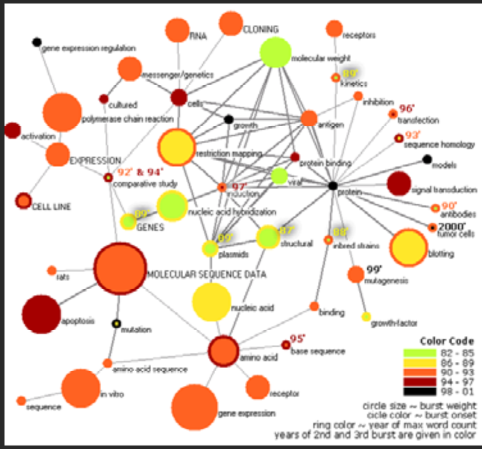
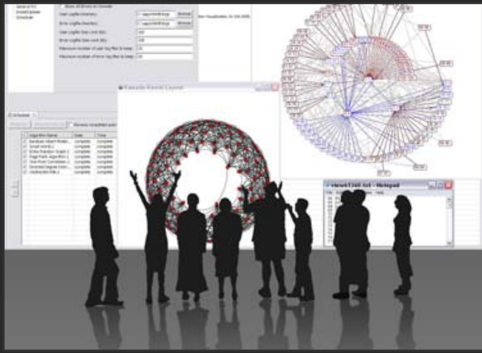
Descriptive and Process Models of Scientific Structure and Evolution

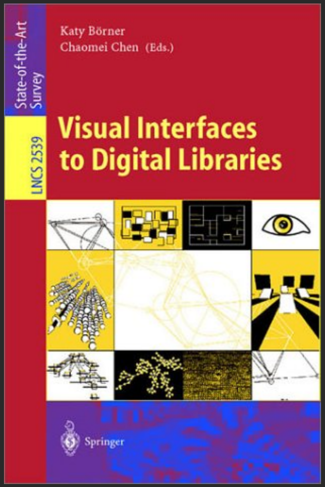
Dr. Katy Börner
Cyberinfrastructure for Network Science Center, Director
Information Visualization Laboratory, Director
School of Library and Information Science
Indiana University, Bloomington, IN
katy@indiana.edu

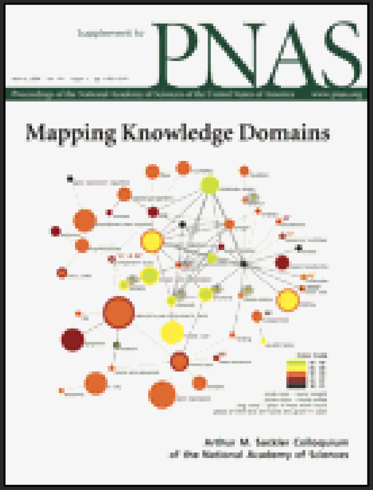
NESCent Seminar
February 20th, 2008





Computational Scientometrics: Studying Science by Scientific Means




PNAS
Mapping Knowledge Domains
Arthur M. Sackler Colloquium of the National Academy of Sciences

- *Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003).* ***Visualizing Knowledge Domains.*** *In Blaise Cronin (Ed.), Annual Review of Information Science & Technology, Medford, NJ: Information Today, Inc./American Society for Information Science and Technology, 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, Sanyal, Soma and Vespignani, Alessandro (2007).* ***Network Science.*** *In Blaise Cronin (Ed.), Annual Review of Information Science & Technology, Information Today, Inc./American Society for Information Science and Technology, Medford, NJ, Volume 41, Chapter 12, pp. 537-607.* http://ivl.slis.indiana.edu/km/pub/2007-borner-arist.pdf

- ***Places & Spaces: Mapping Science*** *exhibit, see also* *http://scimaps.org*.

