Information Visualization of NSWC Crane Innovation Ecosystem as part of Naval Science and Technology

Katy Börner, Leonard E. Cross, Michael Gallant, Shutian Ma, Adam S. Martin, Elizabeth Record, Olga B. Scrivner, and Haici Yang

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

CRANE Distinguished Lecture Series

Club Lakeview, Naval Surface Warfare Center, Crane Division
300 Highway 361, Crane, IN 47522

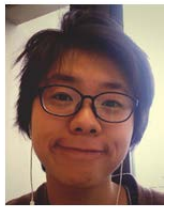
September 25, 2018

@katycns

The CNS Team

- **Katy Börner**, Faculty and CNS Director; Study design, training, guidance, result validation, documentation.
- **Leonard E. Cross**, Senior Interaction Designer; Visualization design optimization.
- **Michael Gallant**, Database Admin; IT Setup, database loading, optimization, query design and execution.
- **Shutian Ma**, Visiting Scholar; Data analysis and visualization.
- **Adam S. Martin**, CNS Assistant; Geocoding, Network extraction.
- **Elizabeth Record**, Associate Director; Grant management, user needs analysis and user studies, documentation.
- **Olga B. Scrivner**, Research Scientist; Query design and execution, user needs analysis and user studies, data analysis and visualization.
- **Haici Yang**, Exchange Student; Data analysis and visualization.







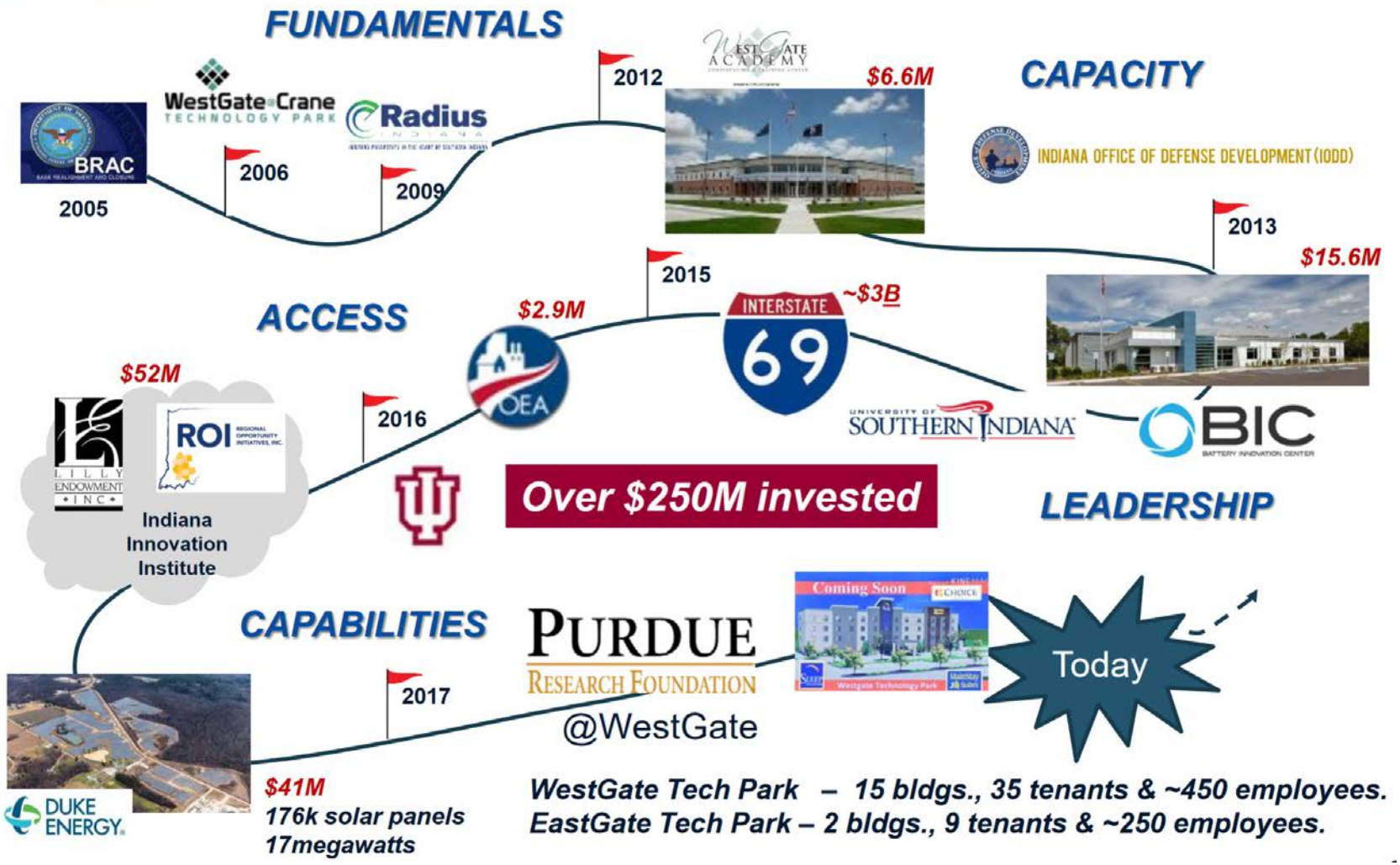
Project Goal

CNS Cyberinfrastructure for Network Science Center

CRANE has a demonstrated need to **allocate resources, both monetary and human, to foster innovations in science and technology.**

Advances in computational power combined with the unprecedented volume and variety of data on science and technology developments (e.g., publications, patents, funding data) create ideal conditions for the advancement of **data analysis and visualization approaches that can be empirically validated and used to simulate and understand the structure and dynamics of STI and to augment human decision making.**

Evolution of the Crane Innovation Ecosystem



Building the Crane Innovation Ecosystem *Dr. Kyle Werner*
http://sites.nationalacademies.org/cs/groups/pgasite/documents/webpage/pga_188467.pdf

Project Goal

CNS Cyberinfrastructure for Network Science Center

CRANE has a demonstrated need to **allocate resources, both monetary and human, to foster innovations in science and technology**

Advances in computational power combined with the unprecedented volume and variety of data on science and technology developments (e.g., publications, patents, funding data) create ideal conditions for the advancement of **data analysis and visualization approaches that can be empirically validated and used to simulate and understand the structure and dynamics of STI and to augment human decision making.**

In ***Phase I***, publication and funding data will be used in Study 1. In ***Phase II***, patent data will be added to support Study 2; course and degree data as well as job advertisement data will be used in support of Study 3.

This talk presents results from Study 1 that

1. conducted a user needs analysis to identify areas of strategic interest to CRANE as well as analyses and visualizations that best support strategic decision making and
2. developed methods for the identification of leading experts and potential collaborators in relevant research areas, analyzing investments by other funding organizations, and tracing the development of emerging research areas.

Stakeholder Needs Analysis

CNS Cyberinfrastructure for Network Science Center

The design of easy-to read and actionable data visualizations requires a **deep understanding of target stakeholder's needs and expertise**. An expert survey was performed to identify stakeholder types, demographics, task types, insight needs, work contexts, and priorities.

Surveyed 11 experts to understand

- how they make strategic decisions
- which data analyses and visualizations would be most beneficial

Overall, participants were quite comfortable with a variety of visualization types.

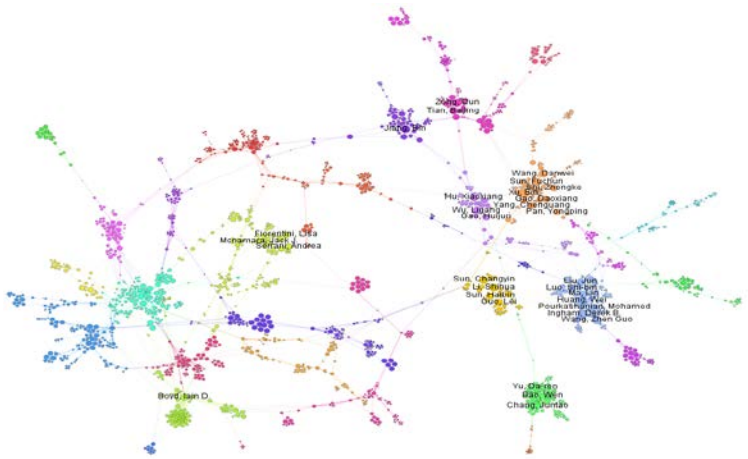
- 11 expressed familiarity with graphs and maps
- 10 expressed familiarity with tree diagrams
- 8 expressed familiarity with network diagrams
- 4 have taken a seminar, course, or training session in visualization.

Expert Survey Results cont.

Understanding **static visualizations** around the topic "**Hypersonic**" including

1. # papers per year (rated most useful by 1)
2. top-10 most frequently acknowledged organizations that fund "hypersonic" research. (rated most useful by 7)
3. Co-author network (rated most useful by 2)

One participant responded that all were equally useful in different ways but were each not as useful on their own.



Expert Survey Results cont.

CNS Cyberinfrastructure for Network Science Center

Understanding **interactive visualizations** around the topic "Hypersonic" including

1. co-authorship geospatial visualization (rated most useful by 6)
2. co-authorship network visualization (rated most useful by 3)
3. science map visualization (rated most useful by 2)

Other comments:

- Fig 1 allows me to 'quickly see **where and with whom, researchers are interacting**. I can **find related researchers and research quickly, specifically groups that I'm unaware of.'**
- In Fig 2, it would help to '**focus on influential authors** which could then lead to other insights such as geospatial or topic maps.'
- Fig 3, the science map is good for '**target growth & engagement of complementary expertise already internal to Crane**. Co-author network is best to '**identify insertion points for new areas of R&D engagement.'**

Several expressed the desire to **modify the time frame** represented in these visualizations. Note that interactive visualizations are planned for Phase II of this project.

Expert Survey cont.

CNS Cyberinfrastructure for Network Science Center

Topical Interests

Topic Area	Number of Experts that Expressed Interest
Advanced electronics	10
Artificial intelligence	10
Sensors and sensor fusion	9
Internet of things	4
Human systems integration	4
Robotics	3
System of systems test and evaluation	3
Power and energy management	2

Other strategic areas included: Atmospheric models, adversary technology progression, quantum sensing, quantum computing, cybersecurity (2), hypersonics (5), energy, MOOC-based systems engineering, signal processing, hardware assurance, radiation-material interactions, model-based systems engineering, quantum science (2), Photonics, Synthetic Biology, Biomimicry, Trusted & assured microelectronics, Trusted & assured quantum electronics.

Picking Key Topics

CNS Cyberinfrastructure for Network Science Center

Given that Study 1 focuses on the usage of **funding** and **publication** data, below Table also shows the number of NSF awards and publications for the top-ranked topics. For many areas, no or few funding awards and publications exist.

Topic	#Experts that expressed interest	#NSF Funding awards active in 1998-2017	WOS Publications
Advanced electronics	10	122	
Artificial intelligence	10	**1,075**	**6,533**
Sensors and sensor fusion	9	145	
Internet of things	4	**349**	**4,898**
Human systems integration	4	0	
Robotics	3	**3,075**	**13,152**
System of systems test and evaluation	3	4	
Power and energy management	2	0	

Picking Key Topics cont.

CNS Cyberinfrastructure for Network Science Center

Recent publications on the importance of **Artificial Intelligence**, **Internet of Things**, and **Robotics** for global defense and security:

- "As the world deals with changes brought about by emerging technologies such as **robotics**, rapid manufacturing, autonomous vehicles, **artificial intelligence**, and at the same time struggles with cybersecurity and biosecurity, the nation's defense and security agencies are perhaps moving even faster to understand what these new technologies will mean in the future" (Tally, 2016).
- **Artificial Intelligence, IoT**, and **Robotics** were named top technologies in 2018 with strong arguments and examples of how these technologies will drive digital innovation and completely transform business models (Matthieu, 2017; Pascu, 2018).
- **AI** and **Robotics** will have a major impact on the future of work and it is of utmost importance that businesses prepare for these transformational changes.

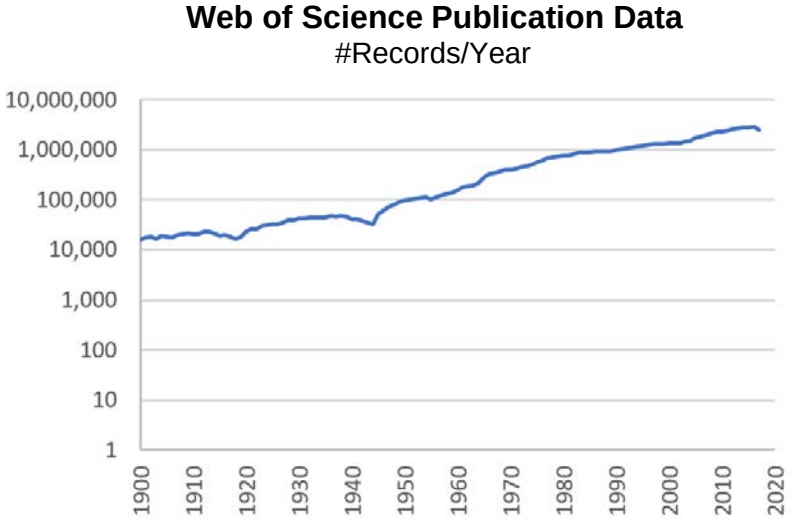
Thus, **Artificial Intelligence**, **Internet of Things**, and **Robotics** were selected for detailed analysis is in Phase I of this project.

