

The Future of Work

Katy Börner @katycns

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

The Future of Learning & Work Workshop
Futurium, Berlin, Germany

March 15, 2022



Map of Scientific Collaborations from 2005-2009



Computed Using Data from Elsevier's Scopus

The Structure of Science

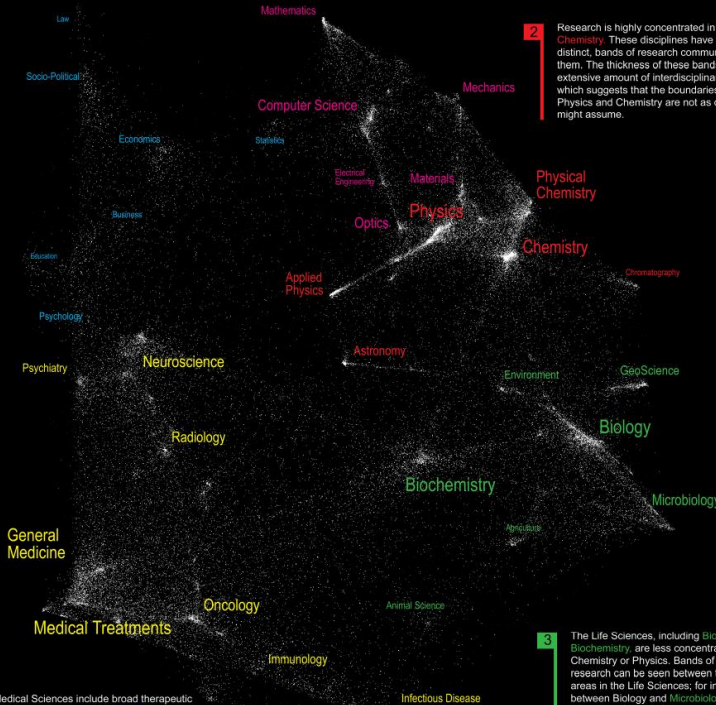
5 The Social Sciences are the smallest and most diffuse of all the sciences. **Psychology** serves as the link between **Medical Sciences (Psychiatry)** and the **Social Sciences, Statistics** serves as the link with **Computer Science** and **Mathematics**.

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.



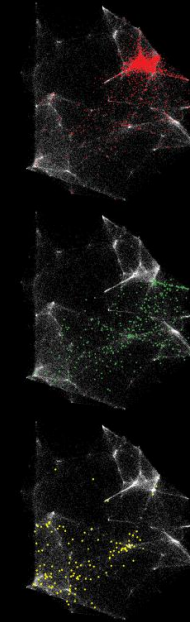
We are all familiar with traditional maps that show the relationships between countries, provinces, states, and cities. Similar relationships exist between the various disciplines and research topics in science. This allows us to map the structure of science.

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

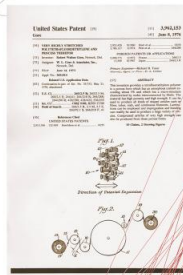
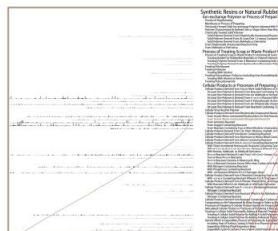
Pharmacogenomics is a relatively new field with most of its activity in **Medicine**. It also has many communities in **Biochemistry** and two communities in the **Social Sciences**.

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 166,523 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.

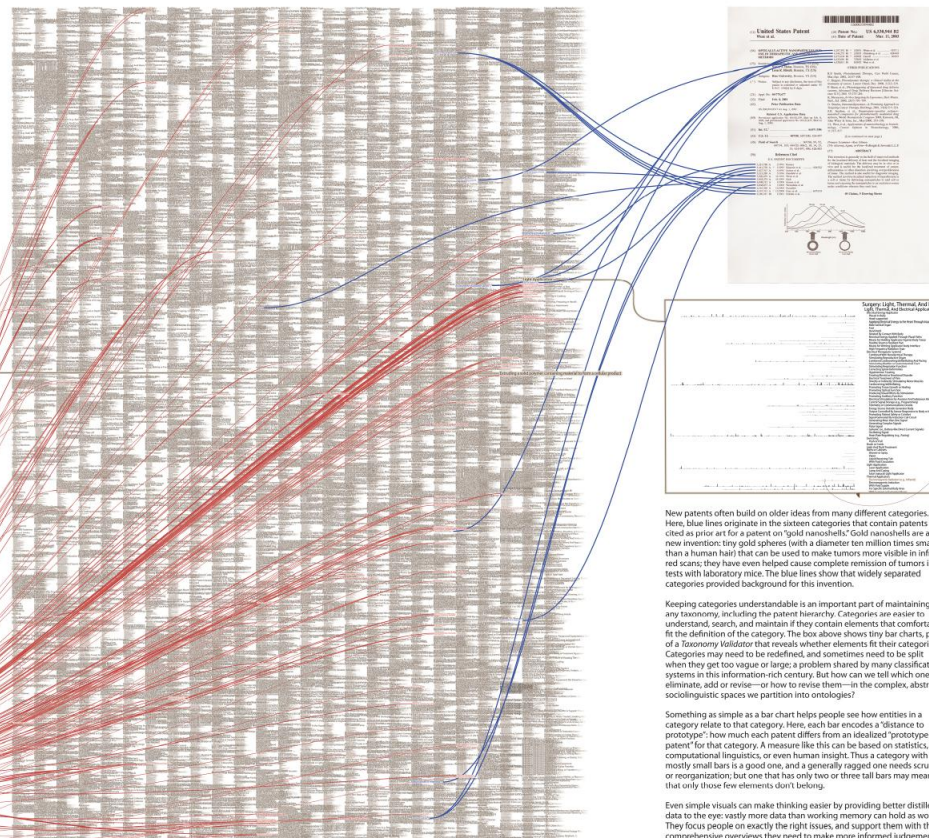
Patent applications are required to be unique and non-obvious, partially by revealing any previous patents that might be similar in nature or provide a foundation for the current invention. In this way we can trace the impact of a single patent, seeing how many patents and categories it affects.

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 1961 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?

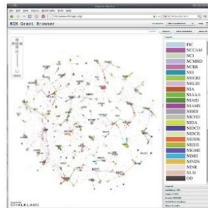
Something as simple as a bar chart helps people see how entities in a category relate to that category. Here, each bar encodes a "distance to prototype": how much each patent differs from an idealized "prototype patent" for that category. A measure like this can be based on statistics, computational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

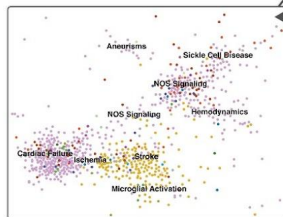
A Topic Map of NIH Grants 2007

Bruce W. Herr II (ChalkLabs & IU), Gully Burns (SI), 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. 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.

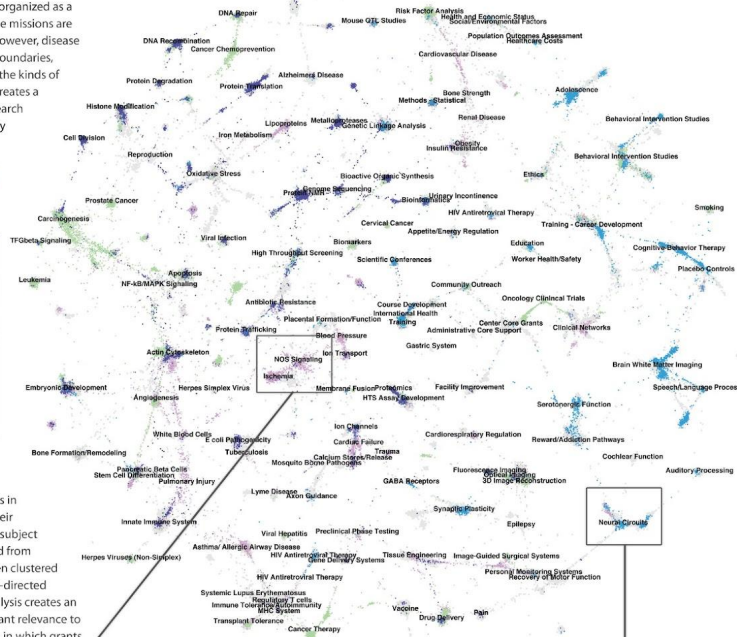
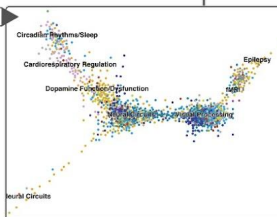


Cardiac Diseases Research

An area of the map focused on 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 Anemias.