Process of Analyzing and Mapping Knowledge Domains

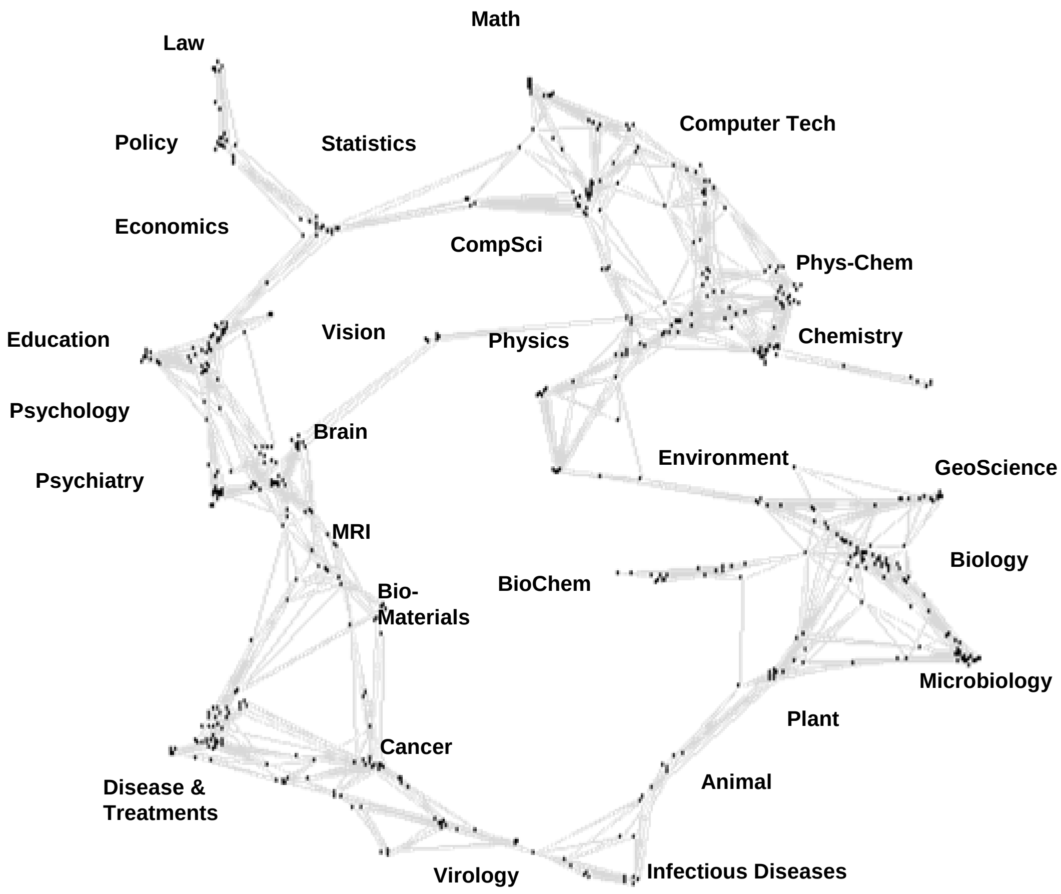
DATA EXTRACTION	UNIT OF ANALYSIS	MEASURES	LAYOUT (often one code does both similarity and ordination steps)		DISPLAY
			SIMILARITY	ORDINATION	
SEARCHES	COMMON	COUNTS/FREQUENCIES	SCALAR (unit by unit matrix)	DIMENSIONALITY REDUCTION	INTERACTION
ISI	CHOICES	Attributes (e.g. terms)	Direct citation	Eigenvector/ Eigenvalue solutions	Browse
INSPEC	Journal	Author citations	Co-citation	Factor Analysis (FA) and	Pan
Eng Index	Document	Co-citations	Combined linkage	Principal Components Analysis (PCA)	Zoom
Medline	Author	By year	Co-word / co-term	Multi-dimensional scaling (MDS)	Filter
ResearchIndex	Term		Co-classification	LSA , Topics	Query
Patents		THRESHOLDS		Pathfinder networks (PFNet)	Detail on demand
etc.		By counts	VECTOR (unit by attribute matrix)	Self-organizing maps (SOM)	
			Vector space model (words/terms)	includes SOM, ET-maps, etc.	ANALYSIS
BROADENING			Latent Semantic Analysis (words/terms)		
By citation			incl. Singular Value Decomp (SVD)	CLUSTER ANALYSIS	
By terms					
			CORRELATION (if desired)	SCALAR	
			Pearson's R on any of above	Triangulation	
				Force-directed placement (FDP)	

Börner, Katy, Chen, Chaomei, and Boyack, Kevin. (2003) Visualizing Knowledge Domains. In Blaise Cronin (Ed.), Annual Review of Information Science & Technology, Volume 37, Medford, NJ: Information Today, Inc./American Society for Information Science and Technology, chapter 5, pp. 179-255.

Latest 'Base Map' of Science

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007). Mapping the Structure and Evolution of Chemistry Research. 11th International Conference on Scientometrics and Informetrics. pp. 112-123.

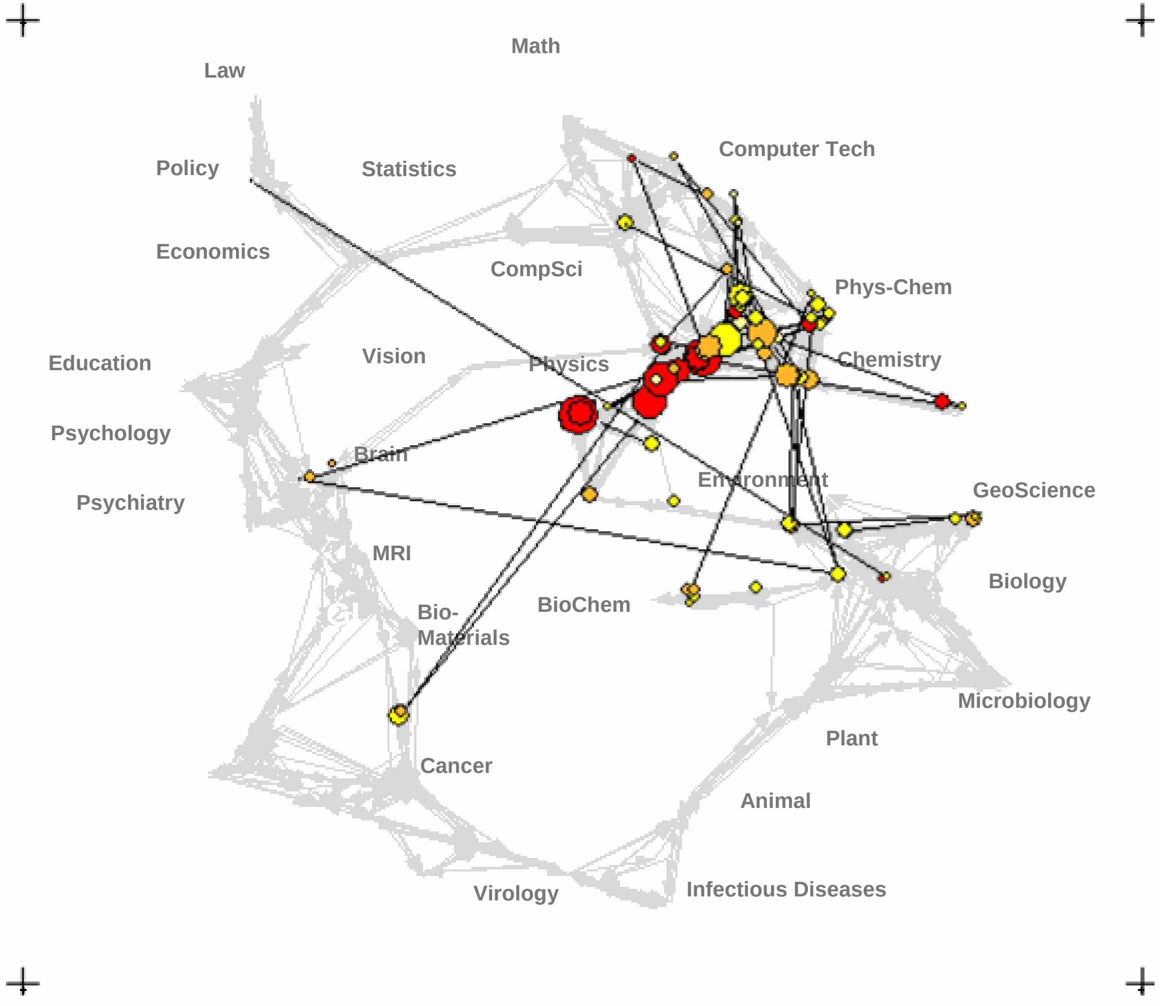
- Uses combined SCI/SSCI from 2002
  - 1.07M papers, 24.5M references, 7,300 journals
  - Bibliographic coupling of papers, aggregated to journals
- Initial ordination and clustering of journals gave 671 clusters
- Coupling counts were reaggregated at the journal cluster level to calculate the
  - (x,y) positions for each journal cluster
  - by association, (x,y) positions for each journal