Data

Publications

Clarivate (formerly Thomson Reuters) Web of Science XML raw data (Web of Knowledge version 5.4) was acquired by the IUNI Science of Science Hub. The number of total publications is 69M and there are more than 1B citation links from 1900 through 2017. Most publications have title, abstract, keyword information that can be used for text mining. Publications have a publication year and author data.

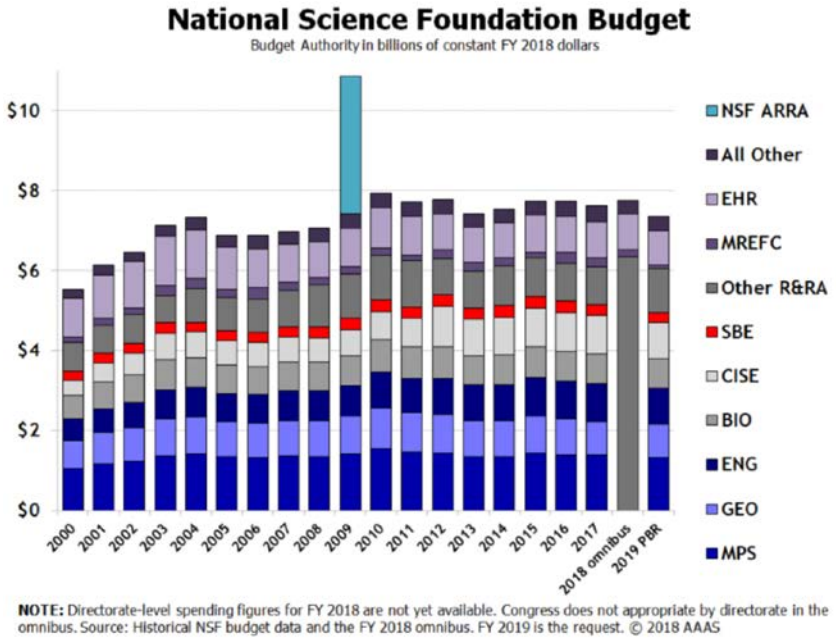
The ER diagram and data dictionary can be found at http://iuni.iu.edu/resources/web-of-science.



Funding

National Science Foundation (NSF) funds research and education in science and engineering. It does this through grants, contracts, and cooperative agreements to more than 2,000 colleges, universities, and other research and/or education institutions in all parts of the United States. NSF accounts for about 20 percent of federal support to academic institutions for basic research.

CNS has access to more than 500K awards covering active, expired, and historical awards for 1952 to today, see also https://nsf.gov/awardsearch for recent data and https://nsf.gov/awardsearch/download.jsp for bulk download.



Top-Funded AI Awards, Active in 1998-2017

Title	NSF Org.	Start Date	Principal Investigator	Organization	Program Manager	End Date	Awarded Amount To Date
BEACON: An NSF Center for the Study of Evolution in Action	DBI	8/1/2010	Erik Goodman	Michigan State University	George W. Gilchrist	7/31/2021	$43,035,209
Center for Research in Cognitive Science	BCS	2/1/1991	Aravind Joshi	University of Pennsylvania	Cecile Mckee	7/31/2002	$20,878,702
Spatial Intelligence and Learning Center (SILC)	SMA	10/1/2011	Nora Newcombe	Temple University	Soo-Siang Lim	9/30/2018	$18,306,816
System Engineering Risk Reduction	CNS	1/1/2008	Brig 'Chip' Elliott	Raytheon BBN Technologies Corp.	Joseph Lyles	12/31/2012	$14,058,434
RII: Enhancing Alabama's Research Capacity in Nano/Bio Science and Sensors	EPS	9/1/2011	Mahesh Hosur	Tuskegee University	Uma D Venkateswaran	8/31/2017	$11,332,243
Collaborative Research: Computational Sustainability: Computational Methods for a Sustainable Environment, Economy, and Society	CNS	8/15/2008	Carla Gomes	Cornell University	Ralph Wachter	7/31/2016	$7,939,359
RII: Enhancing Alabama's Research Capacity in Nano/Bio Science and Sensors	EPS	9/1/2008	Mahesh Hosur	Alabama A&M University	Kelvin Chu	10/31/2011	$6,000,000
MRI-R2: Development of Common Platform for Unifying Humanoids Research	CNS	7/1/2010	Youngmoo Kim	Drexel University	Rita V. Rodriguez	9/30/2015	$5,999,997
PIRE: International Program for the Advancement of Neurotechnology (IPAN)	OISE	11/1/2015	Euisik Yoon	University of Michigan Ann Arbor	Cassandra M. Dudka	10/31/2020	$5,000,000
Digital Government: COPLINK Center: Information and Knowledge Management for Law Enforcement	IIS	7/1/2000	Hsinchun Chen	University of Arizona	Lawrence Brandt	9/30/2006	$3,579,649

Data Preprocessing

CNS Cyberinfrastructure for Network Science Center

All unique **WOS *Author Keywords*** specific to the three topic areas were identified and used to identify relevant terms for NSF funding awards.

Specifically, characters like [ ] { } ” were removed, then MaxMatch was applied to NSF titles and abstracts.

MaxMatch performs word segmentation to improve precision. The algorithm first computes the maximum number of words in the lexical resource (here NSF award titles and abstracts); it then matches long terms first before matching shorter terms. Thus, given text ‘Artificial Intelligence’ and ‘Intelligence’ in a set of relevant terms, and ‘Artificial Intelligence’ in the title and/or abstract of an award, the algorithm returns ‘Artificial Intelligence’ accounting for oversampling of popular, short terms.

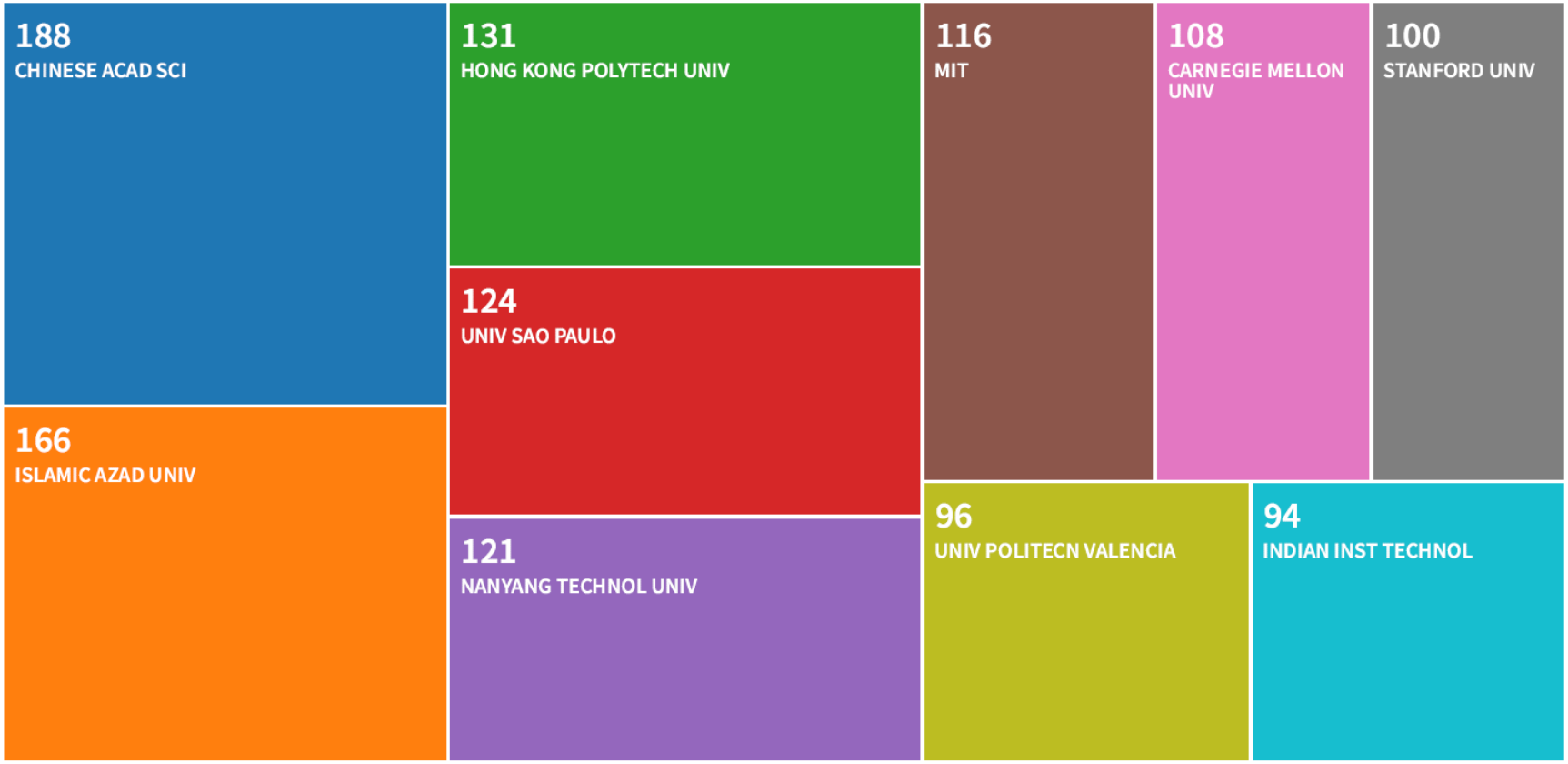
Wong P-K, Chan C (1996) Chinese word segmentation based on maximum matching and word binding force. *Proc 16th Conf Comput Linguist* 1:200–203.

Algorithm Comparison Results

Algorithm Name	Average Precision	Average Recall	Average F_1
ExactMatch	0.419	0.731	0.514
MaxMatch	**0.704**	**0.705**	**0.686**
NgramMatch	0.177	0.250	0.194
StanfordNER	0.473	0.487	0.467
StanfordLLDA3	0.333	0.042	0.073
StandfordlLLDA5	0.200	0.042	0.068

