

# Visualizing Education, Science, and Technology

Katy Börner @katycns

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



Keynote at SocInfo, <https://socinfo2019.qcri.org>  
Doha, Qatar

November 19, 2019



# Overview

## Data Visualization Literacy (DVL)

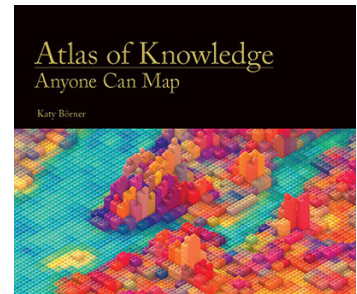
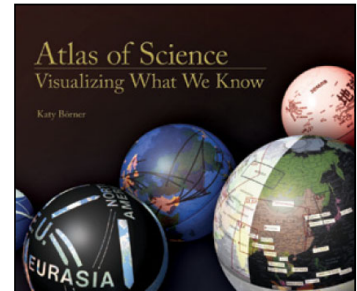
- Börner, Katy, Andreas Bueckle, and Michael Ginda. 2019. [Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments](#). *PNAS*, 116 (6) 1857-1864.
- Börner, Katy. 2015. [Atlas of Knowledge: Anyone Can Map](#). Cambridge, MA: The MIT Press.
- Börner, Katy. 2010. [Atlas of Science: Visualizing What We Know](#). Cambridge, MA: The MIT Press.

## Skill Discrepancies

- Börner, Katy, Olga Scrivner, Michael Gallant, Shutian Ma, Xiaozhong Liu, Keith Chewning, Lingfei Wu, and James Evans. 2018. "Skill Discrepancies Between Research, Education, and Jobs Reveal the Critical Need to Supply Soft Skills for the Data Economy". *PNAS* 115 (50): 12630-12637. doi: 10.1073/pnas.1804247115.

## Scaling-Up: Increase global DVL (<https://ivmooc.cns.iu.edu> & <https://visanalytics.cns.iu.edu>)

## The 15<sup>th</sup> iteration of the *Places & Spaces: Mapping Science* exhibit (<http://scimaps.org>).



# Data Visualization Literacy Framework

Börner, Katy, Andreas Bueckle, and Michael Ginda. 2019. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *PNAS*, 116 (6) 1857-1864.



# Data Visualization Literacy (DVL)

Data visualization literacy (ability to read, make, and explain data visualizations) requires:

- literacy (ability to read and write text in titles, axis labels, legends, etc.),
- visual literacy (ability to find, interpret, evaluate, use, and create images and visual media), and
- mathematical literacy (ability to formulate, employ, and interpret math in a variety of contexts).

Being able to “read and write” data visualizations is becoming as important as being able to read and write text. Understanding, measuring, and improving data and visualization literacy is important to strategically approach local and global issues.



# DVL Framework: Desirable Properties

- Most existing frameworks focus on **READING**. We believe that much expertise is gained from also **CONSTRUCTING** data visualizations.
- Reading and constructing data visualizations needs to take human perception and cognition into account.
- Frameworks should build on and consolidate prior work in cartography, psychology, cognitive science, statistics, scientific visualization, data visualization, learning sciences, etc. in support of a de facto standard.
- Theoretically grounded + practically useful + easy to learn/use.
- Highly modular and extendable.

# DVL Framework: Development Process

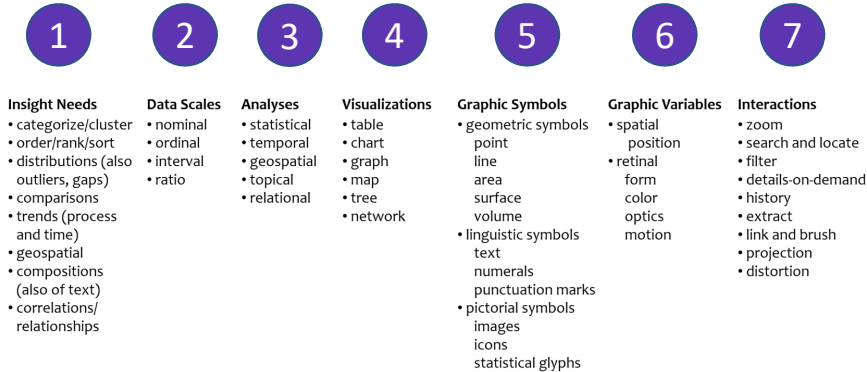
- The initial DVL-FW was developed via an extensive literature review.
- The resulting DVL-FW typology, process model, exercises, and assessments were then tested in the *Information Visualization* course taught for more than 17 years at Indiana University. More than 8,500 students enrolled in the IVMOOC version (<http://ivmooc.cns.iu.edu>) over the last six years.
- The FW was further refined using feedback gained from constructing and interpreting data visualizations for 100+ real-world client projects.
- Data on student engagement, performance, and feedback guided the continuous improvement of the DVL-FW typology, process model, and exercises for defining, teaching, and assessing DVL.
- The DVL-FW used in this course supports the systematic construction and interpretation of data visualizations.

# Data Visualization Literacy Framework (DVL-FW)

Consists of two parts:

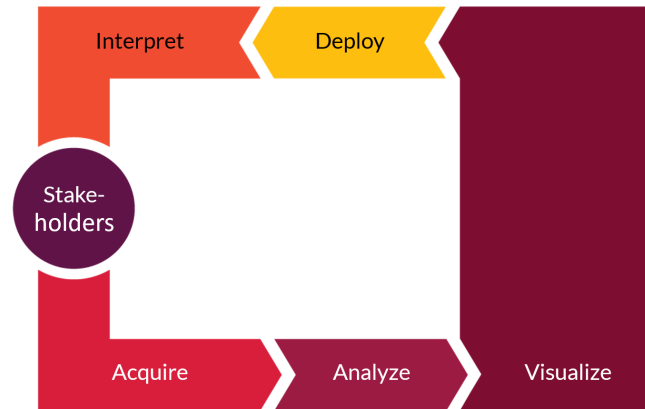
## DVL Typology

Defines 7 types with 4-17 members each.



## DVL Workflow Process

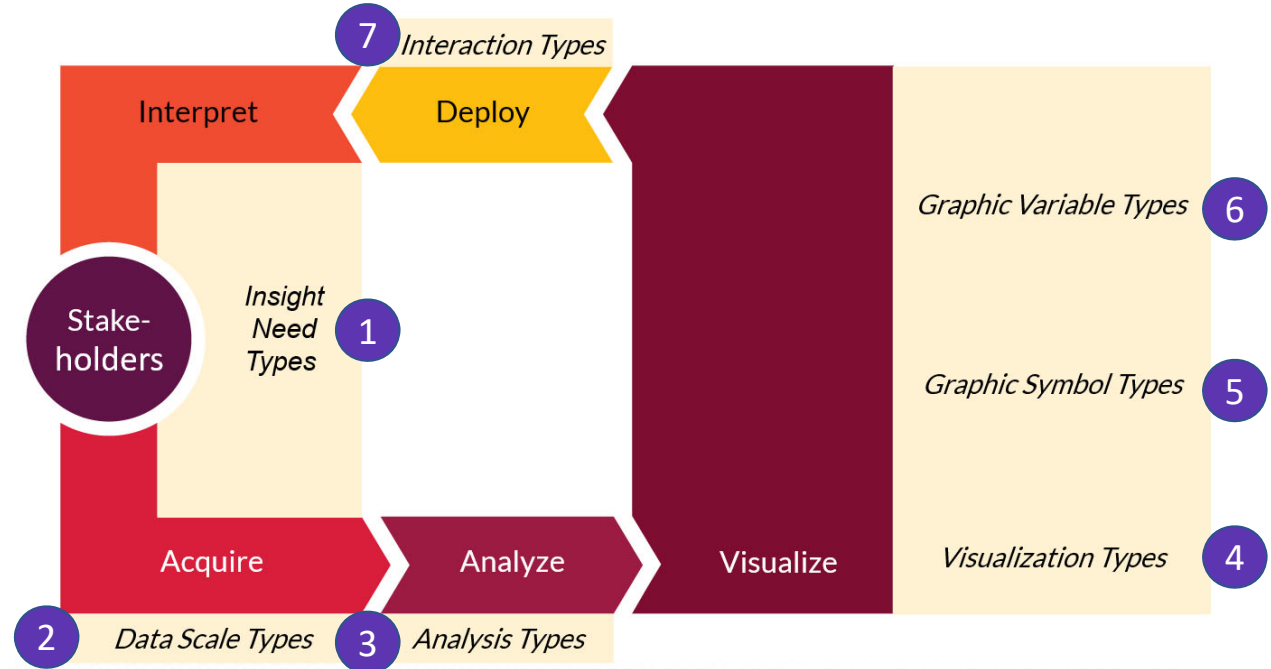
Defines 5 steps required to render data into insights.



# Data Visualization Literacy Framework (DVL-FW)

Consists of two parts that are interlinked:

**DVL Typology +  
DVL Workflow Process**



# Data Visualization Literacy Framework (DVL-FW)

Implemented in Make-A-Vis (MAV) to support learning via horizontal transfer, scaffolding, hands-on learning, etc.

The screenshot displays the Make-A-Vis interface. On the left, the 'Data' section shows two sources: 'ISI Publications: (CSV) Preprocessed-wos' and 'Journals: (from ISI Publications)'. The 'Journals' table is as follows:

Name	#Papers	#Cites	First Year	Last Year
BMC EVOL BIOL	1	7	2006	2006
FEBS J	2	0	2005	2005
NAT PHYS	3	18	2005	2006

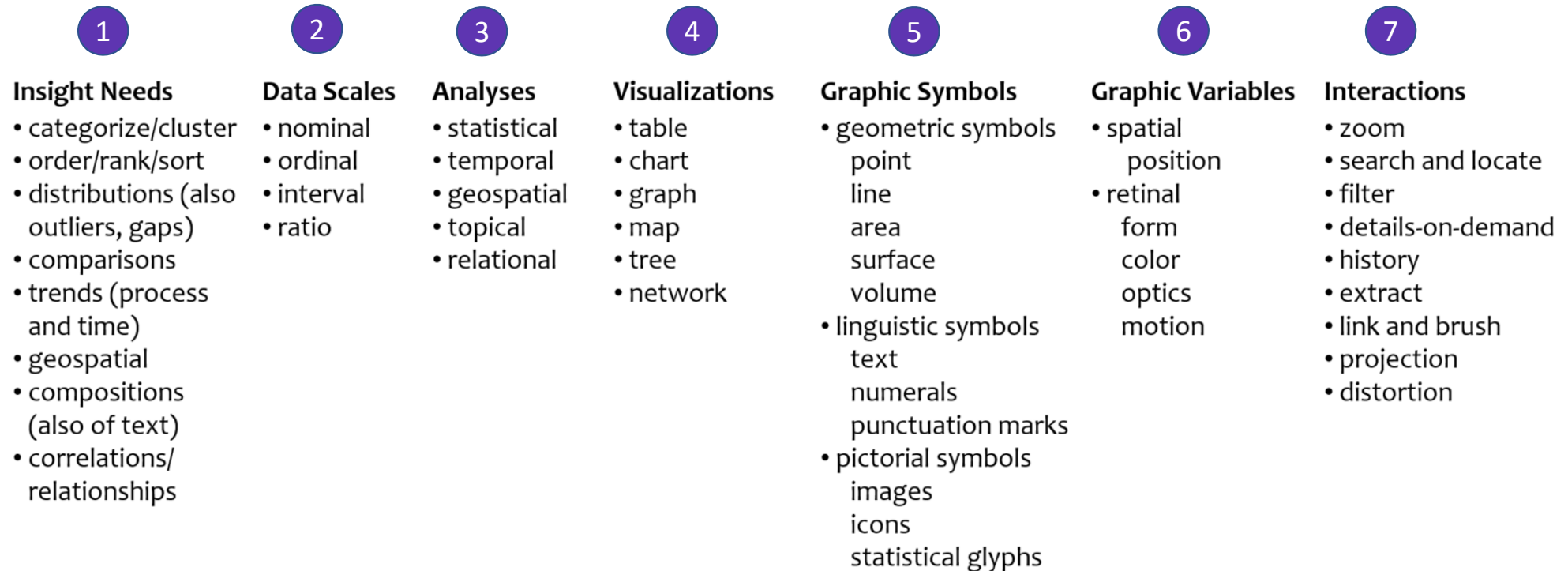
The 'Make Visualization' section offers four visualization types: Scatter Graph, Geomap, Scimap, and Temporal Bar Graph. The 'Temporal Bar Graph' is selected. On the right, the 'Temporal Bar Graph' visualization shows a horizontal bar chart with the x-axis representing years from 1998 to 2017. The y-axis lists various topics. The bars represent the duration of each topic's activity:

- Machine: 1998-2000
- Big Data: 2000-2003
- Education: 2000-2001
- Building: 2000-2001
- Making: 2001-2002
- Computing: 2001-2002
- Web: 2002-2003
- Form: 2002-2003
- Smart: 2003-2004
- Capacity: 2004-2005
- Algebraic Geometry: 2005-2006
- Parts: 2006-2007
- Law: 2007-2008
- Stem: 2008-2009
- Analysis: 2009-2010
- Recovery: 2010-2012
- Geometry: 2012-2014
- Computer: 2013-2014
- Application: 2014-2015
- Robotics: 2015-2017

Numbered callouts 4, 5, and 6 point to the 'Select Visualization Type', 'Select Graphic Symbol Type(s)', and 'Select Graphic Variable Types' options respectively.



# Typology of the Data Visualization Literacy Framework



Börner, Katy. 2015. *Atlas of Knowledge: Anyone Can Map*. Cambridge, MA: The MIT Press. 25.

# Typology of the Data Visualization Literacy Framework

1

## Insight Needs

- categorize/cluster
- order/rank/sort
- distributions (also outliers, gaps)
- comparisons
- trends (process and time)
- geospatial
- compositions (also of text)
- correlations/relationships

## Data Scales

- nominal
- ordinal
- interval
- ratio

## Analyses

- statistical
- temporal
- geospatial
- topical
- relational

## Visualizations

- table
- chart
- graph
- map
- tree
- network

## Graphic Symbols

- geometric symbols
  - point
  - line
  - area
  - surface
  - volume
- linguistic symbols
  - text
  - numerals
  - punctuation marks
- pictorial symbols
  - images
  - icons
  - statistical glyphs

## Graphic Variables

- spatial
  - position
- retinal
  - form
  - color
  - optics
  - motion

## Interactions

- zoom
- search and locate
- filter
- details-on-demand
- history
- extract
- link and brush
- projection
- distortion

Börner, Katy. 2015. *Atlas of Knowledge: Anyone Can Map*. Cambridge, MA: The MIT Press. 26-27.