Science map applications: Identifying core competency

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

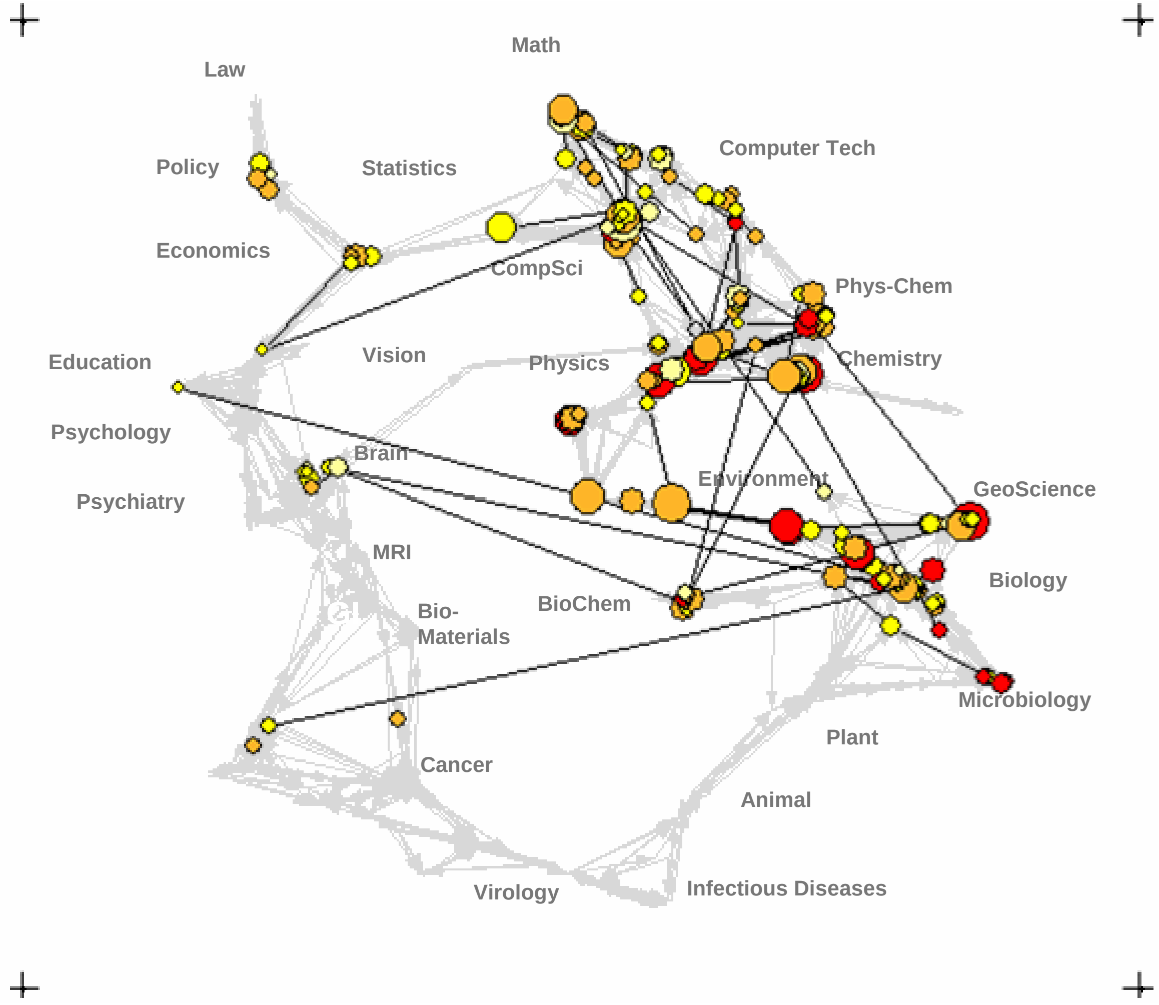
Funding patterns of the US Department of Energy (DOE)