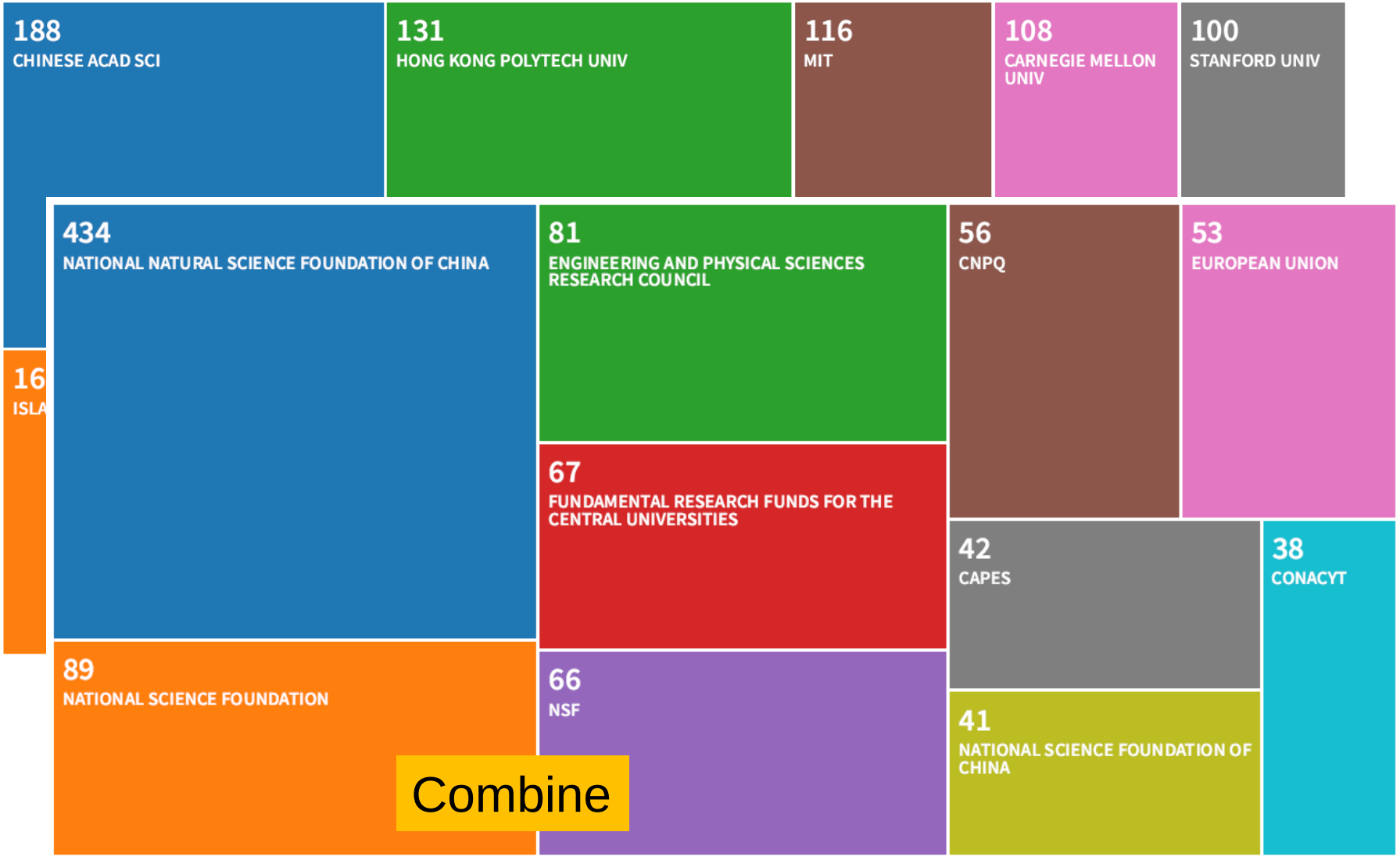
WOS-Top Organizations and Funding: AI

CNS Cyberinfrastructure for Network Science Center



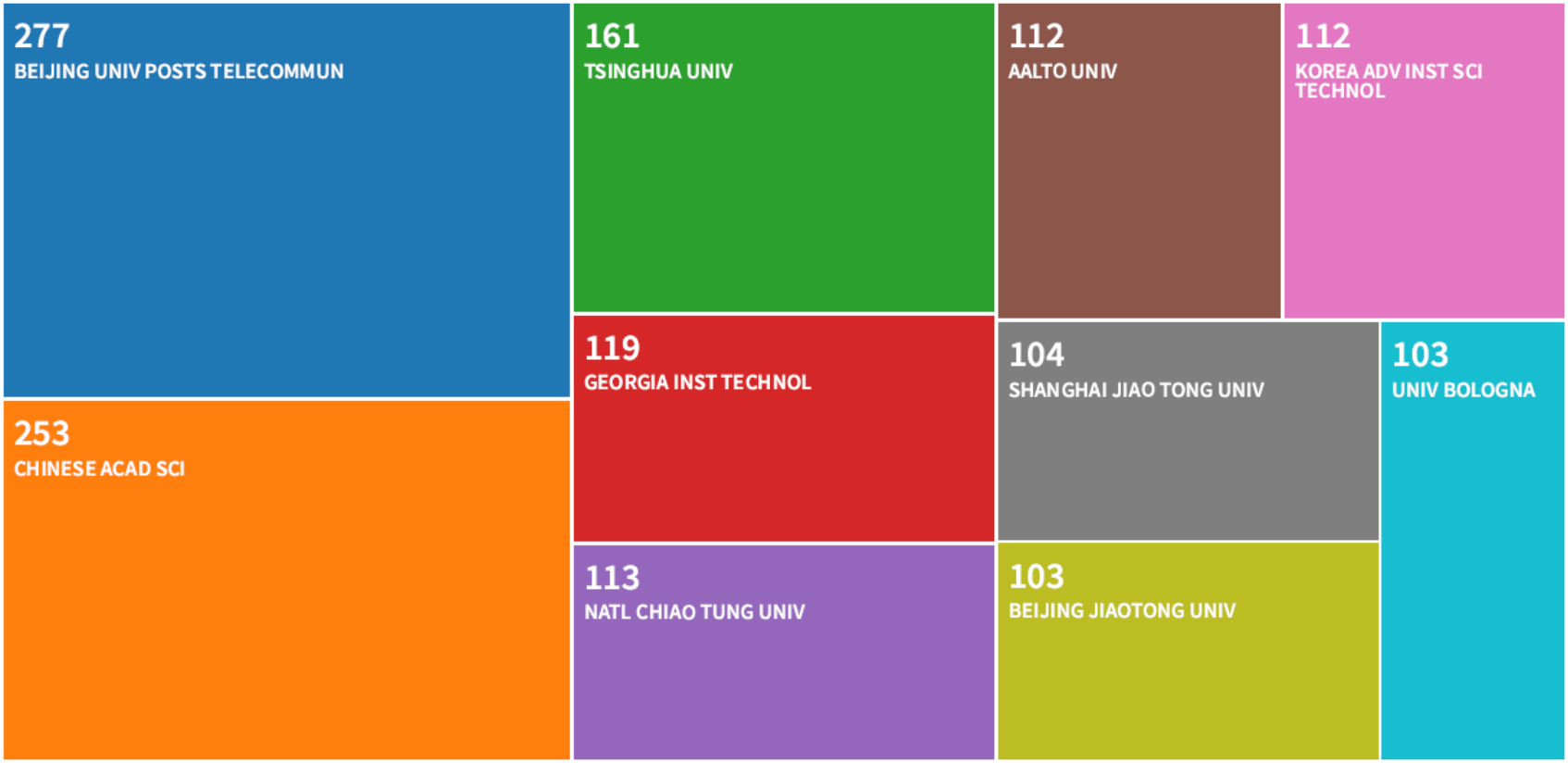
WOS-Top Organizations and Funding: AI

CNS Cyberinfrastructure for Network Science Center



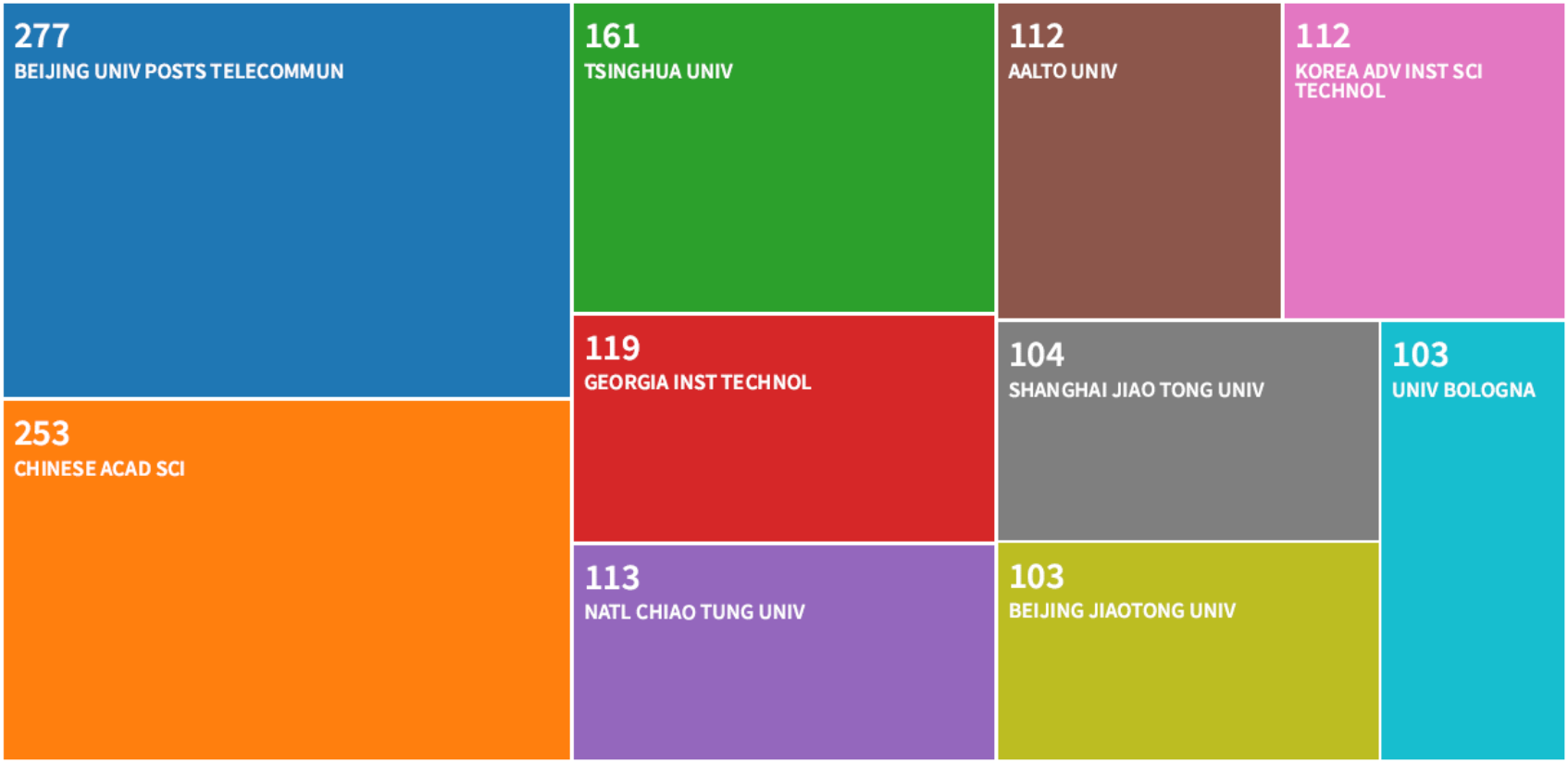
WOS-Top Organizations and Funding: IoT

CNS Cyberinfrastructure for Network Science Center

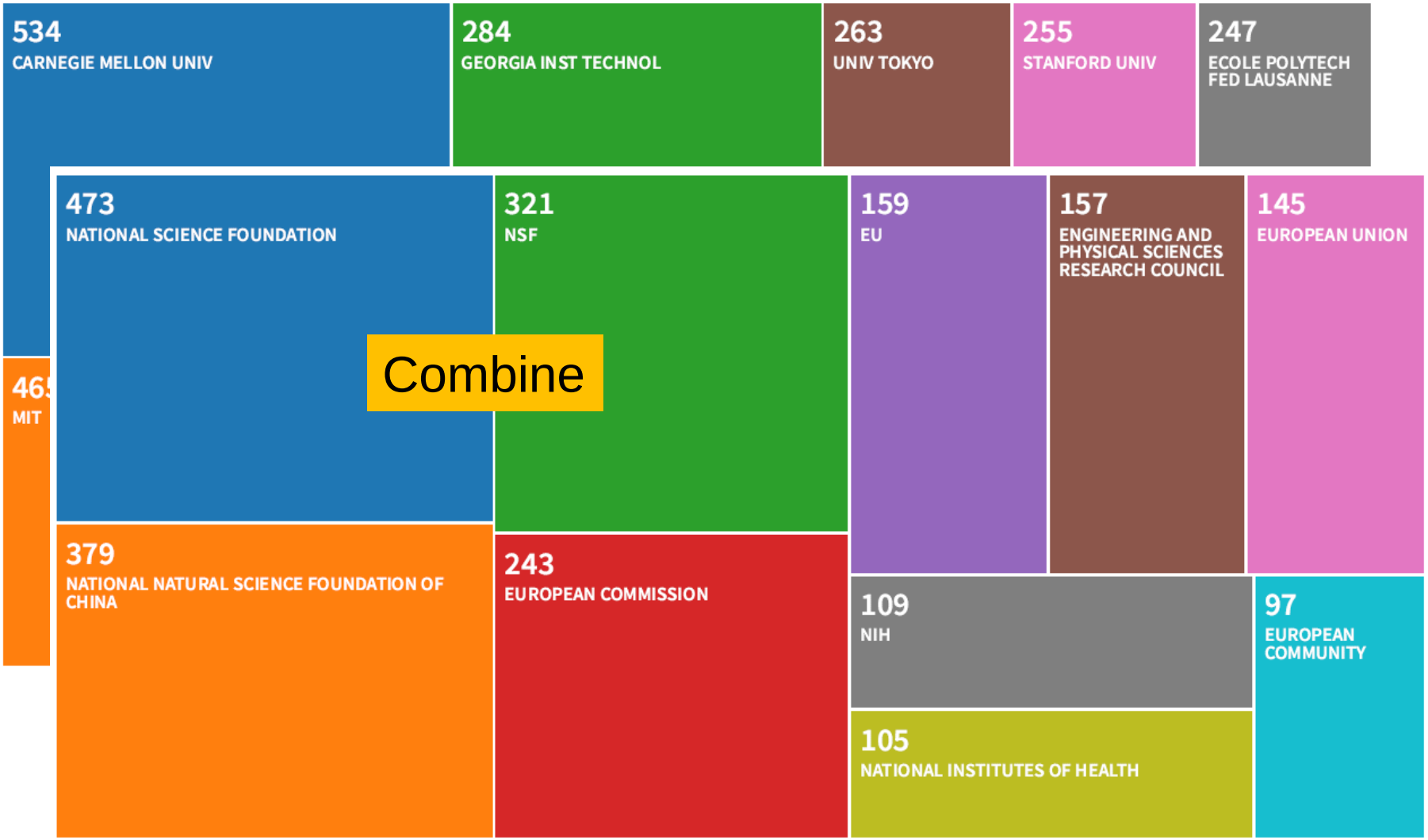


WOS-Top Organizations and Funding: IoT

CNS Cyberinfrastructure for Network Science Center



WOS-Top Organizations and Funding: Robotics



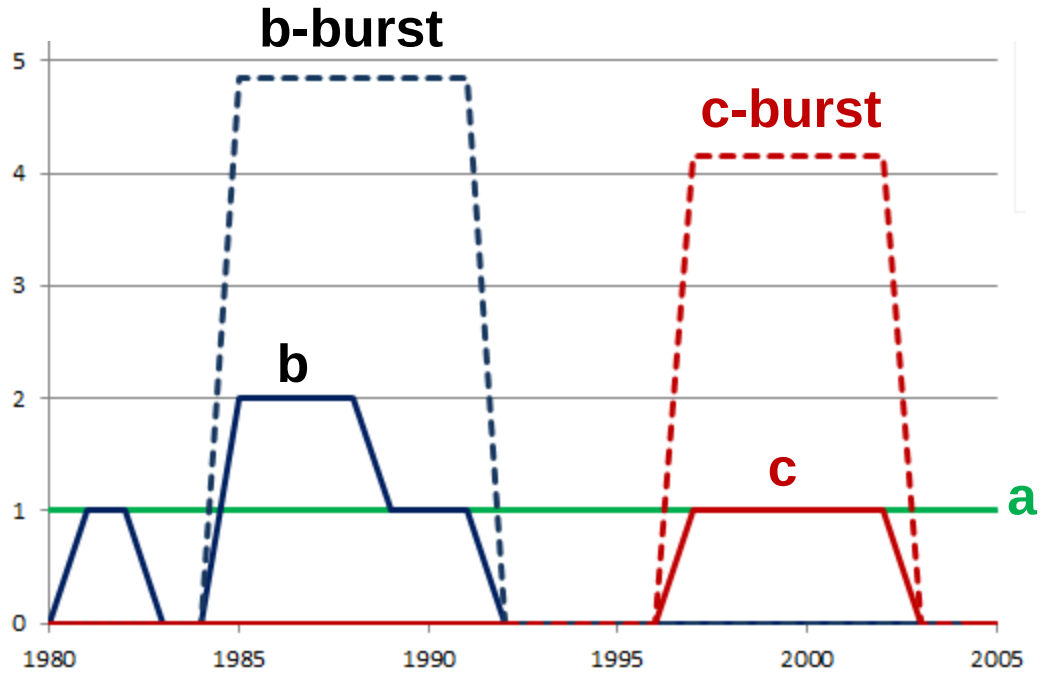
Burst Analysis & Visualization

	A	B
1	Year	Words
2	1980	a
3	1981	a b
4	1982	a b
5	1983	a
6	1984	a
