Modeling and Visualizing Complex Unifiable Systems

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 Network Science Institute (IUNI)
Indiana University, Bloomington, IN, USA

Forum on Complex Unifiable Systems: A Virtual Convocation
National Academy of Engineering

April 28, 2020
The Structure of Science

The Social Sciences are the smallest and most recent of all the sciences. Psychology serves as the link between Medical Sciences and the Social Sciences. Psychology serves as the link with Computer Science and Mathematics.

Mathematics is our starting point, the purest of all sciences. It lies at the outer edge of the map. Computer Science, Electrical Engineering and Photonics are also pure sciences that share open knowledge in Mathematics and Physics. These three disciplines provide a great example of a seamless transition from one pure science (Mathematics) to another (Phyics) through inclusive disciplines. Although applied, these disciplines are highly concentrated with distinct bands of research communicating the link from. Bands indicate interdisciplinary research.

Research is highly concentrated in Physics and Chemistry. These disciplines have far, far fewer distinct bands than the others. The absence of these bands indicates an extensive amount of interdisciplinary research. Nonetheless, Physics and Chemistry are not as distinct as one might assume.

The Medical Sciences include broad therapeutic studies and targeted areas of treatment (e.g. cardiovascular, endocrinologic, gastrointestinal, etc.). Unlike Physics and Chemistry, the medical disciplines are more spread out, indicating a more multi-disciplinary approach to research. The transition into Life Sciences (see related 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 outlined 44 areas of science from the classics (mathematics, 13,800 scientific papers) to the Map of Science (2014) generated at Vlado Mil引力a’s Laboratorium for Advancement of Graph Theory (2014). View this online at stopwatch.mts.illinois.edu. Each dot in the glyph represents one of the 96,000 research communities active in science in 2014. A research community for any topic is a group of papers that are more similar to each other than to any other. Over time, communities can be born, continue, split, merge, or die.

The Map of Science can be used as a tool for students. It is the terrain in which organizations and societies orientate their scientific exploration. Additional information about the scientific and economic impact of each research community allows policy makers to decide which areas to exploit, expand, fund, or ignore.

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

Nonotechnology
Most research communities in nontechnology are concentrated in Physics, Chemistry, and Materials Science. However, many disciplines in the Life and Medical Sciences also have nano-technology applications.

Proteomics
Research communities in proteomics are concentrated biochemistry. In addition, there is a heavy focus in the fields related to chemistry, such as biochemistry.

The balance of the proteomics communities are widely dispersed among the Life and Medical Sciences.

Pharmacogenetics
Pharmacogenetics is a relatively new field with most of its activity in Medicine. It also has many connections with the other communities in the Social Sciences.
The US Patent Hierarchy

Impact

The United States Patent and Trademark Office does not collect and analyze patent data to identify emerging trends or evaluate the impact of patents. Instead, it is concerned with granting patents to inventors. However, there are tools that can be used to analyze patent data, such as the Patent Indexer and the Patent Information System. These tools provide information on the number of patents granted in a particular field, the number of applications in a field, and the number of patents that are cited in other patents.


Prior Art

New patents often build on older ideas from many different categories. Here, the figure shows the patent landscape of a particular category. For example, if you were looking for patents on medical devices, you might find a large number of patents on catheters and other medical tools.

Keeping categories under Maine is an important part of maintaining any library or repository, especially in a large category like patent data. However, if you search for a specific category, such as medical devices, you might find a relatively small number of patents.

Categorizing patents can be done in a number of ways, from simple text-based categorization to complex algorithms that analyze the text of the patents. For example, you might categorize patents based on the inventor or the date the patent was granted.

Overall, the visualization of patent data can be a powerful tool for understanding the landscape of innovation and identifying emerging trends.
A Topic Map of NIH Grants 2007

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 fluid 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.nihmap.org. Institute abbreviations can be found at www.nih.gov/NCI.

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 analytic 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 dysfunctions. Cardiac failure (primarily funded by NHLBI) is topologically clustered near the Starke (NIEHS). Close to the two major medical emergencies associated with ischemia, which results from restricted blood supply. Also located in this area are grants focused on Nitric Oxide (iNOS) Signaling, a major biochemical pathway for vasodilation, and grants on Hemodynamics, Solute-Cell Balance, and Inflammation.

Neural Circuits Research

Also 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 Cardiac/Pulmonary Physiology, primarily funded by NHLBI. Processing, primarily funded by NID, and Hypothalamic/Pituitary, primarily funded by NIMH. For scale, we are in the upper-left throat.
This visualization explains the activity of science, math, and technology (SM&T) related articles in the English-language Wikipedia (http://en.wikipedia.org). The central image shows 455,888 articles (nodes). Overlaid is a 35 x 35 grid of robobirds that visit these nodes.

Blue, green, and yellow circles represent the 3,509 math, 6,634 science, and 3,763 technology-related articles, respectively. The larger the size of a circle the higher the likelihood that it is that type of article. The four corners show activity patterns of SM&T articles.

