Actionable Data Visualizations

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

Innovation Growth Lab Global Conference
Boston, MA

June 13, 2018
Maps of Science, Technology & Innovation

Using advanced data mining and visualization techniques to render large-scale datasets into actionable knowledge.
Maps of Science & Technology
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.

100 maps and 12 macroscopes by 215 experts on display at 354 venues in 28 countries.
Science & Technology Outlook: 2005–2055

MAP THEMES

**Small World**

Over 20 years of basic research and development at the 100-meter scale, the importance of nanotechnology as a source of innovations and new capabilities in everything from materials science to medicine is already well-understood. These trends, however, will define how nanotechnology will unfold, and what impacts it will have. First, nanotechnology is not a single field with a coherent intellectual program; it is an opportunistic hybrid, shaped by the catchment of fundamental research questions, promising technical applications, and venture and state capital. Second, nanotechnology is moving away from the classicalvision of small-scale mechanical engineering in which an assemblage builds mechanical systems from individual atoms—toward one in which molecular biology and biochemistry contribute essential tools. That is, proteins that build remarkably. Finally, nanotechnology will also serve as a model for transdisciplinary science. It will support both fundamental research and commercially oriented innovation, and it will be conducted not just by conventional academic or corporate research departments, but in institutions that encourage collaboration and, socially, at large, is understood.

**Intentional Biology**

For 3.5 billion years, evolution has governed biology on this planet. But today, Nature has not yet investigated: Intensive research to read and write the genetic code will still be the top priority of our ability to manipulate biology from the bottom up. We will only genetically engineer existing life but likely not create new life forms or purpose. Still, we will be left to decide what nature has to offer. Evolution is stochastic, and the smallest scales will be a rich source of inspiration as we build the bio-technical system of the next 50 years.

**Extended Self**

The next 50 years, we will be faced with broad opportunities to rewire our minds and bodies in profoundly different ways. Advances in biotechnology, brain sciences, information technology, and robotics will lead to an array of methods to dramatically alter, enhance, and extend the physical and neural function that nature has bestowed upon us. Working these tools as our tools, humans will begin to define a variety of different "transformed" states that is, ways of being and doing that extend beyond what we consider normal for our species. In the very long term, following these paths could someday lead to an evolutionary leap for humanity.

**Mathematical World**

The ability to produce, manipulate, and ultimately understand patterns in enormous amounts of data will allow decades of previously mysterious processes in everything from biological to social systems. Scientists are learning that the core of many biological phenomena—reproduction, growth, repair, and other—are computational processes that can be identified and simulated. Using techniques of combinatorial science to view such patterns—and those that are physical, biological, or social—will let us see new, increasing share of computing cycles in the next 50 years. Such massive computations will also make simulations widespread. Computer simulation will be used not only to help make decisions about large complex scientific and social problems but also to help individuals make better choices in their daily lives.

**Seascape Transformery**

In the next ten years, physical objects, places, and even human beings themselves will increasingly become entangled with computational devices that can sense, understand, and act upon their environment. They will be able to react to environmental clues about the physical, social, and even emotional state of people and things in their surroundings. As a result, increasing demand for goods and services that can be provided, and other sensory abilities. Information previously encoded as text and numbers will be displayed in other senses—such as graphics, pictures, patterns, sounds, smells, and tactile experiences. This enriched sensory environment will coincide with major breakthroughs in our understanding of the brain—in how we process sensory information and connect various sensory functions.

**Humanities**

Humans will become much more sophisticated in their ability to understand, create, and manage sensory information and abilities to perform such tasks will become key to success.

**Balkanized Infrastructure**

A confluence of new materials and distributed intelligence is pointing the way toward a new kind of infrastructure that will dramatically reshape the economics of moving people, goods, energy, and information. From the molecular level to the macroscale, these new infrastructure designs will employ smaller, smarter, more independent components. These components will be organized into more efficient, more flexible, and more secure systems than the capital-intensive networks of the 20th century. These lightweight infrastructures have the potential to best emerging economies, improve social connectivity, mitigate the environmental impacts of rapid global urbanization, and offer new future paths in energy.

**Democratized Innovation**

Before the 20th century, most of the greatest scientific discoveries and technical inventions were made by relatively small and independent inventors. In the last 100 years, a professional class of scientists and engineers, suggested by universities, industry, and the state, has produced advances as a collective force. At the national scale, the capital-intensive character of scientific research made world-class research the property of a prosperous and innovative few. In the near future, a number of trends and technologies will lower the barriers to participation in science and technology again, both for individuals and for emerging countries. This trend will be a reflection of the serious analyst's, the growth of new scientific and technical centers of excellence in developing countries, and a more global distribution of world-class scientists and technologies.
Map of Scientific Collaborations from 2005-2009

Stream of Scientific Collaborations Between World Cities - Olivier H. Beauchesne - 2012
Examining the Evolution & Distribution of Patent Classifications

Managing Growing Patent Portfolios
Organizations, businesses, and individuals rely on patents to protect their intellectual property and business models. As market competition increases, patenting and innovation become more important.