7	1985	a b b
8	1986	a b b
9	1987	a b b
10	1988	a b b
11	1989	a b
12	1990	a b
13	1991	a b
14	1992	a
15	1993	a
16	1994	a
17	1995	a
18	1996	a
19	1997	a c
20	1998	a c
21	1999	a c
22	2000	a c
23	2001	a c
24	2002	a c
25	2003	a
26	2004	a
27	2005	a

Kleinberg's burst-detection algorithm identifies sudden increases in the frequency of words.

Given time-stamped text, it identifies words that burst.

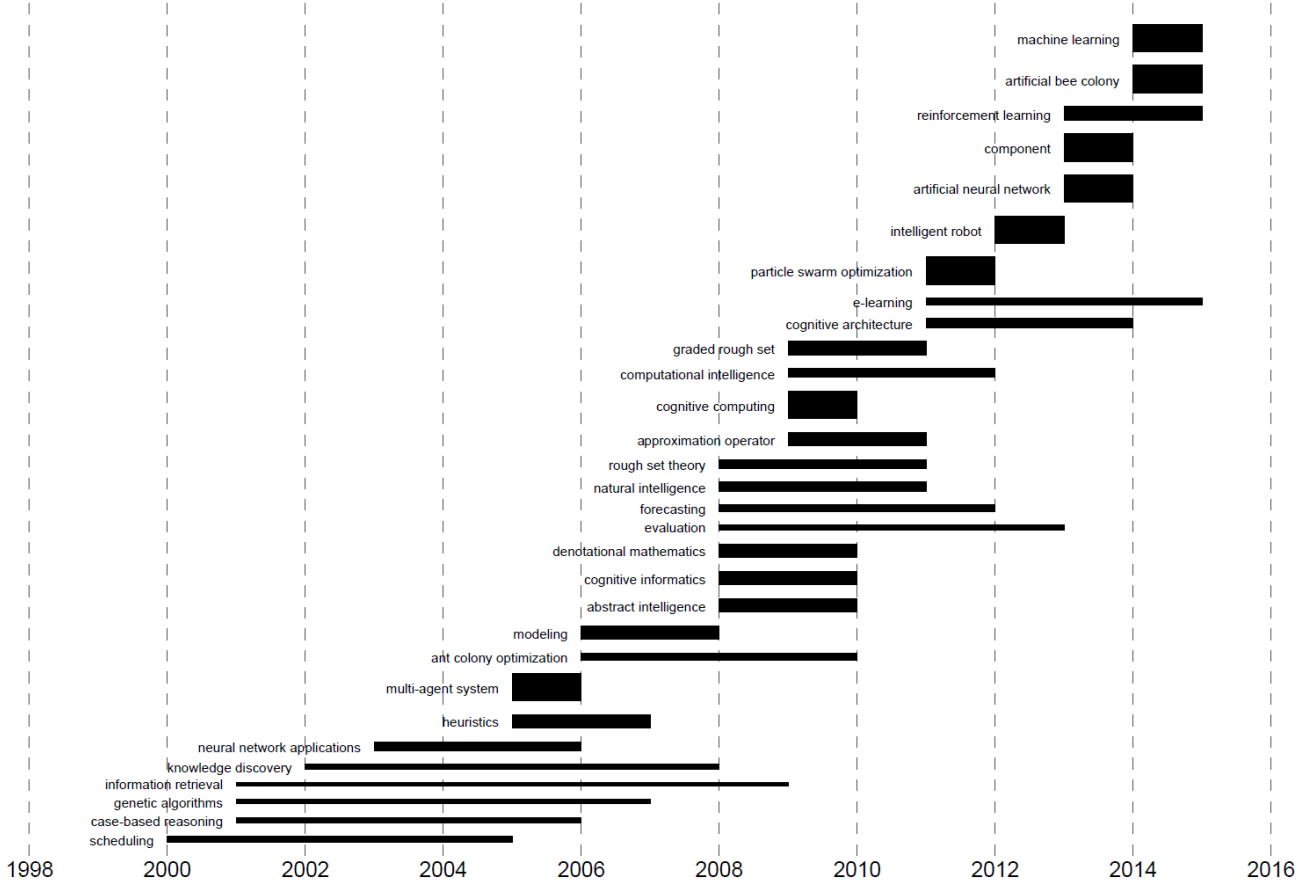
“a” does not burst. “b” bursts more than “c”

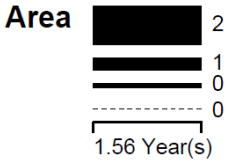


Bursts: Artificial Intelligence Publications

Temporal Visualization

(Generated from Burst detection analysis (Year, AuthorKeywords): maximum burst level 1)
September 25, 2018 | 10:41 AM -04:00