Science map applications: Identifying core competency

Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

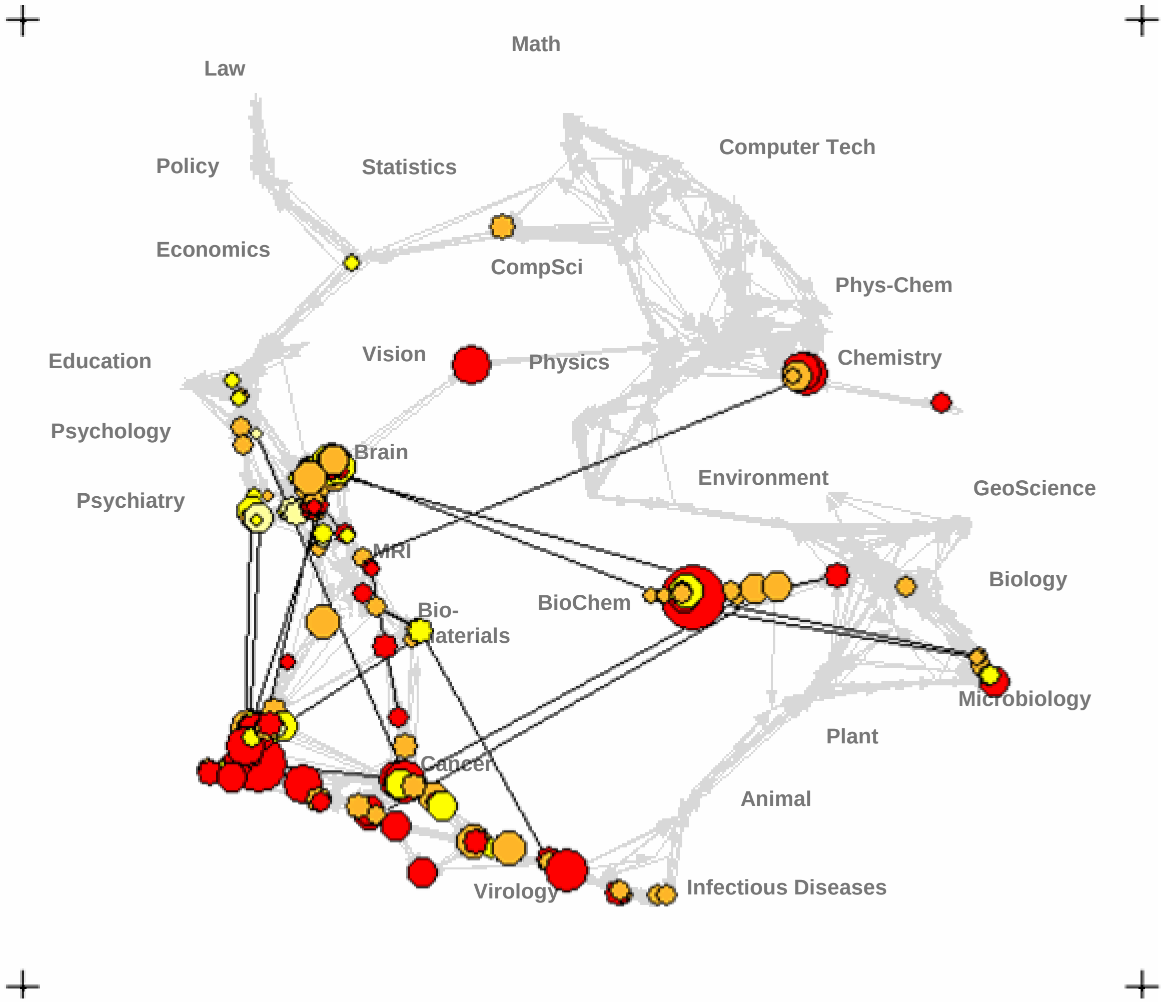
Funding Patterns of the National Science Foundation (NSF)



Science map applications: Identifying core competency

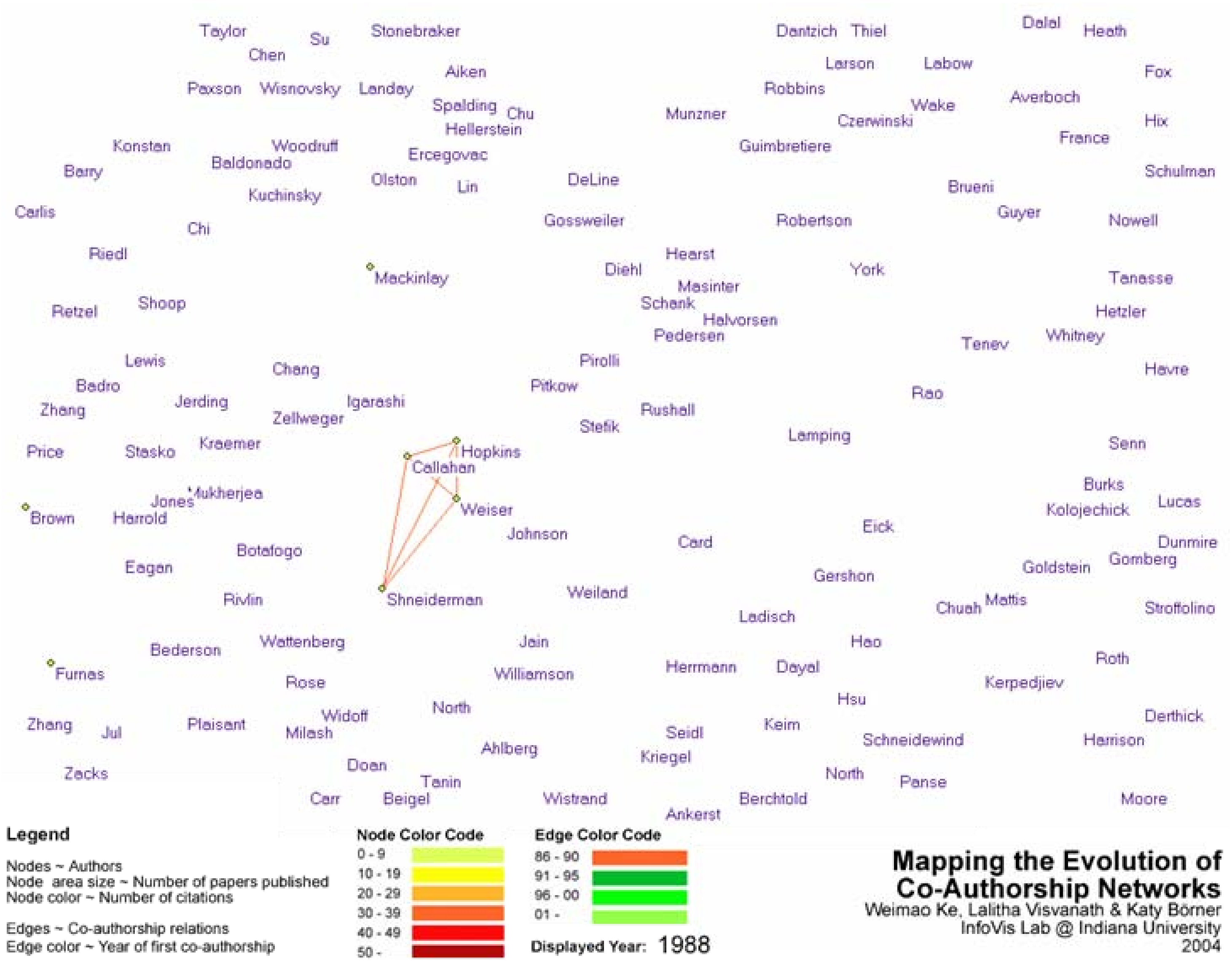
Kevin W. Boyack, Katy Börner, & Richard Klavans (2007).