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

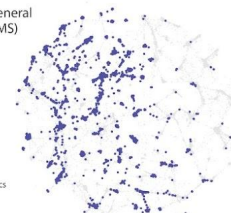
- 1 Oncology Clinical Trials
- 2 Cancer Treatment
- 3 Cancer Therapy
- 4 Carcinogenesis
- 5 Risk Factor Analysis
- 6 Cancer Chemotherapy
- 7 Metastasis
- 8 Leukemia
- 9 Prediction/Prognosis
- 10 Cancer Chemoprevention



National Institute of General Medical Sciences (NIGMS)

TOP 10 TOPICS

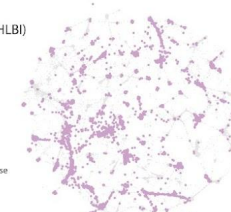
- 1 Bioactive Organic Synthesis
- 2 X-ray Crystallography
- 3 Protein NMR
- 4 Computational Models
- 5 Yeast Biology
- 6 Metalloproteases
- 7 Enzymatic Mechanisms
- 8 Protein Complexes
- 9 Invertebrate/Zebrafish Genetics
- 10 Cell Division



National Heart, Lung, and Blood Institute (NHLBI)

TOP 10 TOPICS

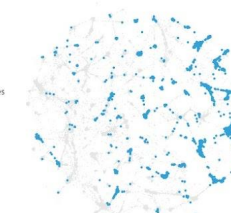
- 1 Cardiac Failure
- 2 Pulmonary Injury
- 3 Genetic Linkage Analysis
- 4 Cardiovascular Disease
- 5 Atherosclerosis
- 6 Hemostasis
- 7 Blood Pressure
- 8 Asthma/ Allergic Airway Disease
- 9 Gene Association
- 10 Lipoproteins



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



Science related Wikipedian ACTIVITY

This visualization explores the activity of science, math, and technology (SMT) related articles in the English-language Wikipedia (<http://en.wikipedia.org>). The central image shows 659,388 articles (circles). Overlaid is a 37 x 37 grid of relevant half-inch sized images.

Blue, green, and yellow circles represent the 3,599 math, 6,874 science, and 3,164 technology related articles respectively. The larger the size of a circle the higher the likelihood it is that type of article. The four corners show activity patterns of SMT articles.

Article Edit Activity

Articles are size coded based on how frequently they have been edited from Feb. 6, 2001 to April 6, 2007. More consideration is given to current and major edits. Larger circles have been edited more frequently than smaller circles.

Article Popularity

Articles are size coded based on the number of Wikipedia articles referencing it. Larger circles are receiving more links from other articles than smaller circles. The highest number of references to an article was 142,602.

2007 Major Edits

Articles are size coded based on how many major edits they received from January 1st, 2007 to April 6th, 2007. Larger circles have received more edits than smaller circles. The highest number of major edits was 2,627.

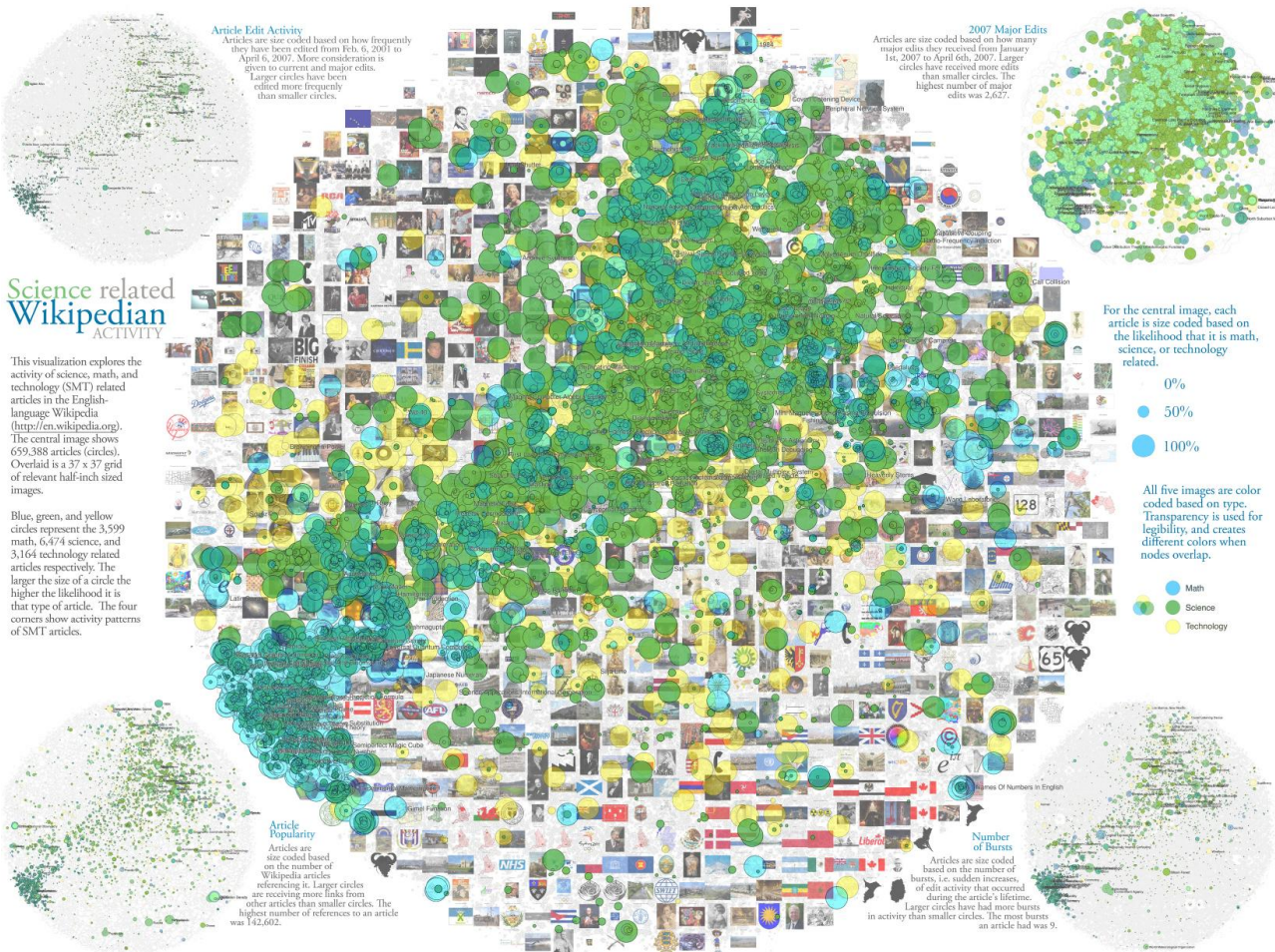
Number of Bursts

Articles are size coded based on the number of bursts, i.e. sudden increases, of edit activity that occurred during the article's lifetime. Larger circles have had more bursts in activity than smaller circles. The most bursts an article had was 9.

For the central image, each article is size coded based on the likelihood that it is math, science, or technology related.



All five images are color coded based on type. Transparency is used for legibility, and creates different colors when nodes overlap.



Diseasome

The Human Disease Network

Explore online at <http://diseasome.eu>

Statistics

of Nodes: 516
 # of Edges: 1188
 Density: 0.0009
 Average Degree: 9.20
 Diameter: 15
 Average Shortest Path: 6.5

Disorder Class

- Cancer
- Endocrine
- Ear, Nose, Throat
- Ophthalmological
- Neurological
- Hematological
- Cardiovascular
- Muscular
- Immunological
- Dermatological
- Nutritional
- Connective Tissue Disorder
- Renal
- Psychiatric
- Metabolic
- Bone
- Skeletal
- Developmental
- Gastrointestinal
- Respiratory
- Multiple
- Unclassified

Top 5 Diseases

1. Deafness
2. Endemia
3. Colon Cancer
4. Retinitis Pigmentosa
5. Diabetes Mellitus

Top 5 Genes

1. TP53
2. PAK3
3. FGFR2
4. RTN1
5. MSH2

Description

This map presents a network of 516 diseases linked by 1188 disorder-disorder associations, indicating the common genetic origin of many diseases.

GENE-BASED LINKS

The map shows a global reference of the genetic links between disorders and a valuable global perspective for physicians, genetic counselors, and biomedical researchers alike. The size (ranked) size lists 15 genetically diverse disorders reported by their affected genes, across the understanding of the causes of disease, and the number of partner genes.