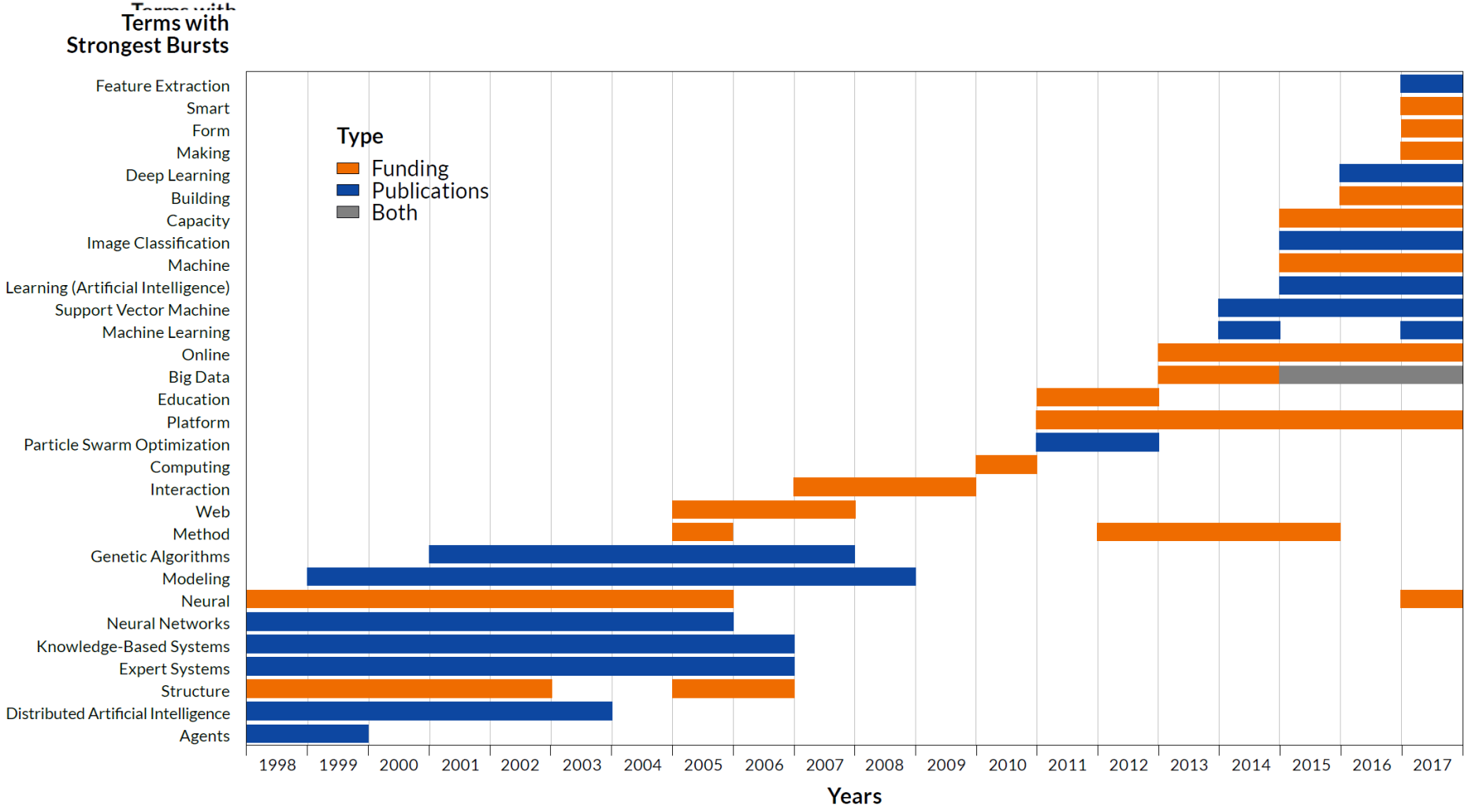




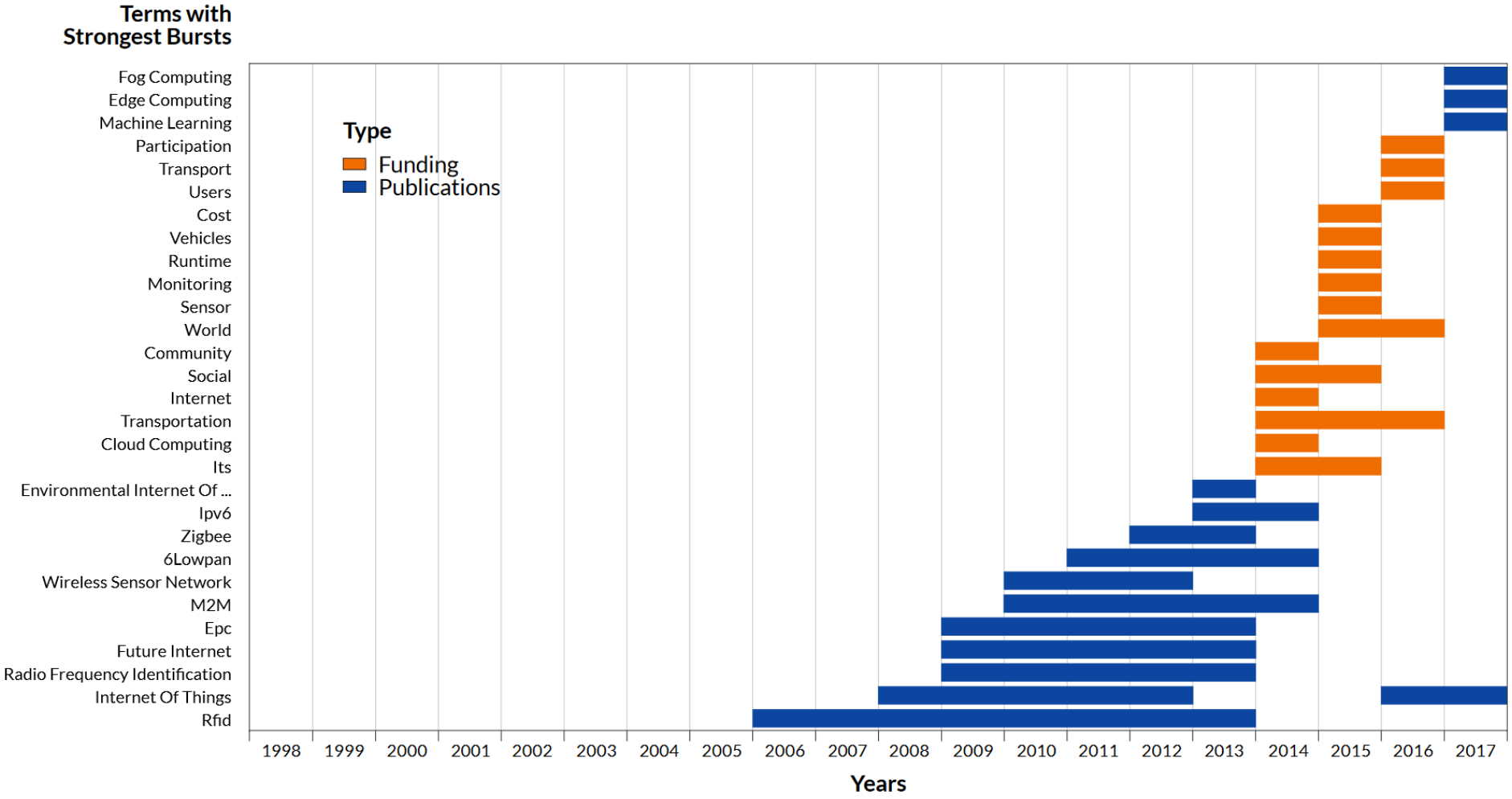
How To Read This Map

This *temporal bar graph* visualization represents each record as a horizontal bar with a specific start and end date and a text label on its left side. The area of each bar encodes a numerical attribute value, e.g., total amount of funding. Bars may be colored to present categorical attribute values of records.

Bursts: Artificial Intelligence (AI)

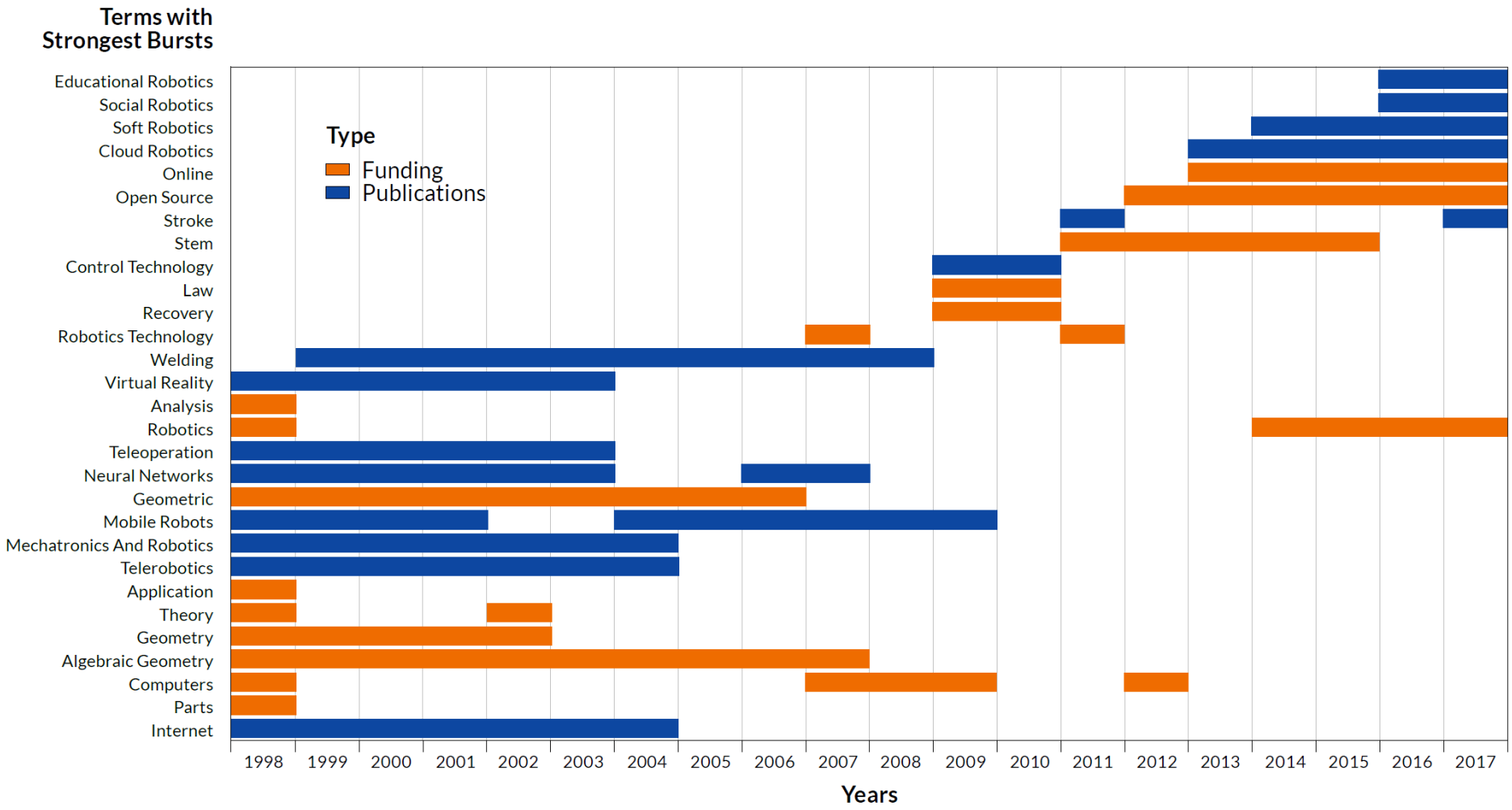


Bursts: Internet of Things (IoT)



Bursts: Robotics

CNS Cyberinfrastructure for Network Science Center

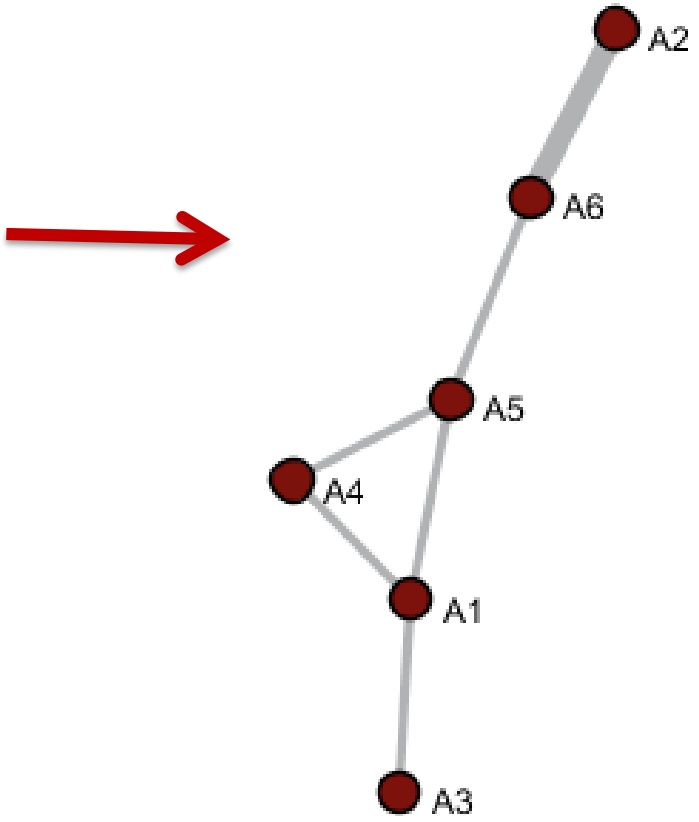


Co-Author Analysis & Visualization

Network Extraction:

Weighted, undirected co-occurrence network

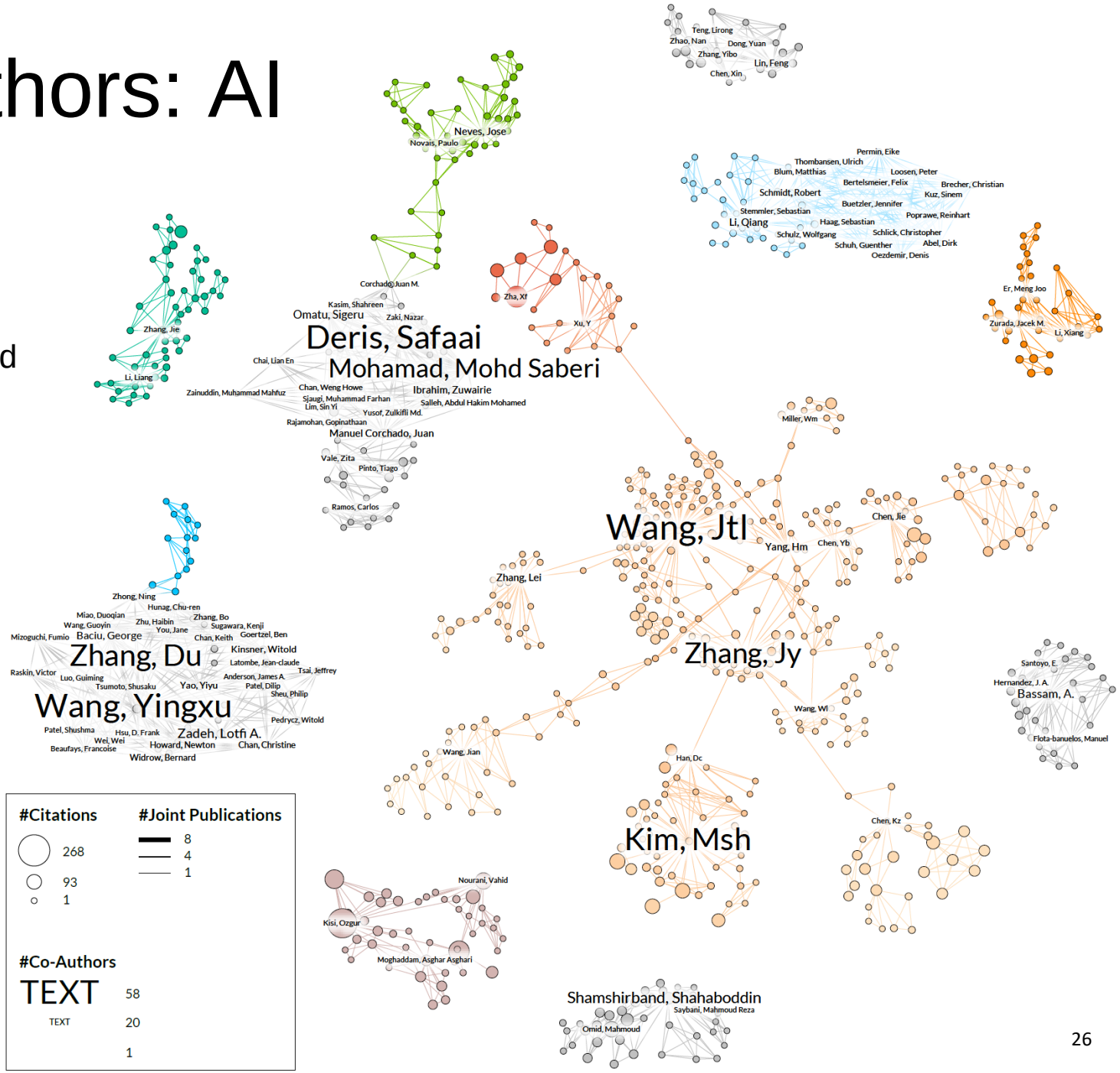
	A	B	C	D
1	**Paper**	**Authors**	**References**	**Year**
2	P1	A1		1970
3	P2	A2;A6	P1	1980
4	P3	A1;A3	P1;P2	1990
5	P4	A1;A4;A5	P2	1995
6	P5	A5;A6	P1;P2;P3;P4	1995
7	P6	A2;A6	P5	2000



Co-Authors: AI

Nodes: 858
Edges: 2,544
No self loops

Weakly connected components 10
Largest has 386 nodes



Topical Analysis & Visualization

Year	Journal Title
2006	BMC EVOLUTIONARY BIOLOGY
2005	FEBS JOURNAL
2005	FEBS JOURNAL
2005	NATURE PHYSICS
2005	EUROPHYSICS LETTERS
2005	NATURE
2005	PNAS
2005	NATURE
2005	SCIENCE
2005	PHYSICAL REVIEW E