Funding Patterns of the National Institutes of Health (NIH)



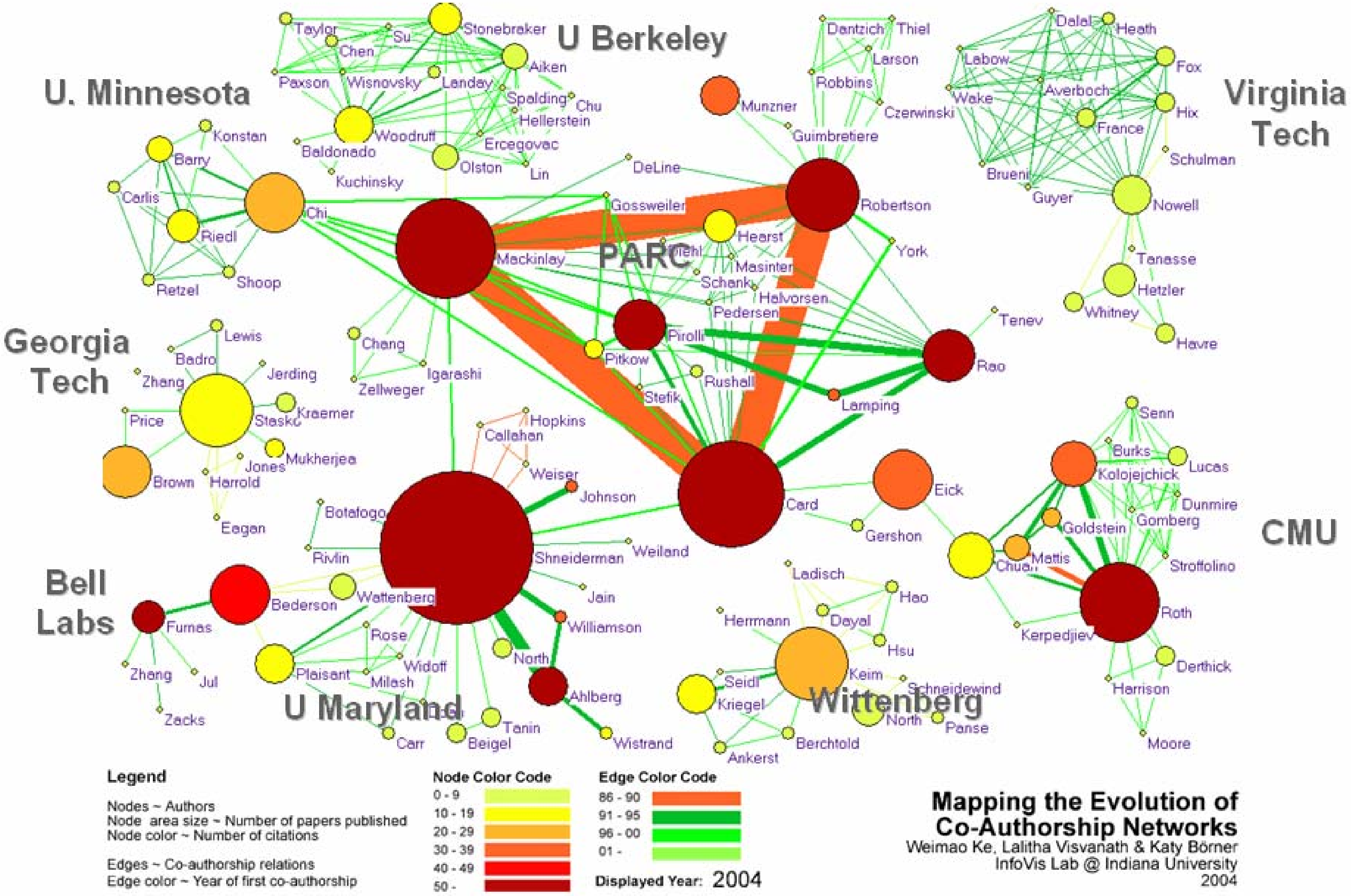
Mapping the Evolution of Co-Authorship Networks

Ke, Visvanath & Börner, (2004) Won 1st price at the IEEE InfoVis Contest.



Mapping the Evolution of Co-Authorship Networks

Ke, Visvanath & Börner, (2004) Won 1st price at the IEEE InfoVis Contest.