Bertin, 1967	Wehrend & Lewis, 1996	Few, 2004	Yau, 2011	Rendgen & Wiedemann, 2012	Frankel, 2012	Tool: Many Eyes	Tool: Chart Chooser	Börner, 2014
selection	categorize			category				categorize/ cluster
order	rank	ranking					table	order/rank/ sort
	distribution	distribution					distribution	distributions (also outliers, gaps)
	compare	nominal comparison & deviation	differences		compare and contrast	compare data values	comparison	comparisons
		time series	patterns over time	time	process and time	track rises and falls over time	trend	trends (process and time)
		geospatial	spatial relations	location		generate maps		geospatial
quantity		part-to- whole	proportions		form and structure	see parts of whole, analyze text	composition	compositions (also of text)
association	correlate	correlation	relationships	hierarchy		relations between data points	relationship	correlations/ relationships

# Typology of the Data Visualization Literacy Framework

6

## Insight Needs

- categorize/cluster
- order/rank/sort
- distributions (also outliers, gaps)
- comparisons
- trends (process and time)
- geospatial
- compositions (also of text)
- correlations/relationships

## Data Scales

- nominal
- ordinal
- interval
- ratio

## Analyses

- statistical
- temporal
- geospatial
- topical
- relational

## Visualizations

- table
- chart
- graph
- map
- tree
- network

## Graphic Symbols

- geometric symbols
  - point
  - line
  - area
  - surface
  - volume
- linguistic symbols
  - text
  - numerals
  - punctuation marks
- pictorial symbols
  - images
  - icons
  - statistical glyphs

## Graphic Variables

- spatial
  - position
- retinal
  - form
  - color
  - optics
  - motion

## Interactions

- zoom
- search and locate
- filter
- details-on-demand
- history
- extract
- link and brush
- projection
- distortion

Börner, Katy. 2015. *Atlas of Knowledge: Anyone Can Map*. Cambridge, MA: The MIT Press. 34-35.

# Graphic Variable Types

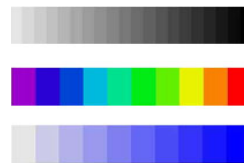
**Position:** x, y; possibly z

**Form:**

- Size
- Shape
- Rotation (Orientation)

**Color:**

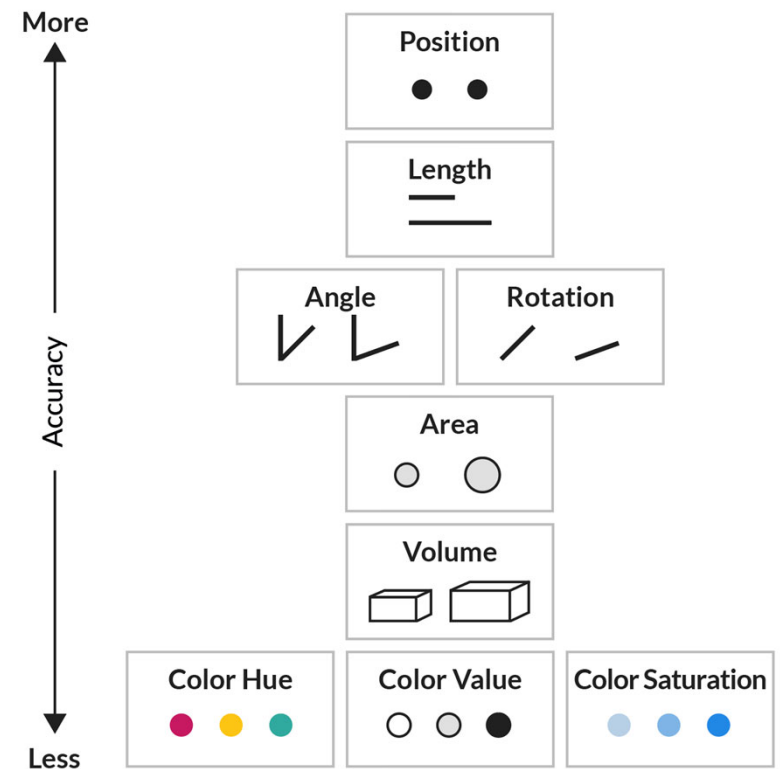
- Value (Lightness)
- Hue (Tint)
- Saturation (Intensity)



**Optics:** Blur, Transparency, Shading, Stereoscopic Depth

**Texture:** Spacing, Granularity, Pattern, Orientation, Gradient

**Motion:** Speed, Velocity, Rhythm





## Graphic Symbol Types

		Geometric Symbols		Linguistic Symbols	Pictorial Symbols	
		Point	Line			
Spatial	Position					
	Form	Size:		Text:		
Retinal	Shape			Text:		
	Color	Value			Text:	
		Hue			Text:	
		Saturation			Text:	
	Texture	Granularity				
Pattern						
Optics	Blur			Text:		
Motion	Speed					

See *Atlas of Knowledge* pages 36-39 for complete table.

### Qualitative

Also called:  
Categorical Attributes  
Identity Channels

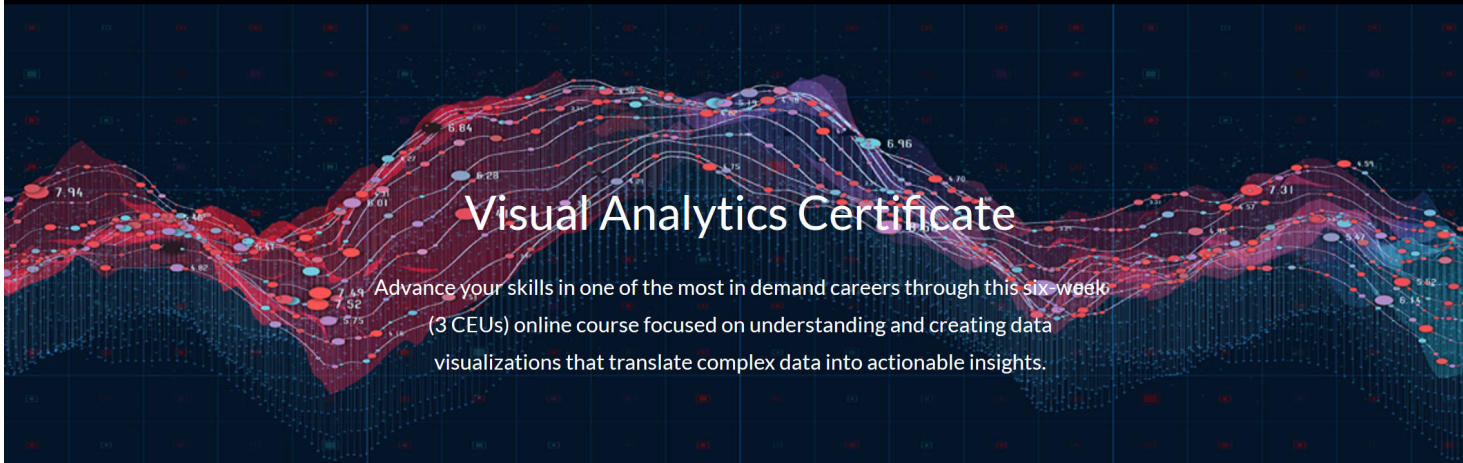
### Quantitative

Also called:  
Ordered Attributes  
Magnitude Channels

# Graphic Variable Types Versus Graphic Symbol Types

		Point	Line	Geometric Symbols		Surface	Volume	Linguistic Symbols Text, Numerals, Punctuation Marks			Pictorial Symbols Images, Icons, Statistical Glyphs	
Spatial	x	quantitative										
	y	quantitative										
	z	quantitative										
Perceptual	Size	quantitative	NA (Not Applicable)									
	Shape	qualitative	NA									
	Rotation	quantitative	NA									
	Curvature	quantitative	NA									
	Angle	quantitative	NA									
	Closure	quantitative	NA									
	Value	quantitative										
	Color	qualitative										
Saturation	quantitative											
Retinal	Spacing	quantitative										
	Granularity	quantitative										
	Pattern	qualitative										
	Orientation	quantitative	NA									
	Gradient	quantitative										
	Blur	quantitative										
	Transparency	quantitative										
	Shading	quantitative										
	Stereoscopic Depth	quantitative	Point in foreground .. background	Line in foreground .. background	Area in foreground .. background	Surface in foreground .. background	Volume in foreground .. background	Text in foreground .. background	Text in foreground .. background	Text in foreground .. background	Text in foreground .. background	Icons in foreground .. background
	Speed	quantitative										
Velocity	quantitative											
Rhythm	quantitative	Blinking point slow .. fast	Blinking line slow .. fast	Blinking area slow .. fast	Blinking surface slow .. fast	Blinking volume slow .. fast	Blinking text slow .. fast	Blinking text slow .. fast	Blinking text slow .. fast	Blinking text slow .. fast	Blinking icons slow .. fast	

See *Atlas of Knowledge* pages 36-39 for complete table.



## Visual Analytics Certificate

Advance your skills in one of the most in demand careers through this six-week (3 CEUs) online course focused on understanding and creating data visualizations that translate complex data into actionable insights.

DOWNLOAD FLYER

REGISTER FOR JAN 20–MAR 1, 2020



### Learn from Experts

Connect with industry professionals and leading researchers.



### Evolve Yourself

Gain forever knowledge and skill-up in powerful data visualization tools.



### Make a Difference

Embrace data-driven decision-making in your personal and professional life.

<https://visanalytics.cns.iu.edu>

# Modelling Science, Technology, Education & Innovation

Börner, Katy, William Rouse, Paul Trunfio, and H. Eugene Stanley. 2018. "Forecasting Innovations in Science, Technology, and Education". *PNAS* 115 (50): 12573-12581. doi: 10.1073/pnas.1818750115.

See also <https://www.pnas.org/modeling>





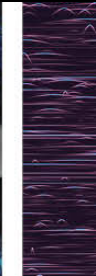
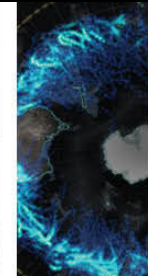


Government, academic, and industry leaders discussed challenges and opportunities associated with using big data, visual analytics, and computational models in STI decision-making.

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







## Modeling and Visualizing Science and Technology Developments

National Academy of Sciences Sackler Colloquium, December 4-5, 2017, Irvine, CA

### Rankings and the Efficiency of Institutions

H. Eugene Stanley | Albert-László Barabási | Lada Adamic | Marta González | Kaye Husbands Fealing | Brian Uzzi | John V. Lombardi

### Higher Education and the Science & Technology Job Market

Katy Börner | Wendy L. Martinez | Michael Richey | William Rouse | Stasa Milojevic | Rob Rubin | David Krakauer

### Innovation Diffusion and Technology Adoption

William Rouse | Donna Cox | Jeff Alstott | Ben Shneiderman | Rahul C. Basole | Scott Stern | Cesar Hidalgo

### Modeling Needs, Infrastructures, Standards

Paul Trunfio | Sallie Keller | Andrew L. Russell | Guru Madhavan | Azer Bestavros | Jason Owen-Smith

[nasonline.org/Sackler-Visualizing-Science](http://nasonline.org/Sackler-Visualizing-Science)



## PROGRAMS

### Sackler Colloquia

- ▶ About Sackler Colloquia
- ▶ Upcoming Colloquia
- ▶ Completed Colloquia
- ▶ Sackler Lectures
- ▶ Video Gallery
- ▶ Connect with Sackler Colloquia
- ▶ Give to Sackler Colloquia

### Cultural Programs

### Distinctive Voices

### Kavli Frontiers of Science

### Keck Futures Initiative

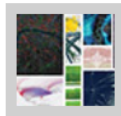
### LabX

### Sackler Forum

### Science & Entertainment Exchange



## Modeling and Visualizing Science and Technology Developments



December 4-5, 2017; Irvine, CA  
Organized by Katy Börner, H. Eugene Stanley, William Rouse and Paul Trunfio

### Overview

This colloquium was held in Irvine, CA on December 4-5, 2017.

This colloquium brought together researchers and practitioners from multiple disciplines to present, discuss, and advance computational models and visualizations of science and technology (S&T). Existing computational models are being applied by academia, government, and industry to explore questions such as: What jobs will exist in ten years and what career paths lead to success? Which types of institutions will likely be most innovative in the future? How will the higher education cost bubble burst affect these institutions? What funding strategies have the highest return on investment? How will changing demographics, alternative economic growth trajectories, and relationships among nations impact answers to these and other questions? Large-scale datasets (e.g., publications, patents, funding, clinical trials, stock market, social media data) can now be utilized to simulate the structure and evolution of S&T. Advances in computational power have created the possibility of implementing scalable, empirically validated computational models. However, because the databases are massive and multidimensional, both the data and the models tend to exceed human comprehension. How can advances in data visualizations be effectively employed to communicate the data, the models, and the model results to diverse stakeholder groups? Who will be the users of next generation models and visualizations and what decisions will they be addressing.

Videos of the talks are available on the [Sackler YouTube Channel](#).

<https://www.pnas.org/modeling>

## Arthur M. Sackler Colloquium on Modeling and Visualizing Science and Technology Developments

✔ **Twin-Win Model: A human-centered approach to research success**

Ben Shneiderman

PNAS December 11, 2018 115 (50) 12590-12594; first published December 10, 2018. <https://doi.org/10.1073/pnas.1802918115>

✔ **Forecasting innovations in science, technology, and education**

FROM THE COVER

Katy Börner, William B. Rouse, Paul Trunfio, and H. Eugene Stanley

PNAS December 11, 2018 115 (50) 12573-12581; first published December 10, 2018. <https://doi.org/10.1073/pnas.1818750115>

✔ **How science and technology developments impact employment and education**

Wendy Martinez

PNAS December 11, 2018 115 (50) 12624-12629; first published December 10, 2018. <https://doi.org/10.1073/pnas.1803216115>

✔ **Scientific prize network predicts who pushes the boundaries of science**

Yifang Ma and Brian Uzzi

PNAS December 11, 2018 115 (50) 12608-12615; first published December 10, 2018. <https://doi.org/10.1073/pnas.1800485115>

✔ **The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms**

C. Jara-Figueroa, Bogang Jun, Edward L. Glaeser, and Cesar A. Hidalgo

PNAS December 11, 2018 115 (50) 12646-12653; first published December 10, 2018. <https://doi.org/10.1073/pnas.1800475115>

## Arthur M. Sackler Colloquium on Modeling and Visualizing Science and Technology Developments

- ✔ **Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy**

Katy Börner, Olga Scrivner, Mike Gallant, Shutian Ma, Xiaozhong Liu, Keith Chewning, Lingfei Wu, and James A. Evans  
PNAS December 11, 2018 115 (50) 12630-12637; first published December 10, 2018. <https://doi.org/10.1073/pnas.1804247115>

- ✔ **Changing demographics of scientific careers: The rise of the temporary workforce**

Staša Milojević, Filippo Radicchi, and John P. Walsh  
PNAS December 11, 2018 115 (50) 12616-12623; first published December 10, 2018. <https://doi.org/10.1073/pnas.1800478115>

- ✔ **The chaperone effect in scientific publishing**

Vedran Sekara, Pierre Deville, Sebastian E. Ahnert, Albert-László Barabási, Roberta Sinatra, and Sune Lehmann  
PNAS December 11, 2018 115 (50) 12603-12607; first published December 10, 2018. <https://doi.org/10.1073/pnas.1800471115>

- ✔ **Modeling research universities: Predicting probable futures of public vs. private and large vs. small research universities**

William B. Rouse, John V. Lombardi, and Diane D. Craig  
PNAS December 11, 2018 115 (50) 12582-12589; first published December 10, 2018. <https://doi.org/10.1073/pnas.1807174115>

and more ...