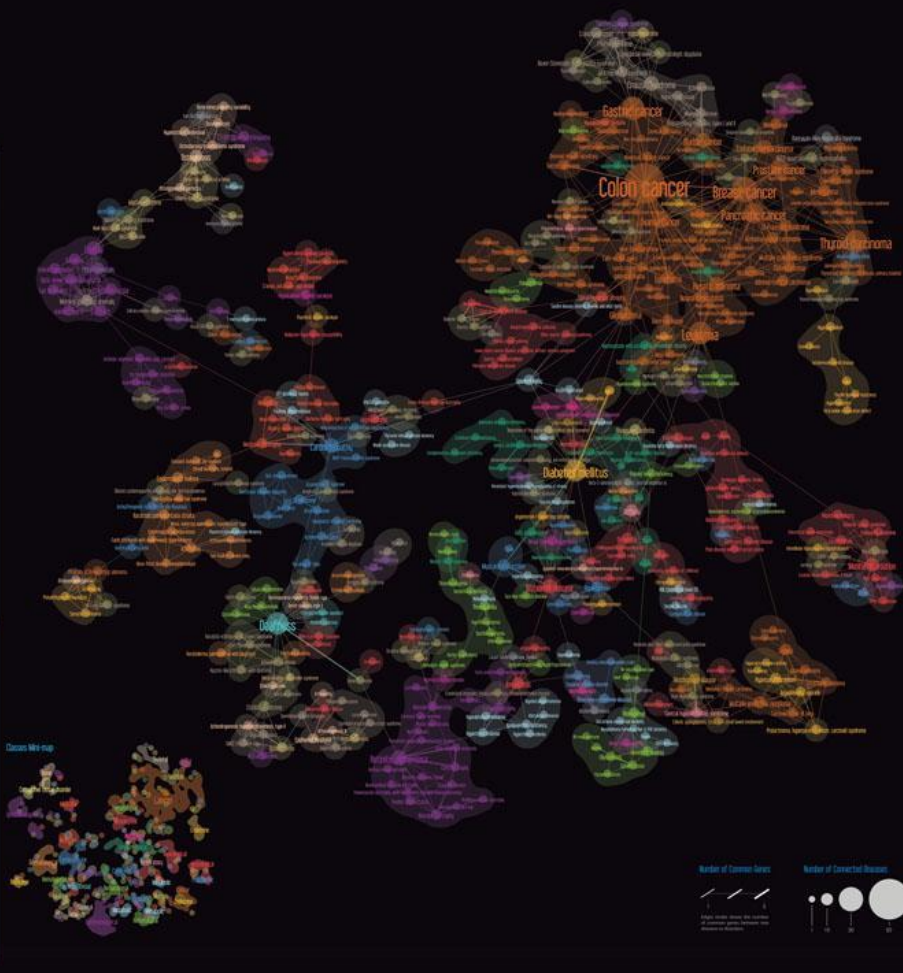
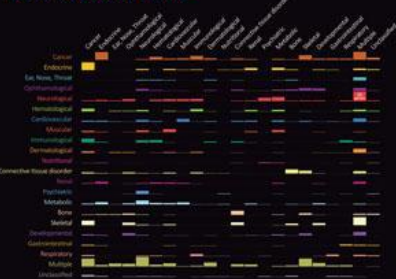
NETWORK VISUALIZATION TOOLS AND APPLS

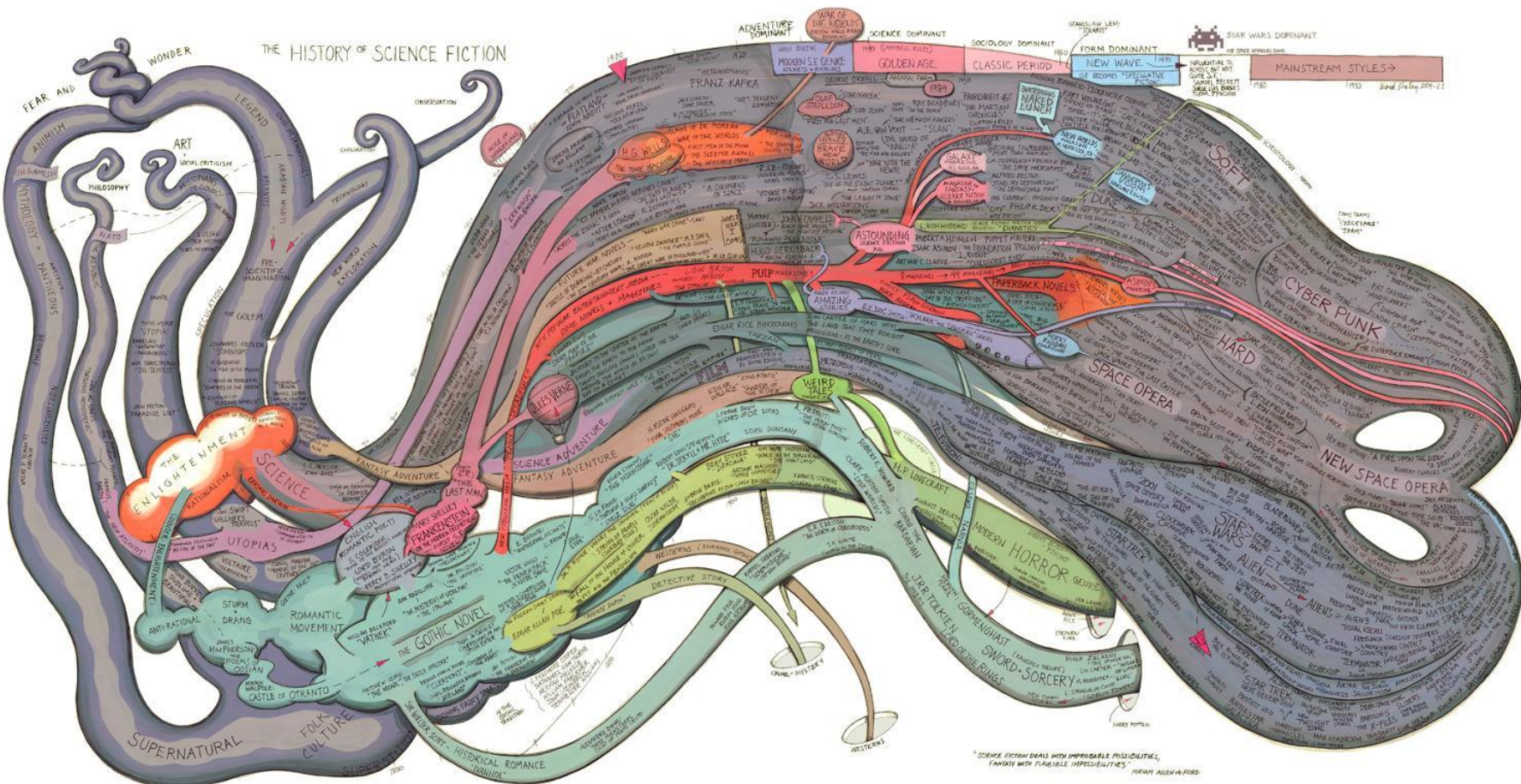
The map was built using the force-directed layout algorithm ForceAtlas2 in Gephi, using color correspondance to the disorder class to which the disorder belongs, and the size to represent the 15 rank degree, the count of number of links, with a size proportional to the number of genes that are reported in both disorders and colored with the average color between those two target classes. Disorders not connected with the average color legend component has been kept. The classes who has shared most connections between common and shared target classes.

The Disorder Class Interaction graph below shows the interaction level between disorder classes, representing the number of shared genes, up to 80.


Mathieu Bastian
 Sébastien Heymann
 10th St South St, Suite 3, Suite 3, 1000 W, Montreal H3A 2G4, Canada
 You can reach us at mathieu@mathieu-bastian.com

Disorder Class Interactions





Check out our **Zoom Maps** online!



VII.10
History of Science Fiction, by Ward Shaffer

BROOKLYN, NY, 2011
Courtesy of Ward Shaffer Studies

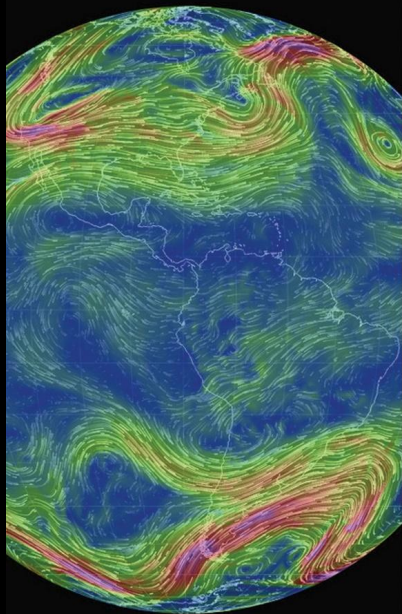
Ward Shaffer is an artist identified with the Williamsburg scene in Brooklyn, New York, about art and culture. This map plots the science fiction literary genre from its rambling origins in the 18th century, through the 19th-century gothic and Victorian eras, to the modernist and postmodernist eras. The map's narrative structure provides and organizes the data in a way that is both visually appealing and easy to understand. The map is a complex, branching tree structure with various sub-sections and text. A zoomed-in inset shows a detailed view of the 'SCIENCE ADVENTURE' and 'FANTASY ADVENTURE' sections, featuring names like 'MARY SHELLEYS FRANKENSTEIN', 'JULIUS VERNE', and 'THE LAST MAN'. The map is set against a dark background with a navigation interface in the top left corner.