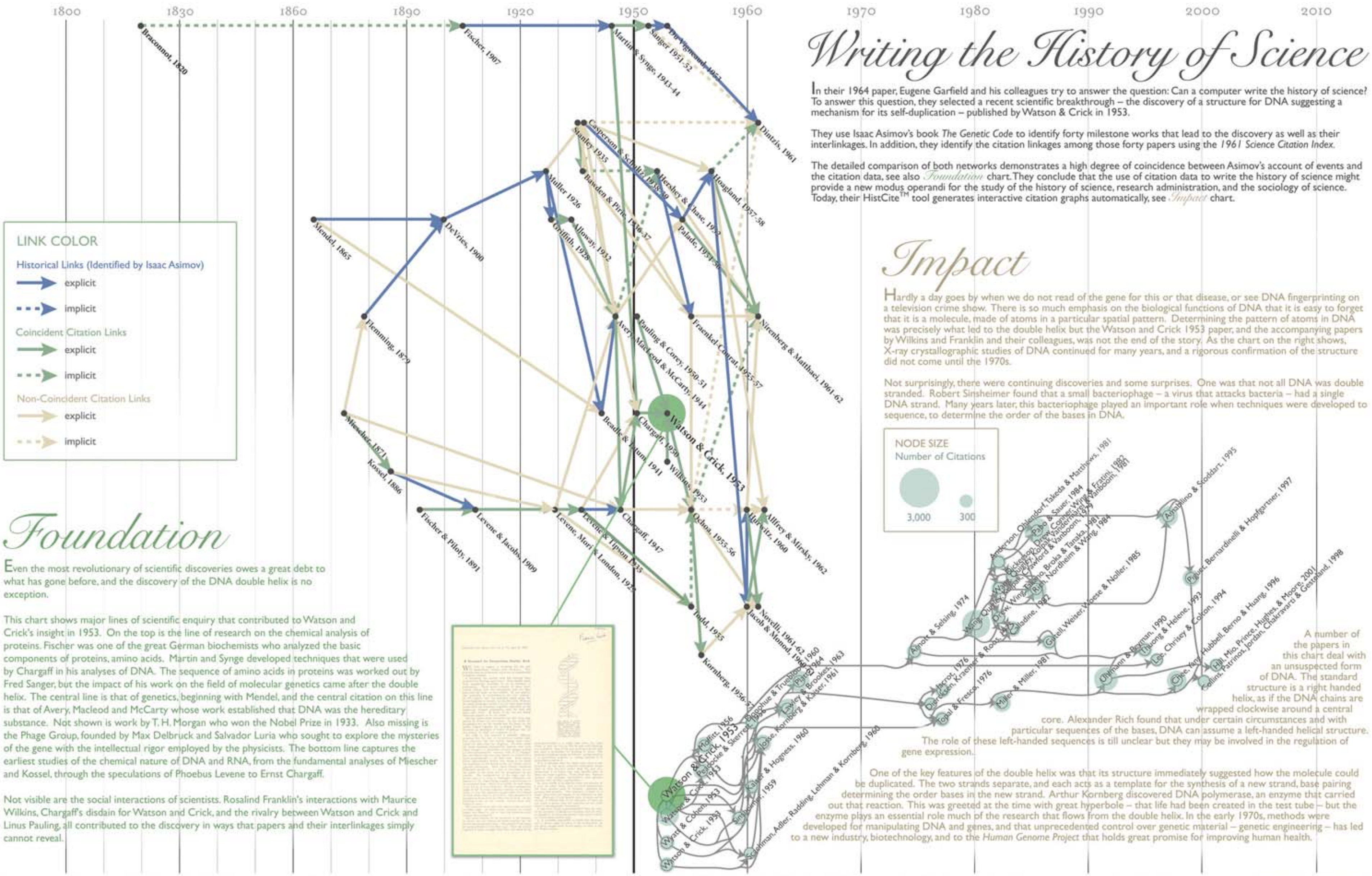


Writing the History of Science

In their 1964 paper, Eugene Garfield and his colleagues try to answer the question: Can a computer write the history of science? To answer this question, they selected a recent scientific breakthrough – the discovery of a structure for DNA suggesting a mechanism for its self-duplication – published by Watson & Crick in 1953.

They use Isaac Asimov's book *The Genetic Code* to identify forty milestone works that lead to the discovery as well as their interlinkages. In addition, they identify the citation linkages among those forty papers using the *1961 Science Citation Index*.

The detailed comparison of both networks demonstrates a high degree of coincidence between Asimov's account of events and the citation data, see also *Foundation* chart. They conclude that the use of citation data to write the history of science might provide a new modus operandi for the study of the history of science, research administration, and the sociology of science. Today, their HistCite™ tool generates interactive citation graphs automatically, see *Impact* chart.



Foundation

Even the most revolutionary of scientific discoveries owes a great debt to what has gone before, and the discovery of the DNA double helix is no exception.

This chart shows major lines of scientific enquiry that contributed to Watson and Crick's insight in 1953. On the top is the line of research on the chemical analysis of proteins. Fischer was one of the great German biochemists who analyzed the basic components of proteins, amino acids. Martin and Synge developed techniques that were used by Chargaff in his analyses of DNA. The sequence of amino acids in proteins was worked out by Fred Sanger, but the impact of his work on the field of molecular genetics came after the double helix. The central line is that of genetics, beginning with Mendel, and the central citation on this line is that of Avery, Macleod and McCarty whose work established that DNA was the hereditary substance. Not shown is work by T. H. Morgan who won the Nobel Prize in 1933. Also missing is the Phage Group, founded by Max Delbruck and Salvador Luria who sought to explore the mysteries of the gene with the intellectual rigor employed by the physicists. The bottom line captures the earliest studies of the chemical nature of DNA and RNA, from the fundamental analyses of Miescher and Kossel, through the speculations of Phoebus Levene to Ernst Chargaff.

Not visible are the social interactions of scientists. Rosalind Franklin's interactions with Maurice Wilkins, Chargaff's disdain for Watson and Crick, and the rivalry between Watson and Crick and Linus Pauling, all contributed to the discovery in ways that papers and their interlinkages simply cannot reveal.