# Skill Discrepancies Between Research, Education, and Jobs Reveal the Critical Need to Supply Soft Skills for the Data Economy

- Data and Crosswalks
- MaxMatch for NLP
- Causal Analyses
- Visualizations

Börner, Katy, Olga Scrivner, Mike Gallant, Shutian Ma, Xiaozhong Liu, Keith Chewning, Lingfei Wu, and James A. Evans. 2018. "Skill Discrepancies Between Research, Education, and Jobs Reveal the Critical Need to Supply Soft Skills for the Data Economy." *PNAS* 115(50): 12630-12637.

PNAS PNAS PNAS

## Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy

Katy Börner<sup>a,b,1</sup>, Olga Scrivner<sup>a</sup>, Mike Gallant<sup>a</sup>, Shutian Ma<sup>a,c</sup>, Xiaozhong Liu<sup>a</sup>, Keith Chewning<sup>d</sup>, Lingfei Wu<sup>e,f,g,h</sup>, and James A. Evans<sup>i,g,i,1</sup>

<sup>a</sup>School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN 47408; <sup>b</sup>Educational Technology/Media Centre, Dresden University of Technology, 01062 Dresden, Germany; <sup>c</sup>Department of Information Management, Nanjing University of Science and Technology, 210094 Nanjing, China; <sup>d</sup>Burning Glass Technologies, Boston, MA 02110; <sup>e</sup>School of Journalism and Communication, Nanjing University, 210008 Nanjing, China; <sup>f</sup>Department of Sociology, University of Chicago, Chicago, IL 60637; <sup>g</sup>Knowledge Lab, University of Chicago, Chicago, IL 60637; <sup>h</sup>Tencent Research Institute, 100080 Beijing, China; and <sup>i</sup>Santa Fe Institute, Santa Fe, NM 87501

Edited by William B. Rouse, Stevens Institute of Technology, Hoboken, NJ, and accepted by Editorial Board Member Pablo G. Debenedetti September 12, 2018 (received for review March 14, 2018)

Rapid research progress in science and technology (S&T) and continuously shifting workforce needs exert pressure on each other and on the educational and training systems that link them. Higher education institutions aim to equip new generations of students with skills and expertise relevant to workforce participation for decades to come, but their offerings sometimes misalign with commercial needs and new techniques forged at the frontiers of research. Here, we analyze and visualize the dynamic skill (mis-)alignment between academic push, industry pull, and educational offerings, paying special attention to the rapidly emerging areas of data science and data engineering (DS/DE). The visualizations and computational models presented here can help key decision makers understand the evolving structure of skills so that they can craft educational programs that serve workforce needs. Our study uses millions of publications, course syllabi, and job advertisements published between 2010 and 2016. We show how courses mediate between research and jobs. We also discover responsiveness in the academic, educational, and industrial system in how skill demands from industry are as likely to drive skill attention in research as the converse. Finally, we reveal the increasing importance of uniquely human skills, such as communication, negotiation, and persuasion. These skills are currently underexamined in research and undersupplied through education for the labor market. In an increasingly data-driven economy, the demand for "soft" social skills, like teamwork and communication, increase with greater demand for "hard" technical skills and tools.

science of science | job market | data mining | visualization | market gap analysis

doors. Some predictions say hundreds or even thousands of colleges and universities will close or merge in the coming years (4). In addition, there seem to be major discrepancies and delays between leading scientific research, job market needs, and educational content. This has been particularly expressed with respect to science, technology, engineering, and mathematics jobs, where scientific and technological progress is rapid. Strategic decision making on what to teach, whom to hire, and what new research to fund benefits from a systematic analysis of the interplay between science and technology (S&T) developments, courses and degrees offered, and job market needs. Specifically, stakeholders in US higher education urgently need answers to the following questions. (i) Students: what jobs might exist in 5–10 years? What educational trajectories will best achieve my dream job? What core and specialized skills are required for what jobs and offered by what schools and programs? (ii) Teachers: what course updates are most needed? What balance of timely vs. timeless knowledge should I teach? How can I innovate in teaching and maintain job security or tenure? (iii) Universities: what programs should be created? What is my competition doing? How do I tailor programs to fit workforce needs? (iv) Science funders: how can S&T investments improve short- and long-term prosperity? Where will advances in knowledge also yield advances in skills and technology (5)? (v) Employers: what skills are needed next year and in 5 and 10 years? Which institutions produce the right talent? What skills are listed in job advertisements by my competition? How do I hire and train

Education has been a critical vehicle of economic growth and social progress throughout the modern era. Higher education

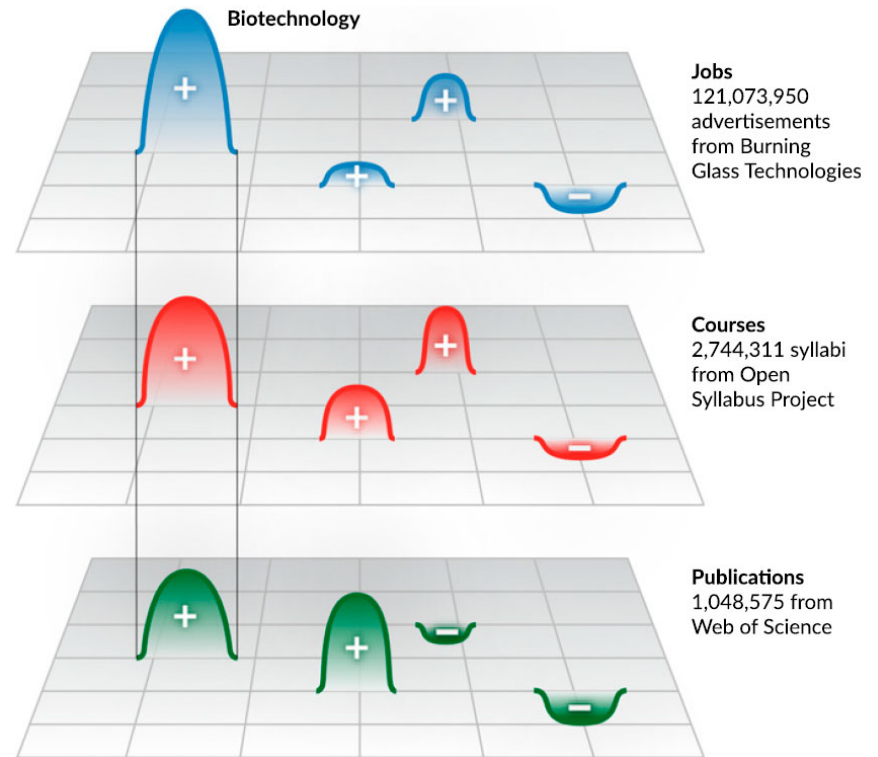
This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Modeling and Visualizing Science and Technology Developments," held December 4–5, 2017, at the Arnold and Mabel Beckman Center of the National Academies of



Study the (mis)match and temporal dynamics of S&T progress, education and workforce development options, and job requirements.

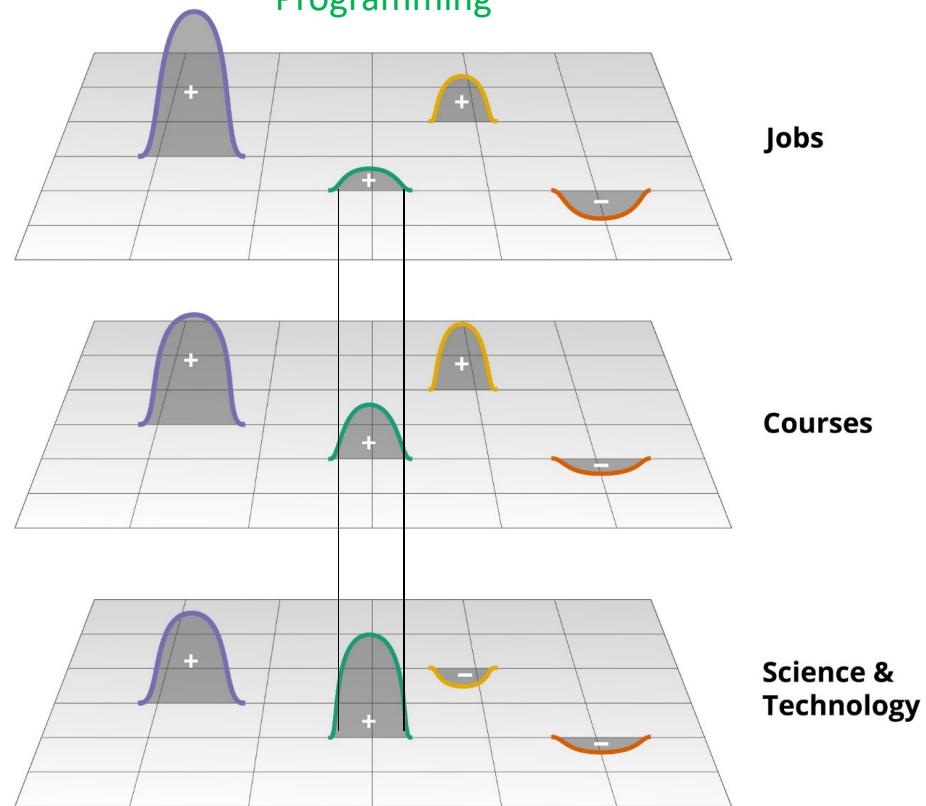
**Challenges:**

- Rapid change of STEM knowledge
- Increase in tools, AI
- Social skills (project management, team leadership)
- Increasing team size



**Fig. 1.** The interplay of job market demands, educational course offerings, and progress in S&T as captured in publications. Color-coded mountains (+) and valleys (-) indicate different skill clusters. For example, skills related to Biotechnology might be mentioned frequently in job descriptions and taught in many courses, but they may not be as prevalent in academic publications. In other words, there are papers that mention these skills, but labor demand and commercial activity might be outstripping publication activity in this area. The numbers of jobs, courses, and publications that have skills associated and are used in this study are given on the right.