PLACES & SPACES

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



MACROSCOPES FOR INTERACTING WITH SCIENCE



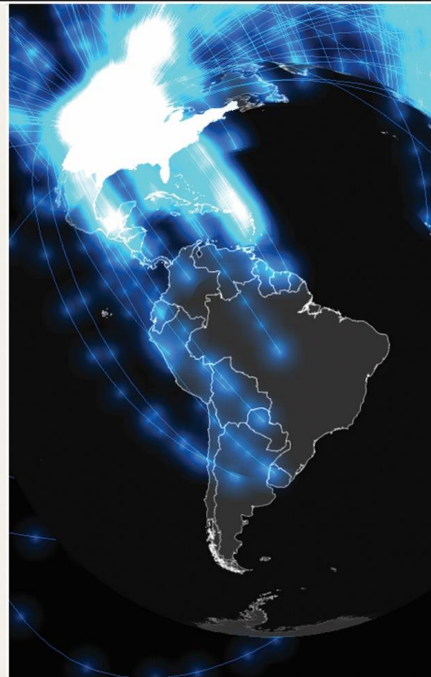
Earth

Weather on a worldwide scale



AcademyScope

Exploring the scientific landscape



Mapping Global Society

Local news from a global perspective

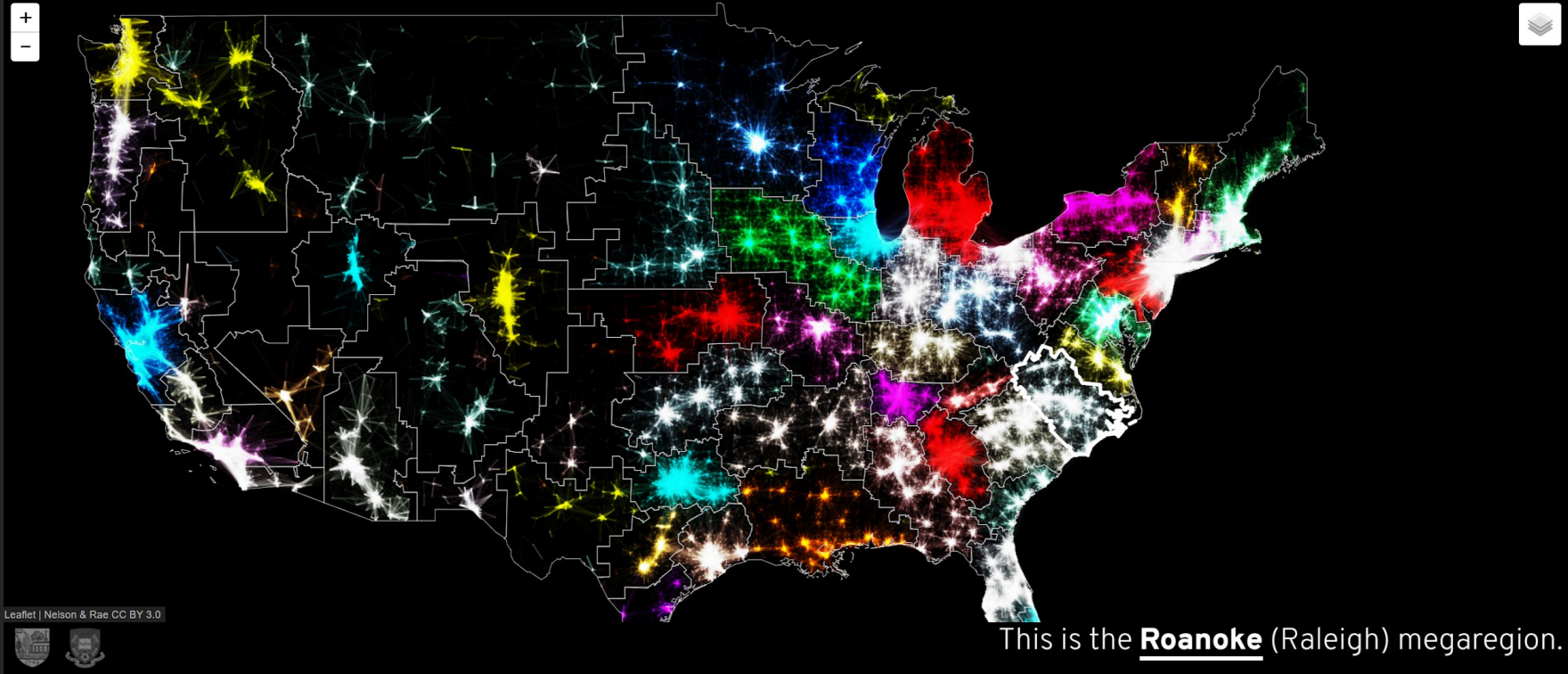


Charting Culture

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 commuteshed.



SMELLY
MAPS



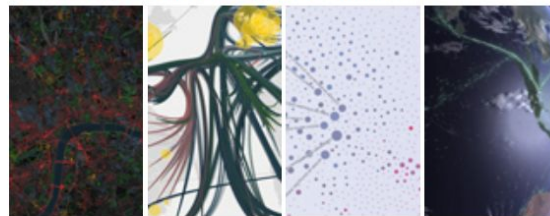
Iteration XI (2015)

Macrosopes for Interacting with Science



Iteration XII (2016)

Macrosopes for Making Sense of Science



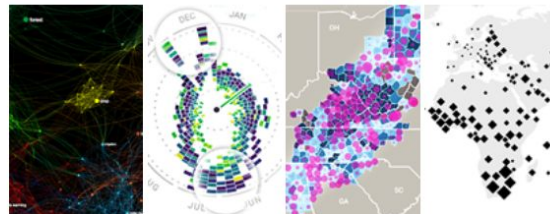
Iteration XIII (2017)

Macrosopes for Playing with Scale



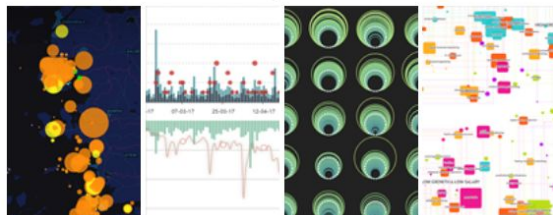
Iteration XIV (2018)

Macrosopes for Ensuring our Well-being



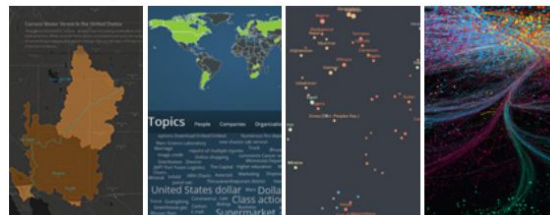
Iteration XV (2019)

Macrosopes for Tracking the Flow of Resources



Iteration XVI (2020)

Macrosopes for Harnessing the Power of Data



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)

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 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^{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^{i,g,i-1}

^aSchool of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN 47408; ^bEducational Technology/Media Centre, Dresden University of Technology, 01062 Dresden, Germany; ^cDepartment of Information Management, Nanjing University of Science and Technology, 210094 Nanjing, China; ^dBurning Glass Technologies, Boston, MA 02110; ^eSchool of Journalism and Communication, Nanjing University, 210008 Nanjing, China; ^fDepartment of Sociology, University of Chicago, Chicago, IL 60637; ^gKnowledge Lab, University of Chicago, Chicago, IL 60637; ^hTencent Research Institute, 100080 Beijing, China; and ⁱSanta 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.

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

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

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 Sciences, Engineering, and Medicine.

