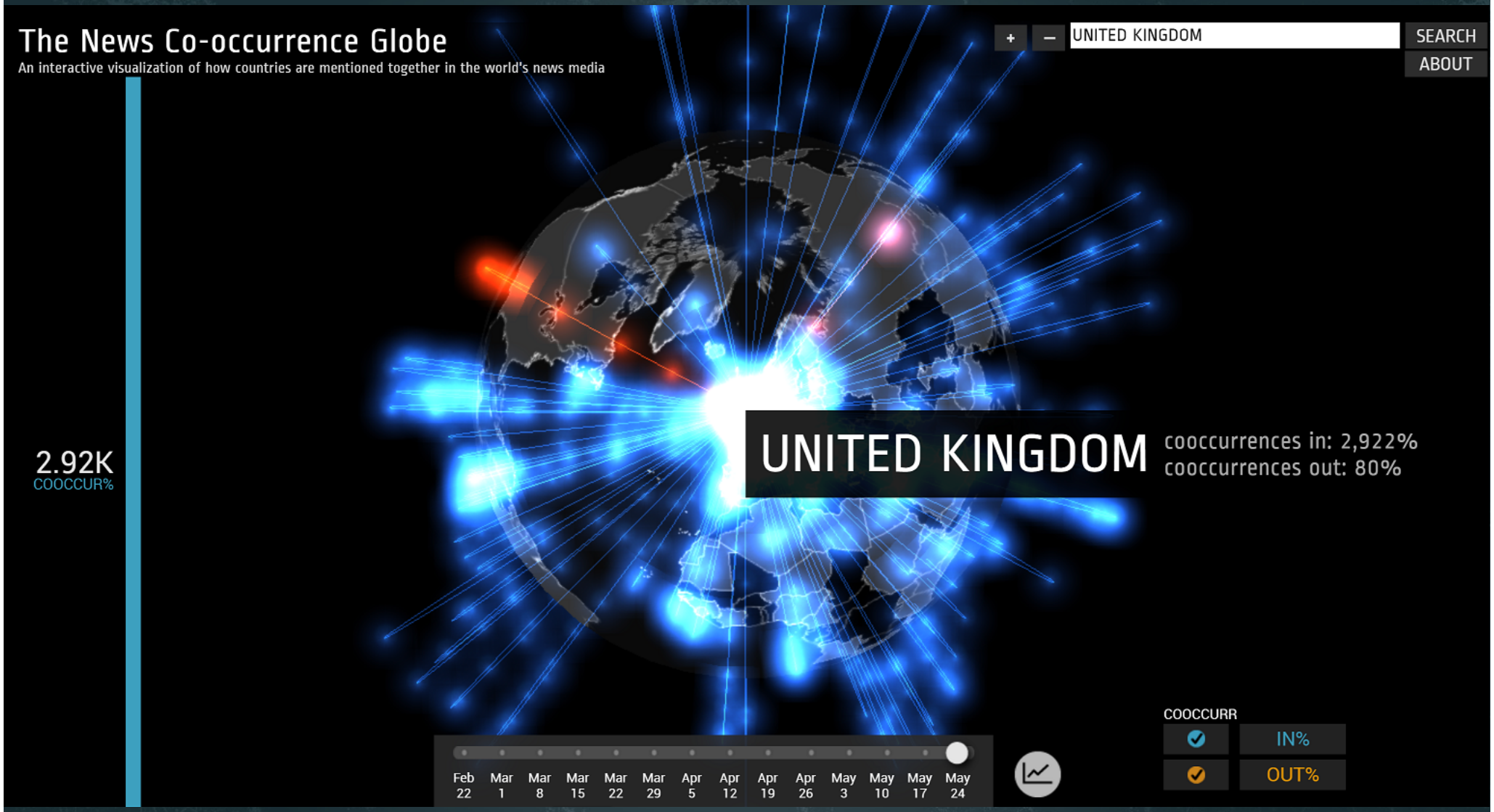
http://scimaps.org/iteration/11



Earth – Cameron Beccario



AcademyScope – National Academy of the Sciences & CNS



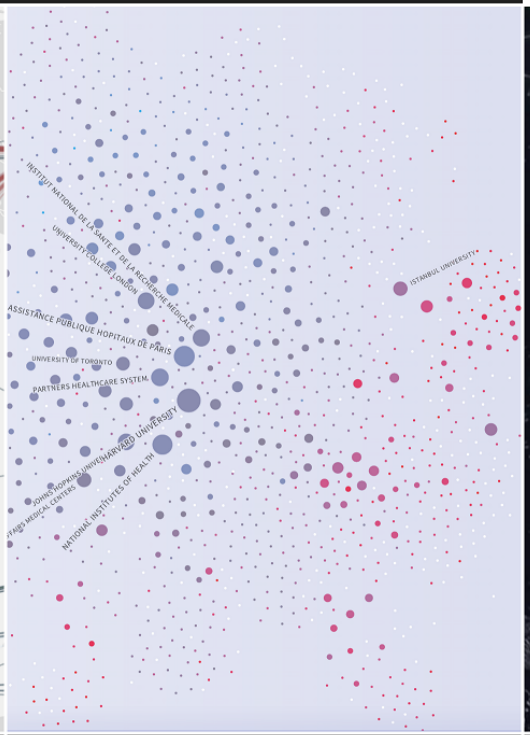
Mapping Global Society –Kalev Leetaru

» Make Sense of Science

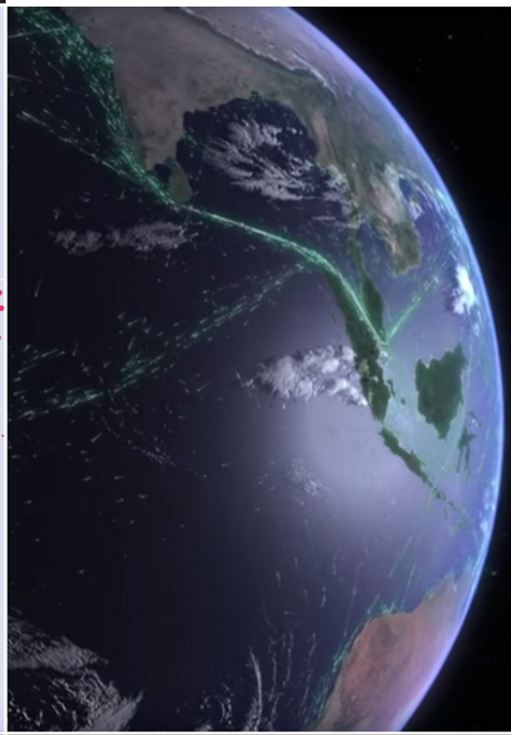
PLACES & SPACES MAPPING SCIENCE scimaps.org

Smelly Maps
Charting urban smellscapes

HathiTrust
Storehouse of knowledge



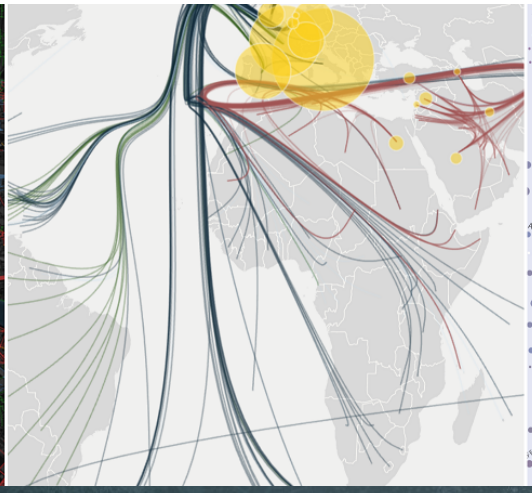
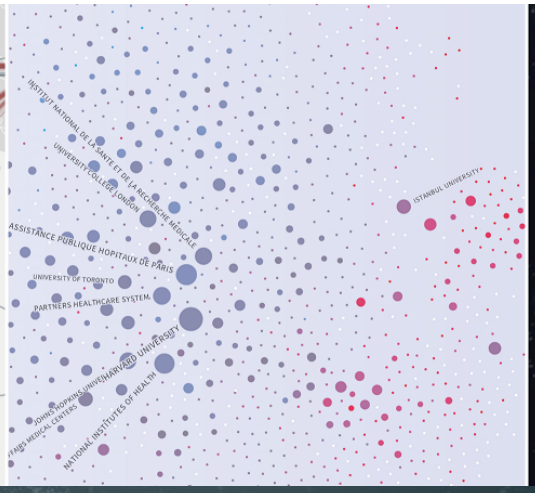
Excellence Networks
Publish or perish together



FleetMon Explorer
Tracking the seven seas

Iteration XII (2016): Macroscopes for Making Sense of Science

http://scimaps.org/iteration/12

1. **Smelly Maps:** Features a "smellscape" of 12 cities mapped by smell using social media
2. **HathiTrust:** Highlights the diversity of publications collected in digital form by HathiTrust.
3. **Excellence Networks:** Compares how research institutions, such as Indiana and Vanderbilt universities, collaborate with one another.
4. **FleetMon:** Shows how the amount of shipping traffic that navigates the Strait of Malacca compared to other major shipping lanes of the world.