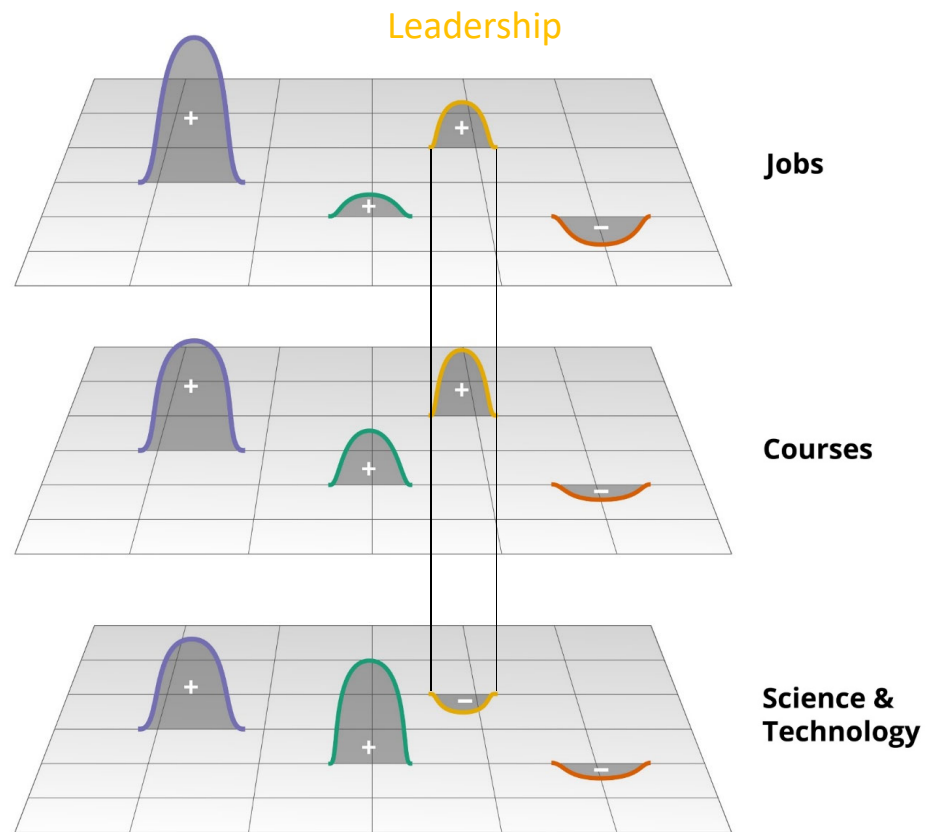
## Programming



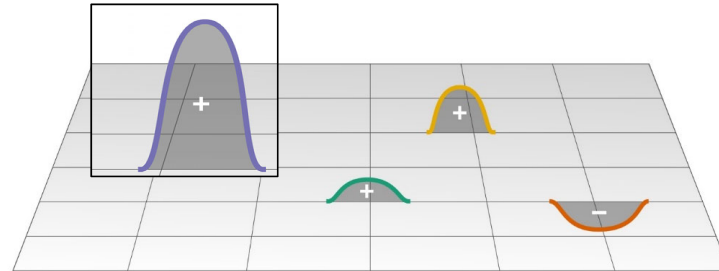
**Jobs**

**Courses**

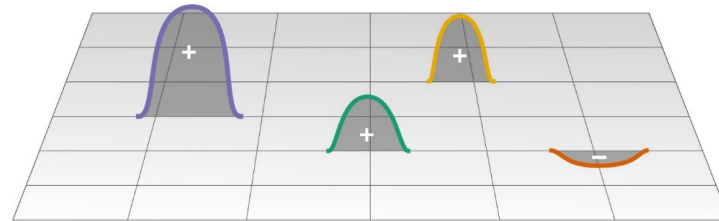
**Science &  
Technology**



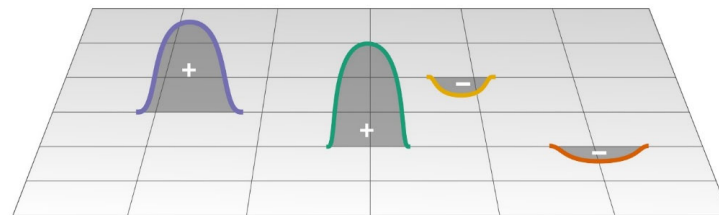
### Biotechnology



**Jobs**

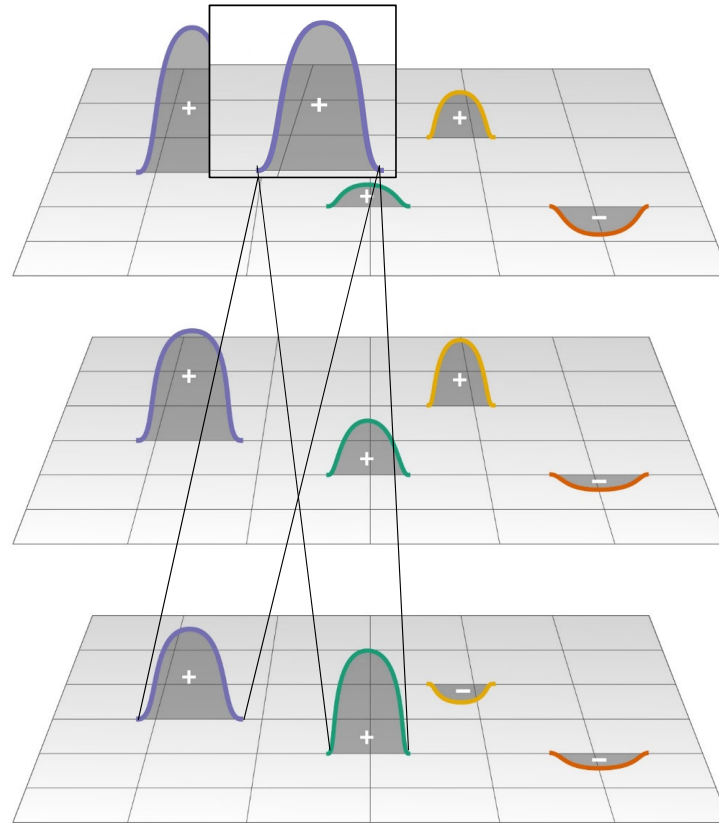


**Courses**



**Science & Technology**

### Biotechnology



**Jobs**

**Courses**

**Science &  
Technology**

# Stakeholders and Insight Needs

- **Students:** What jobs will exist in 1-4 years? What program/learning trajectory is best to get/keep my dream job?
- **Teachers:** What course updates are needed? What balance of timely and timeless knowledge (to get a job vs. learn how to learn) should I teach? How to innovate in teaching and maintain job security or tenure?
- **Universities:** What programs should be created? What is my competition doing? How do I tailor programs to fit local needs?
- **Science Funders:** How can S&T investments improve short- and long-term prosperity? Where will advances in knowledge also yield advances in skills and technology?
- **Employers:** What skills are needed next year and in 5 and 10 years? Which institutions produce the right talent? What skills does my competition list in job advertisements?
- **Economic Developers:** What critical skills are needed to improve business retention, expansion, and recruitment in a region?

**What is ROI of my time, money, compassion?**

# Urgency

- 35% of UK jobs, and 30% in London, are at high risk from automation over the coming 20 years.  
<https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/uk-futures/london-futures-agiletown.pdf>
- The aerospace industry and NASA have a disproportionately large percentage of workers aged 50 and older compared to the national average, and up to **half of the current workforce** will be eligible for retirement within the coming five years.  
Astronautics AIAA (2012) Recruiting, retaining, and developing a world-class aerospace workforce.  
[https://www.aiaa.org/uploadedFiles/Issues\\_and\\_Advocacy/Education\\_and\\_Workforce/Aerospace%20Workforce-%20030112.pdf](https://www.aiaa.org/uploadedFiles/Issues_and_Advocacy/Education_and_Workforce/Aerospace%20Workforce-%20030112.pdf)
- The rise of artificial intelligence will lead to the displacement of **millions of blue-collar as well as white-collar jobs** in the coming decade. Auerswald PE (2017) The Code Economy: A Forty-thousand-year History; Beyer D (2016) The future of machine intelligence: Perspectives from leading practitioners ; Brynjolfsson E, McAfee A (2014) The second machine age: Work, progress, and prosperity in a time of brilliant technologies; Ford M (2015) Rise of the Robots: Technology and the Threat of a Jobless Future.

# Skill Discrepancies Between Research, Education, and Jobs Reveal the Critical Need to Supply Soft Skills for the Data Economy

- Data and Crosswalks
- MaxMatch for NLP
- Causal Analyses
- Visualizations

Börner, Katy, Olga Scrivner, Mike Gallant, Shutian Ma, Xiaozhong Liu, Keith Chewning, Lingfei Wu, and James A. Evans. 2018. "Skill Discrepancies Between Research, Education, and Jobs Reveal the Critical Need to Supply Soft Skills for the Data Economy." *PNAS* 115(50): 12630-12637.

## Skill discrepancies between research, education, and jobs reveal the critical need to supply soft skills for the data economy

Katy Börner<sup>a,b,1</sup>, Olga Scrivner<sup>a</sup>, Mike Gallant<sup>a</sup>, Shutian Ma<sup>a,c</sup>, Xiaozhong Liu<sup>a</sup>, Keith Chewning<sup>d</sup>, Lingfei Wu<sup>e,f,g,h</sup>, and James A. Evans<sup>i,g,i,1</sup>

<sup>a</sup>School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN 47408; <sup>b</sup>Educational Technology/Media Centre, Dresden University of Technology, 01062 Dresden, Germany; <sup>c</sup>Department of Information Management, Nanjing University of Science and Technology, 210094 Nanjing, China; <sup>d</sup>Burning Glass Technologies, Boston, MA 02110; <sup>e</sup>School of Journalism and Communication, Nanjing University, 210008 Nanjing, China; <sup>f</sup>Department of Sociology, University of Chicago, Chicago, IL 60637; <sup>g</sup>Knowledge Lab, University of Chicago, Chicago, IL 60637; <sup>h</sup>Tencent Research Institute, 100080 Beijing, China; and <sup>i</sup>Santa Fe Institute, Santa Fe, NM 87501

Edited by William B. Rouse, Stevens Institute of Technology, Hoboken, NJ, and accepted by Editorial Board Member Pablo G. DeBenedetti September 12, 2018 (received for review March 14, 2018)

Rapid research progress in science and technology (S&T) and continuously shifting workforce needs exert pressure on each other and on the educational and training systems that link them. Higher education institutions aim to equip new generations of students with skills and expertise relevant to workforce participation for decades to come, but their offerings sometimes misalign with commercial needs and new techniques forged at the frontiers of research. Here, we analyze and visualize the dynamic skill (mis-)alignment between academic push, industry pull, and educational offerings, paying special attention to the rapidly emerging areas of data science and data engineering (DS/DE). The visualizations and computational models presented here can help key decision makers understand the evolving structure of skills so that they can craft educational programs that serve workforce needs. Our study uses millions of publications, course syllabi, and job advertisements published between 2010 and 2016. We show how courses mediate between research and jobs. We also discover responsiveness in the academic, educational, and industrial system in how skill demands from industry are as likely to drive skill attention in research as the converse. Finally, we reveal the increasing importance of uniquely human skills, such as communication, negotiation, and persuasion. These skills are currently underexamined in research and undersupplied through education for the labor market. In an increasingly data-driven economy, the demand for "soft" social skills, like teamwork and communication, increase with greater demand for "hard" technical skills and tools.

science of science | job market | data mining | visualization | market gap analysis

doors. Some predictions say hundreds or even thousands of colleges and universities will close or merge in the coming years (4). In addition, there seem to be major discrepancies and delays between leading scientific research, job market needs, and educational content. This has been particularly expressed with respect to science, technology, engineering, and mathematics jobs, where scientific and technological progress is rapid. Strategic decision making on what to teach, whom to hire, and what new research to fund benefits from a systematic analysis of the interplay between science and technology (S&T) developments, courses and degrees offered, and job market needs. Specifically, stakeholders in US higher education urgently need answers to the following questions. (i) Students: what jobs might exist in 5–10 years? What educational trajectories will best achieve my dream job? What core and specialized skills are required for what jobs and offered by what schools and programs? (ii) Teachers: what course updates are most needed? What balance of timely vs. timeless knowledge should I teach? How can I innovate in teaching and maintain job security or tenure? (iii) Universities: what programs should be created? What is my competition doing? How do I tailor programs to fit workforce needs? (iv) Science funders: how can S&T investments improve short- and long-term prosperity? Where will advances in knowledge also yield advances in skills and technology (5)? (v) Employers: what skills are needed next year and in 5 and 10 years? Which institutions produce the right talent? What skills are listed in job advertisements by my competition? How do I hire and train

Education has been a critical vehicle of economic growth and social progress throughout the modern era. Higher education

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Modeling and Visualizing Science and Technology Developments," held December 4–6, 2017, at the Arnold and Mabel Beckman Center of the National Academies of

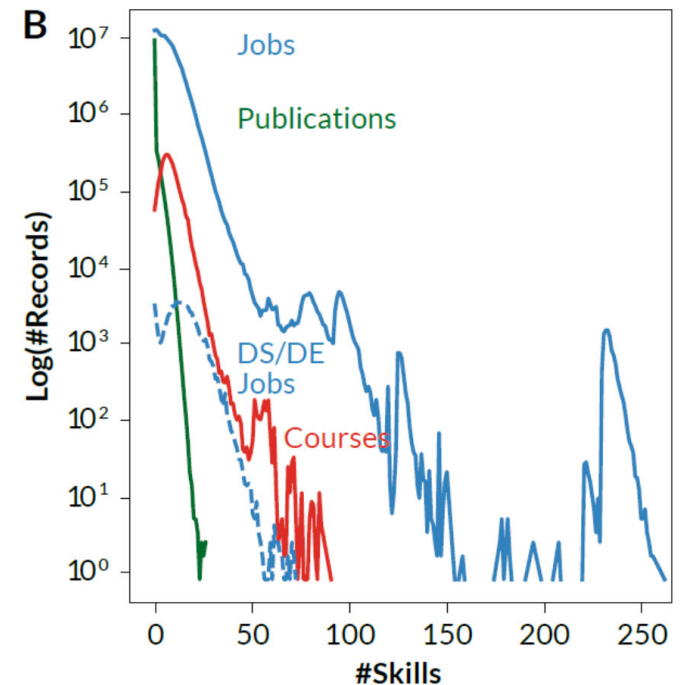
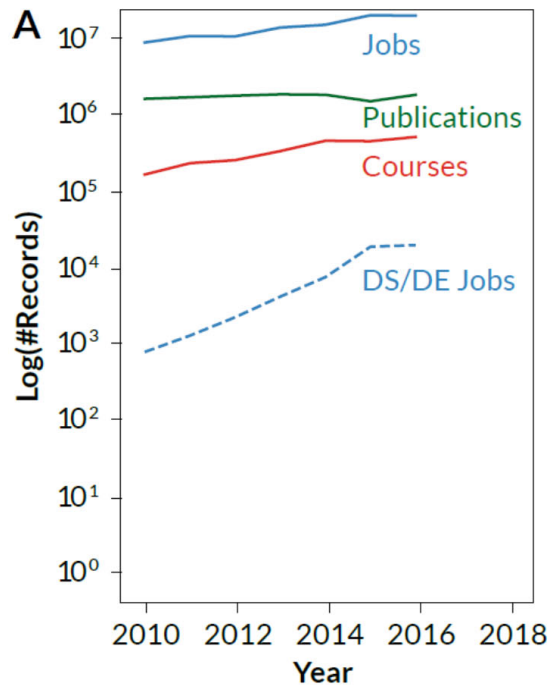


# Datasets Used

Job advertisements by Burning Glass posted between Jan 2010-Dec 2016.

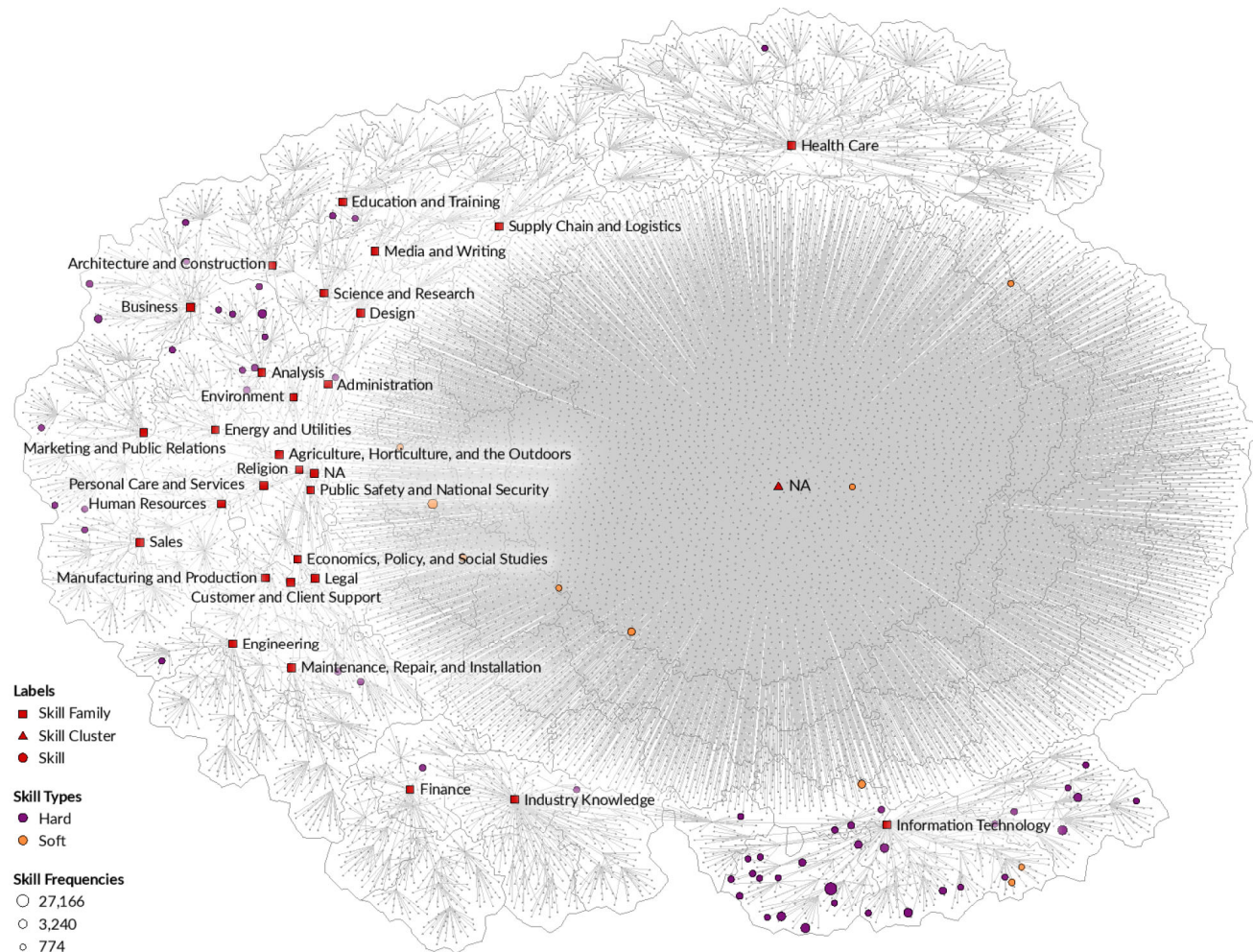
Web of Science publications published Jan 2010-Dec 2016.

Course descriptions from the Open Syllabus Project acquired in June 2018 for courses offered in 2010-2016.