Topical Visualization

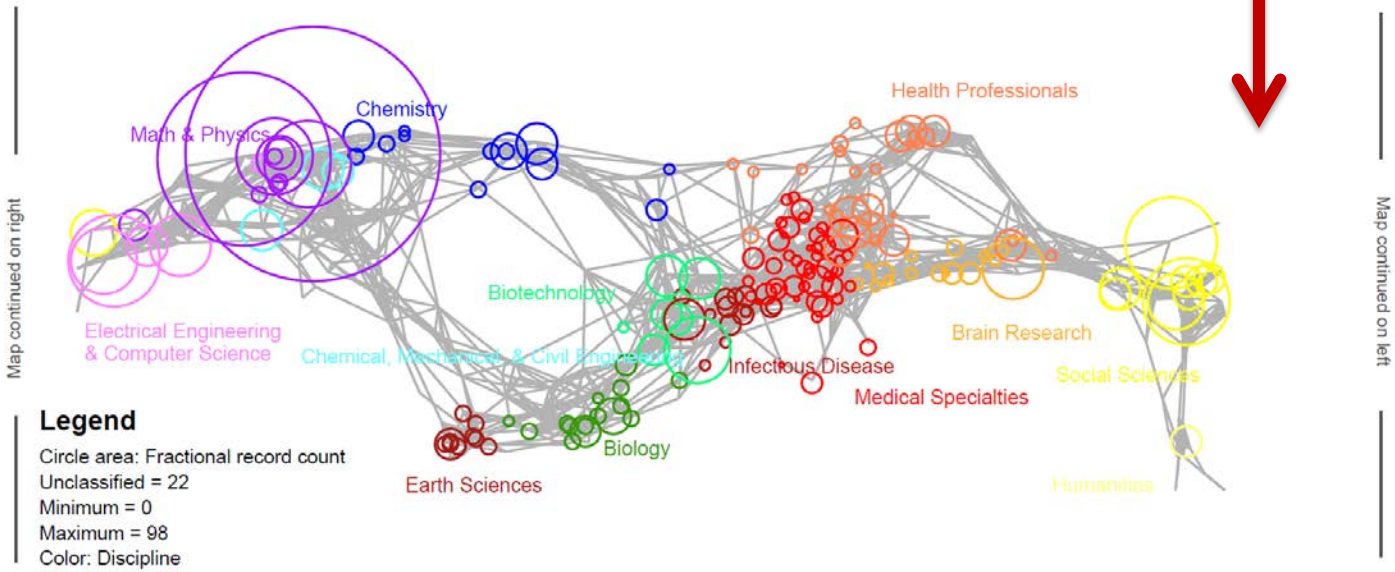
Generated from 361 Unique ISI Records
90 out of 112 records were mapped to 182 subdisciplines and 13 disciplines.
November 21, 2016 | 01:46 PM −05:00

Biology

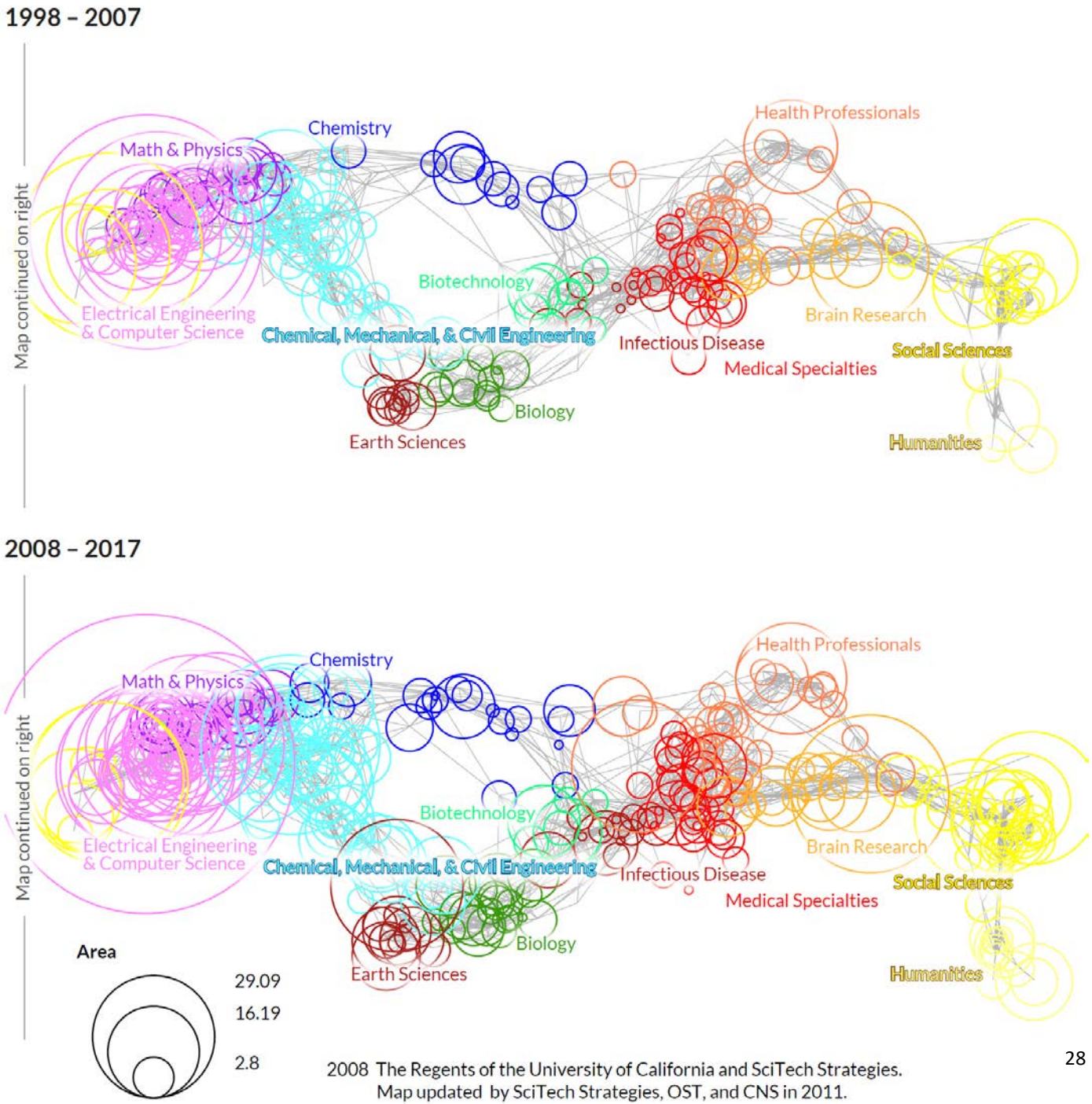
1 bmc evolutionary biology
1 naturwissenschaften

Biotechnology

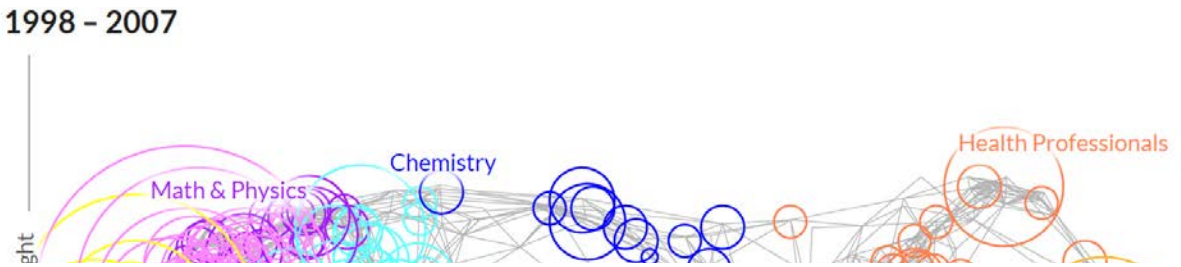
1 bmc bioinformatics
2 febs journal
1 genome research
1 international microbiology
1 nature biotechnology
3 nature genetics