Managing the staggering number of patents demands new tools and methodologies. Grouping patents by their classifications offers an ideal solution for better understanding how intellectual borders are established and change over time.

The charts below show the annual number of patents granted from January 1, 1980 to December 31, 2000 in the United States Patent and Trademark Office (USPTO) patent archive; slow and fast growing patent classes; the top 10 fast growing patent subclasses; and two evolving patent portfolios.

The Structure and Evolution of the Patent Space
The United States Patent and Trademark Office assigns each patent to one of more than 450 classes covering broad application domains. For example, class 514 encompasses all patents dealing with "Drug, Bio-Affecting and Body T赛事ing Compositions." Classes are further broken down by subclasses that have hierarchical associations. As one example, class 455 features subclass 99 entitled "with vehicle."

The top 10 fast growing patent classes for 1998–2002 are listed together with the number of patents granted. Most come from the "Computer and Communications" and the "Drugs and Medical" area.

Top-10 Subclasses

<table>
<thead>
<tr>
<th>Class</th>
<th>Title</th>
<th># of Patents</th>
</tr>
</thead>
<tbody>
<tr>
<td>514</td>
<td>Drug, Bio-Affecting and Body T赛事ing Compositions</td>
<td>18,778</td>
</tr>
<tr>
<td>438</td>
<td>Semiconductor Device Manufacturing/Processing</td>
<td>17,775</td>
</tr>
<tr>
<td>435</td>
<td>Chemistry; Molecular Biology and Microbiology</td>
<td>17,474</td>
</tr>
<tr>
<td>474</td>
<td>Drug, Bio-Affecting and Body T赛事ing Compositions</td>
<td>13,537</td>
</tr>
<tr>
<td>458</td>
<td>Stocks, Shares and Miscellaneous Articles</td>
<td>13,314</td>
</tr>
<tr>
<td>527</td>
<td>Active Solid-State Devices (e.g., Transistors, Solid-State Devices)</td>
<td>12,924</td>
</tr>
<tr>
<td>395</td>
<td>Information Processing System Organization</td>
<td>9,955</td>
</tr>
<tr>
<td>359</td>
<td>Optical Systems and Elements</td>
<td>9,151</td>
</tr>
<tr>
<td>365</td>
<td>Static Information Storage and Retrieval</td>
<td>8,392</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>130,910</td>
</tr>
</tbody>
</table>

Patent Portfolio Analysis
A longitudinal analysis of portfolios reveals different patenting strategies. For each year (since 1980), a treemap of all new patents granted to the assignees is shown. The number of patents is given below each treemap. The same size and color coding as above was used. In addition, yellow indicates that no patent has been granted in that class in the last 5 years.

Apple Computer, Inc.
Apple Computer, Inc.'s portfolio starts in 1980 and increases considerably in size over time. In most years, more than half of Apple Computer's patents were placed into four classes, namely "395 Information Processing System Organization," "345 Computer Graphics Processing, Operator Interface Processing, and Selective Visual Display Systems," "392 Image Analysis," and "707 Data Processing: Database and File Management or Data Structures." These four classes are an integral part of Apple Computer, Inc.'s patent portfolio, receiving patents every year.

Jerome Lemelson
The patent portfolio of Jerome Lemelson shows a very different activity pattern. Starting in 1976, he publishes between 6–20 patents each year. However, the predominance of yellow shows that there is little continuity from previous years in regards to the classes into which patents are filed. No class dominates. Instead, more and more new intellectual space is claimed.
The Product Space

World trade flow data compiled by Feenstra et al. and available at the National Bureau of Economic Research were used to identify the complete co-export matrix of 775 industrial products for 1998-2000. A Minimum Spanning Tree (MST) algorithm was used to reduce the complete co-export matrix to less than 1% of the links. The resulting network, which combines the MST plus all links with a co-export frequency of at least 0.5%, was laid out using a force-directed layout algorithm. Node sizes represent the value of traded products in millions of U.S. dollars. Their color corresponds to ten product groups identified using the Leamer classification. Each product class is labeled by an icon. Link color and width indicate the frequency of joint exports.

The network has a core-periphery structure with higher value product classes, e.g., machinery and chemicals in the core and lower quality classes, e.g., fishing and garments, in the periphery. Products at the core of the network are highly interconnected while products in the periphery are sparsely interlinked.