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

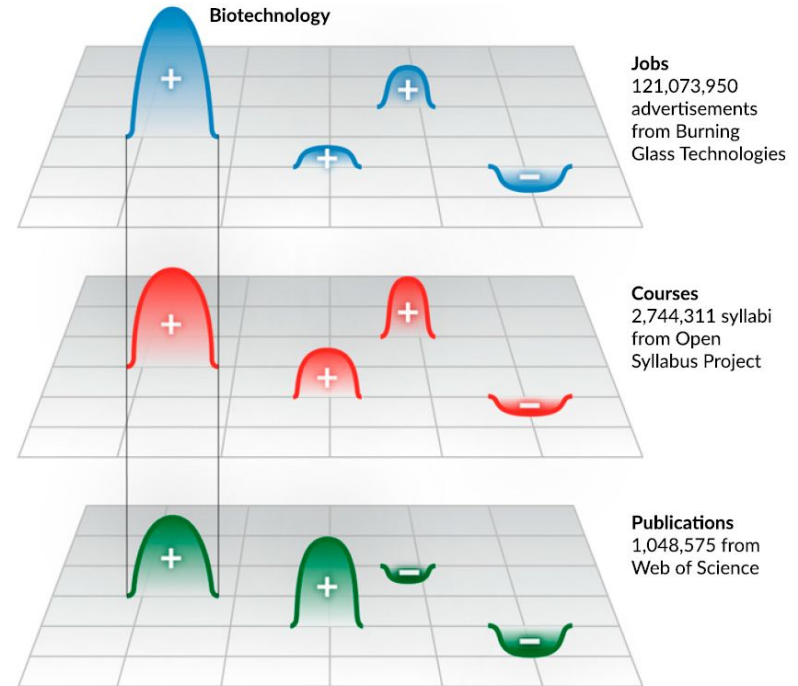
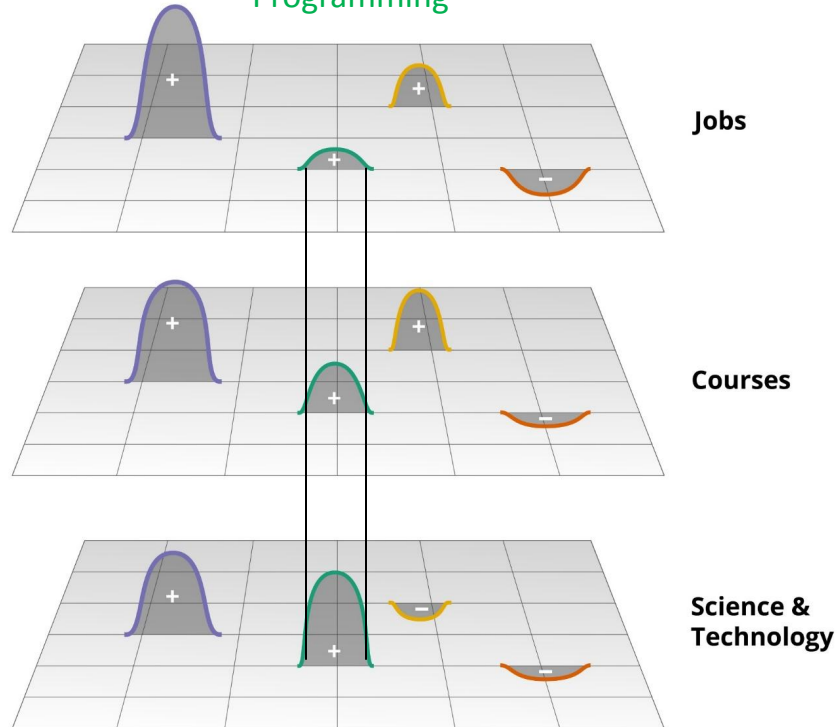
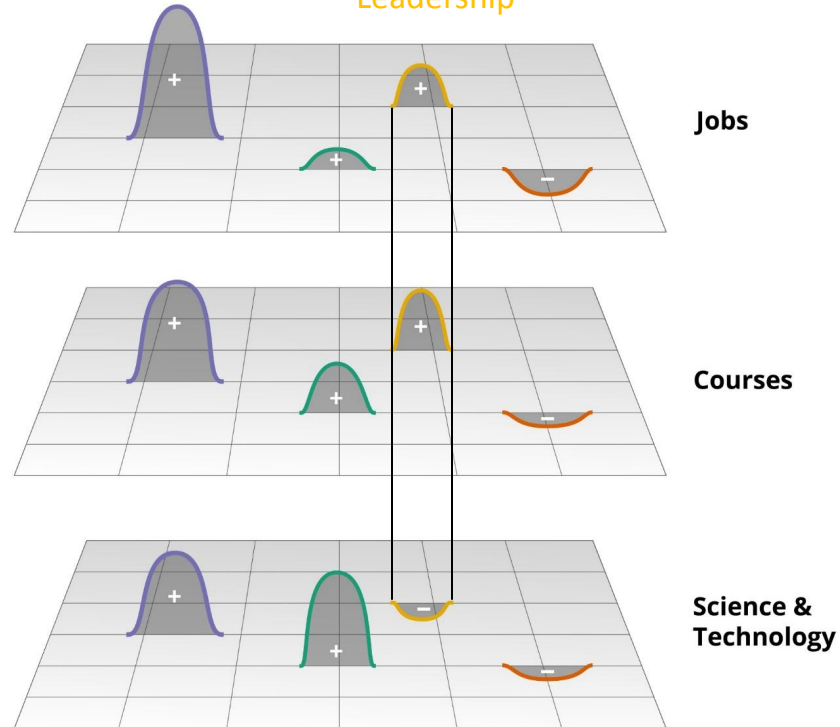


Fig. 1. The interplay of job market demands, educational course offerings, and progress in S&T as captured in publications. Color-coded mountains (+) and valleys (-) indicate different skill clusters. For example, skills related to Biotechnology might be mentioned frequently in job descriptions and taught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



Leadership

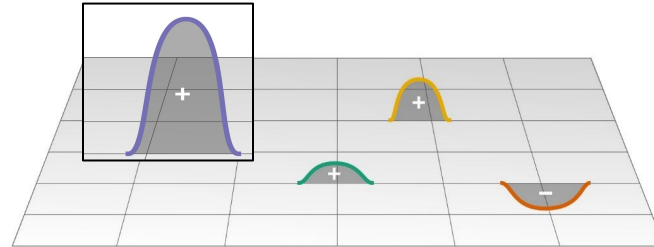


Jobs

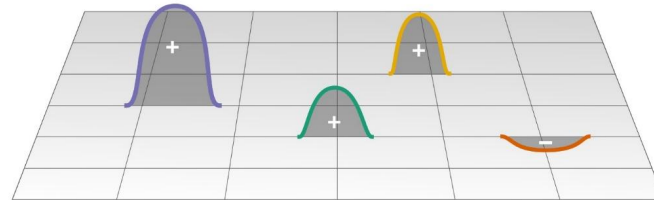
Courses

**Science &
Technology**

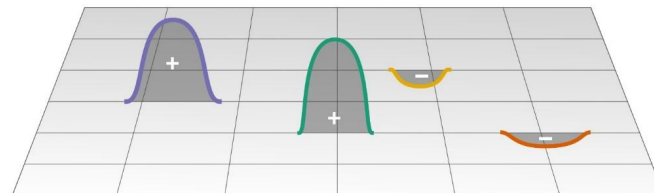
Biotechnology



Jobs

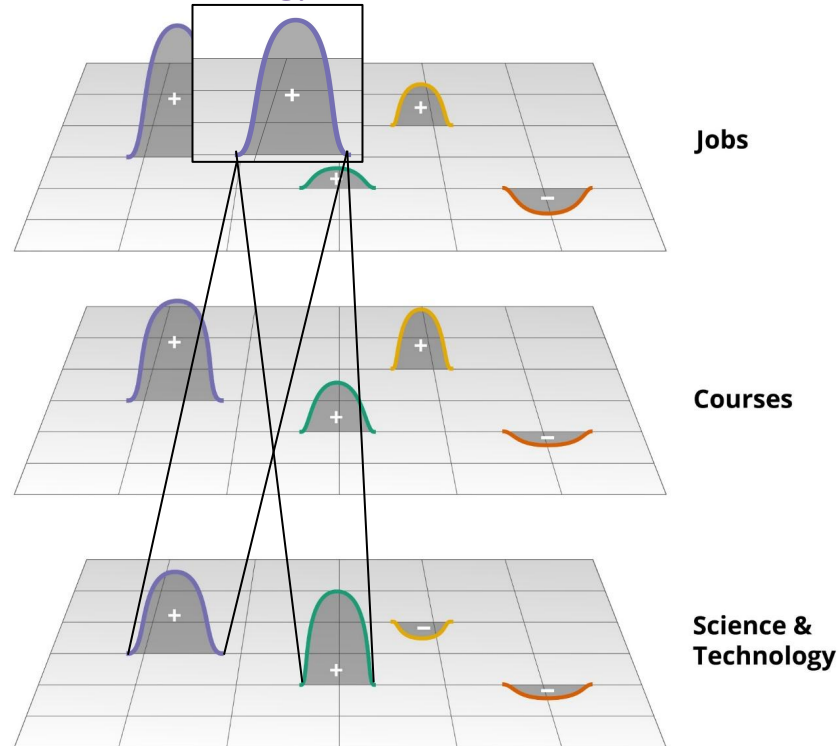


Courses



**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-agile-own.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 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^{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^{i,g,i-1}

^aSchool of Informatics, Computing, and Engineering, Indiana University, Bloomington, IN 47408; ^bEducational Technology/Media Centre, Dresden University of Technology, 01062 Dresden, Germany; ^cDepartment of Information Management, Nanjing University of Science and Technology, 210094 Nanjing, China; ^dBurning Glass Technologies, Boston, MA 02110; ^eSchool of Journalism and Communication, Nanjing University, 210008 Nanjing, China; ^fDepartment of Sociology, University of Chicago, Chicago, IL 60637; ^gKnowledge Lab, University of Chicago, Chicago, IL 60637; ^hTencent Research Institute, 100080 Beijing, China; and ⁱSanta 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.