A visitor explores the macroscope kiosk at the Eskenazi Museum of Art at Indiana University.

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 is designed to open people's hearts and minds to the value, complexity, and beauty of maps of science and technology.

Drawing from across cultures and across scholarly disciplines, the *Places & Spaces: Mapping Science* exhibit demonstrates the

Places & Spaces Exhibit at the David J. Sencer CDC Museum, Atlanta, GA

CDC Opening Event: Maps of Health

Tutorial and Symposium

Research

Mapping Longitudinal Scientific Progress, Collaboration and Impact of the Alzheimer's Disease Neuroimaging Initiative (ADNI)

Xiaohui Yao[1,3,4], Jingwen Yan[1,3,4], Michael Ginda[2,3], Katy Börner[2,3], Andrew J Saykin[1,3], Li Shen[1,3,4], for the Alzheimer's Disease Neuroimaging Initiative*

[1] Center for Neuroimaging, Indiana University School of Medicine

[2] Cyberinfrastructure for Network Science Center, Indiana University Bloomington

[3] Indiana University Network Science Institute

[4] School of Informatics and Computing, Indiana University

*Data used in preparation of this article were obtained from the Alzheimer's disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in data analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

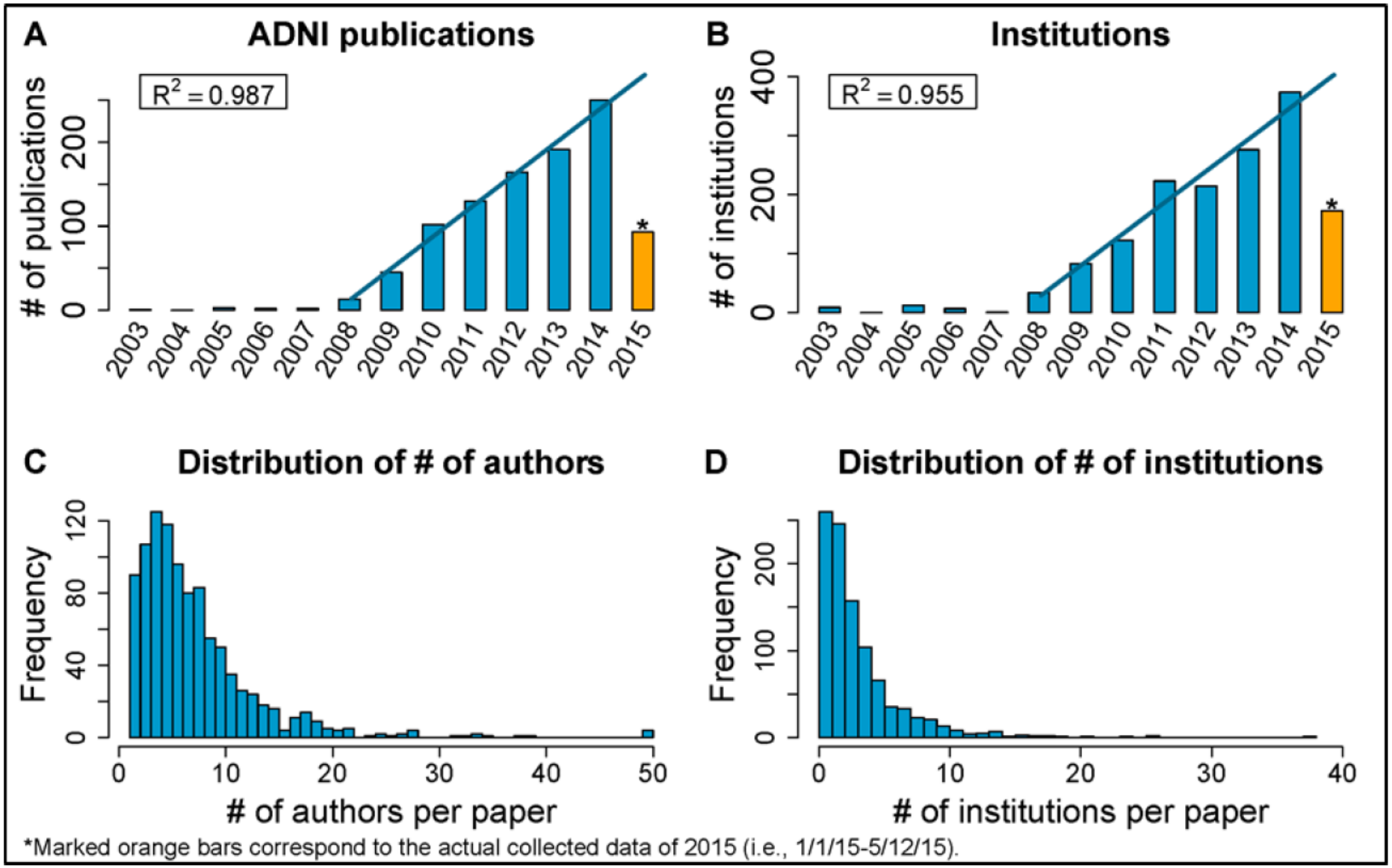


Figure 1. Statistics for ADNI publications between 01/01/2003 and 05/12/2015. (A) Growth of ADNI publications on the year-by-year basis; line indicates a linear regression prediction for the 2015 number using data from 2008 to 2014. (B) Growth of institutions involved in ADNI publications; line indicates a linear regression prediction for the 2015 number using data from 2008 to 2014. (C) Distribution of number of authors per paper. (D) Distribution of number of institutions per paper.