![Visualization of Science-Related Wikipedian Activity](image-url)

For the central image, each article is color-coded based on the likelihood that it is math, science, or technology-related:

- 0%
- 50%
- 100%

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

---

Check out our **Zoom Maps** online!

Visit [scimaps.org](http://scimaps.org) and check out all our maps in stunning detail!
Smelly Maps – Daniele Quercia, Rossano Schifanella, and Luca Maria Aiello – 2015
This is the Roanoke (Raleigh) megaregion.
Iteration XII (2016)
Macroscopes for Making Sense of Science

Iteration XIII (2017)
Macroscopes for Playing with Scale

Iteration XIV (2018)
Macroscopes for Ensuring our Well-being

Iteration XV (2019)
Macroscopes for Tracking the Flow of Resources

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

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

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

The David J. Sencer CDC Museum, Atlanta, GA.

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

Conference slides, recordings, and report are available at http://modsti.cns.iu.edu/report
Modeling and Visualizing Science and Technology Developments
National Academy of Sciences Sackler Colloquium, December 4-5, 2017, Irvine, CA

Rankings and the Efficiency of Institutions
H. Eugene Stanley | Albert-László Barabási | Lada Adamic | Marta González | Kaye Hubsds Fealing | Brian Uzzi | John V. Lombardi

Higher Education and the Science & Technology Job Market
Katy Börner | Wendy L. Martinez | Michael Richey | William Rouse | Stasa Milojevic | Rob Rubin | David Krakauer

Innovation Diffusion and Technology Adoption
William Rouse | Donna Cox | Jeff Alstott | Ben Shneiderman | Rahul C. Basole | Scott Stern | Cesar Hidalgo

Modeling Needs, Infrastructures, Standards
Paul Trunfio | Sallie Keller | Andrew L. Russell | Guru Madhavan | Azer Bestavros | Jason Owen-Smith

nasonline.org/Sackler-Visualizing-Science
Arthur M. Sackler Colloquium on Modeling and Visualizing Science and Technology Developments

🌞 Twin-Win Model: A human-centered approach to research success
Ben Shneiderman

🌞 Forecasting innovations in science, technology, and education
FROM THE COVER
Katy Börner, William B. Rouse, Paul Trufio, and H. Eugene Stanley
PNAS December 11, 2018 115 (50) 12573-12581; first published December 10, 2018. https://doi.org/10.1073/pnas.1818750115

🌞 How science and technology developments impact employment and education
Wendy Martinez

🌞 Scientific prize network predicts who pushes the boundaries of science
Yifang Ma and Brian Uzzi

🌞 The role of industry-specific, occupation-specific, and location-specific knowledge in the growth and survival of new firms
C. Jara-Figueroa, Bogang Jun, Edward L. Glaeser, and Cesar A. Hidalgo
PNAS December 11, 2018 115 (50) 12646-12653; first published December 10, 2018. https://doi.org/10.1073/pnas.1800475115

https://www.pnas.org/modeling
Skill Discrepancies Between Research, Education, and Jobs Reveal the Critical Need to Supply Soft Skills for the Data Economy

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

Challenges:

• Rapid change of STEM knowledge
• Increase in tools, AI, robotics
• Social skills (project management, team leadership) become ever more important
• Increasing team size

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 the 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](https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/uk-futures/london-futures-agiletown.pdf)

• The aerospace industry and NASA have a disproportionately large percentage of workers aged 50 and older compared to the national average, and up to half of the current workforce will be eligible for retirement within the coming five years. Astronautics AIAA (2012) Recruiting, retaining, and developing a world-class aerospace workforce. [https://www.aiaa.org/uploadedFiles/Issues_and_Advocacy/Education_and_Workforce/Aerospace%20Workforce-%20030112.pdf](https://www.aiaa.org/uploadedFiles/Issues_and_Advocacy/Education_and_Workforce/Aerospace%20Workforce-%20030112.pdf)


• COVID-19 economic freeze could cost 47 million jobs and send the unemployment rate past 32%, according to St. Louis Fed projections.
NSF RAISE: C-Accel Pilot: Analytics-Driven Accessible Pathways To Impacts-Validated Education (ADAPTIVE)

**Goal:** Development of data-driven tools to support the tens of millions of US workers whose jobs are being transformed by Artificial Intelligence (AI), automation, COVID-19, and other developments.

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

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

Develop and deploy socio-technical systems that encourage US workers to explore the evolving landscape of new jobs and re/up-skilling opportunities—to not only recover from current risks/crisis but to leap forward into more resilient and more desirable futures.
Data Visualization Literacy

Data Visualization Literacy (DVL)

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

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

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

Data Visualization Literacy Framework (DVL-FW)

Consists of two parts:

DVL Typology
Defines 7 types with 4-17 members each.