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

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

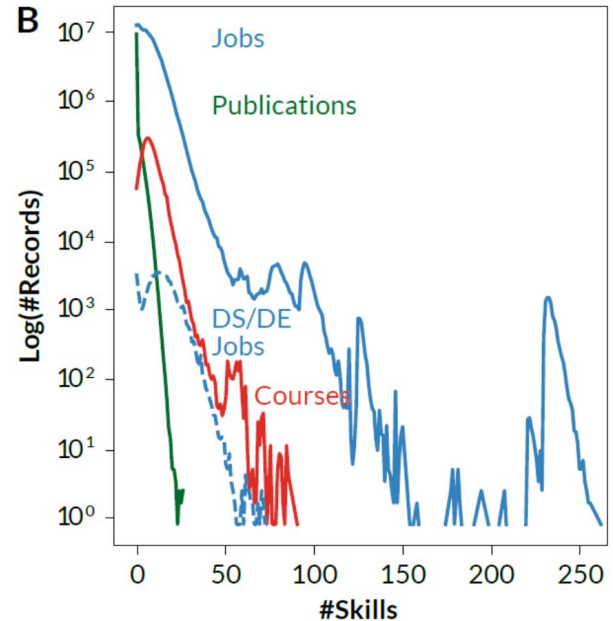
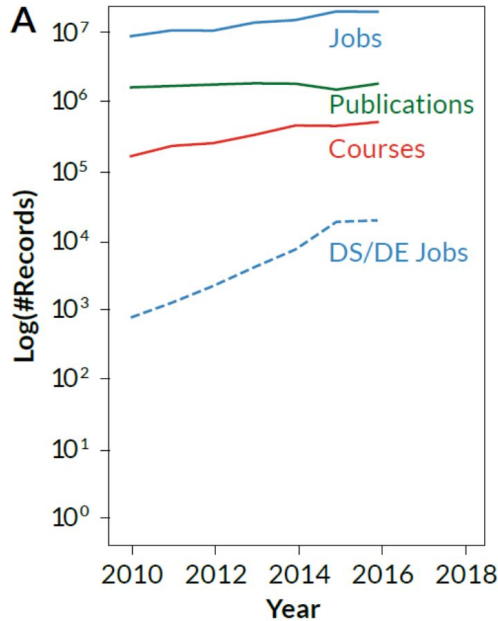
This paper results from the Arthur M. Sackler Colloquium of the National Academy of Sciences, "Modeling and Visualizing Science and Technology Developments," held December 4–5, 2017, at the Arnold and Mabel Beckman Center of the National Academies of Sciences, Engineering, and Medicine.

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.

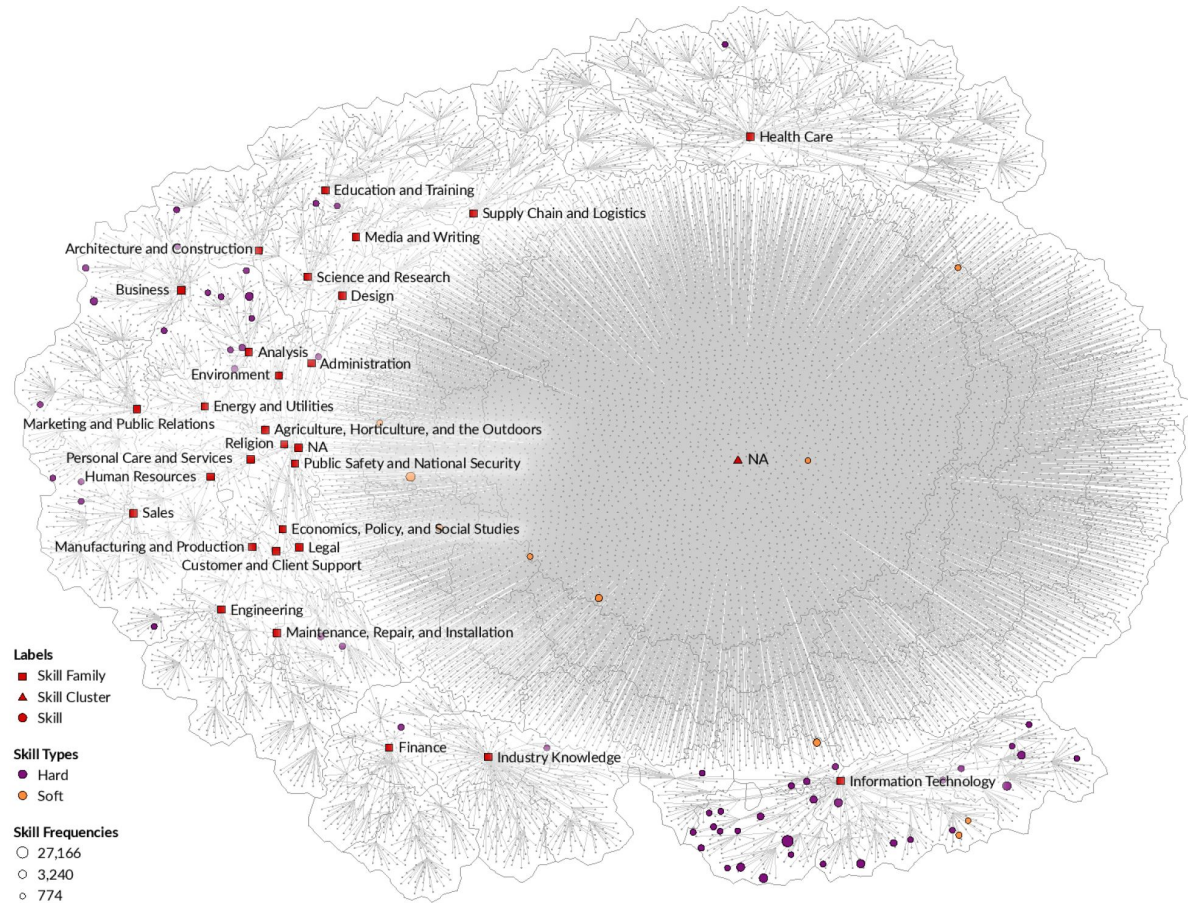
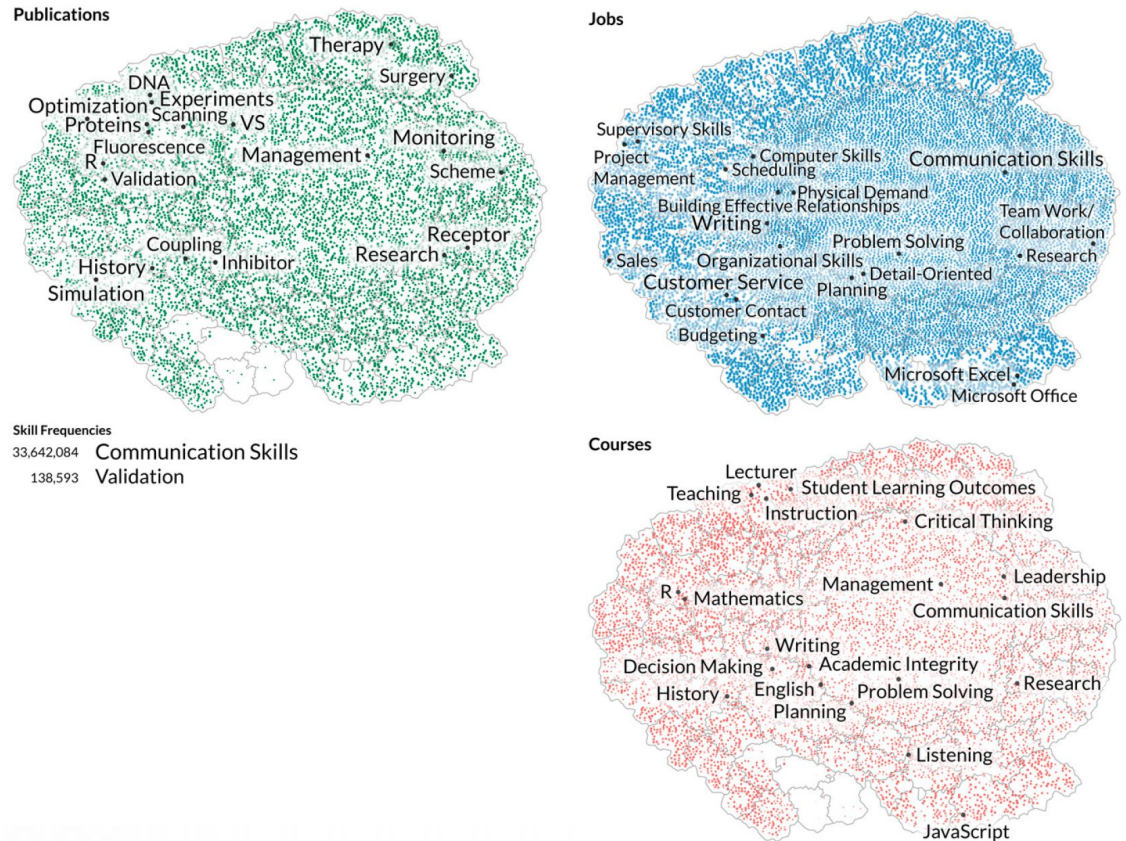