Artificial Intelligence

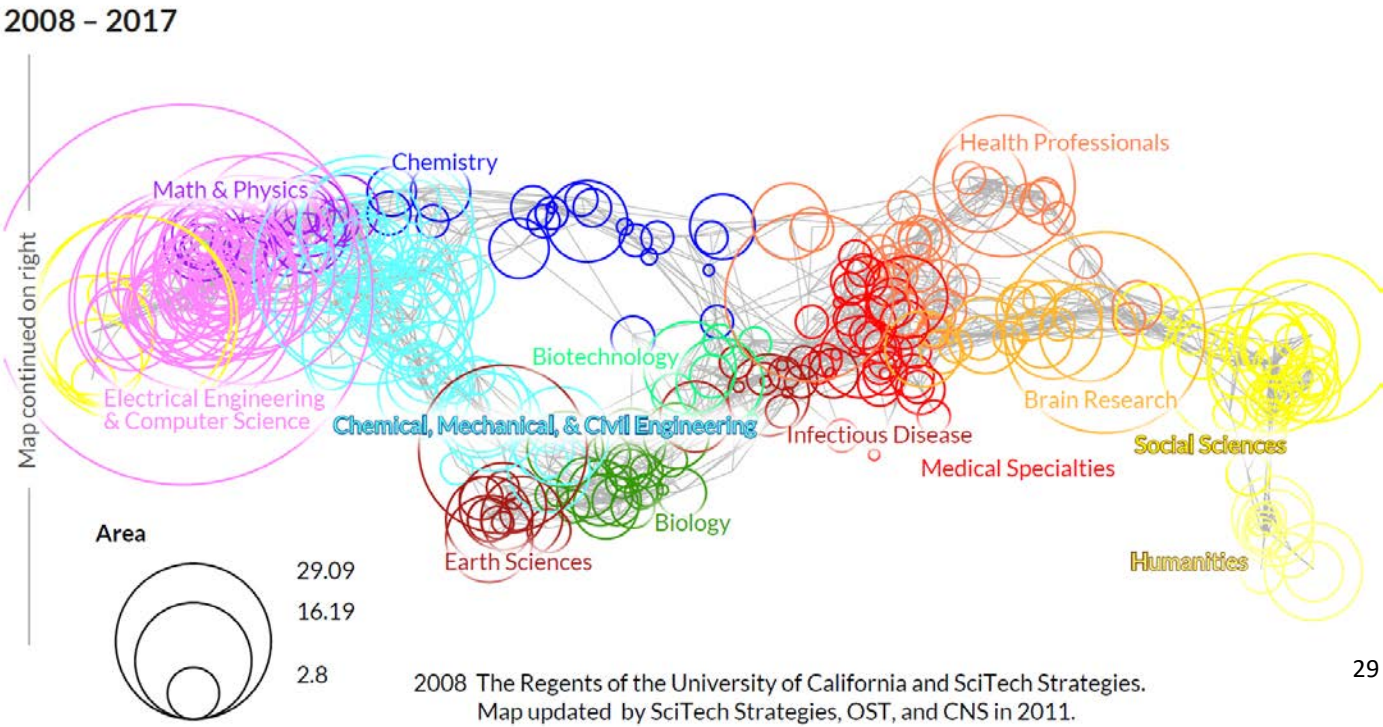


Artificial Intelligence

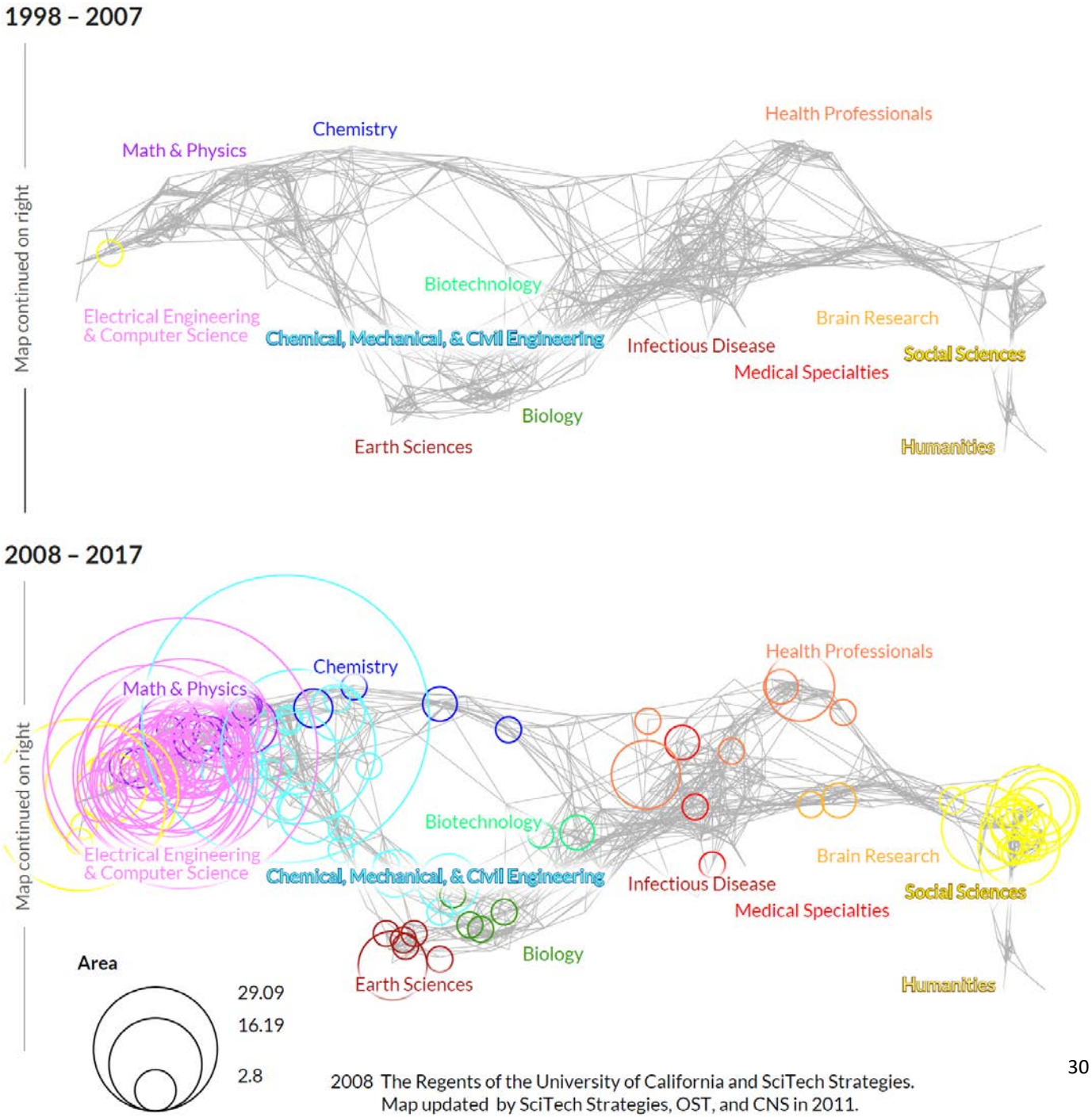


Top-10 most cited

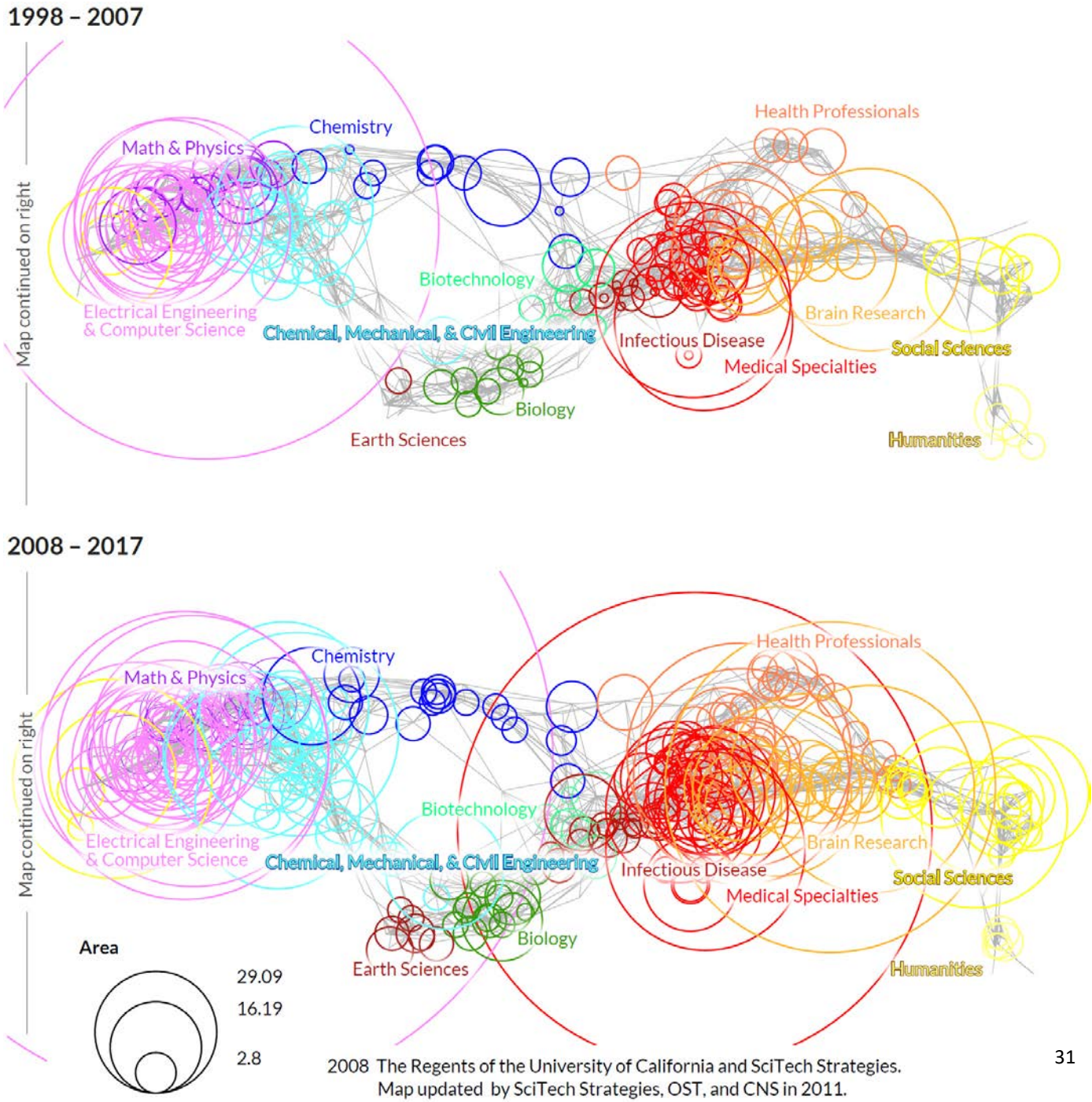
1998	Modeling supply chain dynamics: A multiagent approach	314
1998	Artificial neural networks (the multilayer perceptron) - A review of …	537
2003	Psychological aspects of natural language use: Our words, our …	559
2008	Advances in Diagnostic Techniques for Induction Machines	401
2011	ViBe: A Universal Background Subtraction Algorithm for Video Sequ...	369

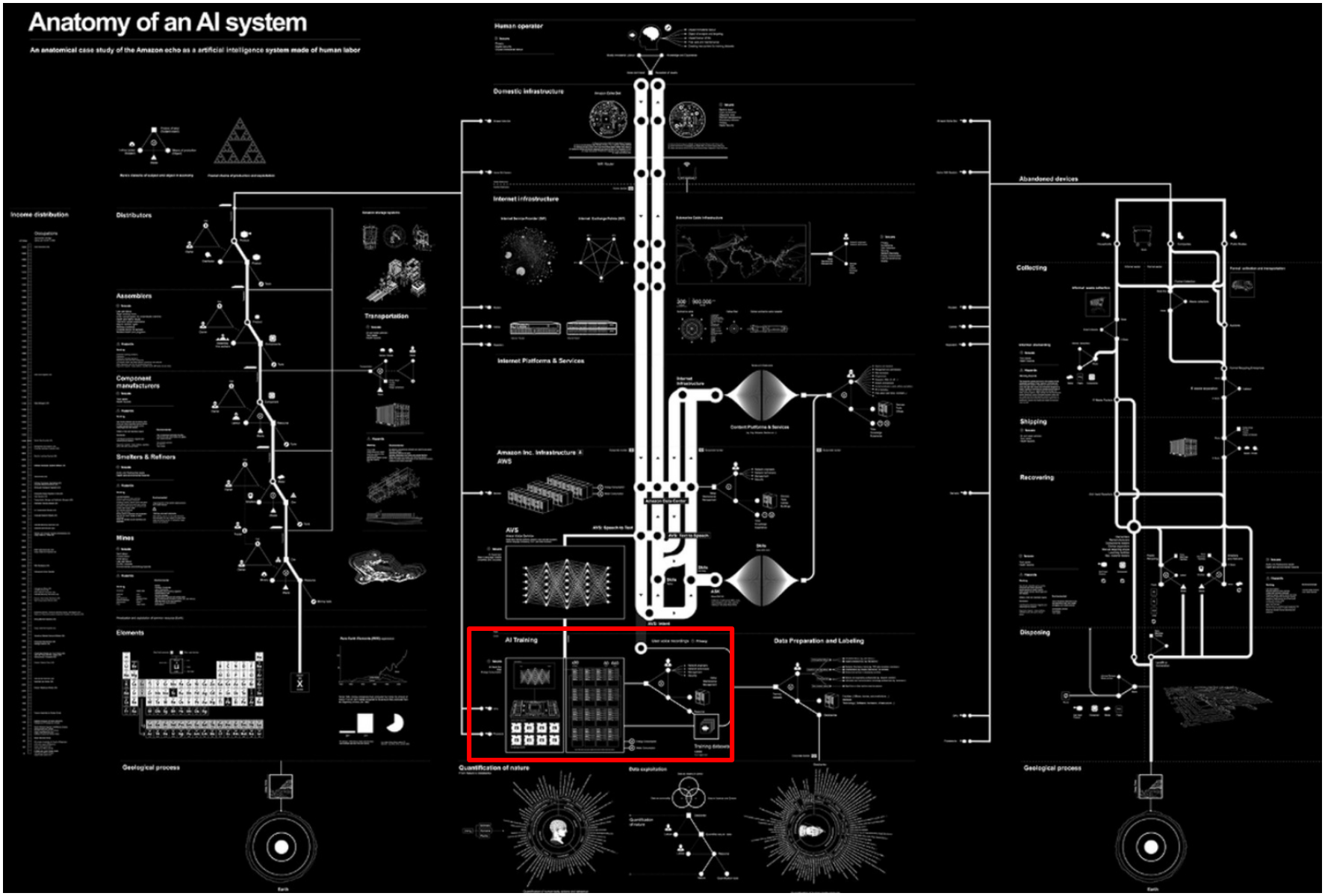


Internet of Things



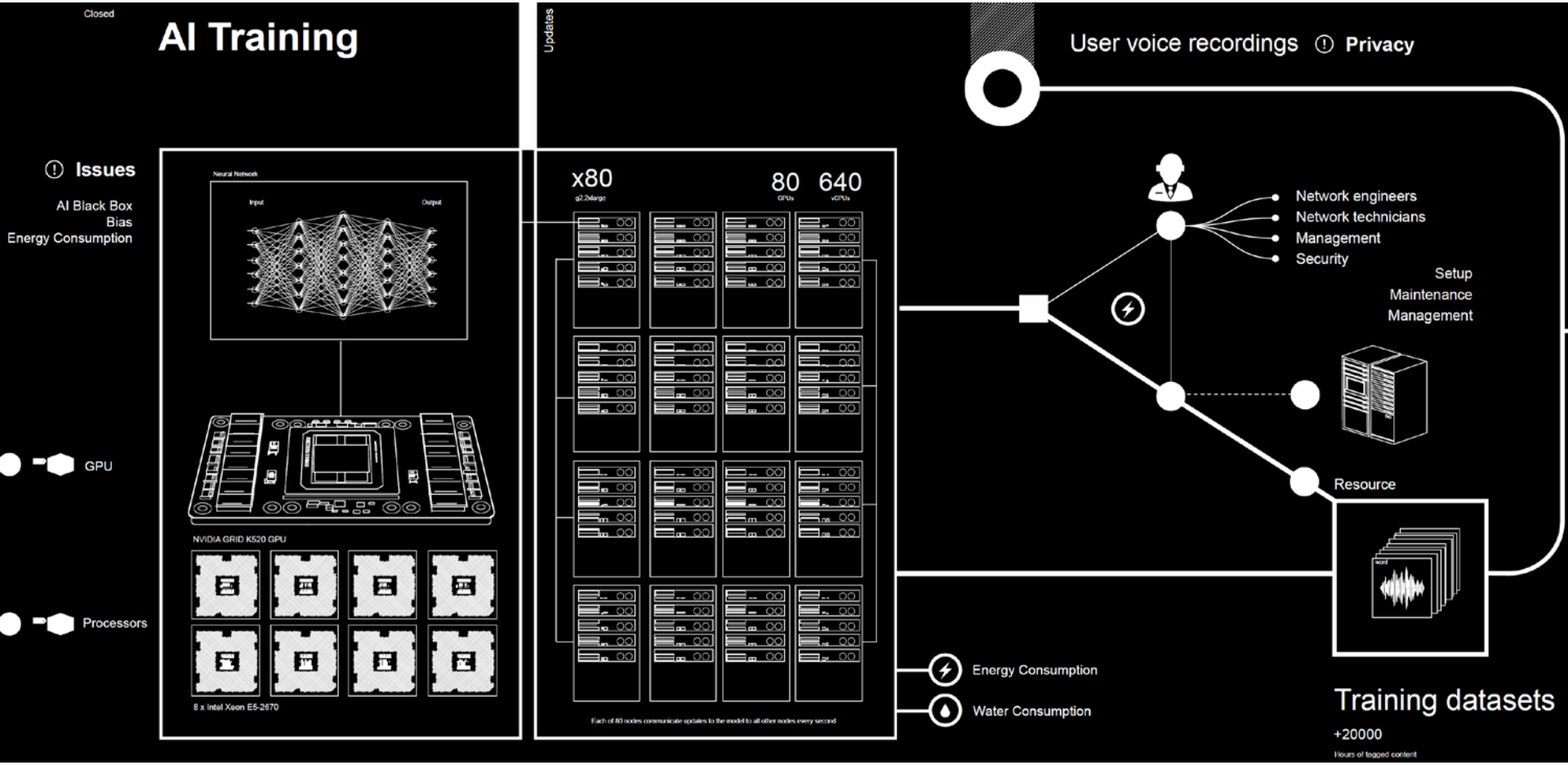
Robotics





Anatomy of an AI system: A map of the many processes—extracting material resources, data, and human labor—that make an Amazon Echo work.

