CNS Cyberinfrastructure for Network Science Center

INDIANA UNIVERSITY SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Data Analytics in Support of Effective Workforce Training

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

InnovationDigi Conference
https://innovationdigi.com

November 5, 2020

Thursday, November 5 • 2:30pm - 3:00pm

Back To Schedule

Data Analytics in Support of Effective Workforce Training

Sign up or log in to save this to your schedule, view media, leave feedback and see who's attending!

https://sched.co/eotR Tweet Share

Data Analytics in support of effective Workforce Training/Katy Borner

Speakers

Katy Börner

Data Analytics, Indiana University

KATY BÖRNER is the Victor H. Yngve Distinguished Professor of Engineering and Information Science in the Departments of Intelligent Systems Engineering and Information Science, School of Informatics, Computing, and Engineering; core faculty of the Cognitive Science Program; and founding... Read More →

Thursday November 5, 2020 2:30pm - 3:00pm EST

Virtual

Advanced Analytics

> Feedback form isn't open yet.

Overview

Identify and Overcome Skill Discrepancies

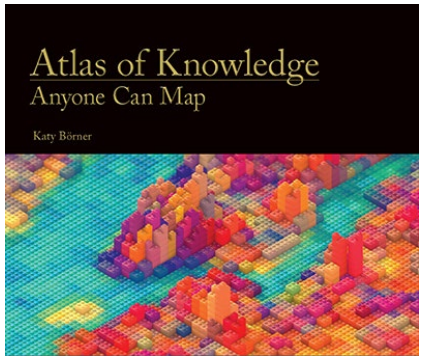
- Börner, Katy, Olga Scrivner, Michael Gallant, Shutian Ma, Xiaozhong Liu, Keith Chewning, Lingfei Wu, and James Evans. 2018. 2018. "Skill Discrepancies Between Research, Education, and Jobs Reveal the Critical Need to Supply Soft Skills for the Data Economy". *PNAS* 115 (50): 12630-12637. doi: 10.1073/pnas.1804247115.

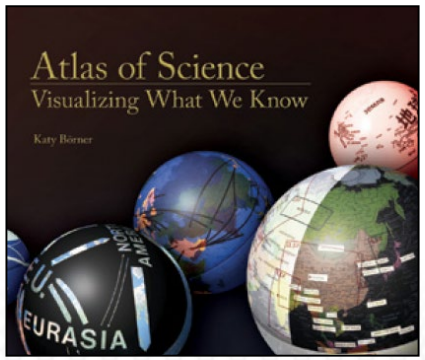
Job Postings in The Substance Use Disorder Treatment Sector

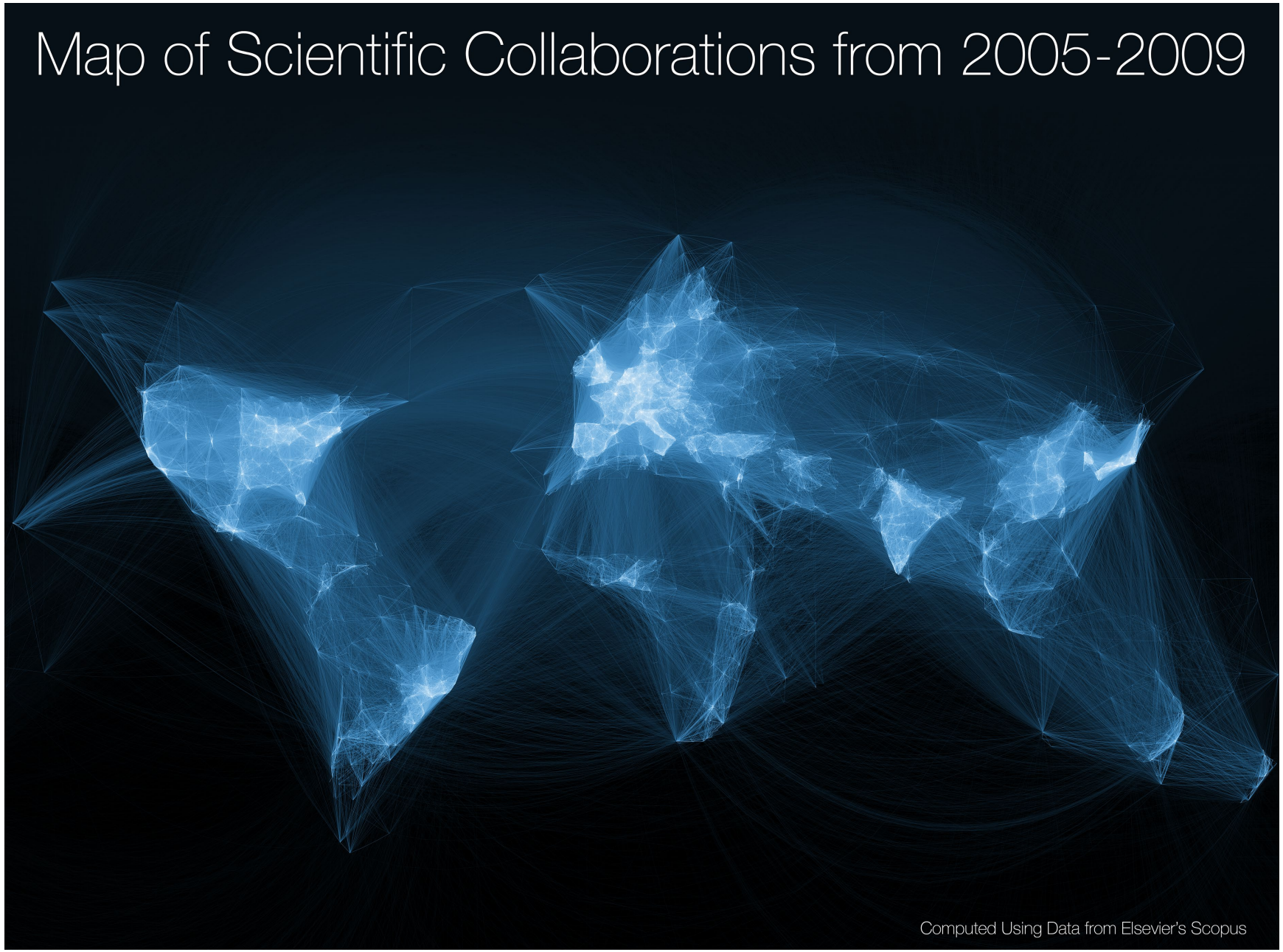
- Scrivner, Olga, Thuy Nguyen, Kosali Simon, Esmé Middaugh, Bledi Taska, and Katy Börner. 2020. "Job postings in the substance use disorder treatment related sector during the first five years of Medicaid expansion". *PLOS One* 15 (1): e0228394. doi: 10.1371/journal.pone.0228394.

Empower Yourself and Others! Take the Visual Analytics Course: (https://visanalytics.cns.iu.edu)

- 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.







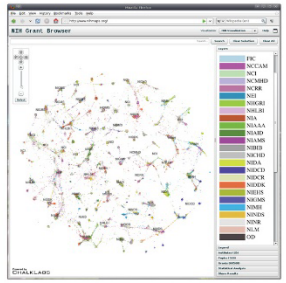
VII.6 Stream of Scientific Collaborations Between World Cities - Olivier H. Beauchesne - 2012

A Topic Map of NIH Grants 2007

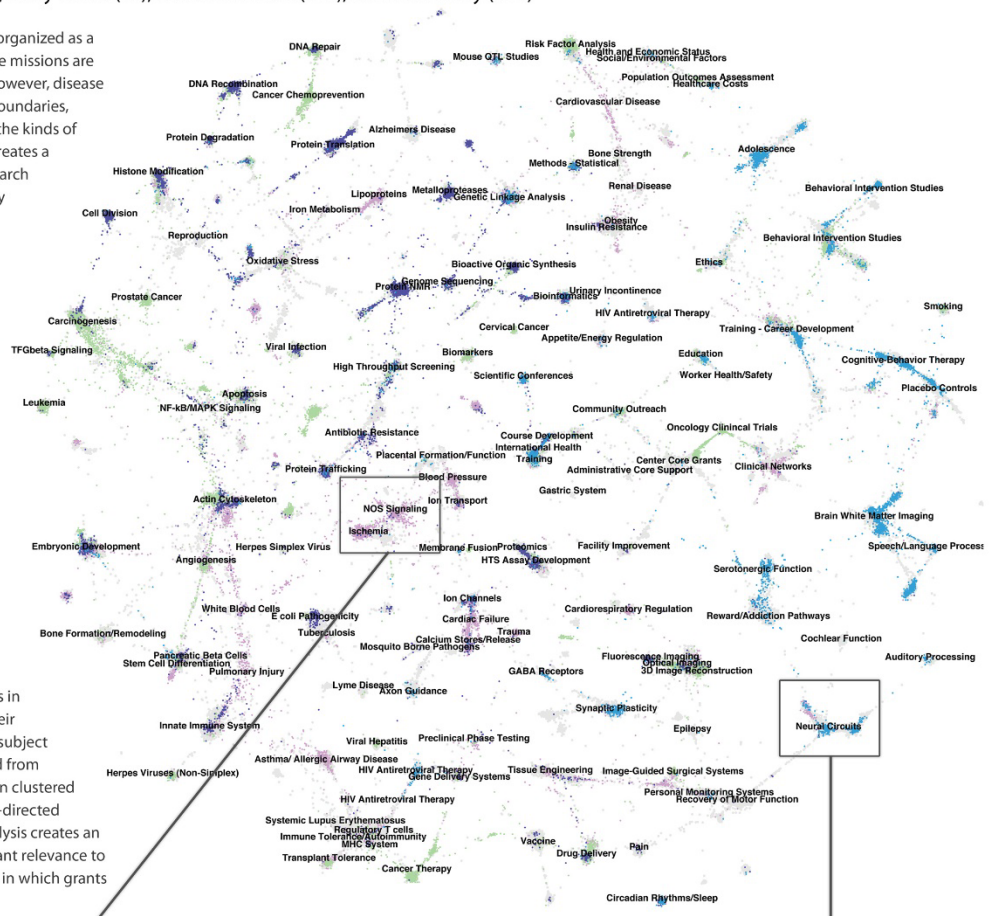
Bruce W. Herr II (Chalklabs & IU), Gully Burns (ISI), David Newman (UCI), Edmund Talley (NIH)

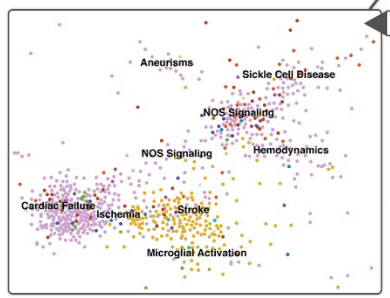
ChalkLabs UCIrvine

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. Institute abbreviations can be found at 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.



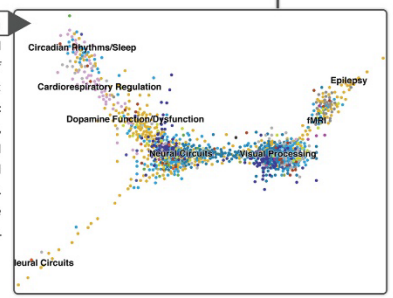


Cardiac Diseases Research

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

Neural Circuits Research

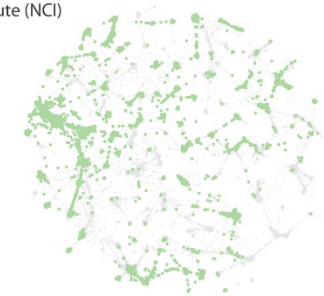
An area of the map focused on neural circuits, which shows the diversity of topics and NIH Institutes that fund research in this area, such as: Cardiorespiratory Regulation, primarily funded by NHLBI; Visual Processing, primarily funded by NEI; and Epilepsy, primarily funded by NINDS. For color coding,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



National Cancer Institute (NCI)

TOP 10 TOPICS

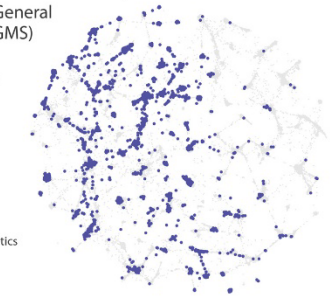
1. Oncology Clinical Trials
2. Cancer Treatment
3. Cancer Therapy
4. Carcinogenesis
5. Risk Factor Analysis
6. Cancer Chemotherapy
7. Metastasis
8. Leukemia
9. Prediction/Prognosis
10. Cancer Chemoprevention



National Institute of General Medical Sciences (NIGMS)