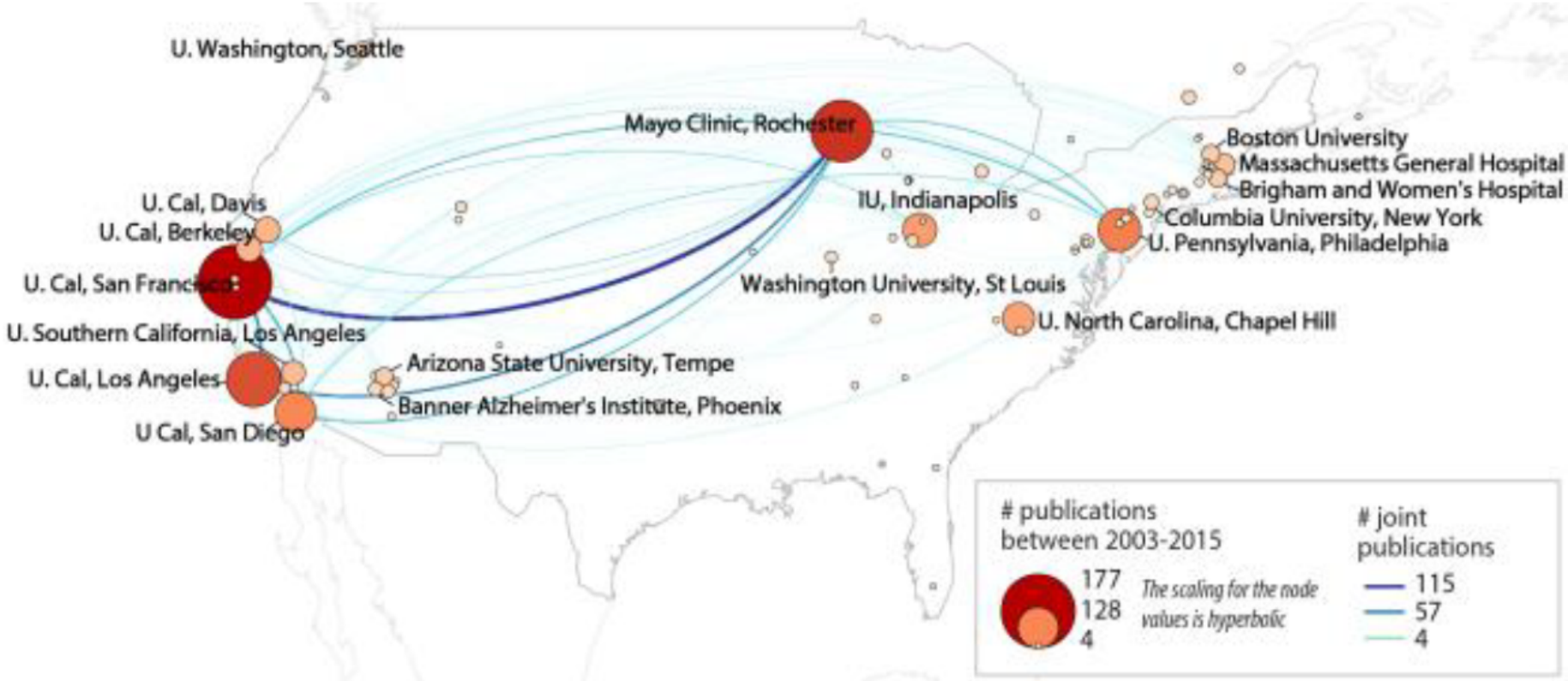
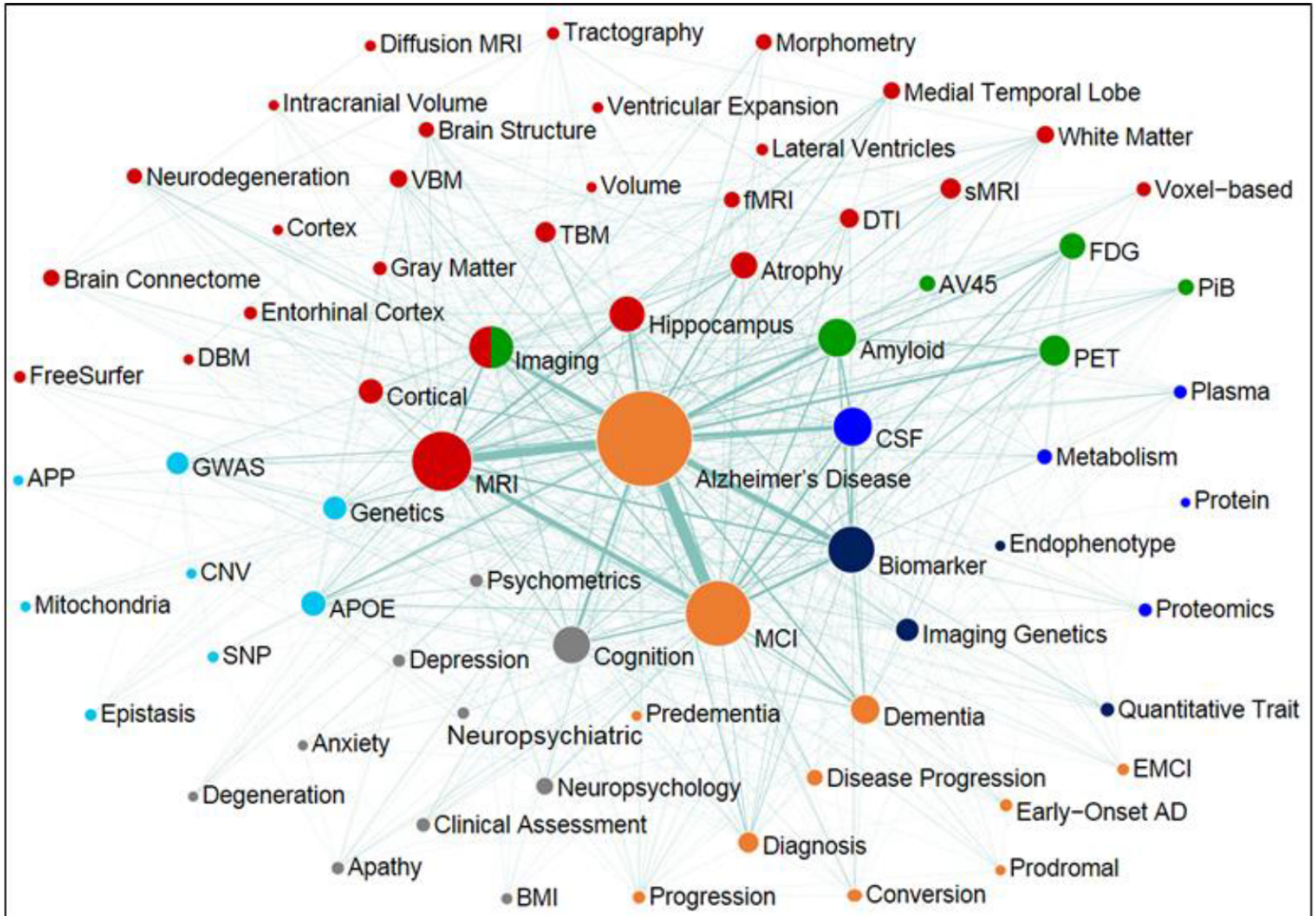


Figure 3: Co-affiliation network overlaid on a geospatial map shows collaborating organizations affiliated with ADNI in North American based on co-authored publications. Only organizations with at least 4 publications are shown; organizations with at least 30 publications or that are a Core ADNI research institution have been labeled in the network. Organization relationships (edges) with four or more co-authorships are shown.