https://www.theverge.com/2018/9/9/17832124/ai-artificial-intelligence-supply-chain-anatomy-of-ai-kate-crawford-interview



Anatomy of an AI system: A map of the many processes—extracting material resources, data, and human labor—that make an Amazon Echo work.
https://www.theverge.com/2018/9/9/17832124/ai-artificial-intelligence-supply-chain-anatomy-of-ai-kate-crawford-interview

User Studies

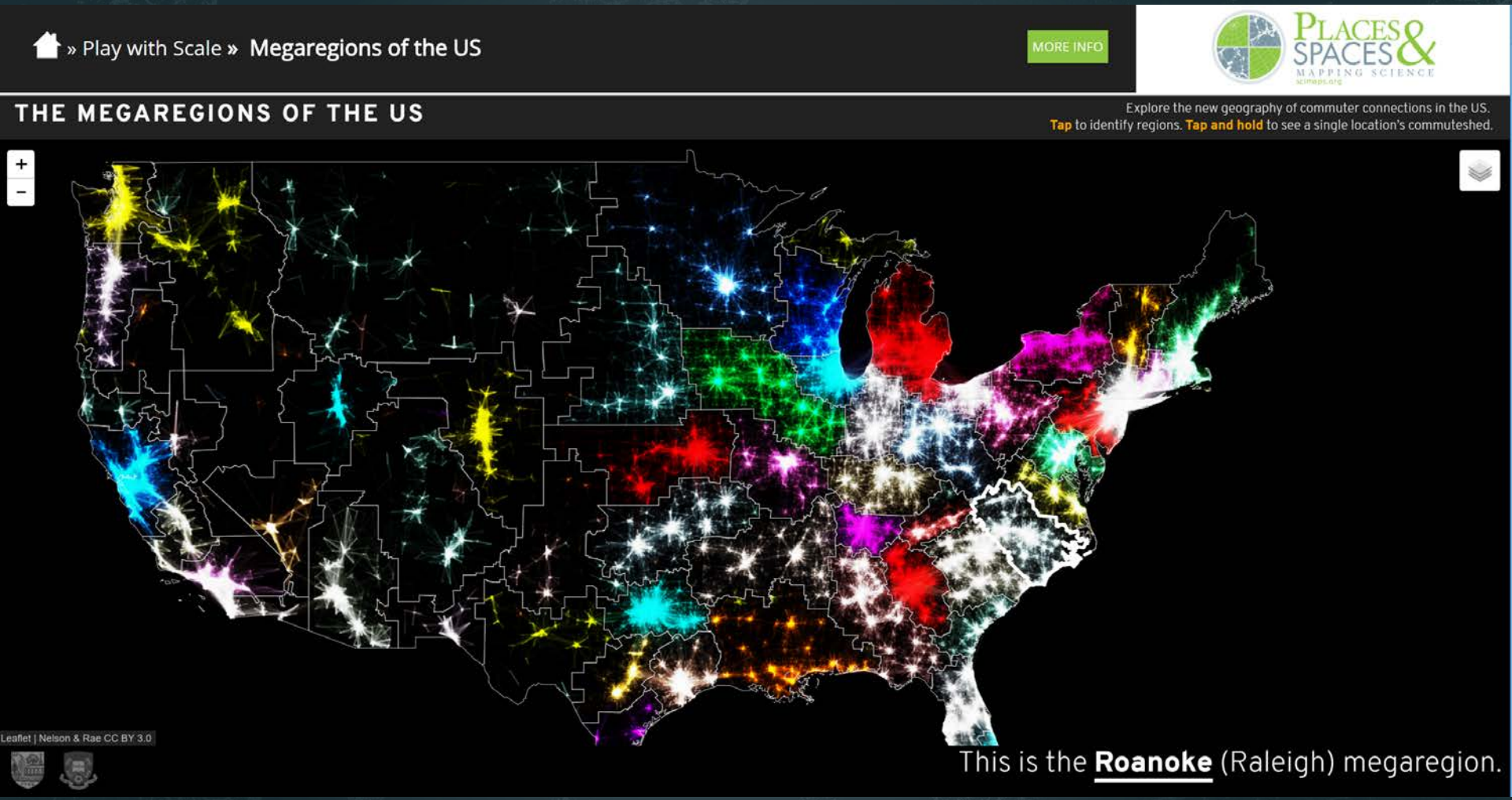
Visualizations were validated and then optimized through a user study that examined readability, memorability, reproducibility, and utility of different data visualizations.

- *readability*: users can easily find certain records, linkages, patterns (e.g., bursts, high degree nodes), pathways;
- *memorability*: users can answer if certain items were present and where they were placed;
- *reproducibility:* users can re-draw key elements of the visualizations after exploration; and
- *utility*: the visualizations provide actionable insights and/or prompt more meaningful questions, or are otherwise useful for human decision-making.

Standard human subject studies procedures were used:

After giving consent, participants

1. completed a pre-test questionnaire that collected demographic info (age, gender, native language, expertise),
2. read an information sheet with basic information on the specific study,
3. performed specific tasks using different visualizations, and
4. completed a post-test questionnaire / interview.



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

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.

https://visanalytics.cns.iu.edu

Visual Analytics Certificate

CNS Cyberinfrastructure for Network Science Center

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*

Starting Date: *Jan 14, 2018*

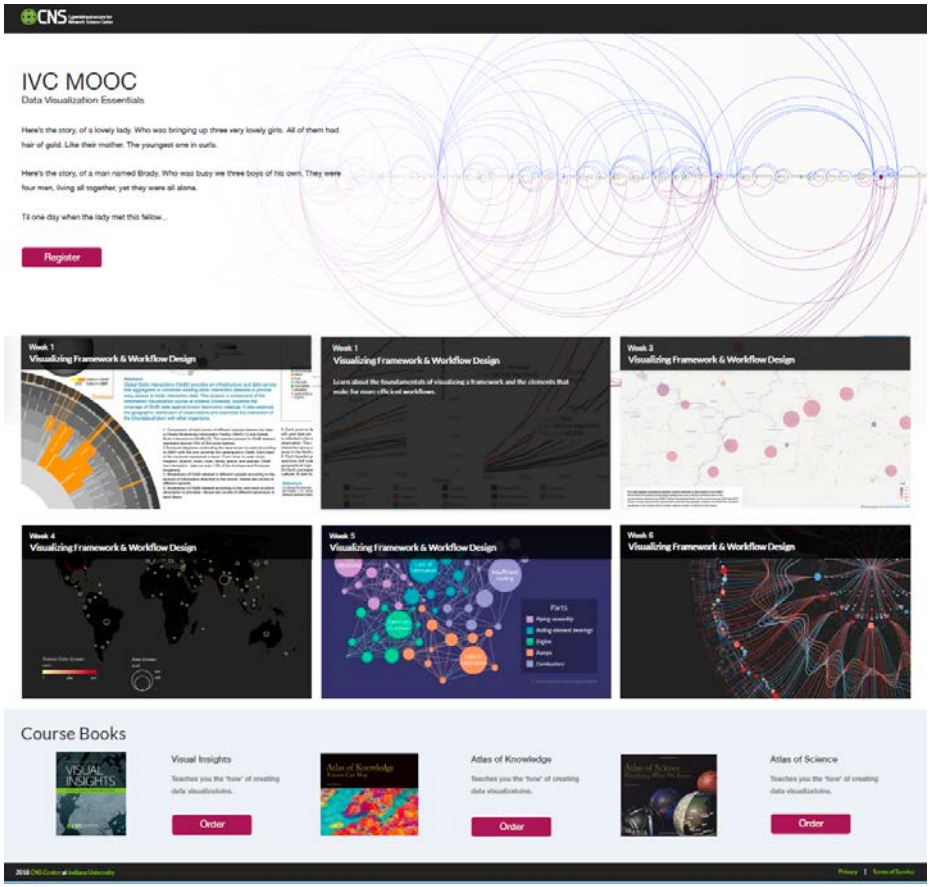
Covers:

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.



https://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.


Monitor S&T Developments


Manage Communication Flows

Optimize Traffic Flows


Manage Customer Feedback


Improve Network Resilience


Develop Workforce

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/VACCaseStudy

https://visanalytics.cns.iu.edu

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

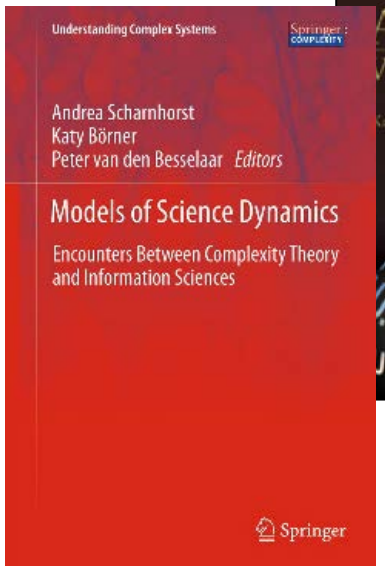
Börner, Katy (2010) **Atlas of Science: Visualizing What We Know**. The MIT Press. http://scimaps.org/atlas

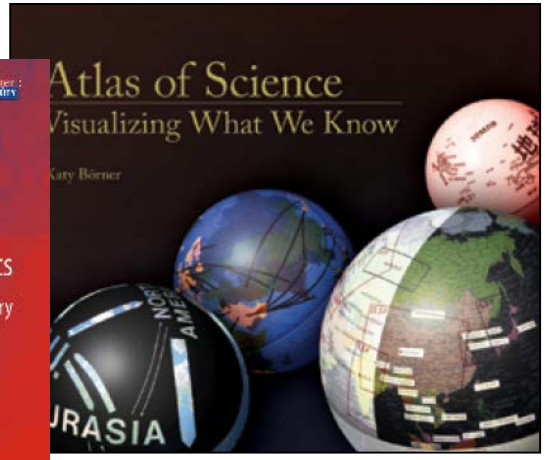
Scharnhorst, Andrea, Börner, Katy, van den Besselaar, Peter (2012) **Models of Science Dynamics**. Springer Verlag.

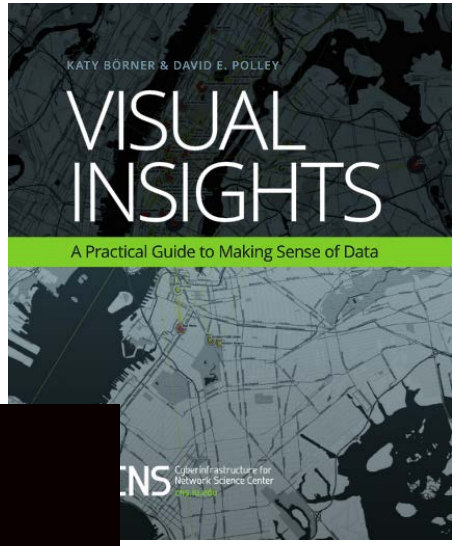
Katy Börner, Michael Conlon, Jon Corson-Rikert, Cornell, Ying Ding (2012) **VIVO: A Semantic Approach to Scholarly Networking and Discovery**. Morgan & Claypool.

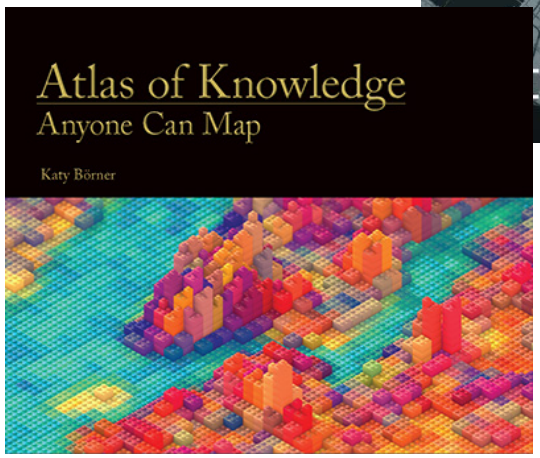
Katy Börner and David E Polley (2014) **Visual Insights: A Practical Guide to Making Sense of Data**. The MIT Press.

Börner, Katy (2015) **Atlas of Knowledge: Anyone Can Map**. The MIT Press. http://scimaps.org/atlas2









SICE CAMPUS PARTNER

OPEN HOUSE

September 28, 3–4 p.m.

Luddy Hall, 700 N. Woodlawn Avenue

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