Each country has a certain product export footprint. Relevant exports are termed "Industrialized Countries," "East Asia Pacific," and "Latin America." The data for these countries are given on the right.

Traditional growth theory assumes that there is always a more sophisticated product within reach. However, given the core-periphery structure of the product space, the distances between products differ considerably.

Countries that operate at the core have capabilities to develop and manufacture a wide range of products. Yet, countries that mostly operate in the periphery of the product space have much fewer opportunities for diversification. A country's current footprint and the structure of the product space have a major impact on a country's future development.
Pulse of the Nation: U.S. Mood Variations Inferred From Twitter

Mood Variations
A number of interesting trends can be observed in the data. First, overall daily variations can be seen in the graph, with the early morning and late evening having the highest level of tweets. Second, geographic variations can be observed across the graph, with a higher density near the coasts, suggesting that online posting is influenced by the time of day.

Weekly Variations
Weekly trends can be observed as well, with weekends having a much higher rate than weekdays.

About the Data and Visualization
The posts were calculated using over 300 million tweets (Sep 2009 - Aug 2009) collected by MALIS researchers, represented as a density-preserving cartogram. The mood of each tweet was inferred using NRC Word-List (Bradley, M. M. & Lang, P. J. Affective norms for English words, NRC: Sentic, instructions manual) and affective ratings. The Center for Research on Psychological and Socially Relevant Interactions. The daily area data was taken from the U.S. Census Bureau at https://www.census.gov and the border data was taken from NetaCensus (National). All locations were inferred using the Google Maps API and mapped onto counties using PostGIS and U.S. county maps from the U.S. National Atlas. Mood colors were selected using Color Brewer 2.

About Cartograms
A cartogram is a map in which the mapping variable (in this case, the number of tweets) is substituted for the true land area. Thus, the geometry of the actual map is altered, such that the shape of each region is maintained as much as possible, but the area is scaled in order to be proportional to the number of tweets that originated in that region. The result is a density-equalizing map. The cartograms in this work were generated using the Carto software by Mark E. J. Newman.

Northeastern University
College of Computer and Information Science
Center for Complex Network Research

Harvard University

IX.4 Pulse of the Nation - Alan Mislove, Sune Lehmann, Yong-Yeol Ahn, Jukka-Pekka Onnela, and James Niels Rosenquist - 2010
The EMERGENCE of NANOtechnology

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

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

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

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

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

References

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

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

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

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

2001
The journal Science is relevant in the citation impact environment. Other journals in nanoscience and technology begin to emerge, and the bridging role of the journal Nanotechnology gradually declines. Nano Letters and the Journal of Nanoscience and Nanotechnology join the new field of nanotechnology.

2002
Other journals in nanoscience and technology begin to emerge, and the bridging role of the journal Nanotechnology gradually declines. Nano Letters and the Journal of Nanoscience and Nanotechnology join the new field of nanotechnology.
Chemical Research & Development Powers the U.S. Innovation Engine
Macroeconomic Implications of Public and Private R&D Investments in Chemical Sciences

INVESTMENT IN CHEMICAL SCIENCE R&D

FEDERAL GOVERNMENT

$1 Billion
FEDERAL FUNDING

$8 Billion
TAXES

$5 Billion
INDUSTRY FUNDING

CHEMICAL INDUSTRY

$1B
$1B + $5 Billion

$10 Billion
CHEMICAL INDUSTRY OPERATING INCOME

$40 Billion
GROWTH IN GNP + 600,000 JOBS CREATED

U.S. ECONOMY

TIMELINE FROM CONCEPTION TO COMMERCIALIZATION

4-5 YRS
FOUNDATIONAL RESEARCH

9-11 YEARS
INVENTION DEVELOPMENT

> 5 YEARS
TECHNOLOGY COMMERCIALIZATION

20 YEARS
COMMERCIALIZATION

The Council for Chemical Research (CCR) has provided the U.S. Congress and government policy makers with important results regarding the impact of Federal Research & Development (R&D) investments on U.S. innovation and global competitiveness through its commissioned 5-year two phase study. To take full advantage of typically brief access to policy makers, CCR developed the graphic below as a communication tool that distills the complex data produced by these studies in direct, concise, and clear terms.

The design shows that an input of $1B in federal investment, leveraged by $5B in industry investment, brings new technologies to market and results in $10B of operating income for the chemical industry. $40B of growth in the Gross National Product (GNP) and further impacts the US economy by generating approximately 600,000 jobs, along with a return of $8B in taxes. Additional details, also reported in the CCR studies, are depicted in the map to the left. This map clearly shows the two R&D investment cycles; the shorter industry investment at the innovation stage to commercialization cycle; and the longer federal investment cycle which begins in basic research and culminates in national economic and job growth along with the increases in tax base that in turn is available for investment in basic research.
MAPS

VS.