Fig. 3. Basemap of 13,218 skills with overlays of skill frequency in jobs, courses, and publications. This figure substantiates the conceptual drawing in Fig. 1 using millions of data records. Jobs skills are plotted in blue, courses are in red, and publications are in green. Node area size coding indicates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



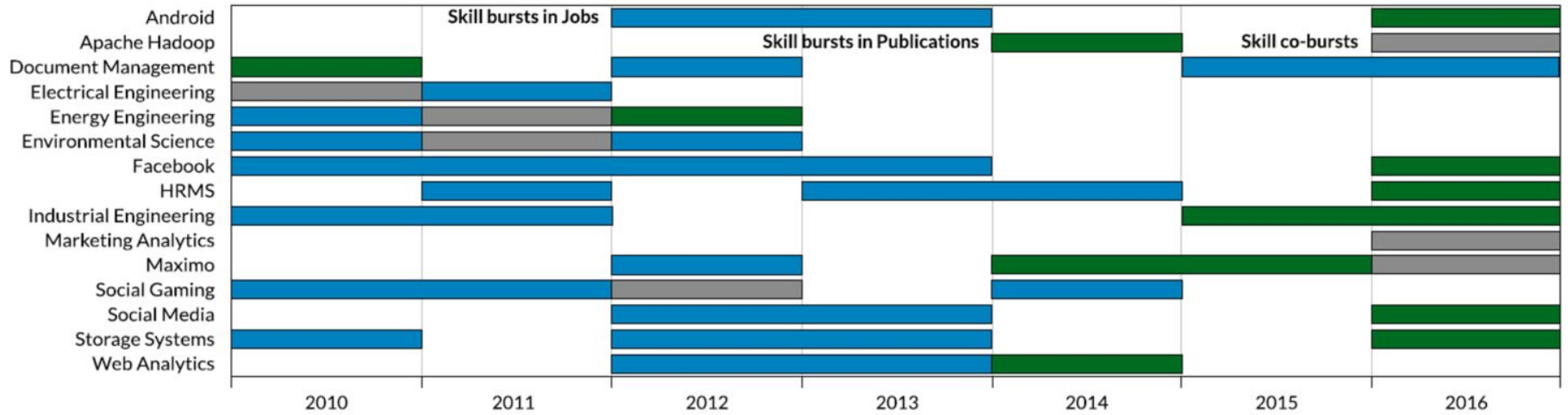


Fig. 4. Burst of activity in DS/DE skills in jobs and publications. Each burst is rendered as a horizontal bar with a start and an end date; skill term is shown on the left. Skills that burst in jobs are blue; skills bursting in publications are green. Seven skills burst in both datasets during the same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

Kullback-Liebler Divergence

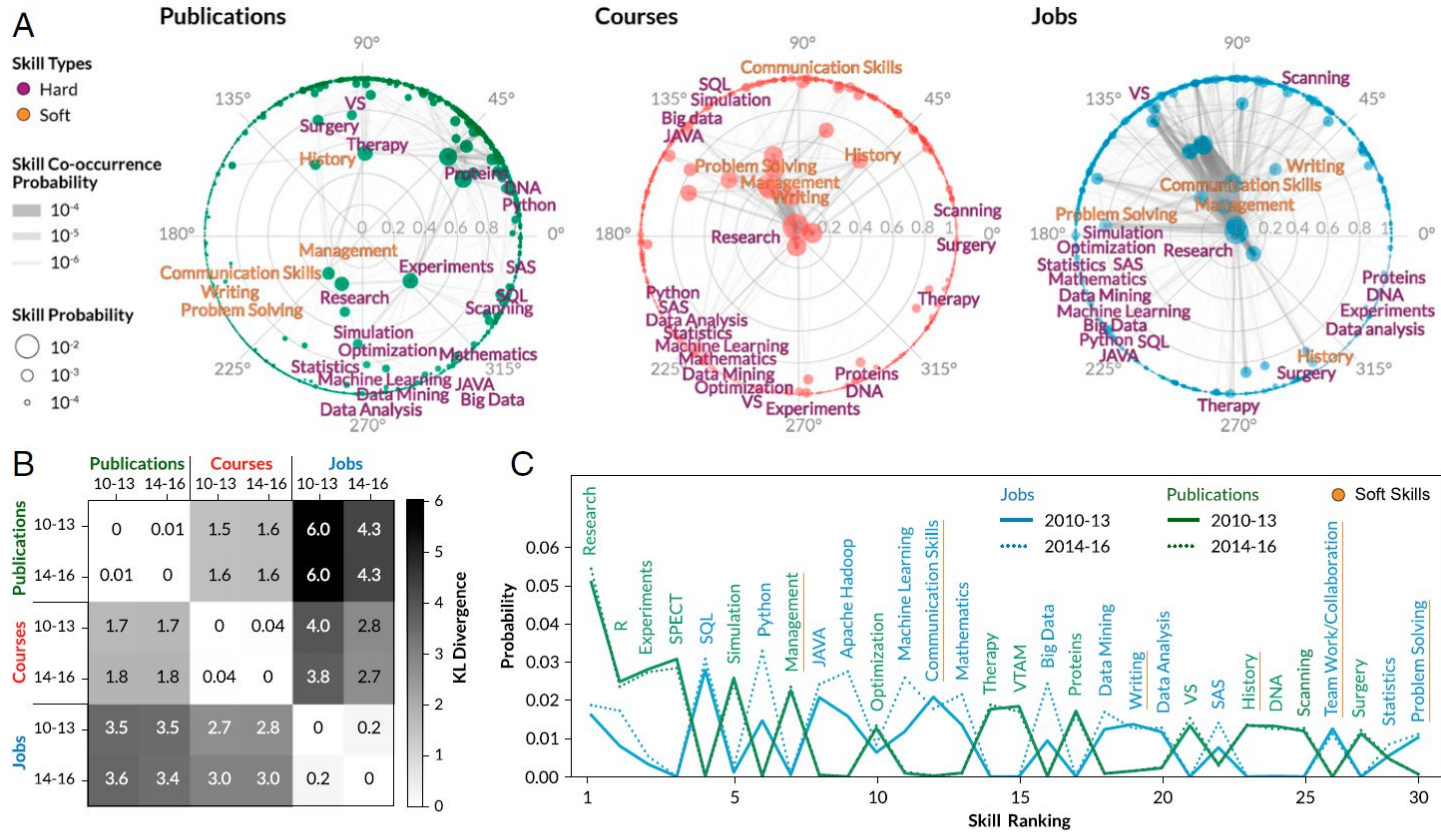


Fig. 5. Structural and dynamic differences between skill distributions in jobs, courses, and publications for 2010–2013 and 2014–2016. (A) Poincaré disks comparing the centrality of soft skills (orange) and hard skills (purple) across jobs, courses, and publications. (B) KL divergence matrix for jobs,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