TOP 10 TOPICS

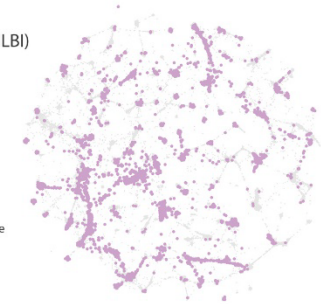
1. Bioactive Organic Synthesis
2. X-ray Crystallography
3. Protein NMR
4. Computational Models
5. Yeast Biology
6. Metalloproteases
7. Enzymatic Mechanisms
8. Protein Complexes
9. Invertebrate/Zebrafish Genetics
10. Cell Division



National Heart, Lung, and Blood Institute (NHLBI)

TOP 10 TOPICS

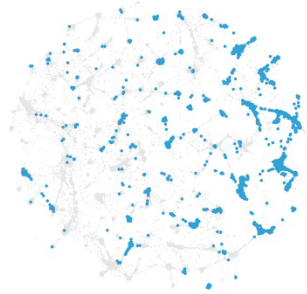
1. Cardiac Failure
2. Pulmonary Injury
3. Genetic Linkage Analysis
4. Cardiovascular Disease
5. Atherosclerosis
6. Hemostasis
7. Blood Pressure
8. Asthma/ Allergic Airway Disease
9. Gene Association
10. Lipoproteins



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



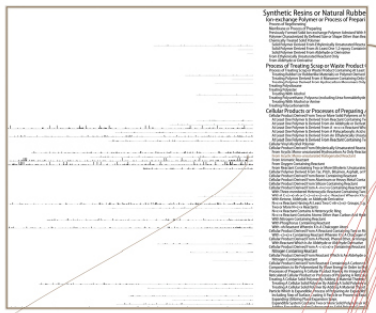
The US Patent Hierarchy

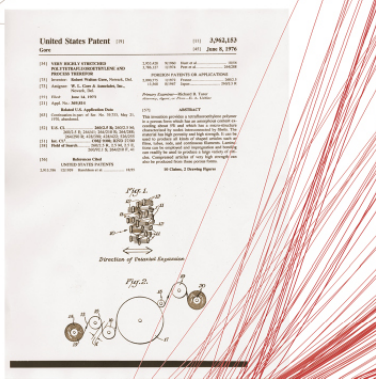
Impact

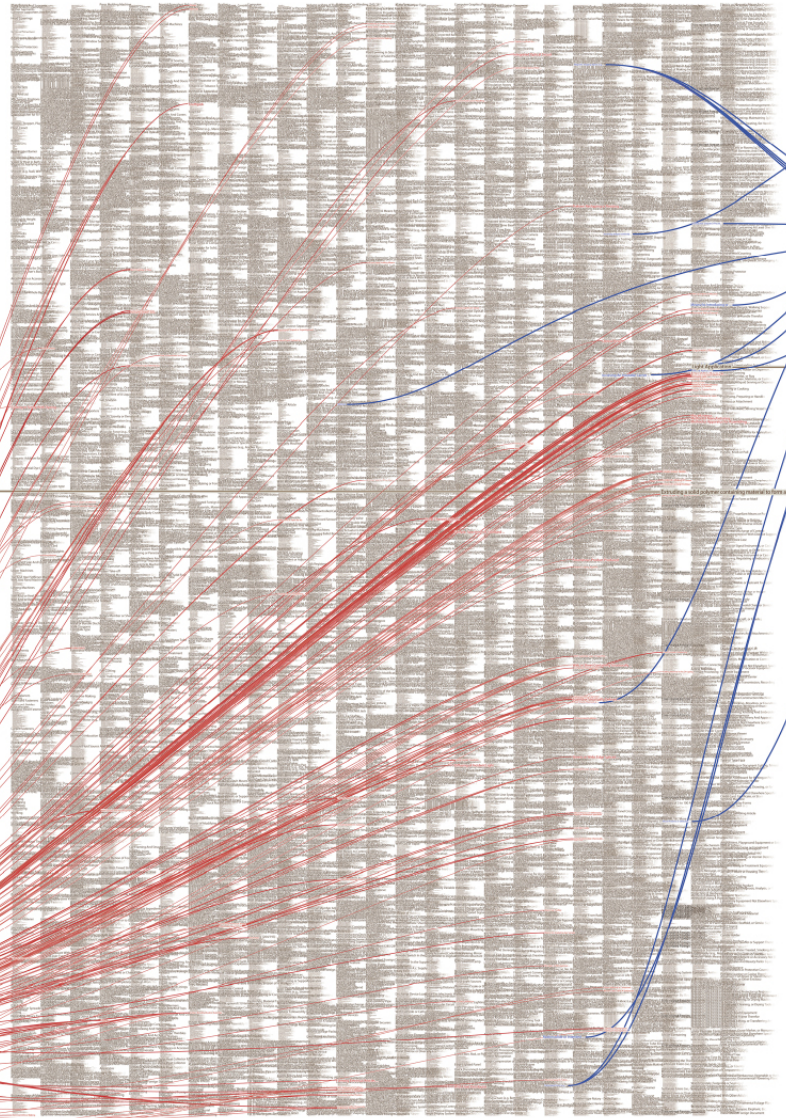
The United States Patent and Trademark Office does scientists and industry a great service by granting patents to protect inventions. Inventions are categorized in a taxonomy that groups patents by industry or use, proximate function, effect or product, and structure. At the time of this writing there 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.

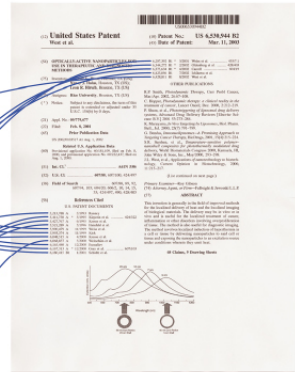
The patent on Goretex—a lightweight, durable synthetic fiber—is an example of one that has had significant impact. The box below enlarges the section of the hierarchy where it is filed, and the red lines (arranged to start along a time line from 1981 to 2006) point to the 130 categories that contain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

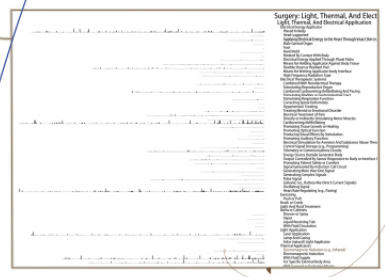






Prior Art





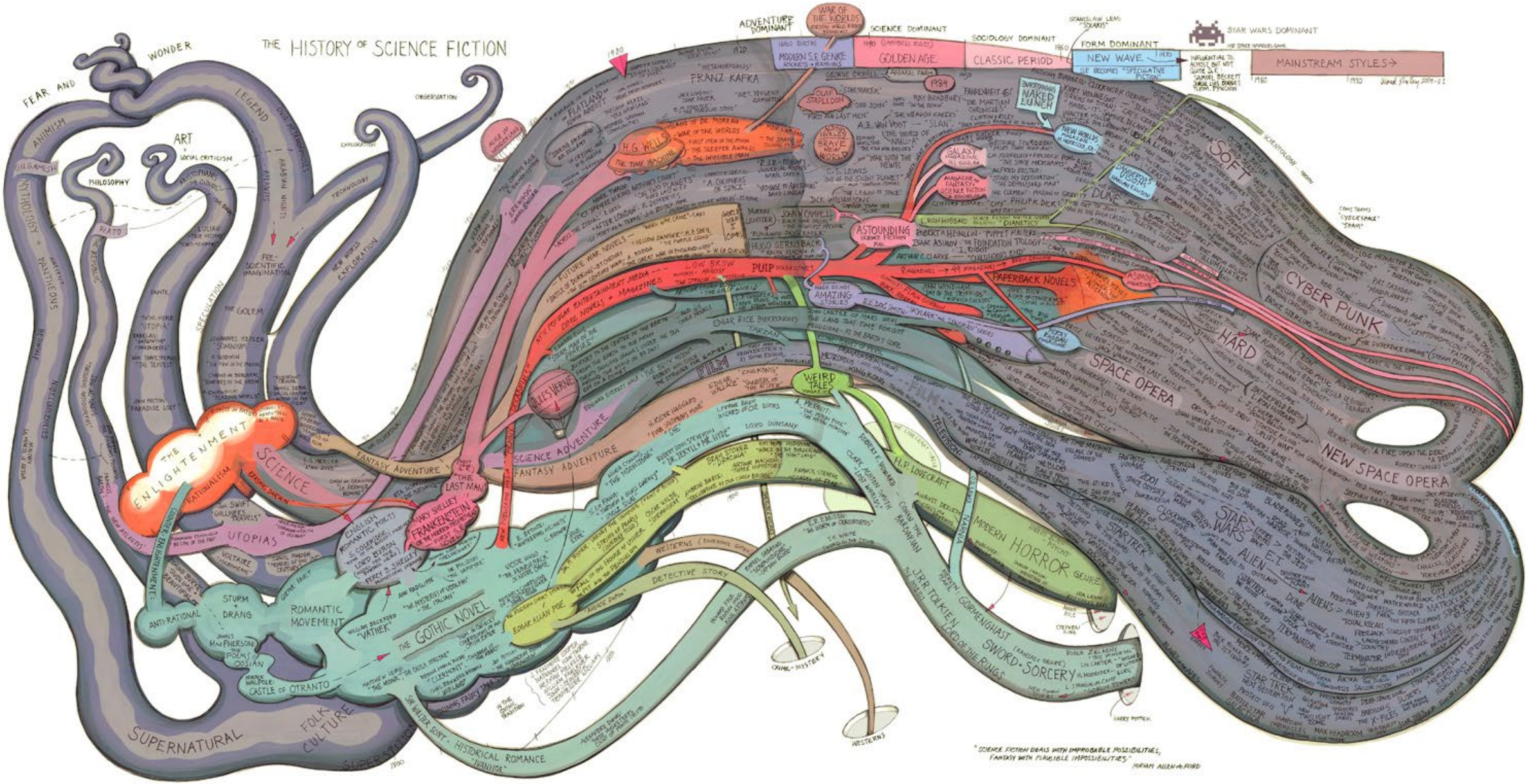
New patents often build on older ideas from many different categories. Here, blue lines originate in the sixteen categories that contain patents cited as prior art for a patent on "gold nanoshells." Gold nanoshells are a new invention: tiny gold spheres (with a diameter ten million times smaller than 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?

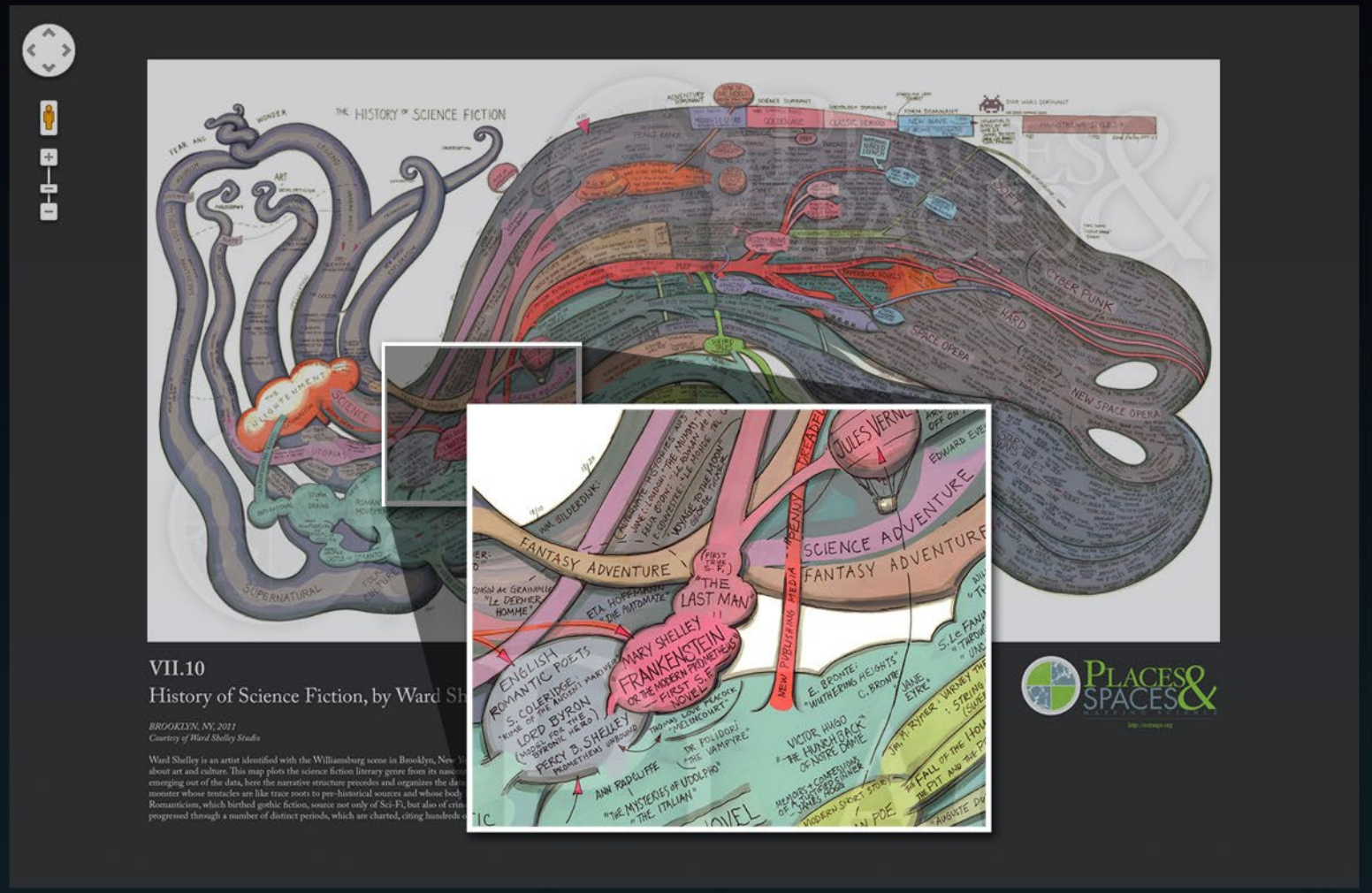
Something as simple as a bar chart helps people see how entities in a category relate to that category. Here, each bar encodes a "distance to prototype": how much each patent differs from an idealized "prototype patent" for that category. A measure like this can be based on statistics, computational 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 judgements.