Supplemental Figure 5: Keyword co-occurrence network focused on major ADNI themes. Nodes represent keywords relevant to major ADNI themes, including MRI, PET, other biological biomarkers, clinical and neuropsychological assessment, genetics, and disease and progression. Edges denote the joint appearance of keywords in a publication. Nodes are colored based on the themes they belonged to, and those across three or more themes are colored in dark blue. Both nodes and edges were scaled proportionally based on Bezier curve. Only nodes with degree > 2 are shown.

Co-Author Network BEFORE Precision Health Initiative

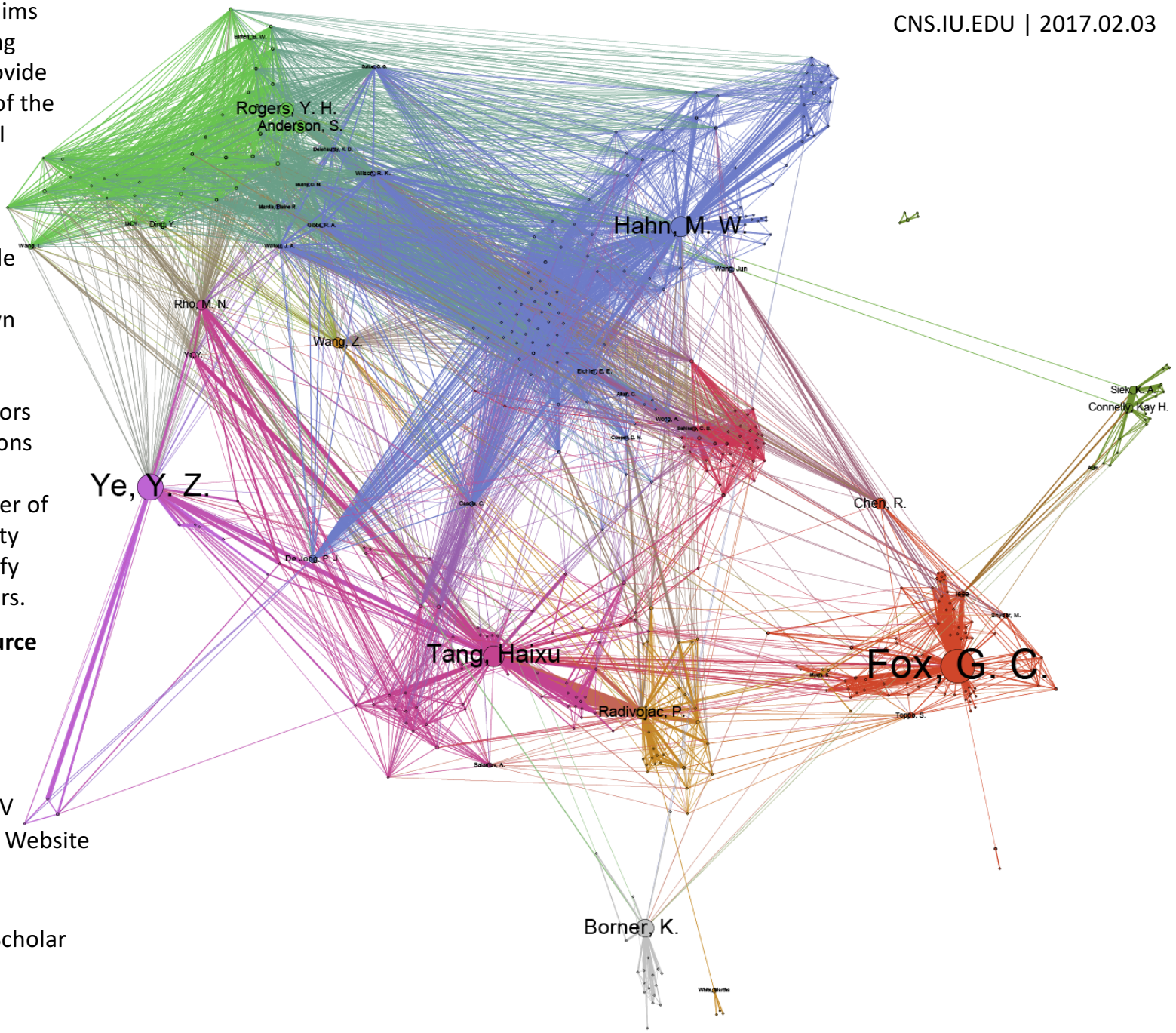
CNS.IU.EDU | 2017.02.03

PHI Impact Assessment work aims to ease the evaluation/reporting burden for PHI leads and to provide a more holistic understanding of the expertise and impact of the PHI team effort. High quality and high coverage data about project activity and outcomes is required to provide actionable insights.

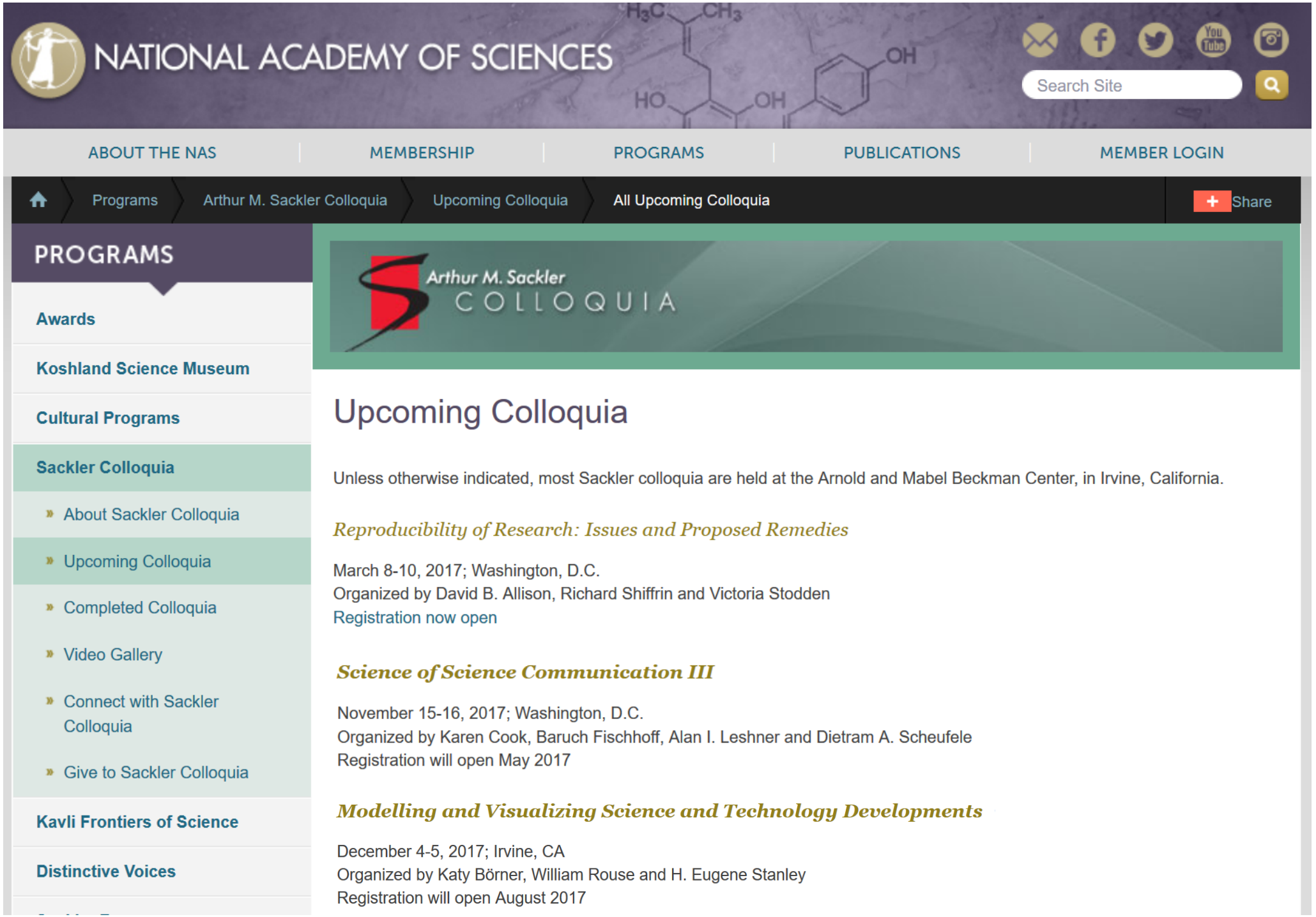
The co-author network shown here was extracted from 800+ papers gathered 10/31 to 11/28/2016. It shows 414 authors with more than three publications and 5,808 collaboration edges.