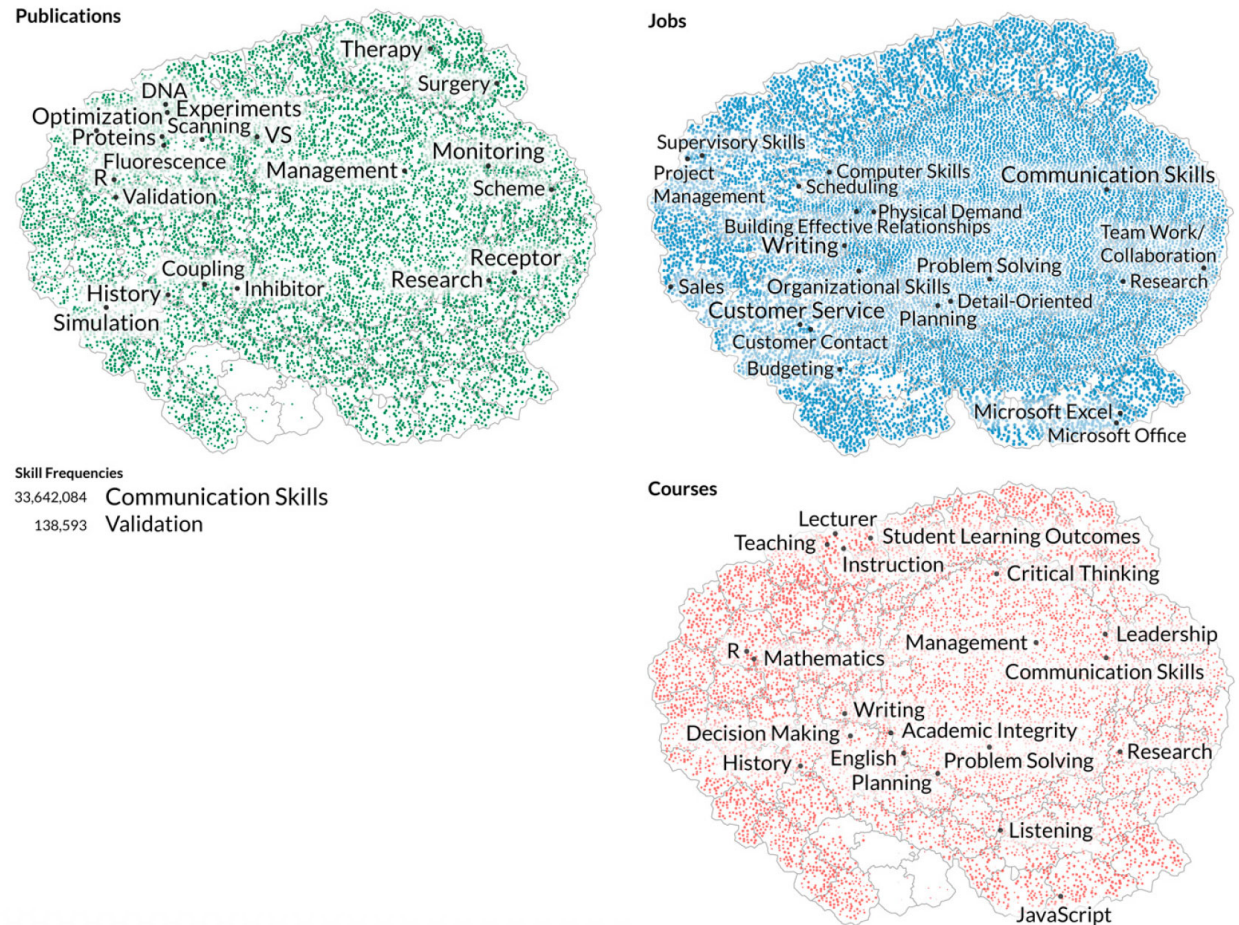
Data Type	#Records	#Records with skills	#Records without skills
All Courses	3,062,277	2,744,311	54,733
All Jobs	132,011,926	121,073,950	10,937,976
DSDE Jobs	69,405	65,944	3,461
All Publications	15,691,162	1,048,575	14,642,587
DSDE Publications	1,048,575	807,756	240,819

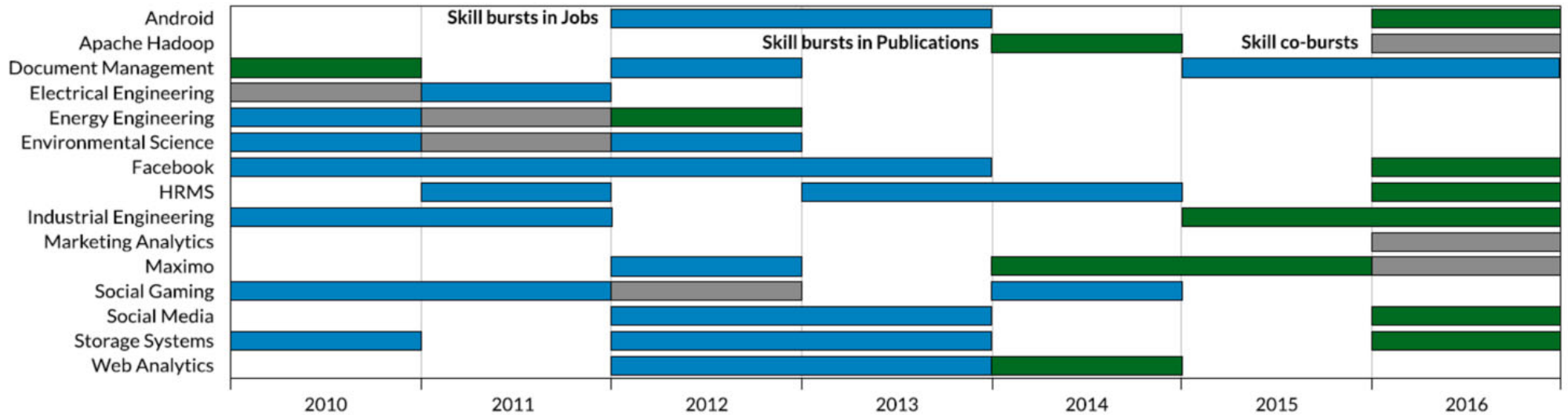
**Fig. 2.** Basemap of 13,218 skills. In this map, each dot is a skill, triangles identify skill clusters, and squares represent skill families from the Burning Glass (BG) taxonomy. Labels are given for all skill family nodes and for the largest skill cluster (NA) to indicate placement of relevant subtrees. Additionally, hard and soft skills are overlaid using purple and orange nodes, respectively; node area size coding indicates base 10 log of skill frequency in DS/DE jobs. Skill area computation uses Voronoi tessellation.





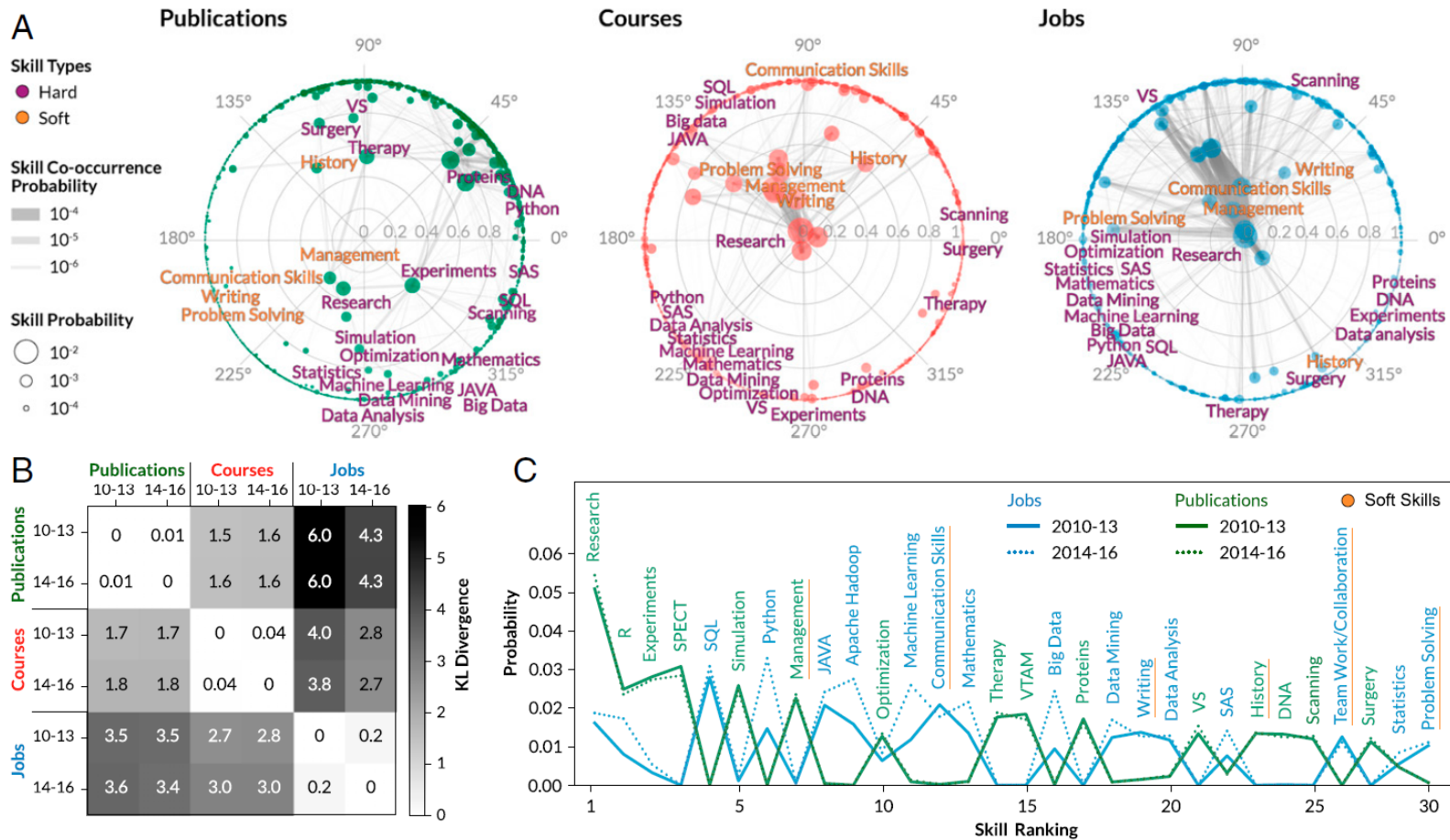
**Fig. 3.** Basemap of 13,218 skills with overlays of skill frequency in jobs, courses, and publications. This figure substantiates the conceptual drawing in Fig. 1 using millions of data records. Jobs skills are plotted in blue, courses are in red, and publications are in green. Node area size coding indicates base 10 log of skills frequency. The top 20 most frequent skills are labeled, and label sizes denote skill frequency.





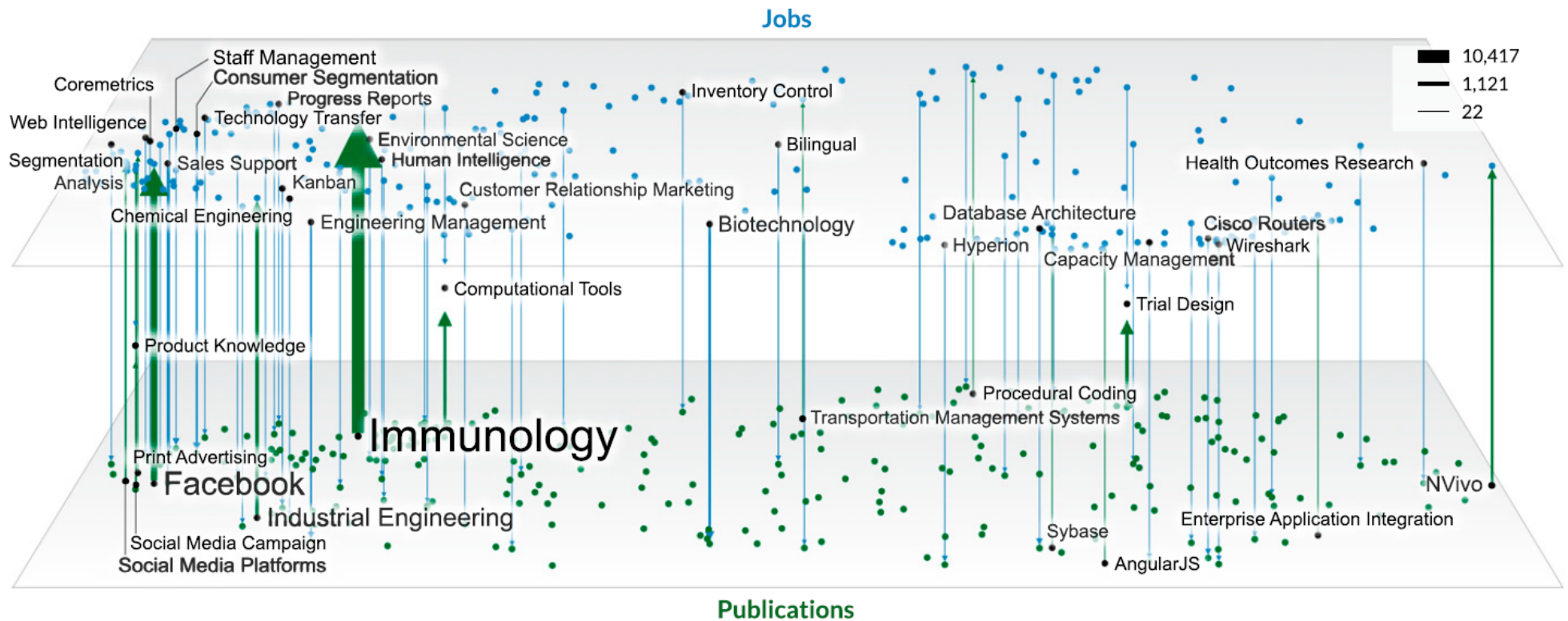
**Fig. 4.** Burst of activity in DS/DE skills in jobs and publications. Each burst is rendered as a horizontal bar with a start and an end date; skill term is shown on the left. Skills that burst in jobs are blue; skills bursting in publications are green. Seven skills burst in both datasets during the same years and are shown in gray. HRMS stands for human resources management system, and Maximo is an IBM system for managing physical assets.

# Kullback-Leibler divergence



**Fig. 5.** Structural and dynamic differences between skill distributions in jobs, courses, and publications for 2010–2013 and 2014–2016. (A) Poincaré disks comparing the centrality of soft skills (orange) and hard skills (purple) across jobs, courses, and publications. (B) KL divergence matrix for jobs, courses, and publications in 2010–2013 and 2014–2016. (C) The most surprising skills in publications and jobs; *R* is a scripting language, VTAM refers to the IBM Virtual Telecommunication Access Method application, VS is the integrated development environment Visual Studio, and SAS is a data analytics software.