II.8 *Taxonomy Visualization of Patent Data* - Katy Börner, Elisha F. Hardy, Bruce W. Herr II, Todd Holloway, and W. Bradford Paley - 2006


THE HISTORY OF SCIENCE FICTION
WONDER
FEAR AND
ANIMISM
GILGAMESH
MYTHOLOGY + PANTHEONS
BEOWULF
NIBELUNGENLIED
ART
SOCIAL CRITICISM
PHILOSOPHY
PLATO
LEGEND
ARABIAN NIGHTS
ATLANTIS
TECHNOLOGY
OBSERVATION
EXPLORATION
NEW WORLD EXPLORATION
SPECULATION
THE GOLEM
PRE-SCIENTIFIC IMAGINATION
THE ENLIGHTENMENT
RATIONALISM
SCIENCE
UTOPIAS
VOLTAIRE
ANTI-RATIONAL
STURM DRANG
ROMANTIC MOVEMENT
THE GOTHIC NOVEL
CASTLE OF OTRANTO
SUPERNATURAL
FOLK CULTURE
FANTASY ADVENTURE
SCIENCE ADVENTURE
THE LAST MAN
FRANKENSTEIN
JULES VERNE
WEIRD TALES
H.G. WELLS
THE TIME MACHINE
WAR OF THE WORLDS
ADVENTURE DOMINANT
SCIENCE DOMINANT
SOCIOLOGY DOMINANT
FORM DOMINANT
STAR WARS DOMINANT
MODERN S.F. GENRE
GOLDEN AGE
CLASSIC PERIOD
NEW WAVE
MAINSTREAM STYLES
FRANZ KAFKA
FLATLAND
OLAF STAPLEDON
STARMAKER
1984
FAHRENHEIT 451
THE MARTIAN CHRONICLES
NAKED LUNCH
DUNE
ASTOUNDING
PULP MAGAZINES
AMAZING STORIES
PAPERBACK NOVELS
FILM
TELEVISION
EDGAR RICE BURROUGHS
TARZAN
H.P. LOVECRAFT
MODERN HORROR GENRE
DETECTIVE STORY
WESTERNS
SWORD-SORCERY
J.R.R. TOLKIEN
LORD OF THE RINGS
SOFT
CYBER PUNK
HARD
SPACE OPERA
NEW SPACE OPERA
STAR WARS
STAR TREK
ALIEN
E.T.
DUNE
TERMINATOR
ROBOCOP
THE X-FILES
BATTLESTAR GALACTICA
"SCIENCE FICTION DEALS WITH IMPROBABLE POSSIBILITIES, FANTASY WITH FEASIBLE IMPOSSIBILITIES." MIRIAM ALLEN deFORD

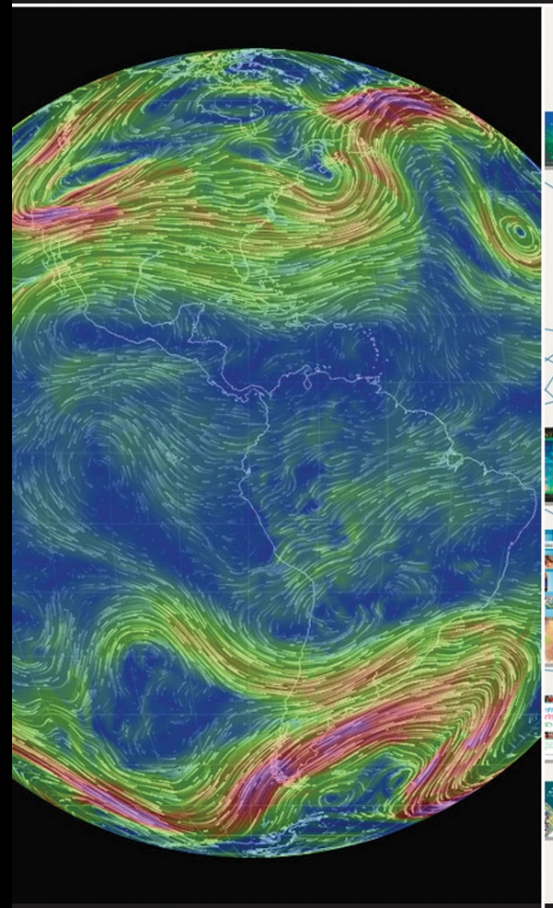
Check out our Zoom Maps online!



Visit scimaps.org and check out all our maps in stunning detail!

MACROSCOPES FOR INTERACTING WITH SCIENCE

PLACES & SPACES
MAPPING SCIENCE
scimaps.org

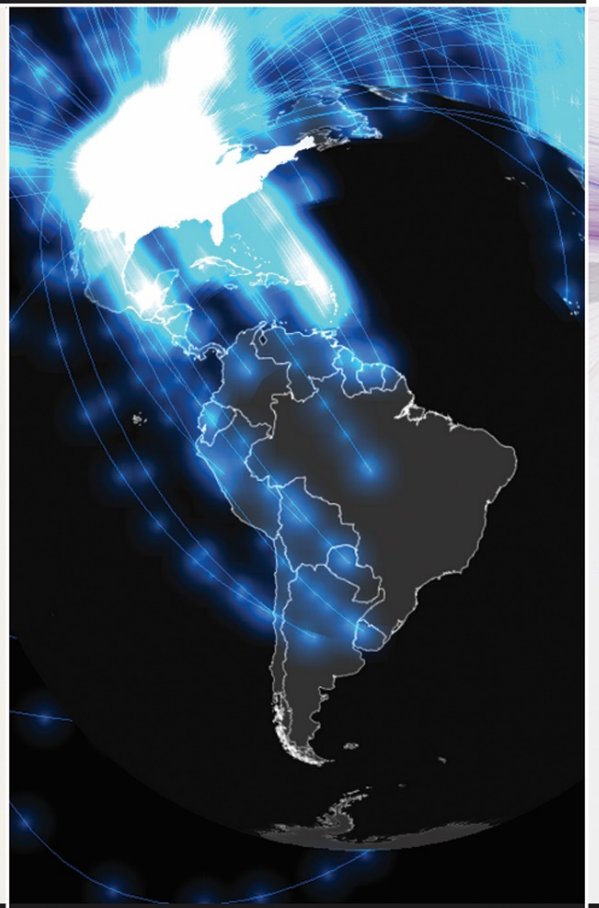


Earth

Weather on a worldwide scale

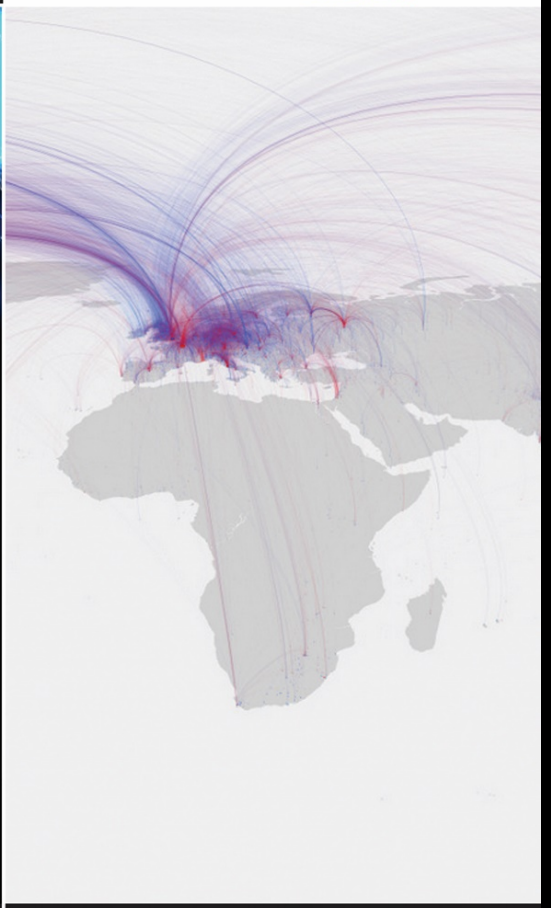
AcademyScope

Exploring the scientific landscape



Mapping Global Society

Local news from a global perspective



Charting Culture

2,600 years of human history in 5 minutes

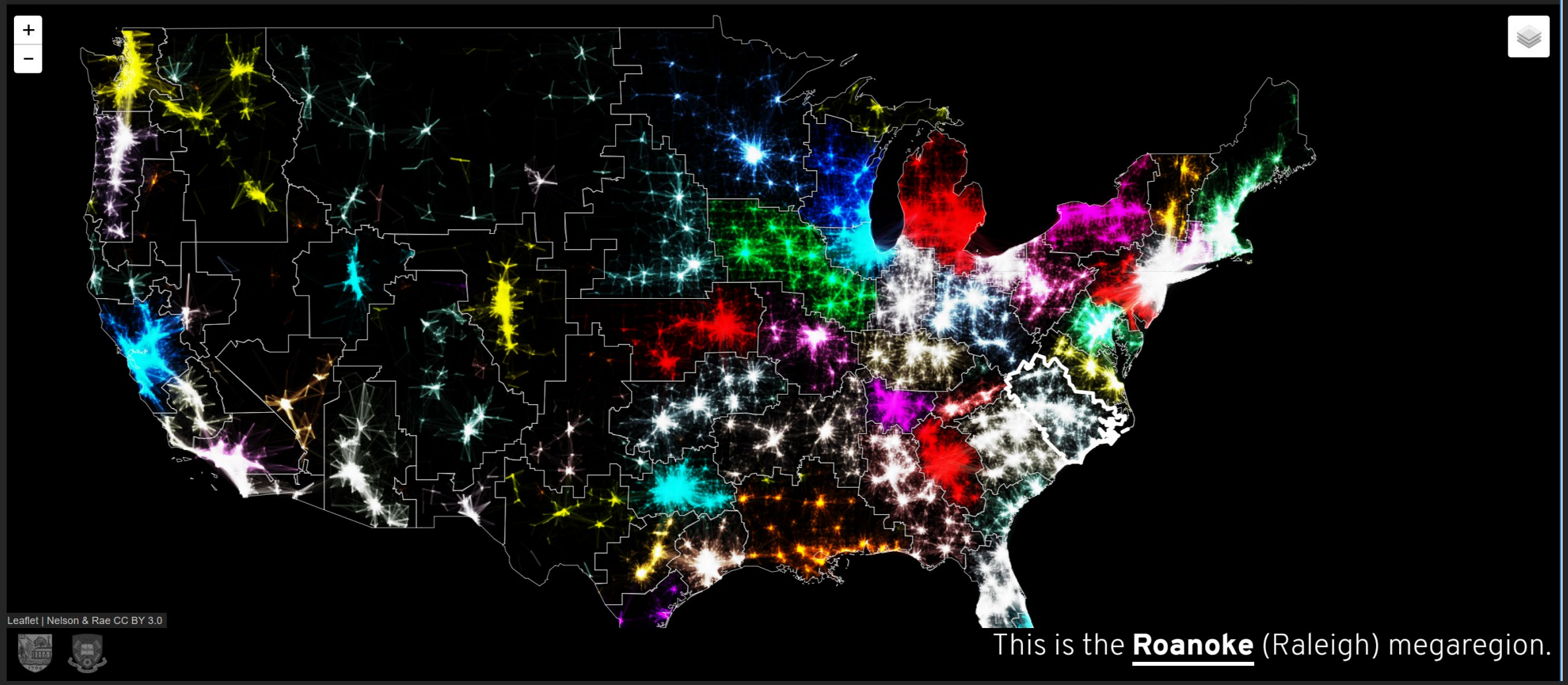
» Play with Scale » **Megaregions of the US**

MORE INFO

PLACES & SPACES MAPPING SCIENCE scimaps.org

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 commuteshed.