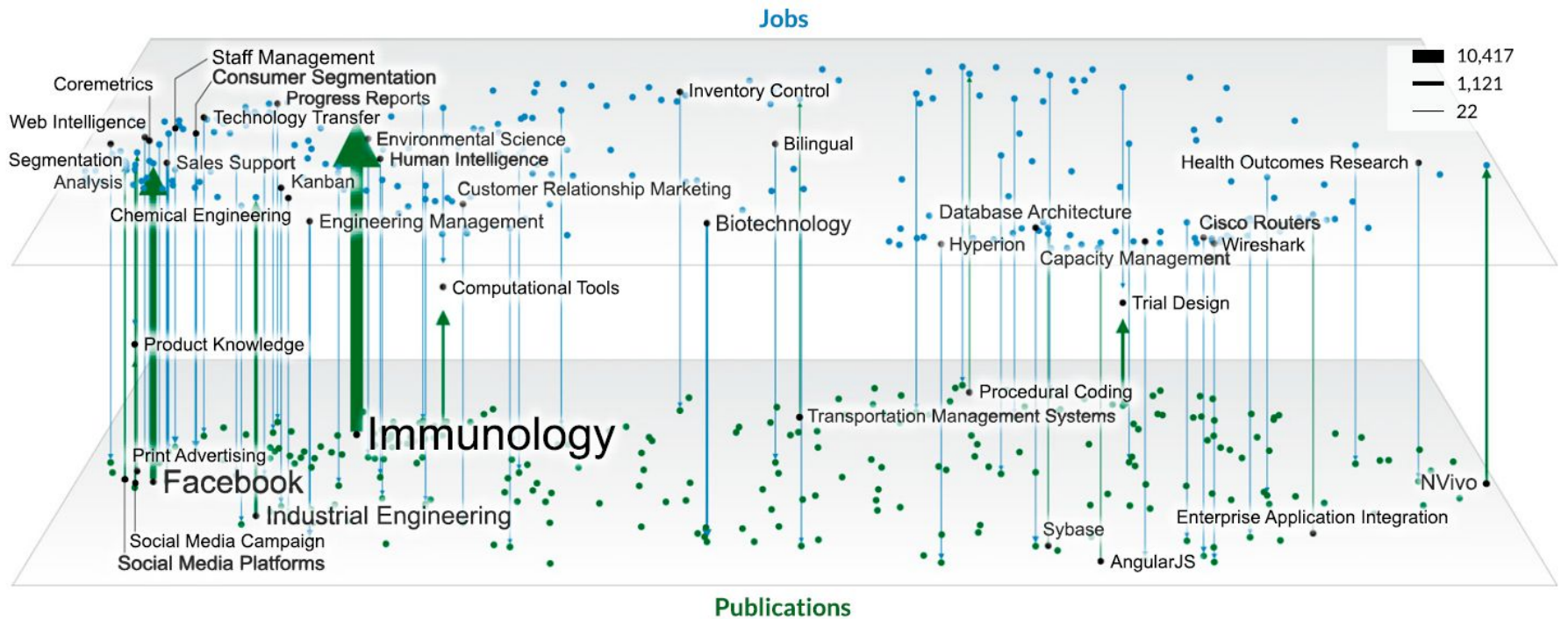
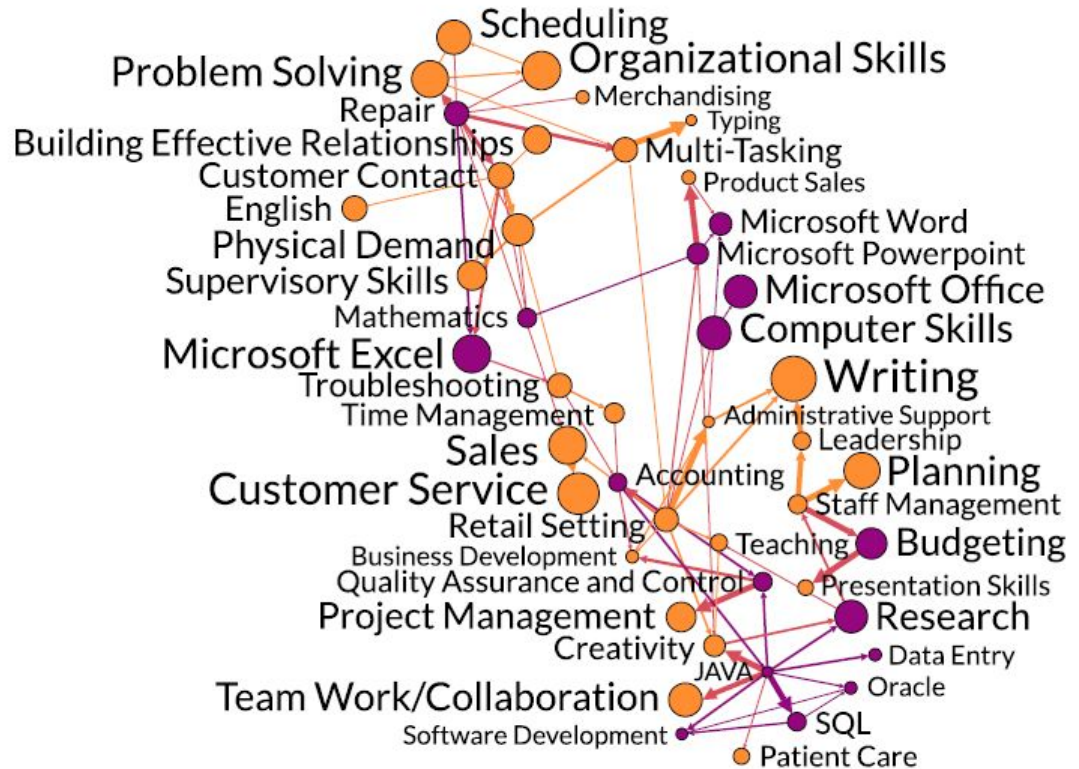


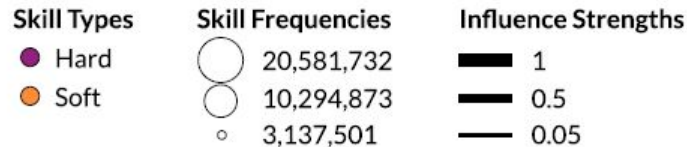
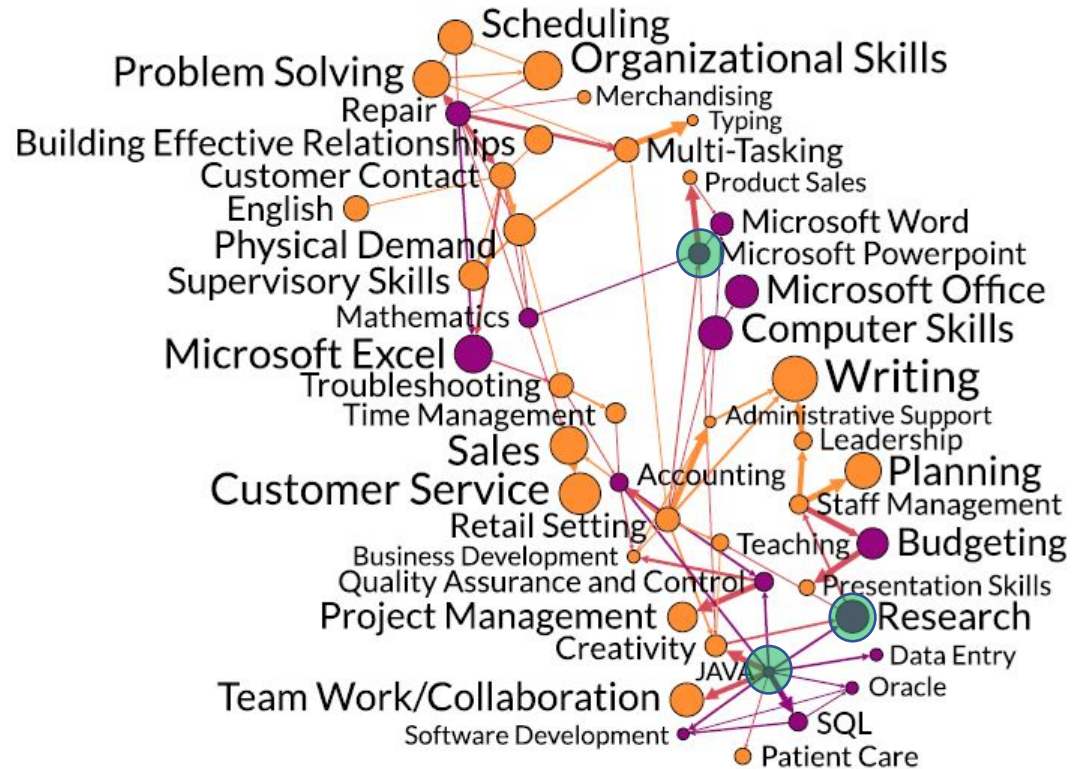
Fig. 6. Strength of influence mapping. Top 200 most frequent skills in jobs (blue) and in publications (green) plotted on the skills basemap from Fig. 2. Arrows represent skills with significant Granger causality (P value < 0.05). Line thickness and label size indicate skill frequency. The direction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



Skill Types	Skill Frequencies	Influence Strengths
● Hard	○ 20,581,732	█ 1
● Soft	○ 10,294,873	█ 0.5
	○ 3,137,501	█ 0.05

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.

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>

Börner, Katy (2021) **Atlas of Forecasts: Modeling and Mapping Desirable Futures**. The MIT Press.