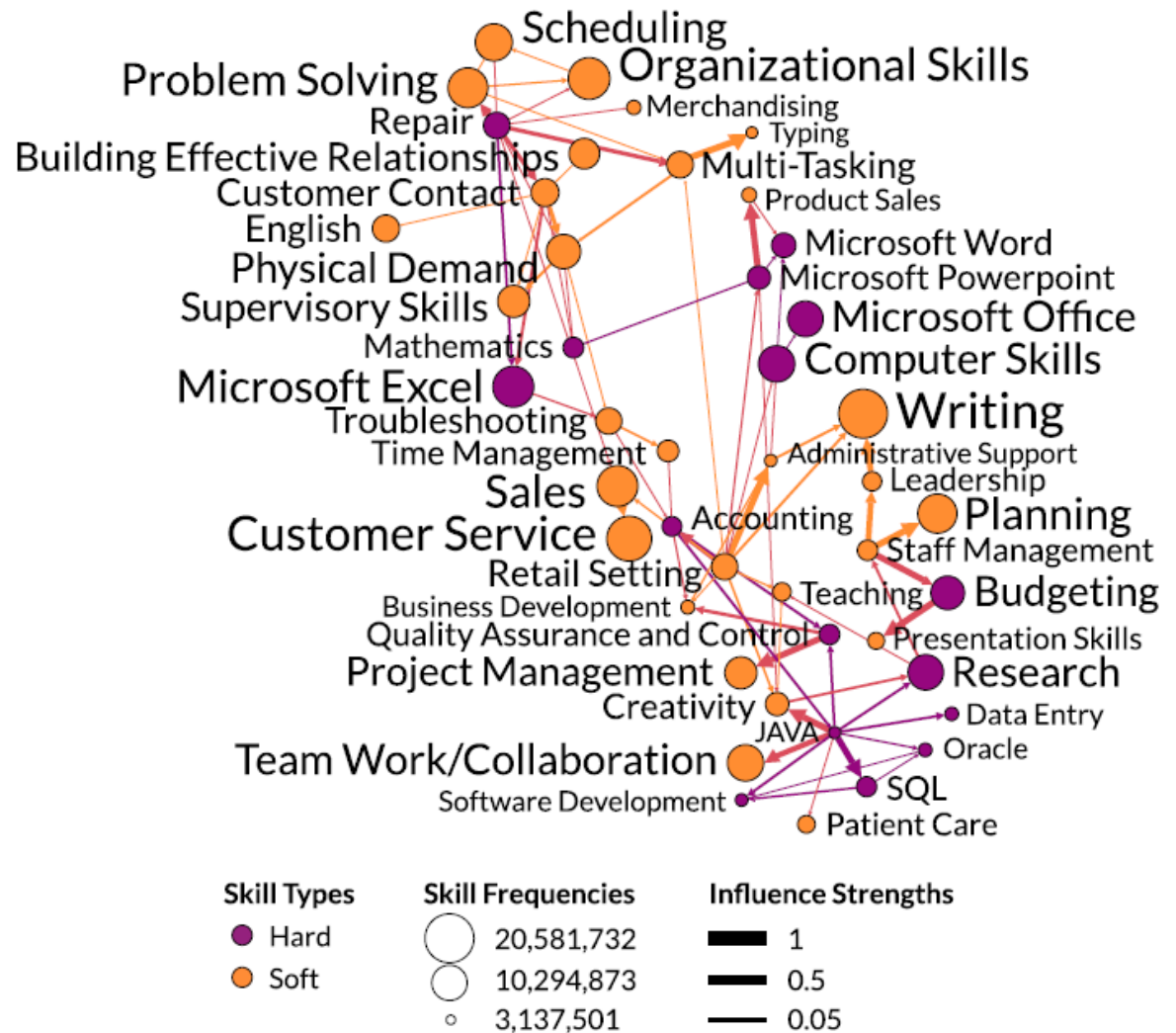




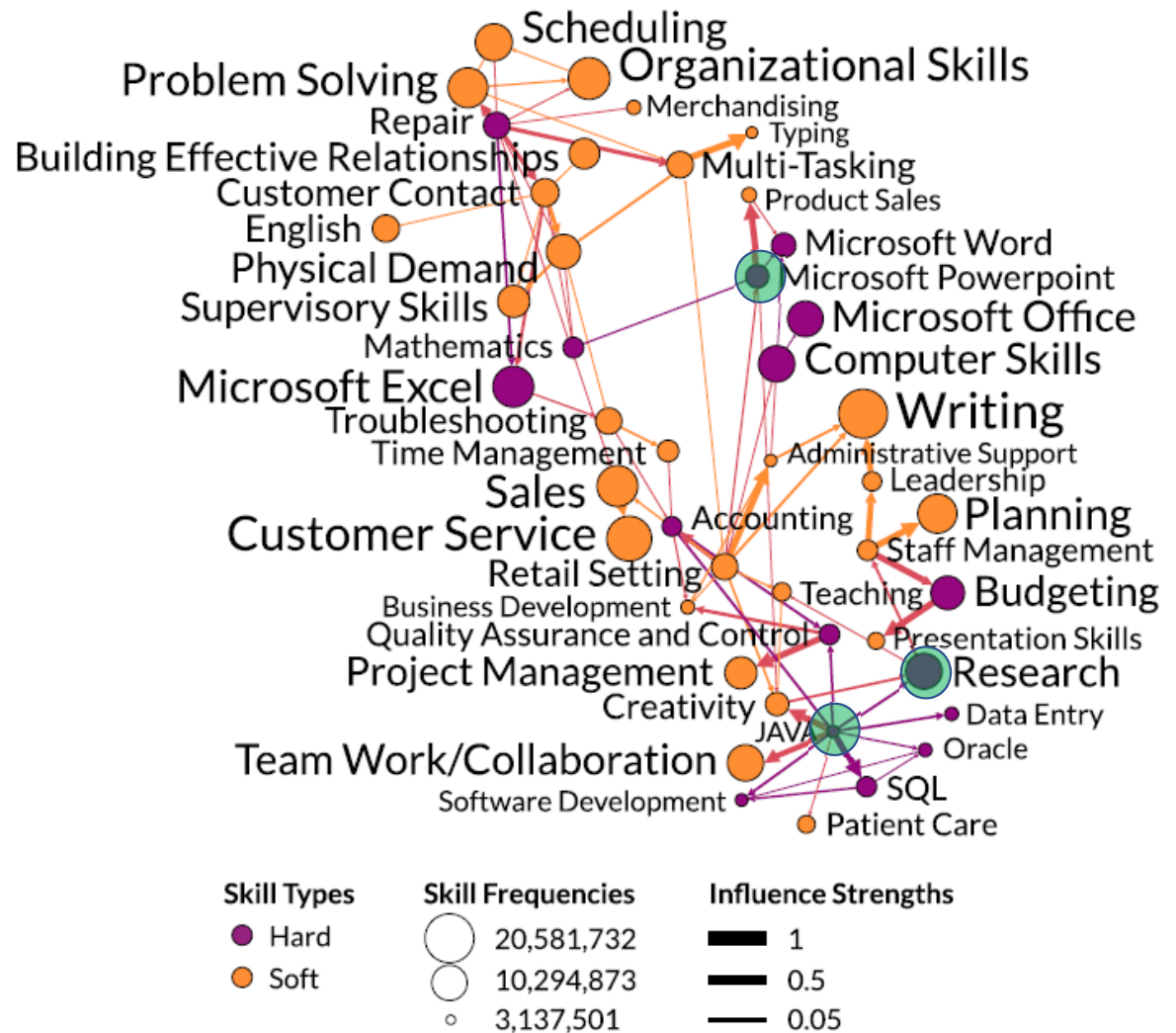
**Fig. 6.** Strength of influence mapping. Top 200 most frequent skills in jobs (blue) and in publications (green) plotted on the skills basemap from Fig. 2. Arrows represent skills with significant Granger causality ( $P$  value  $< 0.05$ ). Line thickness and label size indicate skill frequency. The direction and thickness of each arrow indicate the  $F$ -value strength and direction.



**Fig. 7.** Multivariate Hawkes Process influence network of DS/DE skills within job advertisements 2010–2016. Each of the 45 nodes represents a top-frequency skill (29 soft and 16 hard skills) with a strong influence edge from/to other skill(s) in job advertisements between 2010 and 2016. Node and label size correspond to the number of times that the skill appeared in a job advertisement. Thickness of the 75 directed edges indicates influence strength.



**Fig. 7.** Hawkes influence network of DS/DE skills within job advertisements 2010–2016. Each of the 45 nodes represents a top-frequency skill (29 soft and 16 hard skills) with a strong influence edge from/to other skill(s) in job advertisements between 2010 and 2016. Node and label size correspond to the number of times that the skill appeared in a job advertisement. Thickness of the 75 directed edges indicates influence strength.



# Results

- Novel cross-walk for mapping publications, course offerings, and job via skills.
- Timing and strength of burst of activity for skills (e.g., Oracle, Customer Service) in publications, course offerings, and job advertisements.
- Uniquely human skills such as communication, negotiation, and complex service provision are currently underexamined in research and undersupplied through education for the labor market in an increasingly automated and AI economy.
- The same pattern manifests in the domain of DS/DE where teamwork and communication skills increase in value with greater demand for data analytics skills and tools.
- Skill demands from industry are as likely to drive skill attention in research as the converse.

# NSF RAISE: C-Accel Pilot - Track B1: Analytics-Driven Accessible Pathways To Impacts-Validated Education (ADAPTIVE)

**Goal:** Development of data-driven tools to support the tens of millions of US workers whose jobs are being transformed by Artificial Intelligence (AI) and automation.

The project will demonstrate how labor market and course syllabi data, learning analytics, and insights on transferability of learned skills can be combined and visualized in novel ways to support a learner's decision-making about, sustained engagement in, and application to their job of professional skills acquired through education and job-related training.

 **Team B-6656:** Katy Börner, Indiana University, Ariel Anbar, Arizona State University, Kemi Jona, Northeastern University, Martin Storksdieck and Heather Fischer, Oregon State University







101st Annual Meeting of the Association of American Geographers, Denver, CO.  
April 5th - 9th, 2005 (First showing of Places & Spaces)



University of Miami, Miami, FL.  
September 4 - December 11, 2014.



Duke University, Durham, NC.  
January 12 - April 10, 2015



The David J. Sencer CDC Museum, Atlanta, GA.  
January 25 - June 17, 2016.

<http://scimaps.org>

# Places & Spaces: Mapping Science Exhibit

1<sup>st</sup> Decade (2005-2014)

## Maps



2<sup>nd</sup> Decade (2015-2024)

## Macroscopes

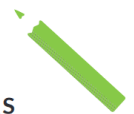


100

MAPS  
in large format, full color, and high resolution.

248

MAPMAKERS  
from fields as disparate as art, urban planning, engineering, and the history of science.



43



MACROSCOPE MAKERS  
including one whose job title is "Truth and Beauty Operator."

20

MACROSCOPES  
for touching all kinds of data.

382

DISPLAY VENUES  
from the Cannes Film Festival to the World Economic Forum.

354

PRESS ITEMS  
including articles in *Nature*, *Science*, *USA Today*, and *Wired*.



<http://scimaps.org>



# Map of Scientific Collaborations from 2005-2009



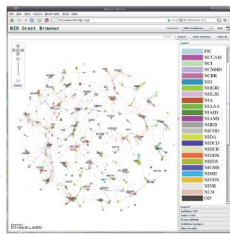
Computed Using Data from Elsevier's Scopus

# A Topic Map of NIH Grants 2007

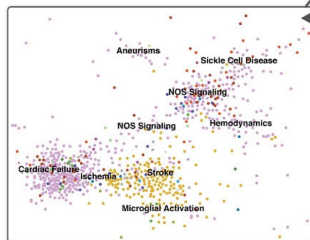
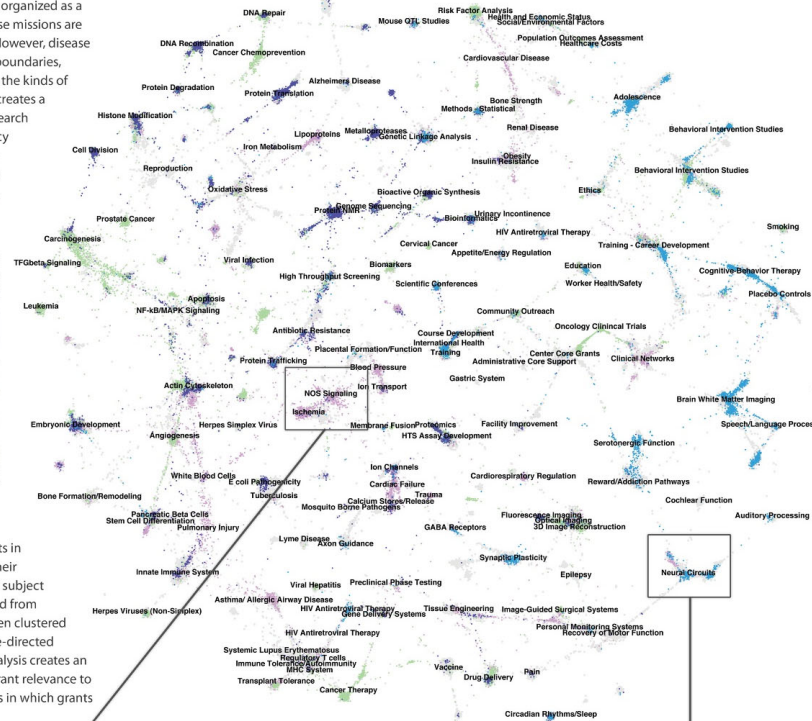
ChalkLabs UCIrvine

Bruce W. Herr II (Chalklabs & IU), Gully Burns (ISI), David Newman (UCI), Edmund Talley (NIH)

The National Institutes of Health (NIH) is organized as a multitude of Institutes and Centers whose missions are primarily focused on distinct diseases. However, disease etiologies and therapies flout scientific boundaries, and thus there is tremendous overlap in the kinds of research funded by each institute. This creates a daunting landscape for decisions on research directions, funding allocations, and policy formulations. Shown here is devised an interactive topic map for navigating this landscape, online at [www.nihmaps.org](http://www.nihmaps.org). Institute abbreviations can be found at [www.nih.gov/icd](http://www.nih.gov/icd).

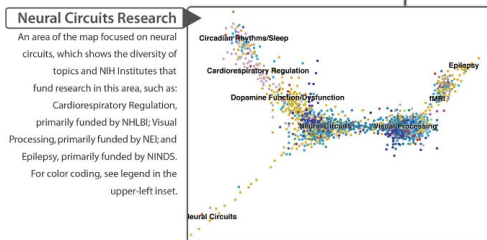


Topic modeling, a statistical technique that automatically learns semantic categories, was applied to assess projects in terms used by researchers to describe their work, without the biases of keywords or subject headings. Grant similarities were derived from their topic mixtures, and grants were then clustered on a two-dimensional map using a force-directed simulated annealing algorithm. This analysis creates an interactive environment for assessing grant relevance to research categories and to NIH Institutes in which grants are localized.