DVL Workflow Process
Defines 5 steps required to render data into insights.
## Typology of the Data Visualization Literacy Framework

<table>
<thead>
<tr>
<th>Insight Needs</th>
<th>Data Scales</th>
<th>Analyses</th>
<th>Visualizations</th>
<th>Graphic Symbols</th>
<th>Graphic Variables</th>
<th>Interactions</th>
</tr>
</thead>
<tbody>
<tr>
<td>• categorize/cluster</td>
<td>• nominal</td>
<td>• statistical</td>
<td>• table</td>
<td>• geometric symbols</td>
<td>• spatial</td>
<td>• zoom</td>
</tr>
<tr>
<td>• order/rank/sort</td>
<td>• ordinal</td>
<td>• temporal</td>
<td>• chart</td>
<td>point</td>
<td>position</td>
<td>search and locate</td>
</tr>
<tr>
<td>• distributions (also outliers, gaps)</td>
<td>• interval</td>
<td>• geospatial</td>
<td>• graph</td>
<td>line</td>
<td></td>
<td>filter</td>
</tr>
<tr>
<td>• comparisons</td>
<td>• ratio</td>
<td>• topical</td>
<td>• map</td>
<td>area</td>
<td></td>
<td>details-on-demand</td>
</tr>
<tr>
<td>• trends (process and time)</td>
<td></td>
<td>• relational</td>
<td>• tree</td>
<td>surface</td>
<td></td>
<td>history</td>
</tr>
<tr>
<td>• geospatial</td>
<td></td>
<td></td>
<td>• network</td>
<td>volume</td>
<td></td>
<td>extract</td>
</tr>
<tr>
<td>• compositions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>link and brush</td>
</tr>
<tr>
<td>(also of text)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>projection</td>
</tr>
<tr>
<td>• correlations/relationships</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>distortion</td>
</tr>
</tbody>
</table>

---

Data Visualization Literacy Framework (DVL-FW)

Consists of two parts that are interlinked:

DVL Typology +
DVL Workflow Process

1. Stakeholders
2. Acquire
3. Analyze
4. Visualize
5. Graphic Symbol Types
6. Graphic Variable Types
7. Interaction Types

Insight Need Types

Data Scale Types

Analysis Types
Data Visualization Literacy Framework (DVL-FW)

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

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

**Form:**
- Size
- Shape
- Rotation (Orientation)

**Color:**
- Value (Lightness)
- Hue (Tint)
- Saturation (Intensity)

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

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

**Motion:** Speed, Velocity, Rhythm
### Graphic Symbol Types

<table>
<thead>
<tr>
<th>Spatial Position</th>
<th>Geometric Symbols</th>
<th>Linguistic Symbols</th>
<th>Pictorial Symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Point</td>
<td>Line</td>
<td></td>
</tr>
<tr>
<td>Size</td>
<td>![Size Symbol]</td>
<td>![Size Symbol]</td>
<td>![Size Symbol]</td>
</tr>
<tr>
<td>Shape</td>
<td>![Shape Symbol]</td>
<td>![Shape Symbol]</td>
<td>![Shape Symbol]</td>
</tr>
<tr>
<td>Value</td>
<td>![Value Symbol]</td>
<td>![Value Symbol]</td>
<td>![Value Symbol]</td>
</tr>
<tr>
<td>Color</td>
<td>![Color Symbol]</td>
<td>![Color Symbol]</td>
<td>![Color Symbol]</td>
</tr>
<tr>
<td>Hue</td>
<td>![Hue Symbol]</td>
<td>![Hue Symbol]</td>
<td>![Hue Symbol]</td>
</tr>
<tr>
<td>Saturation</td>
<td>![Saturation Symbol]</td>
<td>![Saturation Symbol]</td>
<td>![Saturation Symbol]</td>
</tr>
<tr>
<td>Granularity</td>
<td>![Granularity Symbol]</td>
<td>![Granularity Symbol]</td>
<td>![Granularity Symbol]</td>
</tr>
<tr>
<td>Texture</td>
<td>![Texture Symbol]</td>
<td>![Texture Symbol]</td>
<td>![Texture Symbol]</td>
</tr>
<tr>
<td>Pattern</td>
<td>![Pattern Symbol]</td>
<td>![Pattern Symbol]</td>
<td>![Pattern Symbol]</td>
</tr>
<tr>
<td>Blur</td>
<td>![Blur Symbol]</td>
<td>![Blur Symbol]</td>
<td>![Blur Symbol]</td>
</tr>
<tr>
<td>Speed</td>
<td>![Speed Symbol]</td>
<td>![Speed Symbol]</td>
<td>![Speed Symbol]</td>
</tr>
</tbody>
</table>

**Qualitative**
- Also called: Categorical Attributes
- Identity Channels

**Quantitative**
- Also called: Ordered Attributes
- Magnitude Channels

Visual Analytics Certificate

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

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

Next course starts Sept 7, 2020
References


http://www.pnas.org/content/vol101/suppl_1