Megaregions of the US –Garrett Dash Nelson and Alasdair Rae – 2016

Oct 1-Dec 3, 2020: Exhibit on display at the Dimension Mill in Bloomington, IN on https://dimensionmill.org

CNS Cyberinfrastructure for Network Science Center

INDIANA UNIVERSITY SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Identify and Overcome Skill Discrepancies

Börner, Katy, Olga Scrivner, Michael Gallant, Shutian Ma, Xiaozhong Liu, Keith Chewning, Lingfei Wu, and James Evans. 2018. 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.

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

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 Wue, 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[a,b,1], Olga Scrivner[a], Mike Gallant[a], Shutian Ma[a,c], Xiaozhong Liu[a], Keith Chewning[d], Lingfei Wu[e,f,g,h], and James A. Evans[f,g,i,1]

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

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

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Modeling and Visualizing Science and Technology Developments," held Decem-

CNS Cyberinfrastructure for Network Science Center

Study the **(mis)match** and **temporal dynamics** of science and technology (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

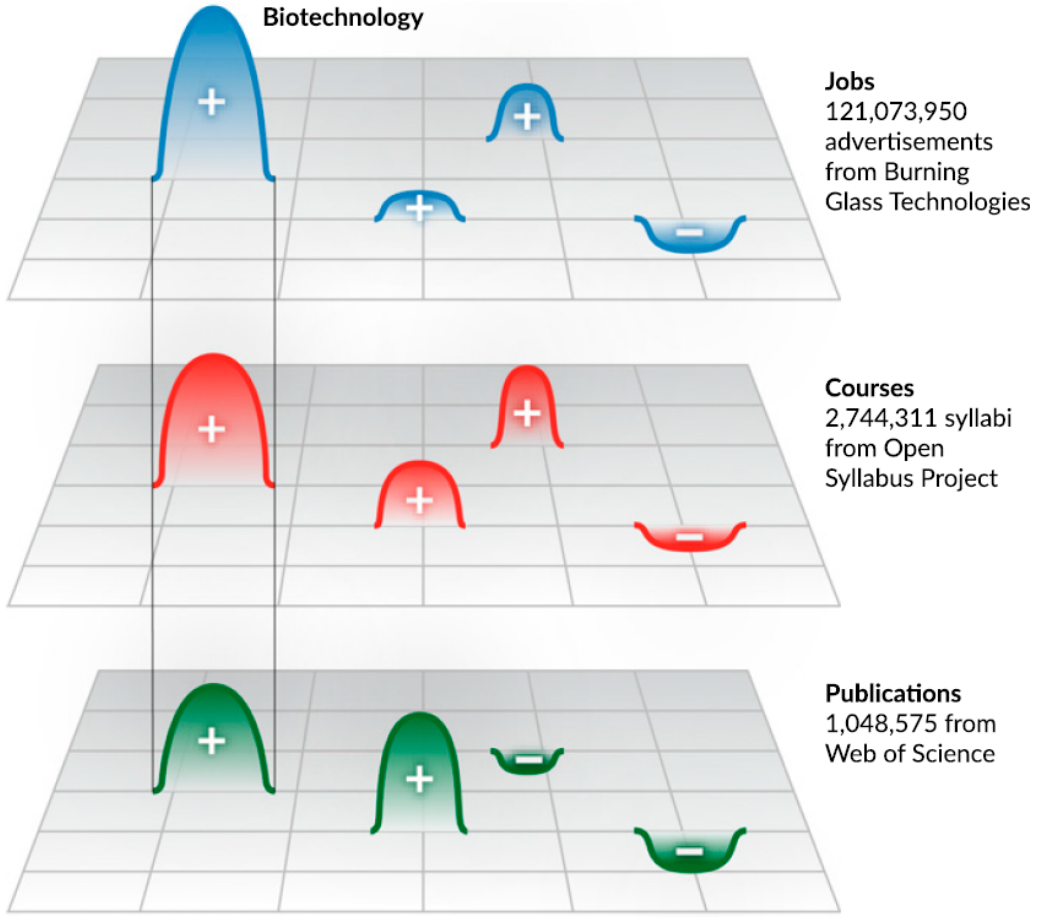
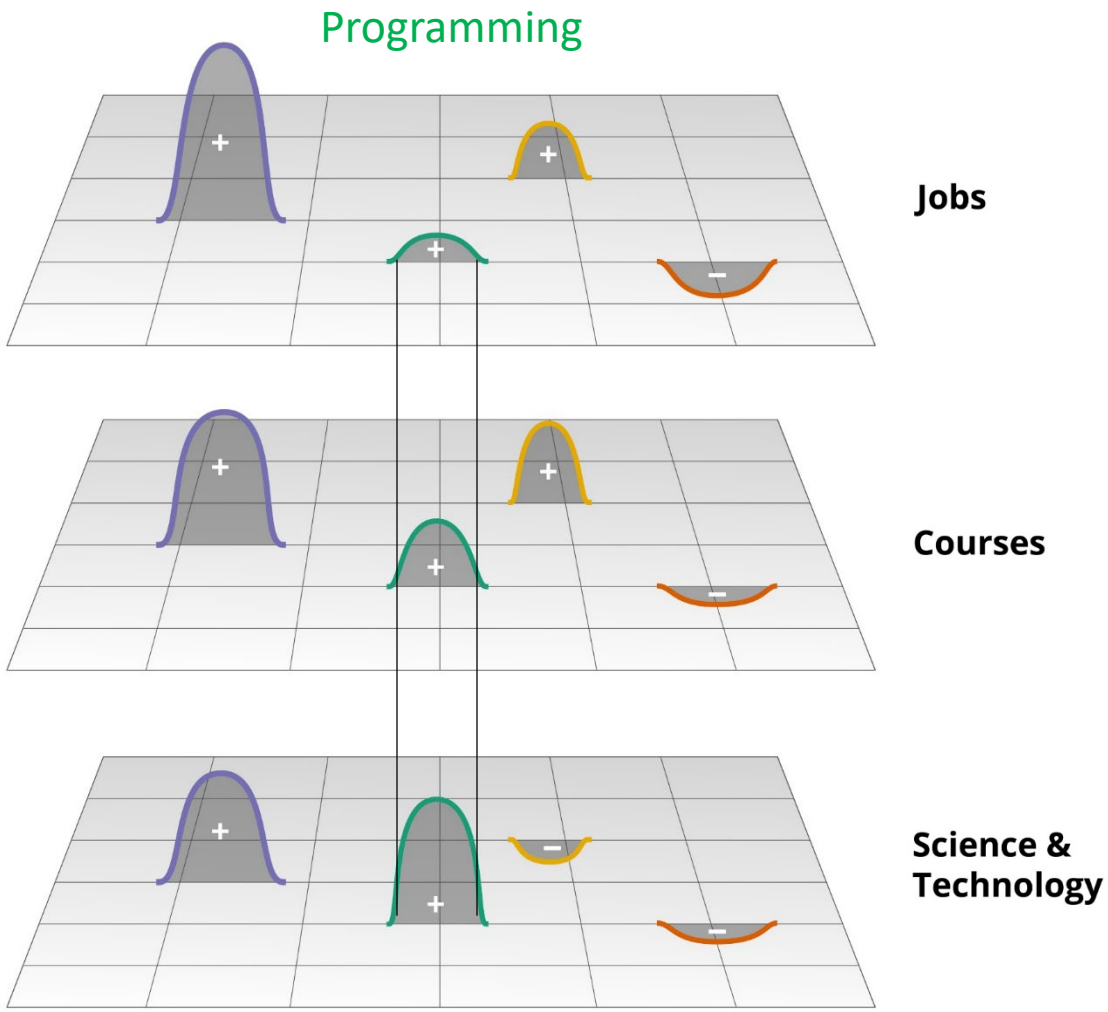
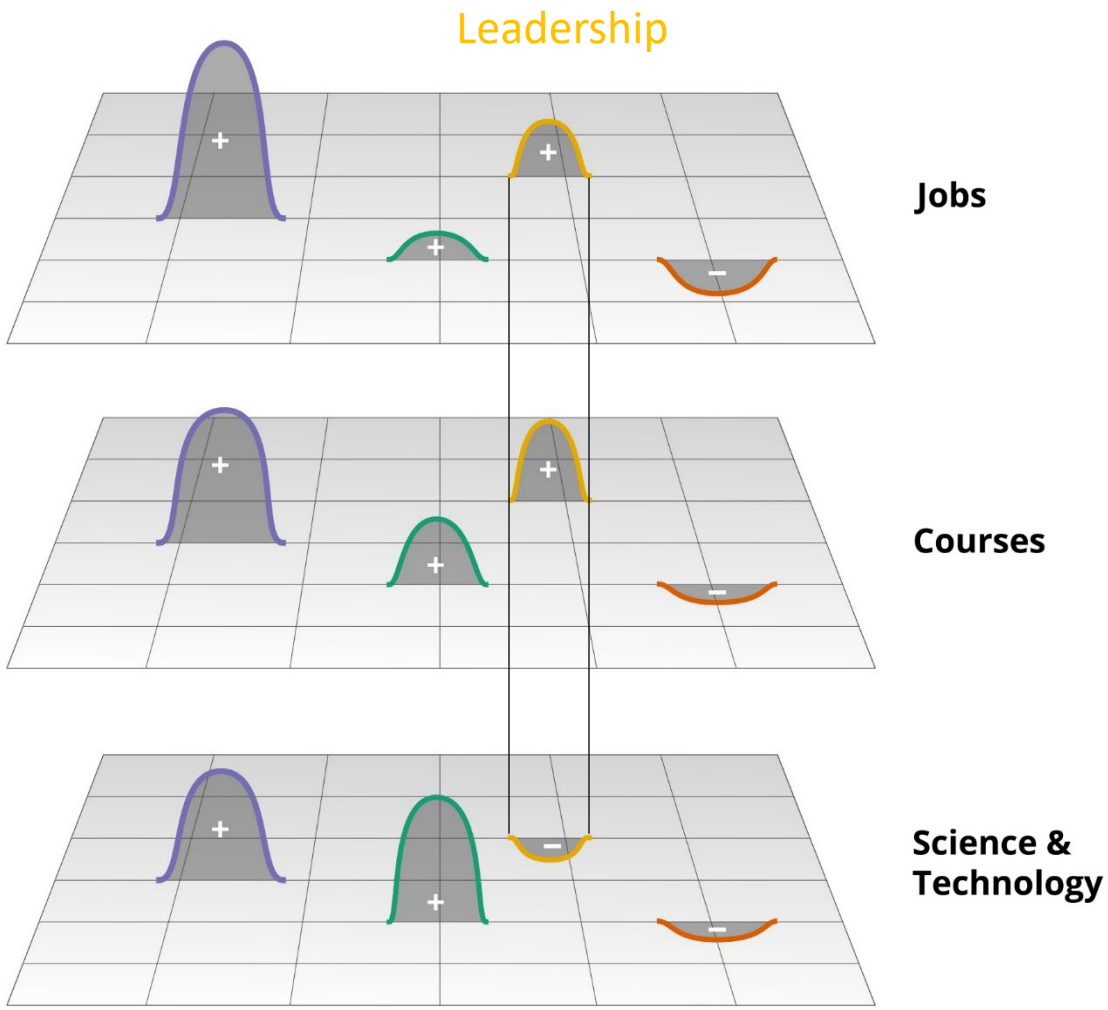
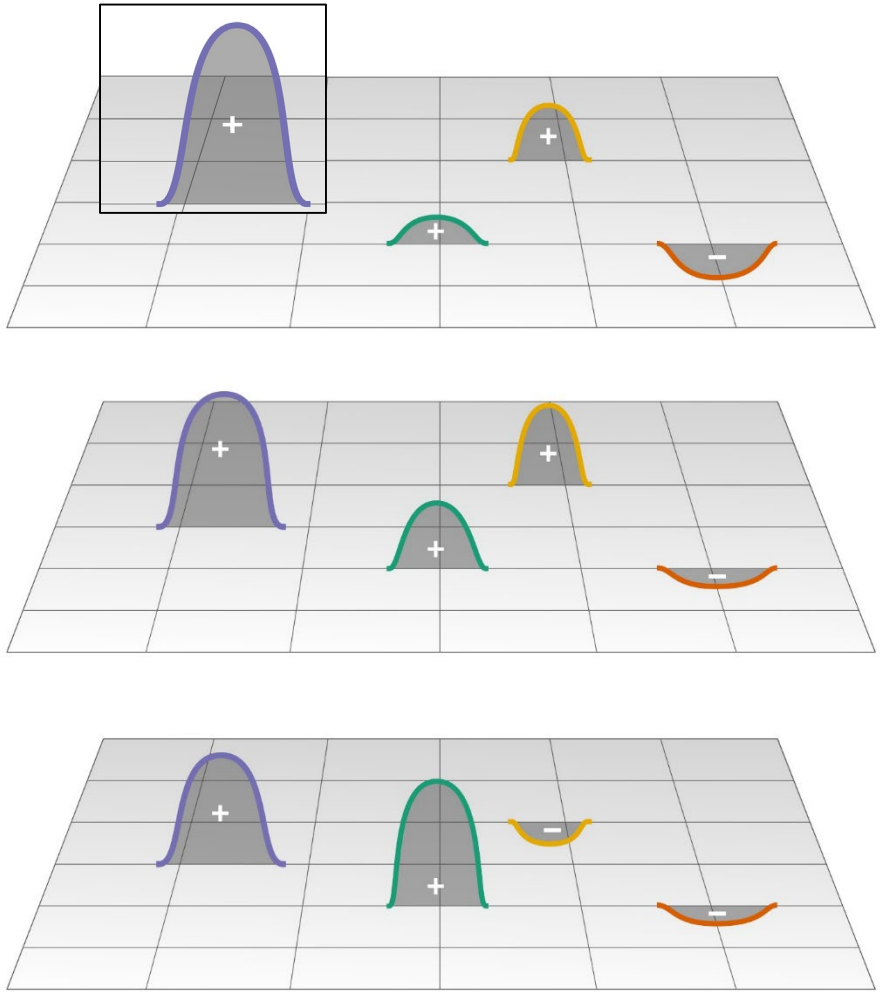