Impact

Hardly a day goes by when we do not read of the gene for this or that disease, or see DNA fingerprinting on a television crime show. There is so much emphasis on the biological functions of DNA that it is easy to forget that it is a molecule, made of atoms in a particular spatial pattern. Determining the pattern of atoms in DNA was precisely what led to the double helix but the Watson and Crick 1953 paper, and the accompanying papers by Wilkins and Franklin and their colleagues, was not the end of the story. As the chart on the right shows, X-ray crystallographic studies of DNA continued for many years, and a rigorous confirmation of the structure did not come until the 1970s.

Not surprisingly, there were continuing discoveries and some surprises. One was that not all DNA was double stranded. Robert Sinsheimer found that a small bacteriophage – a virus that attacks bacteria – had a single DNA strand. Many years later, this bacteriophage played an important role when techniques were developed to sequence, to determine the order of the bases in DNA.

A number of the papers in this chart deal with an unsuspected form of DNA. The standard structure is a right handed helix, as if the DNA chains are wrapped clockwise around a central core. Alexander Rich found that under certain circumstances and with particular sequences of the bases, DNA can assume a left-handed helical structure. The role of these left-handed sequences is till unclear but they may be involved in the regulation of gene expression..

One of the key features of the double helix was that its structure immediately suggested how the molecule could be duplicated. The two strands separate, and each acts as a template for the synthesis of a new strand, base pairing determining the order bases in the new strand. Arthur Kornberg discovered DNA polymerase, an enzyme that carried out that reaction. This was greeted at the time with great hyperbole – that life had been created in the test tube – but the enzyme plays an essential role much of the research that flows from the double helix. In the early 1970s, methods were developed for manipulating DNA and genes, and that unprecedented control over genetic material – genetic engineering – has led to a new industry, biotechnology, and to the *Human Genome Project* that holds great promise for improving human health.

1947 A. Mirsky & I. Goodman

1947 J. Monod

1947 E. Chargaff

1948 A. Mirsky & P.C. Koller

1950 R. Franklin

1953 J. D. Watson & F. H. C. Crick

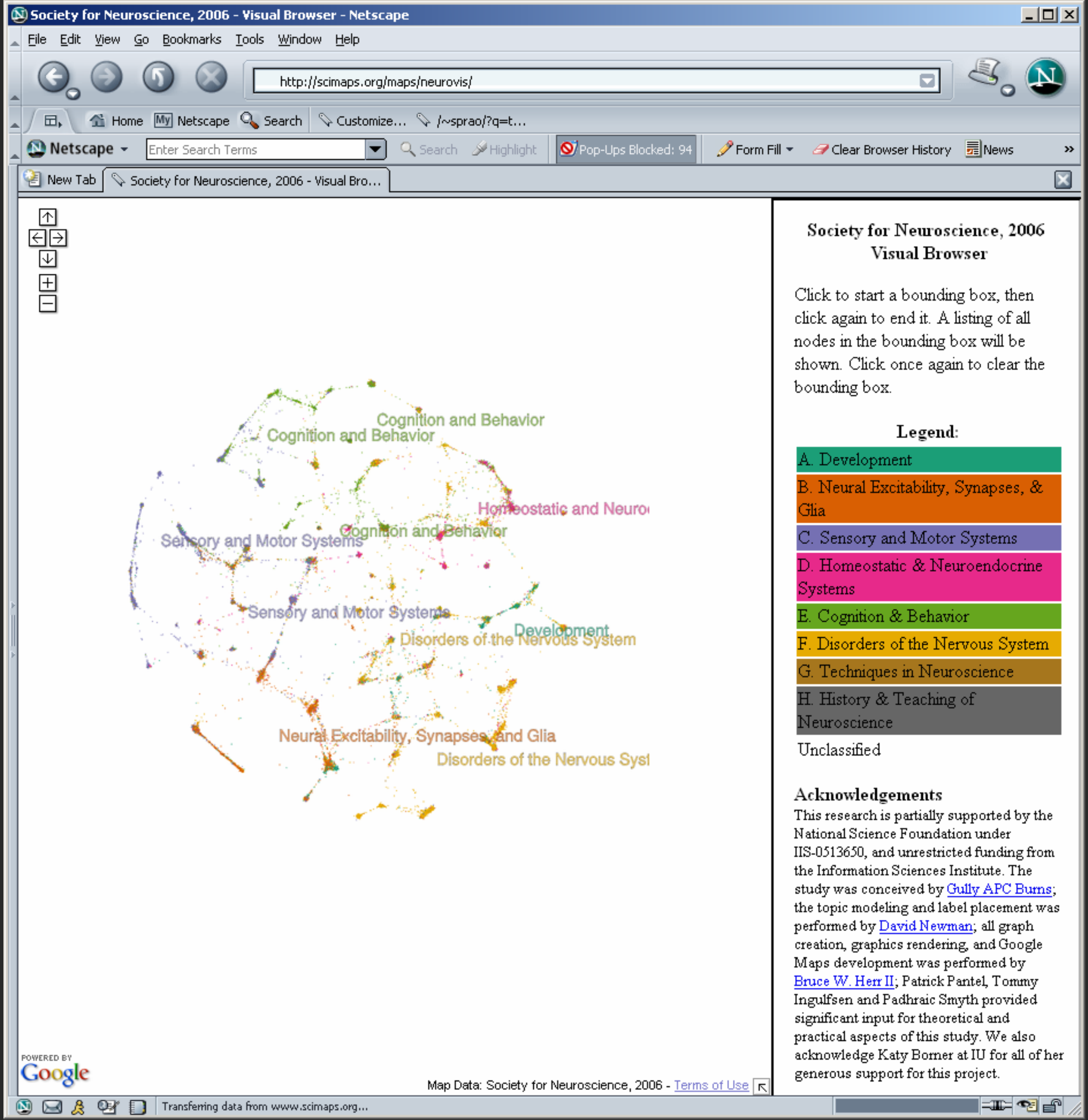
1963 V. Ingram, M. Nirenberg & M. Staehelin