### Cardiac Diseases Research

An area of the map focused on cardiovascular function and dysfunction. Cardiac Failure (primarily funded by NHLBI) is typically clustered next to Stroke (NINDS), since these are the two major medical emergencies associated with ischemia, which results from a restricted blood supply. Also localized in this area are grants focused on Nitric Oxide (NOS) Signaling, a major biochemical pathway for vasodilation, and grants on Hemodynamics, Sickle Cell Disease, and Aneurysms.



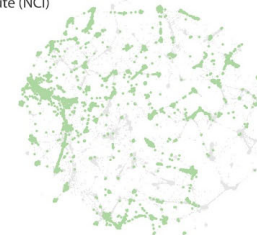
### Neural Circuits Research

An area of the map focused on neural circuits, which shows the diversity of topics and NIH Institutes that fund research in this area, such as: Cardiorespiratory Regulation, primarily funded by NHLBI; Visual Processing, primarily funded by NEI; and Epilepsy, primarily funded by NINDS. For color coding, see legend in the upper-left inset.

### National Cancer Institute (NCI)

#### TOP 10 TOPICS

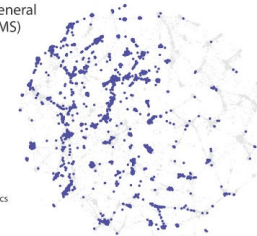
- 1 Oncology Clinical Trials
- 2 Cancer Treatment
- 3 Cancer Therapy
- 4 Carcinogenesis
- 5 Risk Factor Analysis
- 6 Cancer Chemotherapy
- 7 Metastasis
- 8 Leukemia
- 9 Prediction/Prognosis
- 10 Cancer Chemoprevention



### National Institute of General Medical Sciences (NIGMS)

#### TOP 10 TOPICS

- 1 Bioactive Organic Synthesis
- 2 X-ray Crystallography
- 3 Protein NMR
- 4 Computational Models
- 5 Yeast Biology
- 6 Metalloproteases
- 7 Enzymatic Mechanisms
- 8 Protein Complexes
- 9 Invertebrate/Zebrafish Genetics
- 10 Cell Division



### National Heart, Lung, and Blood Institute (NHLBI)

#### TOP 10 TOPICS

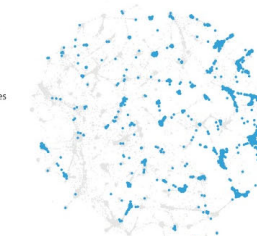
- 1 Cardiac Failure
- 2 Pulmonary Injury
- 3 Genetic Linkage Analysis
- 4 Cardiovascular Disease
- 5 Atherosclerosis
- 6 Hemostasis
- 7 Blood Pressure
- 8 Asthma/ Allergic Airway Disease
- 9 Gene Association
- 10 Lipoproteins



### National Institute of Mental Health (NIMH)

#### TOP 10 TOPICS

- 1 Mood Disorders
- 2 Schizophrenia
- 3 Behavioral Intervention Studies
- 4 Mental Health
- 5 Depression
- 6 Cognitive-Behavior Therapy
- 7 AIDS Prevention
- 8 Genetic Linkage Analysis
- 9 Adolescence
- 10 Childhood





# The Structure of Science

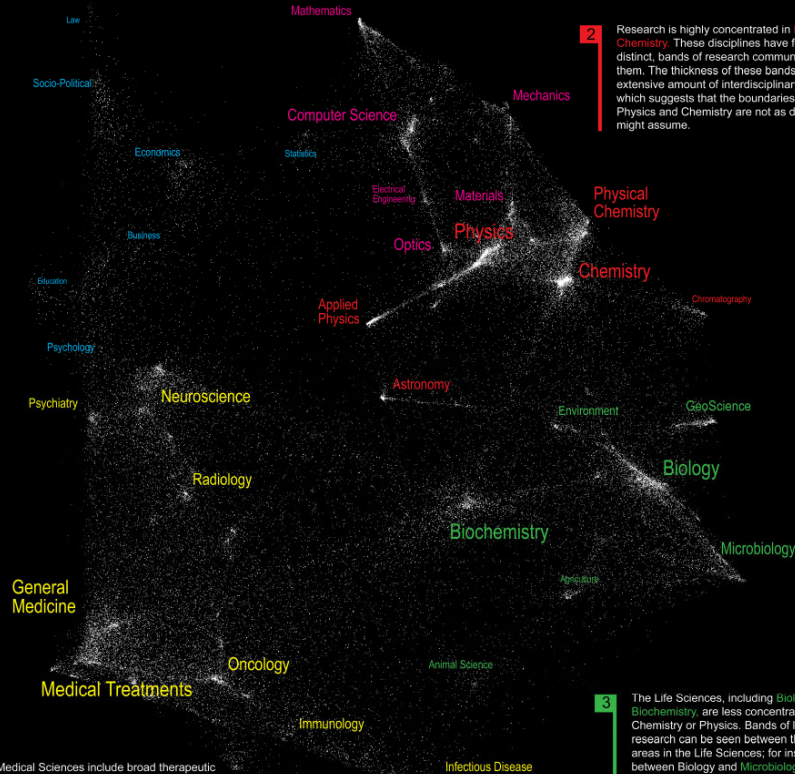
**5** The Social Sciences are the smallest and most diffuse of all the sciences. **Psychology** serves as the link between Medical Sciences (Psychiatry) and the Social Sciences. **Statistics** serves as the link with Computer Science and Mathematics.

**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.

**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.

**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.



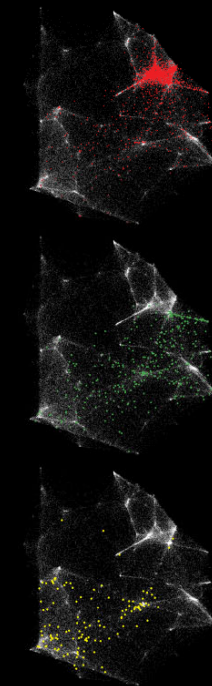
We are all familiar with traditional maps that show the relationships between countries, provinces, states, and cities. Similar relationships exist between the various disciplines and research topics in science. This allows us to map the structure of science.

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

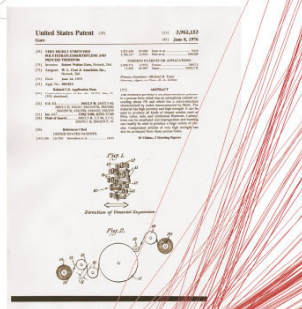
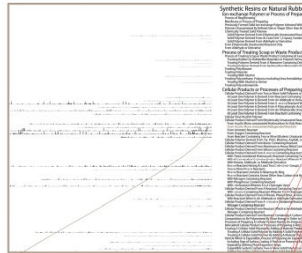
Pharmacogenomics is a relatively new field with most of its activity in **Medicine**. It also has many communities in **Biochemistry** and two communities in the Social Sciences.

## Impact

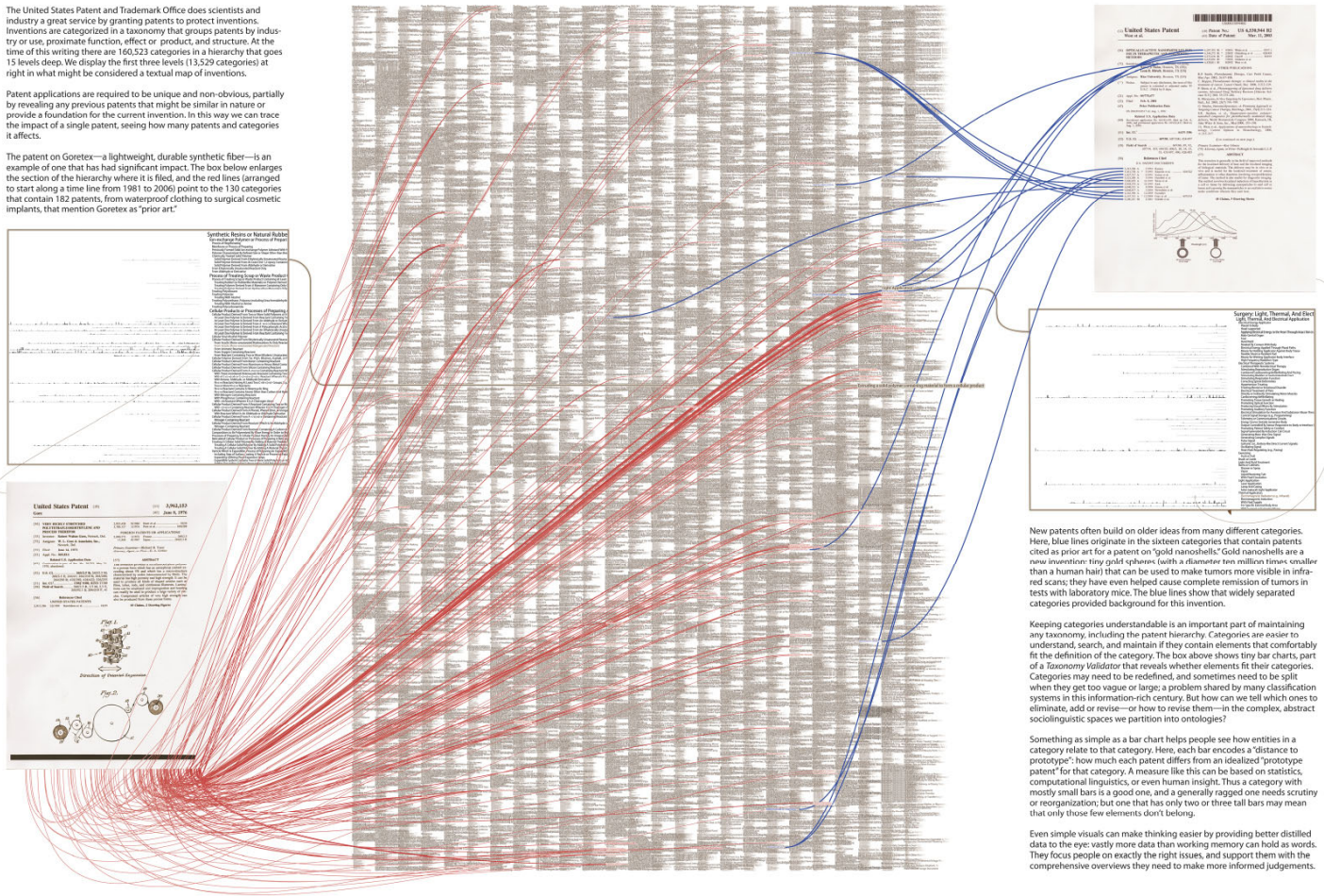
The United States Patent and Trademark Office does scientists and industry a great service by granting patents to protect inventions. Inventions are categorized in a taxonomy that groups patents by industry or use, proximate function, effect or product, and structure. At the time of this writing there are 160,523 categories in a hierarchy that goes 15 levels deep. We display the first three levels (13,529 categories) at right in what might be considered a textual map of inventions.

Patent applications are required to be unique and non-obvious, partially by revealing any previous patents that might be similar in nature or provide a foundation for the current invention. In this way we can trace the impact of a single patent, seeing how many patents and categories it affects.

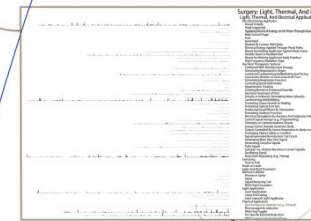
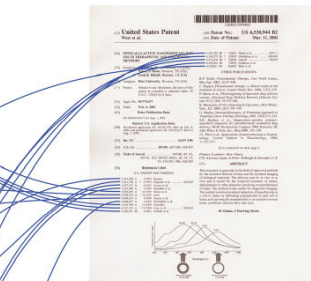
The patent on Goretex—a lightweight, durable synthetic fiber—is an example of one that has had significant impact. The box below enlarges the section of the hierarchy where it is filed, and the red lines (arranged to start along a time line from 1981 to 2006) point to the 130 categories that contain 182 patents, from waterproof clothing to surgical cosmetic implants, that mention Goretex as “prior art.”



# The US Patent Hierarchy



## Prior Art



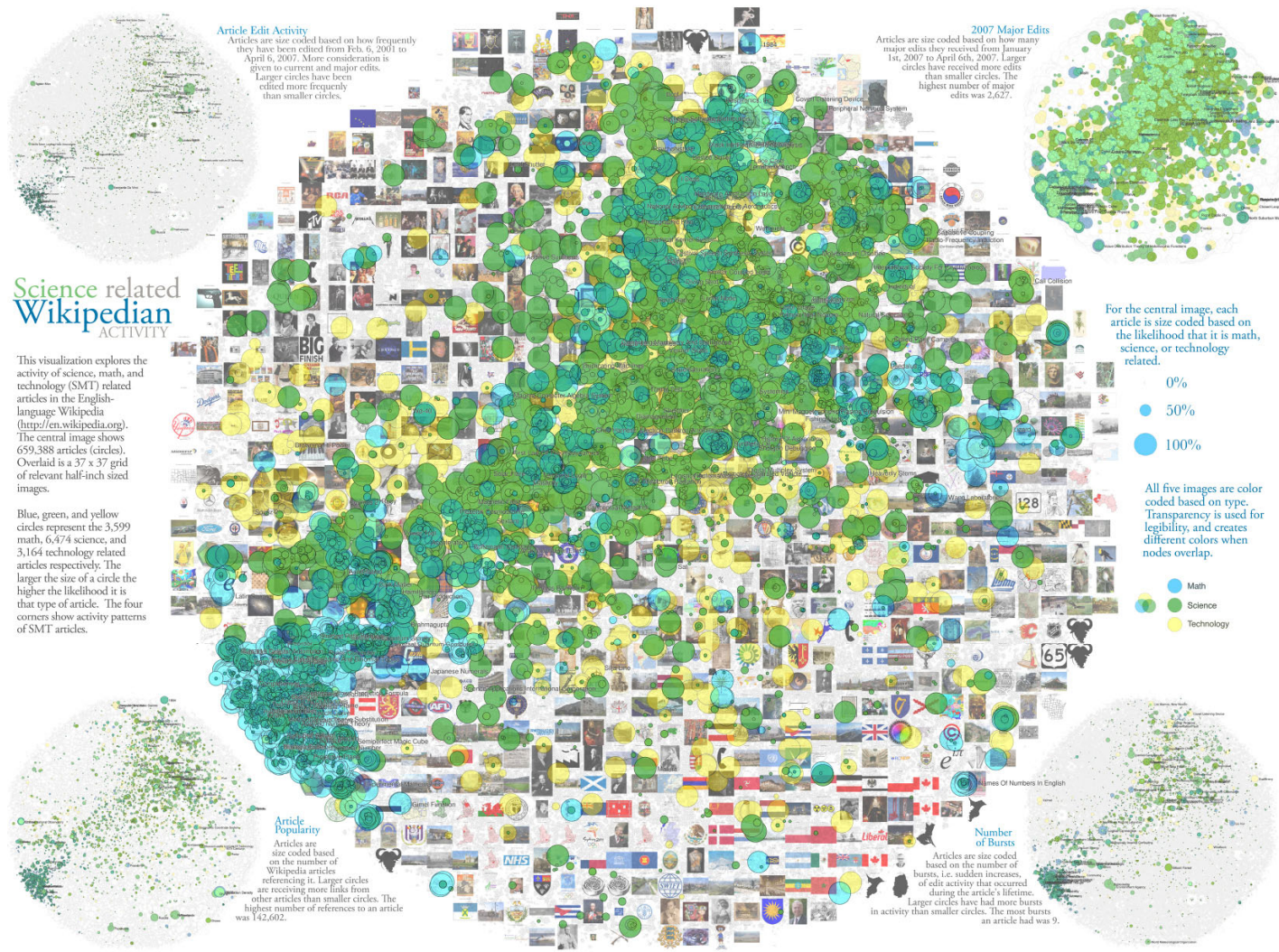
New patents often build on older ideas from many different categories. Here, blue lines originate in the sixteen categories that contain patents cited as prior art for a patent on “gold nanoshells.” Gold nanoshells are a new invention: tiny gold spheres (with a diameter ten million times smaller than a human hair) that can be used to make tumors more visible in infrared scans; they have even helped cause complete remission of tumors in tests with laboratory mice. The blue lines show that widely separated categories provided background for this invention.