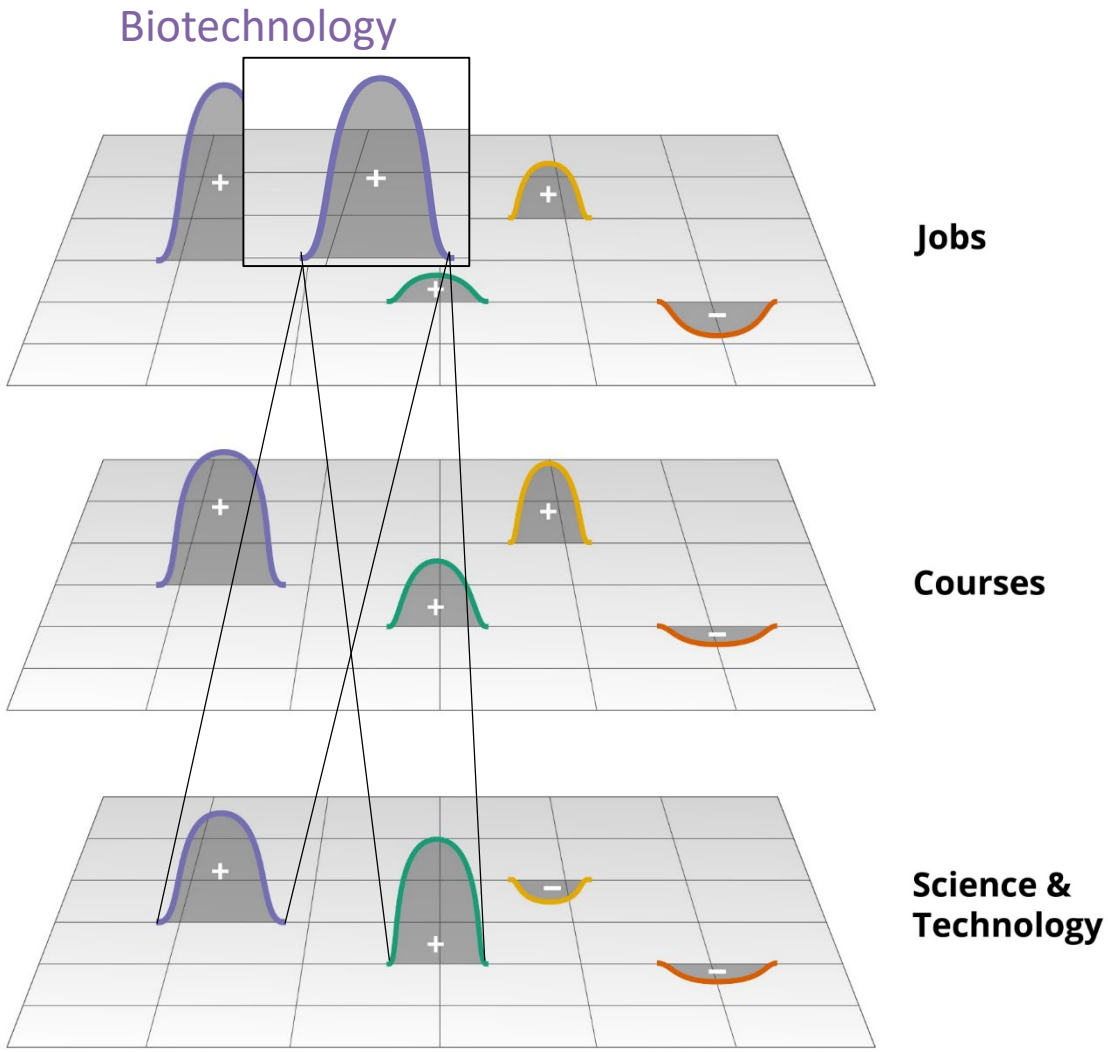
Fig. 1. The interplay of job market demands, educational course offerings, and progress in S&T as captured in publications. Color-coded mountains (+) and valleys (–) indicate different skill clusters. For example, skills related to Biotechnology might be mentioned frequently in job descriptions and taught 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


Leadership
Jobs
Courses
Science &
Technology

Biotechnology




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 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.
- The pandemic is speeding up automation, and 85 million jobs are on the line.
  https://www.cnn.com/2020/10/20/business/wef-future-of-jobs-report/index.html

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 Wue, 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[a,b,1], Olga Scrivner[a], Mike Gallant[a], Shutian Ma[a,c], Xiaozhong Liu[a], Keith Chewning[d], Lingfei Wu[e,f,g,h], and James A. Evans[f,g,i,1]

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

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

This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Modeling and Visualizing Science and Technology Developments," held Decem-

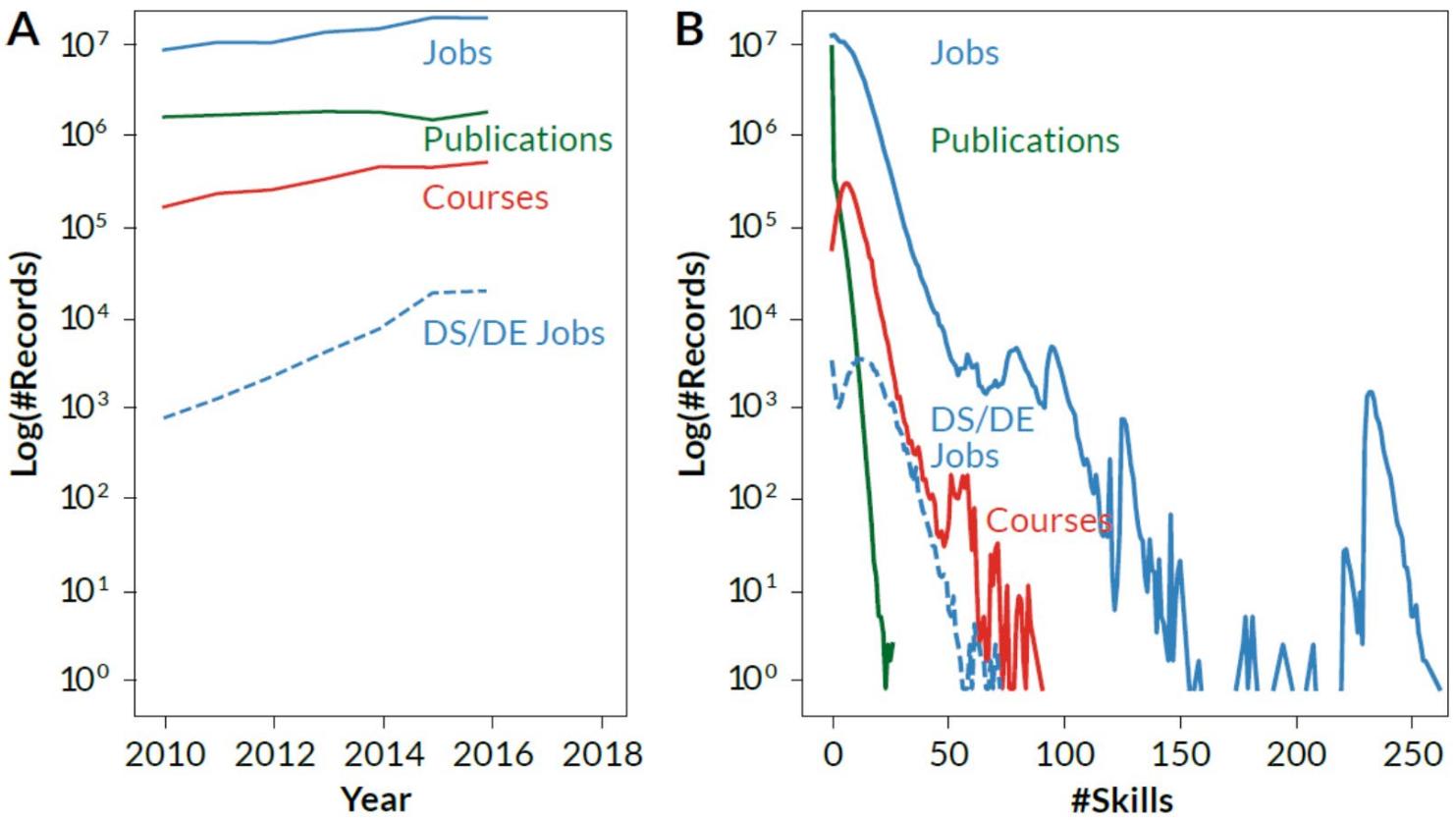
CNS Cyberinfrastructure for Network Science Center

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.