MACROSCOPES
Microscopes & Telescopes vs. MACROSCOPES
Iteration XI (2015): Macrosopes for Interacting with Science
http://scimaps.org/iteration/11
earth ≡

*Earth – Cameron Beccario*
The News Co-occurrence Globe
An interactive visualization of how countries are mentioned together in the world's news media

2.92K COOCCUR%

UNITED KINGDOM
cooccurrences in: 2,922%
cooccurrences out: 80%

Mapping Global Society – Kalev Leetaru
Iteration XII (2016): Macroscopes for Making Sense of Science

http://scimaps.org/iteration/12
Smelly Maps – Daniele Quercia, Rossano Schifanella, and Luca Maria Aiello – 2015
Models of Science, Technology, and Innovation

Using large scale datasets, advanced data mining, modeling, and visualization techniques, and substantial computing resources.

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

Conference slides, recordings, and report are available via http://modsti.cns.iu.edu/report
Modeling and Visualizing Science and Technology Developments

December 4-5, 2017; Irvine, CA

Overview

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

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

Videos of the talks are available on the Sackler YouTube Channel.
Making Science & Technology Visualizations

Using a theoretically grounded visualization framework that defines key terminology and processes together with valid workflows and data mappings.
Register for free: http://ivmooc.cns.iu.edu
### Tasks

#### LEVELS

<table>
<thead>
<tr>
<th>Level</th>
<th>Description</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>MICRO: Individual Level</td>
<td>about 1–1,000 records</td>
<td>page 6</td>
</tr>
<tr>
<td>MESO: Local Level</td>
<td>about 1,001–100,000 records</td>
<td>page 8</td>
</tr>
<tr>
<td>MACRO: Global Level</td>
<td>more than 100,000 records</td>
<td>page 10</td>
</tr>
</tbody>
</table>

#### TYPES

<table>
<thead>
<tr>
<th>Type</th>
<th>Tasks</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical Analysis</td>
<td>Knowledge Cartography page 135, Productivity of Russian life sciences research teams page 105</td>
<td></td>
</tr>
<tr>
<td>WHEN: Temporal Analysis</td>
<td>Visualizing decision-making processes page 95, Key events in the development of the video tape recorder page 85</td>
<td></td>
</tr>
<tr>
<td>WHERE: Geospatial Analysis</td>
<td>Cell phone usage in Milan, Italy page 109, Victorian poetry in Fumpa page 157</td>
<td></td>
</tr>
<tr>
<td>WHAT: Topical Analysis</td>
<td>Evolving patent holdings of Apple Computer, Inc. and Jerome Lemelson page 50, Evolving journal networks in nanotechnology page 159</td>
<td></td>
</tr>
<tr>
<td>WITH WHOM: Network Analysis</td>
<td>World Finance Corporation network page 57, Electronic and new media art networks page 133</td>
<td></td>
</tr>
</tbody>
</table>

See Atlas of Science: Anyone Can Map, page 5
Visual Analytics Certificate

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

Register: tinyurl.com/VACRegister

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.

http://visanalytics.cns.iu.edu
Visual Analytics Certificate

Instructor: Victor H. Yngve Distinguished Professor Katy Börner & CNS Team, ISE, SICE, IUB
Duration: 6 weeks x 5 hours = 30 hours (3 CEUs)
Format: Online | Theory and Hands-on Instruction, Concept Questions, Graded Assignments, Case Studies, Discussions
Start: Sept 15, 2018

Covers:
Temporal, geospatial, topical (linguistic), network analyses and 60+ visualization types

Tools: Tableau, Gephi, BI

Real world case studies such as
• Acting on customer complaints data.
• Improving communication/traffic flows.
• Understanding web page usage.
• Visualizing online shopping behavior.
• Optimizing supply chains.
• Reducing customer/supplier churn.
• Monitoring emerging R&D areas.
• Workforce development planning.

http://visanalytics.cns.iu.edu
Case Studies: Solving Real-World Challenges

Apply your new knowledge and skills in projects that require you to identify user needs and priorities; select the best data, algorithms, and workflows for temporal, geospatial, topical, and network case studies; communicate actionable insights using standard terminology; and deliver high-quality results on time and on budget.

Submit a Case Study

Collaborate with us on solutions that make a difference to you. Submit a project description, we will then implement efficient solutions and provide training materials that will speed adoption by key decision-makers.

Submit: tinyurl.com/VACCcaseStudy

http://visanalytics.cns.iu.edu
References


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


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

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