Keeping categories understandable is an important part of maintaining any taxonomy, including the patent hierarchy. Categories are easier to understand, search, and maintain if they contain elements that comfortably fit the definition of the category. The box above shows tiny bar charts, part of a Taxonomy Validator that reveals whether elements fit their categories. Categories may need to be redefined, and sometimes need to be split when they get too vague or large; a problem shared by many classification systems in this information-rich century. But how can we tell which ones to eliminate, add or revise—or how to revise them—in the complex, abstract sociolinguistic spaces we partition into ontologies?

Something as simple as a bar chart helps people see how entities in a category relate to that category. Here, each bar encodes a “distance to prototype”: how much each patent differs from an idealized “prototype patent” for that category. A measure like this can be based on statistics, computational linguistics, or even human insight. Thus a category with mostly small bars is a good one, and a generally ragged one needs scrutiny or reorganization; but one that has only two or three tall bars may mean that only those few elements don’t belong.

Even simple visuals can make thinking easier by providing better distilled data to the eye: vastly more data than working memory can hold as words. They focus people on exactly the right issues, and support them with the comprehensive overviews they need to make more informed judgments.





III.8 Science-Related Wikipedian Activity - Bruce W. Herr II, Todd M. Holloway, Elisha F. Hardy, Katy Börner, and Kevin Boyack - 2007



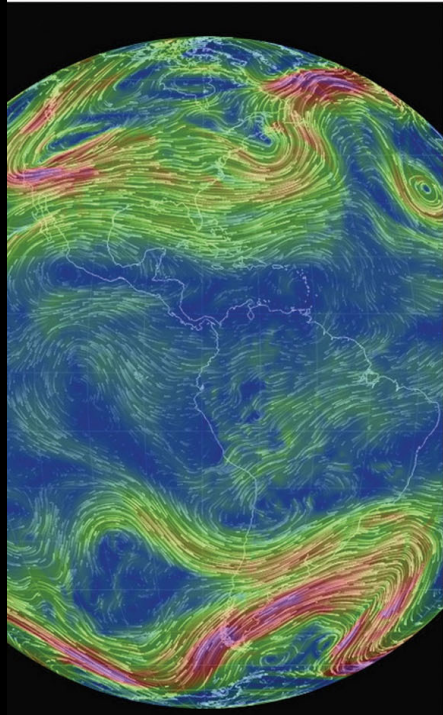








# MACROSCOPES FOR INTERACTING WITH SCIENCE



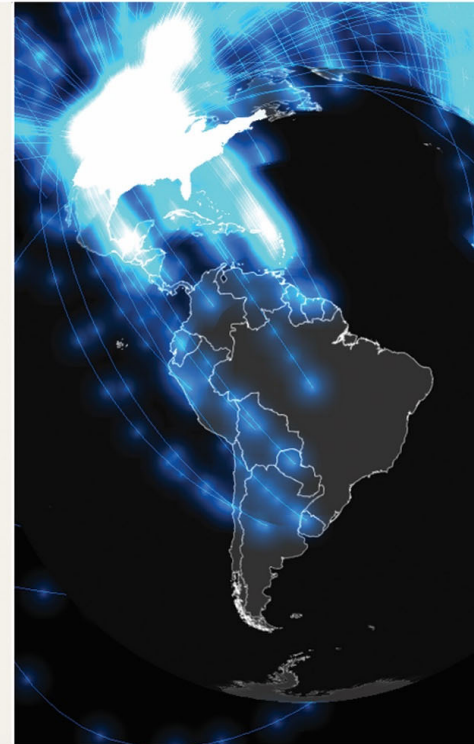
## Earth

*Weather on a worldwide scale*



## AcademyScope

*Exploring the scientific landscape*



## Mapping Global Society

*Local news from a global perspective*



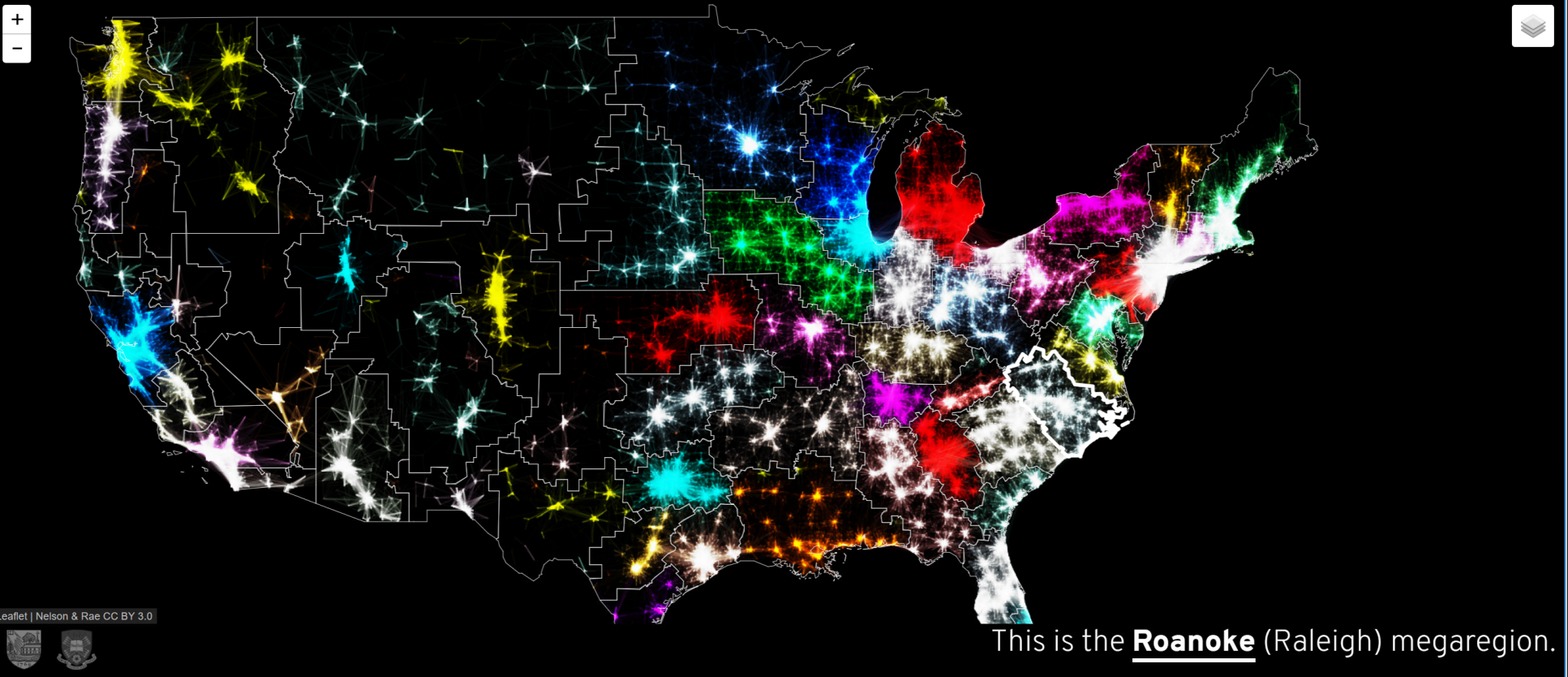
## Charting Culture

*2,600 years of human history in 5 minutes*



## THE MEGAREGIONS OF THE US

Explore the new geography of commuter connections in the US.  
Tap to identify regions. Tap and hold to see a single location's commuted.





SMELLY MAPS

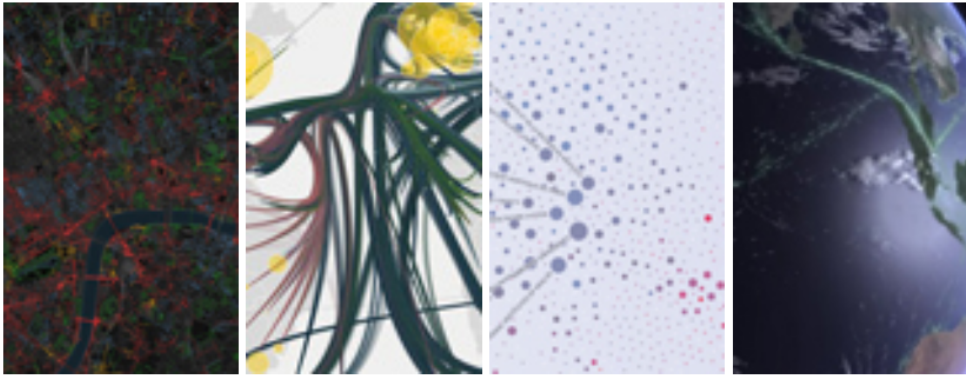


Smelly Maps – Daniele Quercia, Rossano Schifanella, and Luca Maria Aiello – 2015



## Iteration XII (2016)

Macrosopes for Making Sense of Science



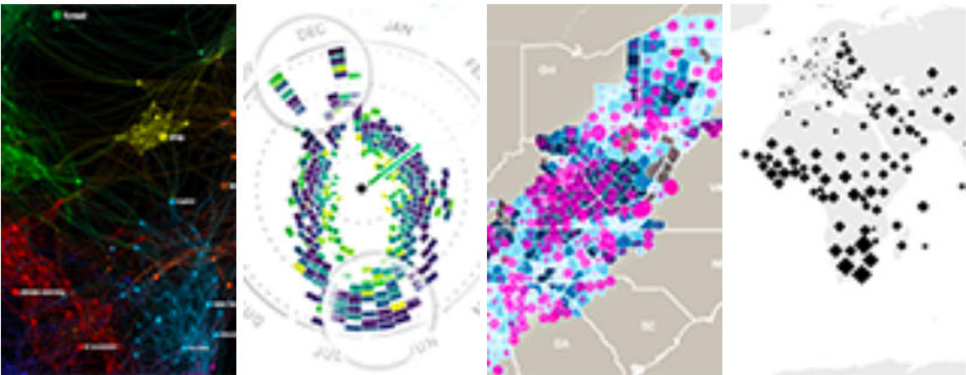
## Iteration XIII (2017)

Macrosopes for Playing with Scale



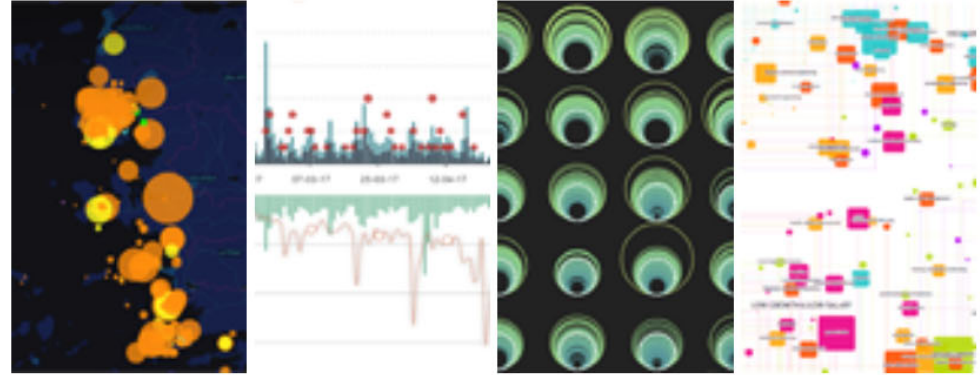
## Iteration XIV (2018)

Macrosopes for Ensuring our Well-being



## Iteration XV (2019)

Macrosopes for Tracking the Flow of Resources





# Acknowledgments

## Exhibit Curators



The exhibit team: Lisel Record, Katy Börner, and Todd Theriault.

<http://scimaps.org>

Plus, we thank the more than 250 authors of the 100 maps and 16 interactive macroscopes.

## Exhibit Advisory Board



**Gary Berg-Cross**  
Cognitive psychologist (PhD, SUNY-Stony Brook). Potomac, MD, USA



**André Skupin**  
Associate Professor of Geography at San Diego State University, California



**Donna J. Cox, MFA, Ph.D.**  
Director of the *Advanced Visualization Laboratory* at the National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, IL, USA



**Moritz Stefaner**  
Freelance designer on the crossroads of data visualization, information aesthetics, and user interface design in Germany



**Bonnie DeVarco**  
*Media X* Distinguished Visiting Scholar at Stanford University, Palo Alto, CA, USA



**Olga Subirós**  
Curator of Big Bang Data and Founder of *Olga Subirós Studio* in Barcelona, Spain



**Peter A. Hook**  
Head of Digital and Scholarly Services and LawArXiv Administrator, Cornell Law Library. Ithaca, NY, USA



**Stephen Uzzo**  
Vice President of Science and Technology for the *New York Hall of Science*



**Francis Harvey**  
Professor of Visual Communication in Geography at the Leibnitz Institute for Regional Geography, Leipzig University, Germany



**Benjamin Wiederkehr**  
Founding Partner and Managing Director of *Interactive Things* in Zürich, Switzerland



**Lev Manovich**  
Professor, *The Graduate Center*, City University of New York; Director, *Software Studies Initiative* (big data, digital humanities, visualization)



# Interdisciplinary Training in Complex Networks and Systems

THE PROGRAM

RESEARCH

HOW TO APPLY

STUDENTS

COLLOQUIA

NEWS



## The program

Share Tweet Email

Understanding complex networked systems is key to solving some of the most vexing problems confronting humankind, from discovering how dynamic brain connections give rise to thoughts and behaviors, to detecting and preventing the spread of misinformation or unhealthy behaviors across a population. Graduate training, however, typically occurs in one of two dimensions: experimental and observational methods in a specific area such as biology and sociology, or in general methodologies such as machine learning and data science.



**CNS NRT**

<https://cns-nrt.indiana.edu>





Indiana University Bloomington will host the  
**International Society of Scientometrics & Informetrics  
Conference (ISSI)**  
in Summer 2023



# 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](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>