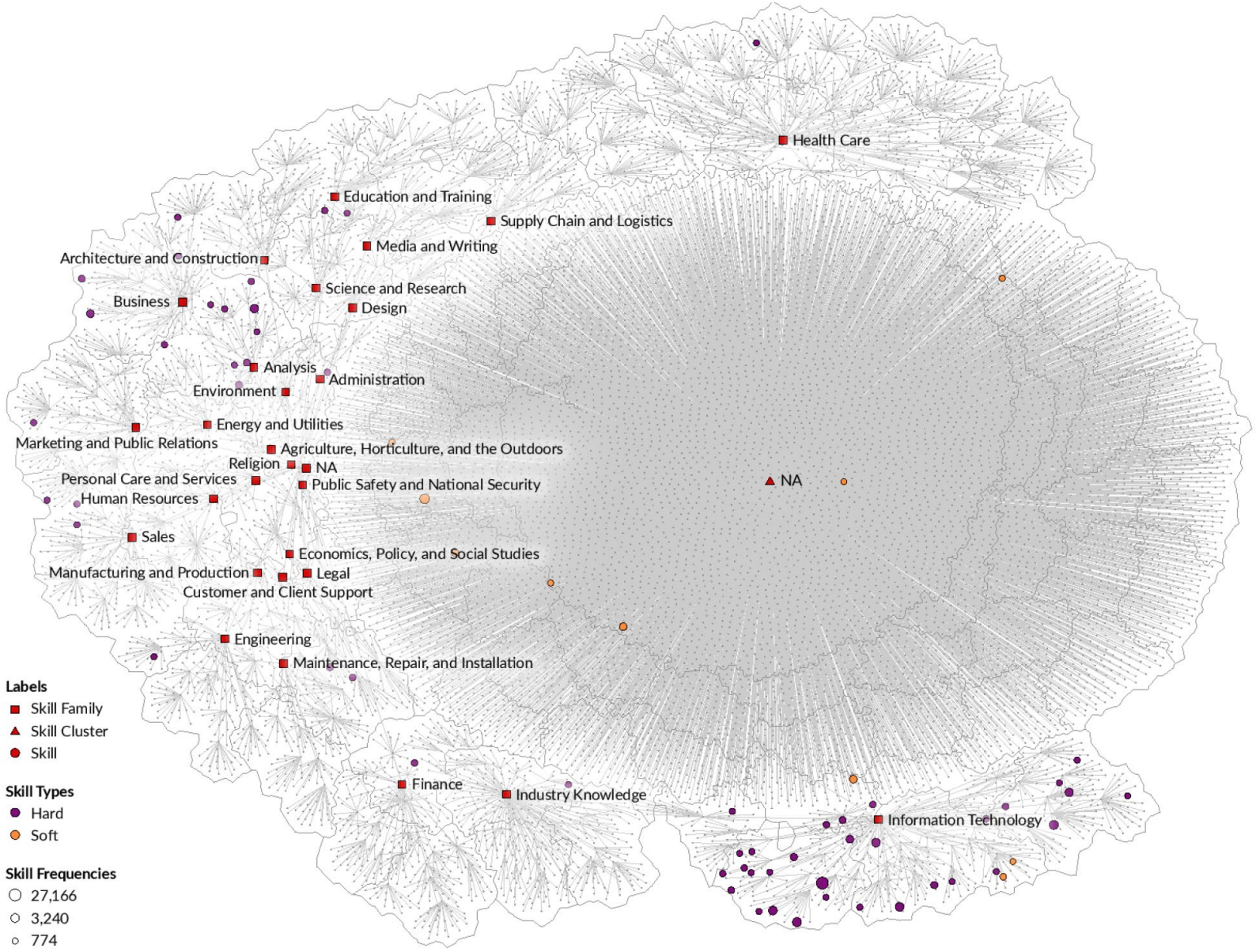
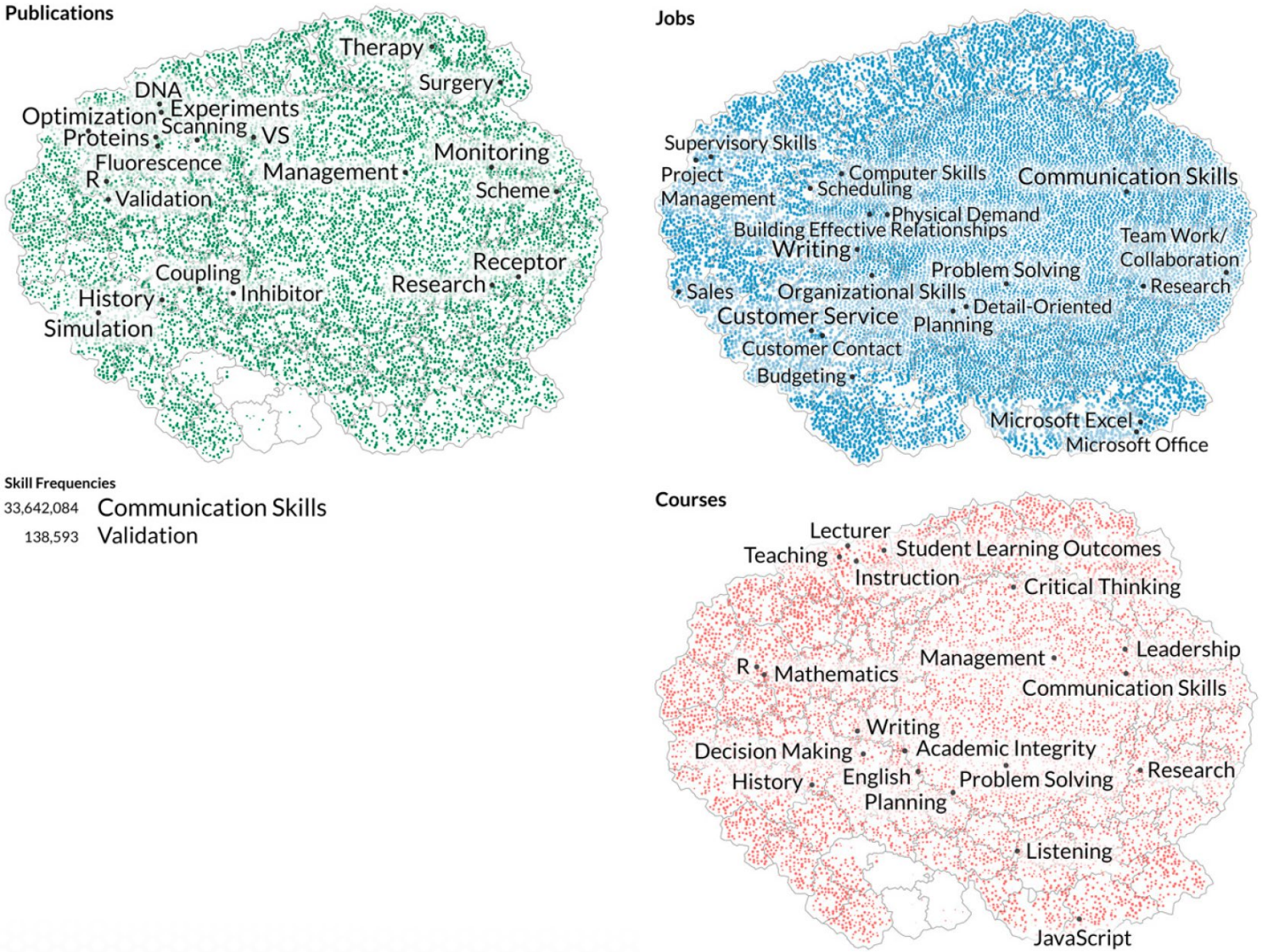


Fig. 3. Basemap of 13,218 skills with overlays of skill frequency in jobs, courses, and publications. This figure substantiates the conceptual drawing in Fig. 1 using millions of data records. Jobs skills are plotted in blue, courses are in red, and publications are in green. Node area size coding indicates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



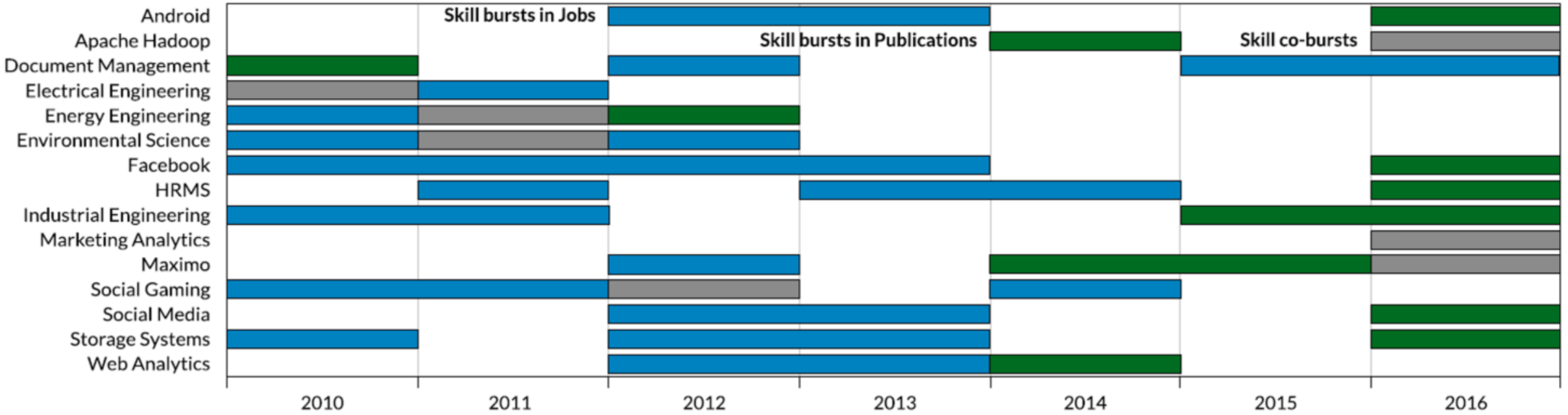


Fig. 4. Burst of activity in DS/DE skills in jobs and publications. Each burst is rendered as a horizontal bar with a start and an end date; skill term is shown on the left. Skills that burst in jobs are blue; skills bursting in publications are green. Seven skills burst in both datasets during the same 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

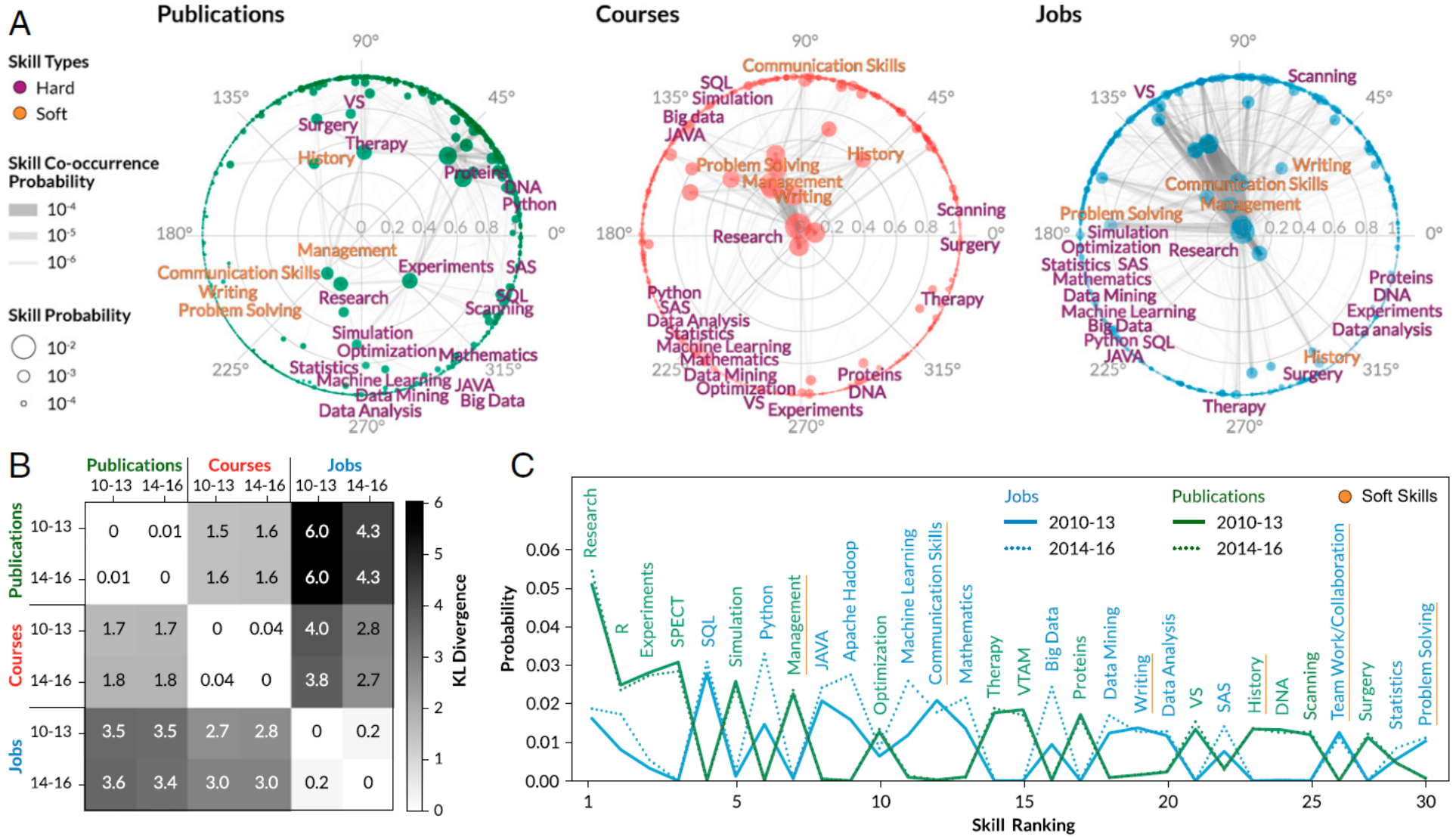


Fig. 5. Structural and dynamic differences between skill distributions in jobs, courses, and publications for 2010–2013 and 2014–2016. (*A*) Poincaré disks comparing the centrality of soft skills (orange) and hard skills (purple) across jobs, courses, and publications. (*B*) KL divergence matrix for 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.