Node size denotes the number of publications. Blondel community detection was applied to identify (and color) collaboration clusters.

Name	# Pubs	Data Source
Borner	108	WoS
Connelly	32	CV
Fox	180	WoS
Hahn	123	WoS
Myers	29	WoS & CV
Natarajan	15	Personal Website
Predrag	60	WoS
Sahinalp	68	PubMed
Siek	32	Google Scholar
Tang	102	PubMed
White	15	CV
Ye	71	WoS



Outlook



http://www.nasonline.org/programs/sackler-colloquia/upcoming-colloquia

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

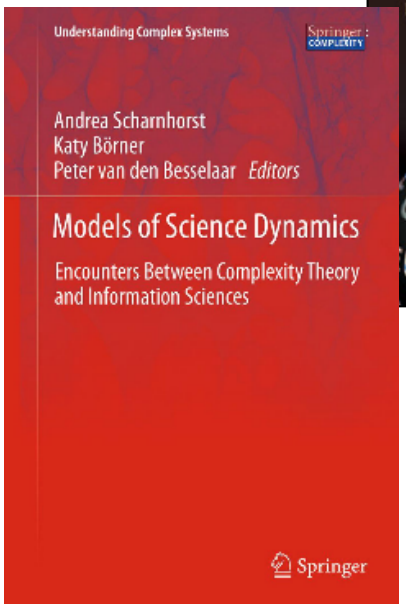
Börner, Katy (2010) **Atlas of Science: Visualizing What We Know**. The MIT Press. http://scimaps.org/atlas

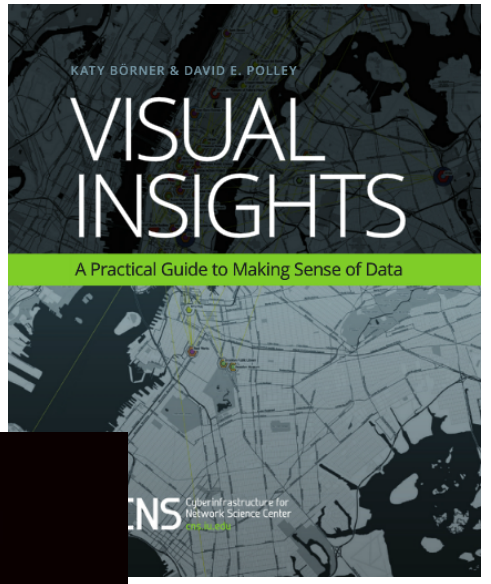
Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2012) **Models of Science Dynamics**. Springer Verlag.

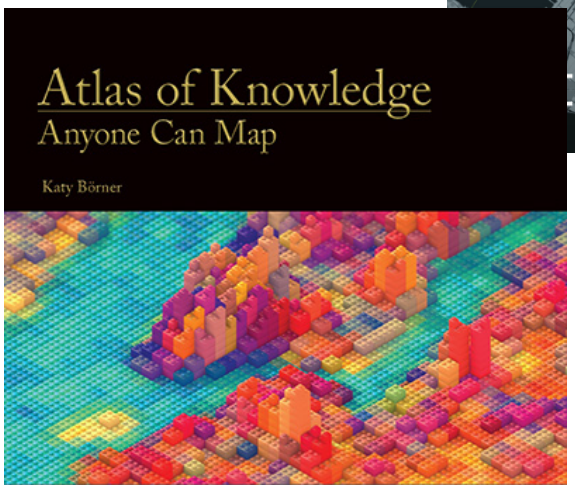
Katy Börner, Michael Conlon, Jon Corson-Rikert, Cornell, Ying Ding (2012) **VIVO: A Semantic Approach to Scholarly Networking and Discovery**. Morgan & Claypool.

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







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

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