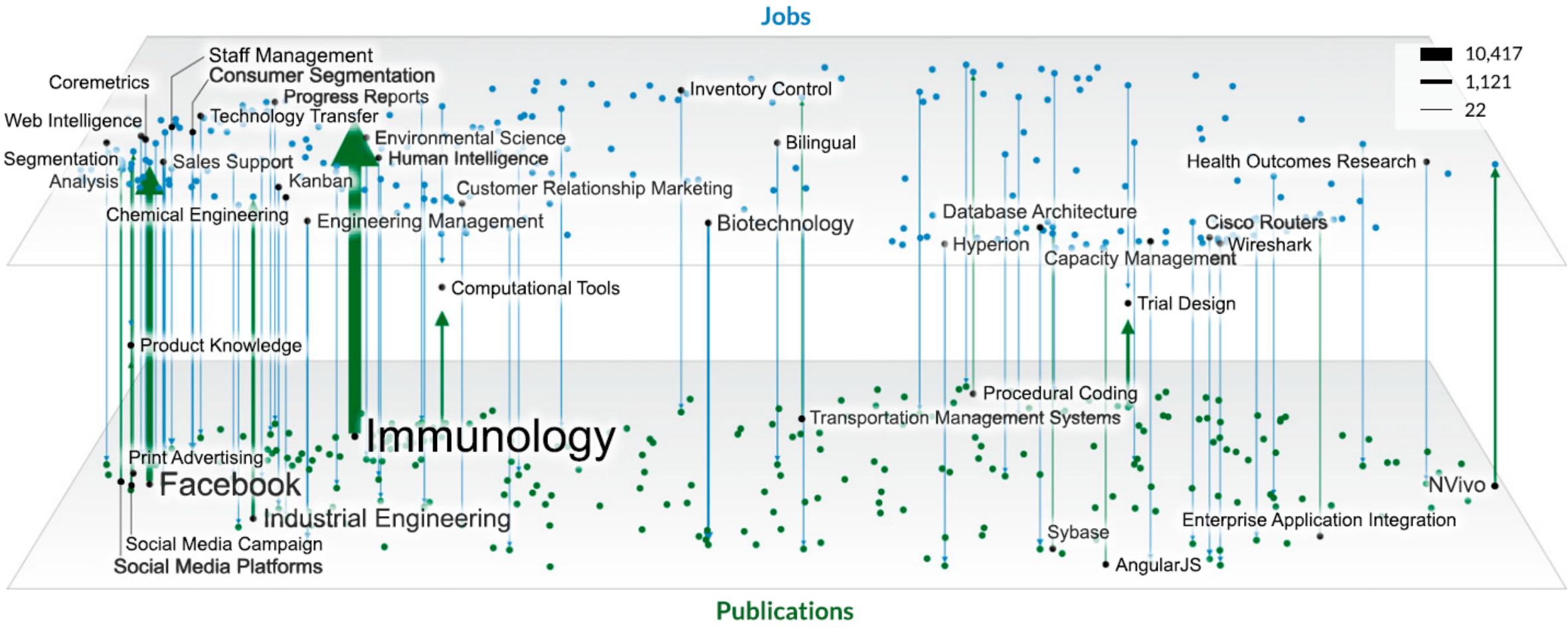


Fig. 6. Strength of influence mapping. Top 200 most frequent skills in jobs (blue) and in publications (green) plotted on the skills basemap from Fig. 2. Arrows represent skills with significant Granger causality (P value < 0.05). Line thickness and label size indicate skill frequency. The direction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

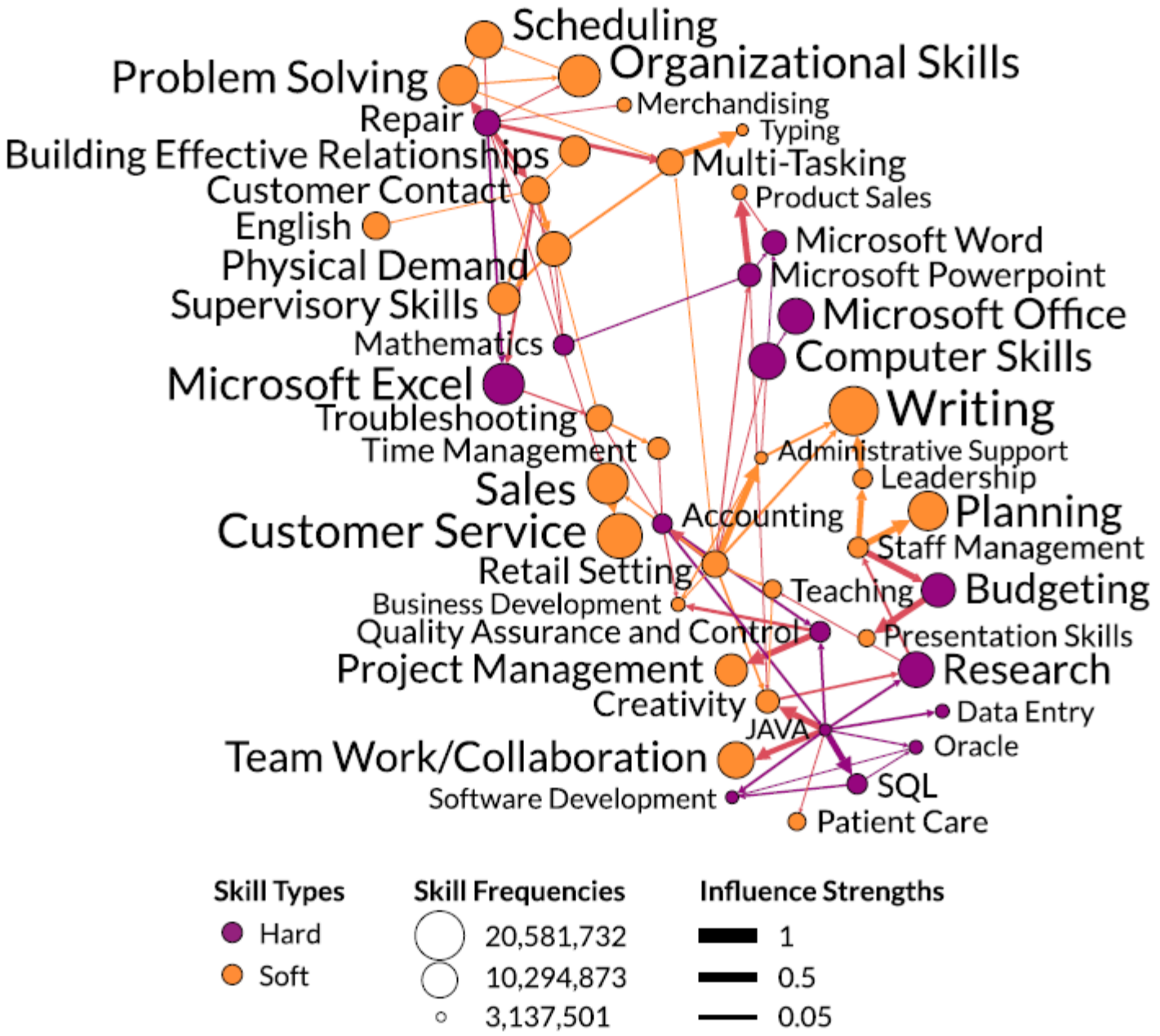
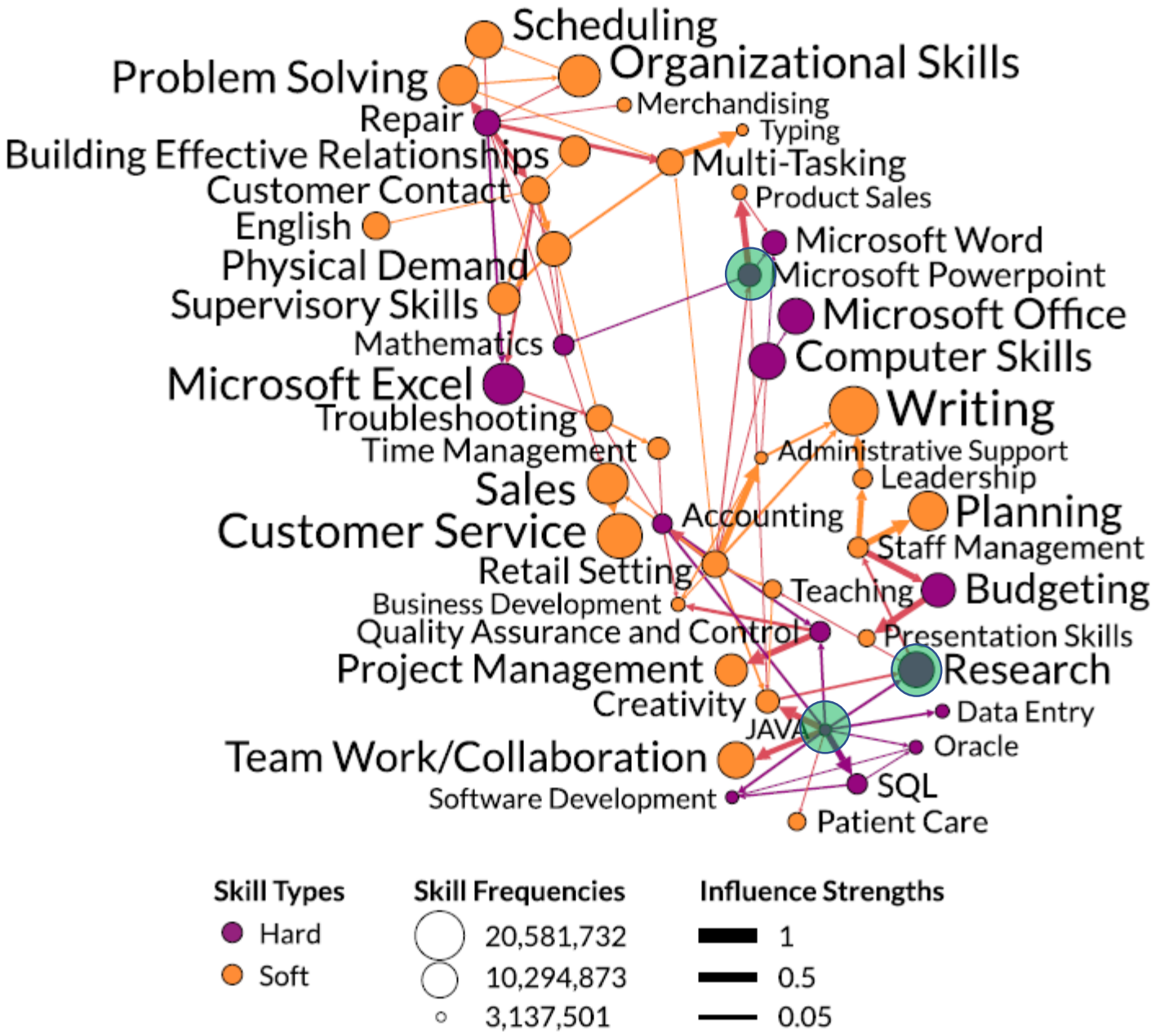


Fig. 7. Hawkes influence network of DS/DE skills within job advertisements 2010–2016. Each of the 45 nodes represents a top-frequency skill (29 soft and 16 hard skills) with a strong influence edge from/to other skill(s) in job advertisements between 2010 and 2016. Node and label size correspond to the 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.

CNS Cyberinfrastructure for Network Science Center

INDIANA UNIVERSITY SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Job Postings in The Substance Use Disorder Treatment Sector

Scrivner, Olga, Thuy Nguyen, Kosali Simon, Esmé Middaugh, Bledi Taska, and Katy Börner. 2020. "Job postings in the substance use disorder treatment related sector during the first five years of Medicaid expansion". *PLOS One* 15 (1): e0228394. doi: 10.1371/journal.pone.0228394.

Background

Effective treatment strategies exist for substance use disorder (SUD), however severe hurdles remain in ensuring adequacy of the SUD treatment (SUDT) workforce as well as improving SUDT affordability, access and stigma. Although evidence shows recent increases in SUD medication access from expanding Medicaid availability under the Affordable Care Act, it is yet unknown whether these policies also led to a growth in hiring in the SUDT related workforce, partly due to poor data availability. Our study uses novel data to shed light on recent trends in a fast-evolving and policy-relevant labor market, and contributes to understanding data sources to track the SUDT related workforce and the effect of recent state healthcare policies on the supply side of this sector.

Methods and data

We examine hiring attempts in the SUDT and related behavioral health sector over 2010-2018 to estimate the causal effect of the 2014-and-beyond state Medicaid expansions on these outcomes through "difference-in-difference" econometric models. We use Burning Glass Technologies (BGT) data covering virtually all U.S. job postings by employers.

Findings

Nationally, we find little growth in the sector's hiring attempts in 2010-2018 relative to the rest of the economy or to health care as a whole. However, this masks heterogeneity in the bimodal trend in SUDT job postings, with some increases in most years but a decrease in 2014 and in 2017, as well as a shift in emphasis between different occupational categories. Medicaid expansion, however, is not associated with any statistically significant change in overall hiring attempts in the SUDT related sector during this time period, although there is moderate evidence of increases among primary care physicians.

PLOS ONE

RESEARCH ARTICLE

Job postings in the substance use disorder treatment related sector during the first five years of Medicaid expansion

Olga Scrivner[1☯*], Thuy Nguyen[2☯], Kosali Simon[2,3☯], Esmé Middaugh[1☯], Bledi Taska[4☯], Katy Börner[1☯]

1 Luddy School of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN, United States of America, **2** O'Neill School of Public and Environmental Affairs, Indiana University, Bloomington, IN, United States of America, **3** National Bureau of Economic Research, Cambridge, Massachusetts, United States of America, **4** Burning Glass Technologies, Boston, Massachusetts, United States of America

☯ These authors contributed equally to this work.
* obscrivn@indiana.edu

OPEN ACCESS

Citation: Scrivner O, Nguyen T, Simon K, Middaugh E, Taska B, Börner K (2020) Job postings in the substance use disorder treatment related sector during the first five years of Medicaid expansion. PLoS ONE 15(1): e0228394. https://doi.org/10.1371/journal.pone.0228394

Editor: Benjamin Cook, Cambridge Health Alliance, UNITED STATES

Received: July 15, 2019

Accepted: January 14, 2020

Published: January 30, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0228394

Data Availability Statement: Job Postings data are available from Burning Glass Technologies upon request. An interested researcher should send a request to info@burning-glass.com. The remaining

Abstract

Background

Effective treatment strategies exist for substance use disorder (SUD), however severe hurdles remain in ensuring adequacy of the SUD treatment (SUDT) workforce as well as improving SUDT affordability, access and stigma. Although evidence shows recent increases in SUD medication access from expanding Medicaid availability under the Affordable Care Act, it is yet unknown whether these policies also led to a growth in hiring in the SUDT related workforce, partly due to poor data availability. Our study uses novel data to shed light on recent trends in a fast-evolving and policy-relevant labor market, and contributes to understanding data sources to track the SUDT related workforce and the effect of recent state healthcare policies on the supply side of this sector.

Methods and data

We examine hiring attempts in the SUDT and related behavioral health sector over 2010-2018 to estimate the causal effect of the 2014-and-beyond state Medicaid expansions on these outcomes through "difference-in-difference" econometric models. We use Burning Glass Technologies (BGT) data covering virtually all U.S. job postings by employers.

Findings

Nationally, we find little growth in the sector's hiring attempts in 2010-2018 relative to the rest of the economy or to health care as a whole. However, this masks heterogeneity in the bimodal trend in SUDT job postings, with some increases in most years but a decrease in 2014 and in 2017, as well as a shift in emphasis between different occupational categories. Medicaid expansion, however, is not associated with any statistically significant change in overall hiring attempts in the SUDT related sector during this time period, although there is moderate evidence of increases among primary care physicians.

Introduction & Motivation

- Worldwide, the direct burden of illicit drug dependence increased to 20 million disability-adjusted life years in 2010. Examples of these illicit drugs are opioids, cocaine, amphetamines, and cannabis.
- In the US, mental health and SUD together became the leading cause of disease burden in 2015, while nearly 3% of Americans aged 12 years or older reported SUDs in the same year.
- The most effective SUD treatment (SUDT) is a combination of long-acting medications (usually methadone or buprenorphine) administered as part of a cognitive behavioral approach (such as counseling, family therapy, and peer support programs).
- In 2017, there were 13,857 treatment facilities in the U.S. with over 1,356,015 clients enrolled (The National Survey of Substance Abuse Treatment Services; NSSATS).
- The SUDT workforce is deemed inadequate by almost any measure. Workforce shortages and barriers have played a prominent role in limiting treatment access among those suffering from SUDs.

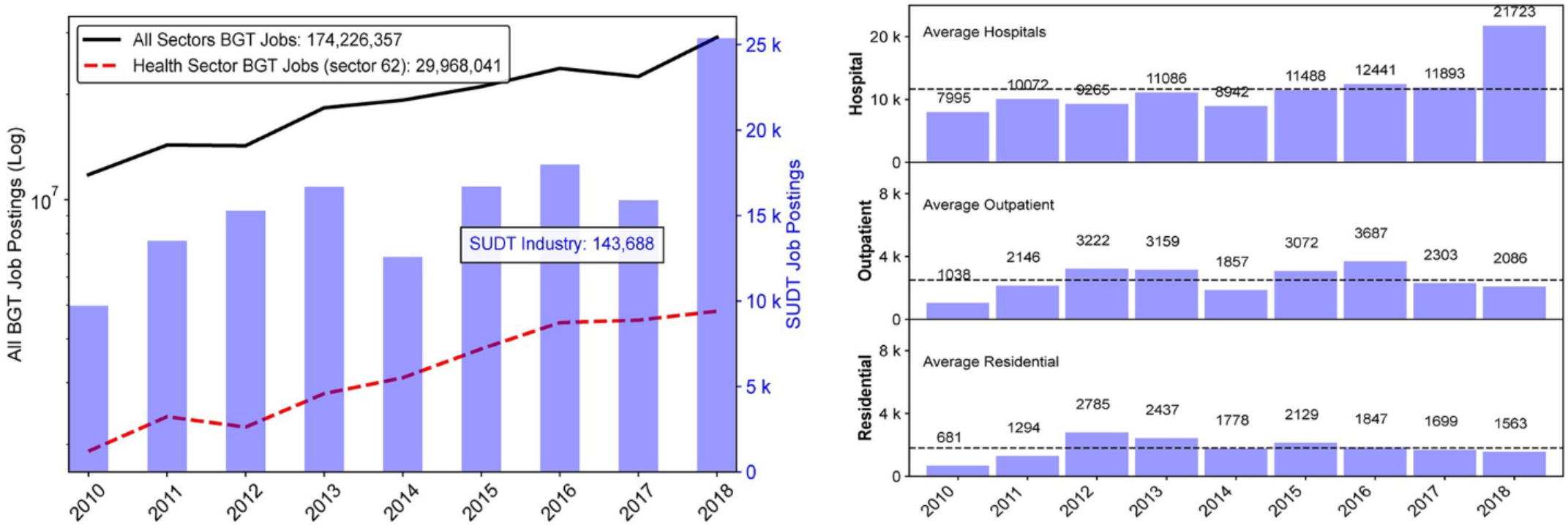


Fig 1. BGT online job postings. (A) BGT Job postings for all industries (black), Healthcare industry (red) and SUDT industries (light blue). The aggregated amount for all job postings is calculated for the period from 2010 through 2018. The healthcare sector is identified by the NAICS code '62'. The SUDT facilities are identified by three NAICS codes '6222', '6214','6232' filtered at 6-digit level. The left y-axis corresponds to the logarithmic trend lines for the total of all BGT job postings (black solid line) and the total of BGT healthcare sector (red dashed line). The y-right axis represents the SUDT sector values, shown as bar graphs. (B) Break down of job postings for three SUDT sectors. Three SUDT sectors are represented by their number of annual online job postings. Average line is calculated for each SUDT sector. Data Source: Burning Glass Technologies. 2019.

https://doi.org/10.1371/journal.pone.0228394.g001

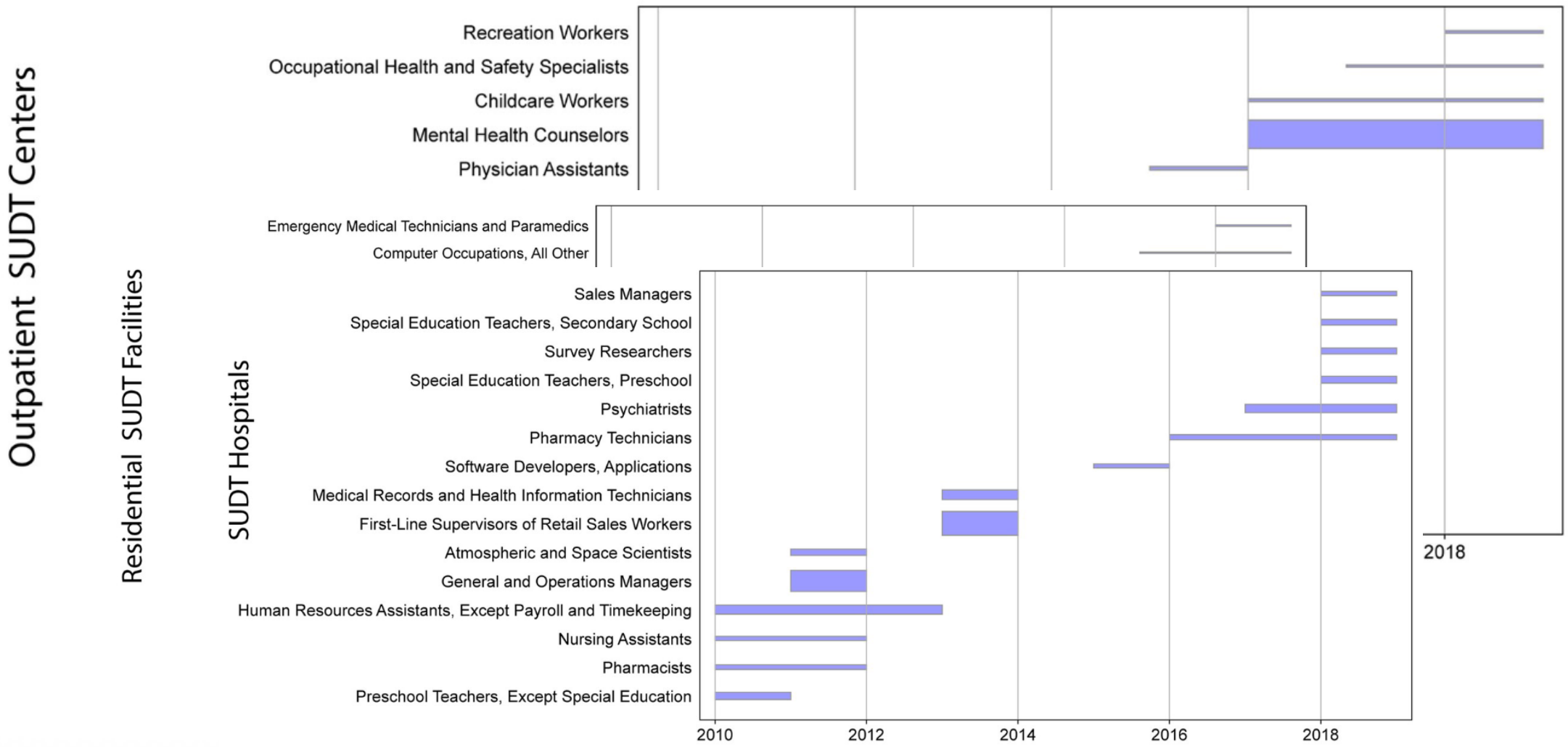


Fig 2. Bursting top-15 SUDT occupations during 2010-2018. Each spike in demand is shown as a horizontal bar with a start and an end date. The length of the bar corresponds to the duration of the hiring burst, the width of the bar shows the burst strength, measured as weight (e.g., in the top panel, the Mental Health Counselor occupation has the strongest and the longest burst in the years 2016–2018).

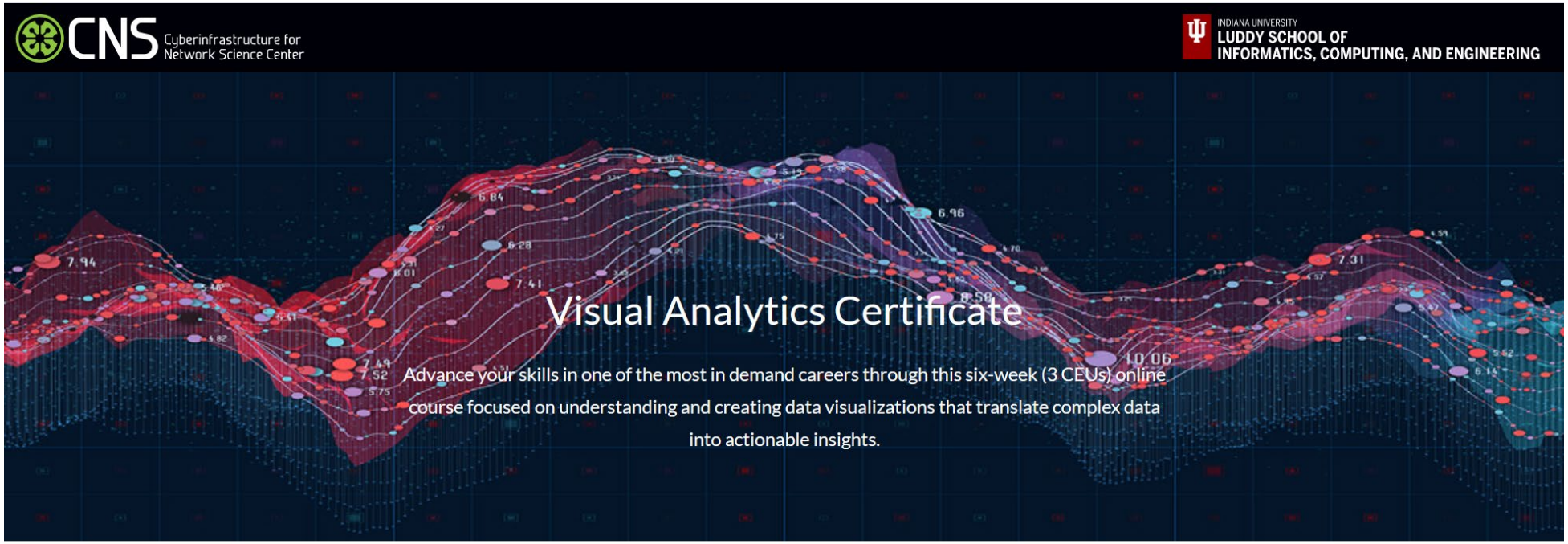
https://doi.org/10.1371/journal.pone.0228394.g002

CNS Cyberinfrastructure for Network Science Center

INDIANA UNIVERSITY SCHOOL OF INFORMATICS, COMPUTING, AND ENGINEERING

Empower Yourself and Others! Data Visualization Literacy

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



FLYER | REGISTER FOR JAN 18–FEB 28, 2021 | FAQS

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

US Employers which have sent students include
The Boeing Company, Eli Lilly, DOE, CDC, NSWC Crane.

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