1963 J. Speyer & M. Nirenberg

1968 C. Thomas & A. Kornberg

1974 Participants

1978 A. Kornberg


Society for Neuroscience, 2006 - Visual Browser - Netscape
File Edit View Go Bookmarks Tools Window Help
http://scimaps.org/maps/neurovis/
Home My Netscape Search Customize... /~sprao/?q=t...
Netscape Enter Search Terms Search Highlight Pop-Ups Blocked: 94 Form Fill Clear Browser History News
New Tab Society for Neuroscience, 2006 - Visual Bro...
Cognition and Behavior
Cognition and Behavior
Homeostatic and Neuro
Sensory and Motor Systems
Cognition and Behavior
Sensory and Motor Systems
Development
Disorders of the Nervous System
Neural Excitability, Synapses, and Glia
Disorders of the Nervous Syst
POWERED BY Google
Map Data: Society for Neuroscience, 2006 - Terms of Use
Transferring data from www.scimaps.org...

Society for Neuroscience, 2006 Visual Browser

Click to start a bounding box, then click again to end it. A listing of all nodes in the bounding box will be shown. Click once again to clear the bounding box.

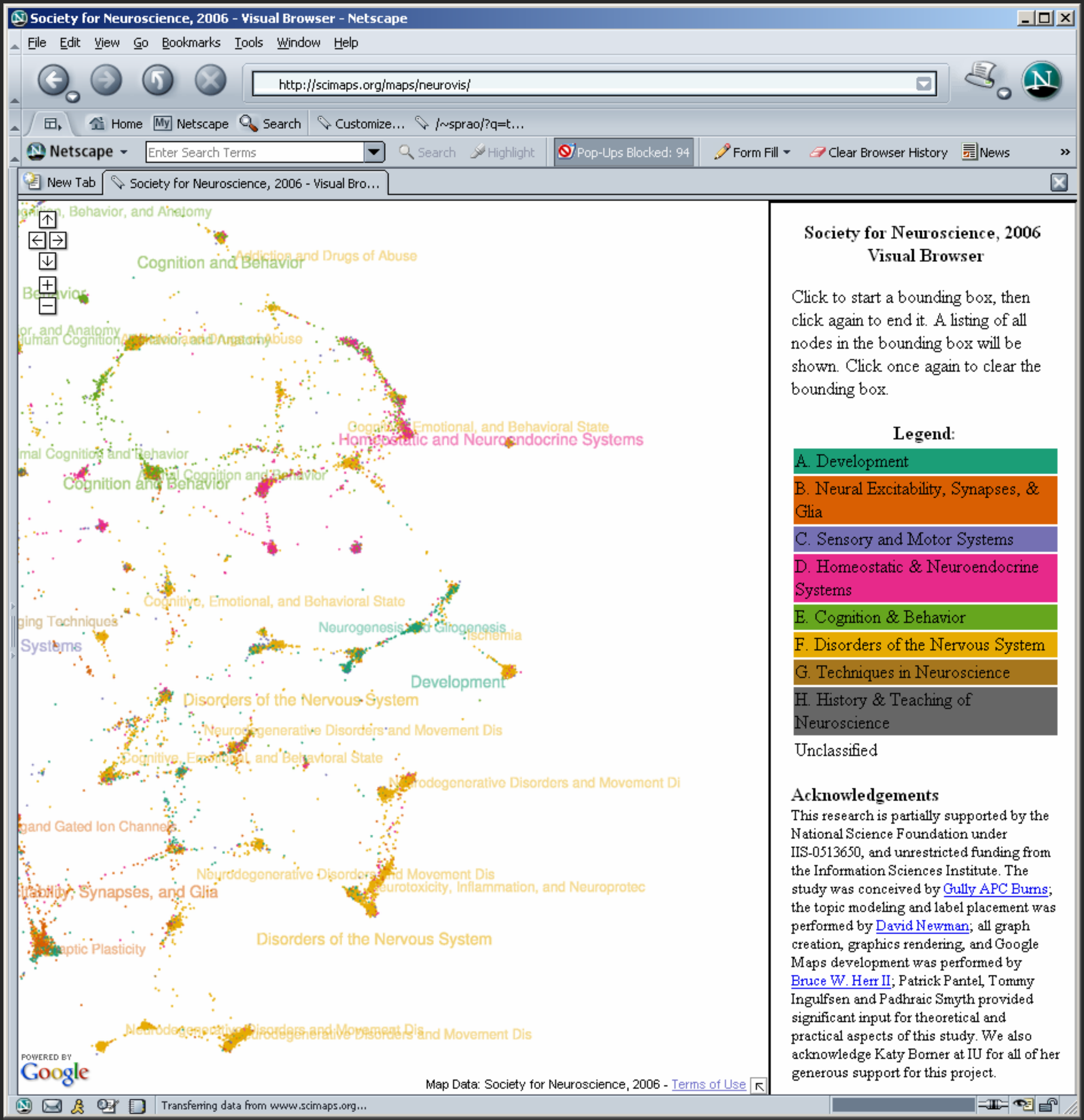
Legend:

- A. Development
- B. Neural Excitability, Synapses, & Glia
- C. Sensory and Motor Systems
- D. Homeostatic & Neuroendocrine Systems
- E. Cognition & Behavior
- F. Disorders of the Nervous System
- G. Techniques in Neuroscience
- H. History & Teaching of Neuroscience
- Unclassified

Acknowledgements

This research is partially supported by the National Science Foundation under IIS-0513650, and unrestricted funding from the Information Sciences Institute. The study was conceived by Gully APC Burns; the topic modeling and label placement was performed by David Newman; all graph creation, graphics rendering, and Google Maps development was performed by Bruce W. Herr II; Patrick Pantel, Tommy Ingulfsen and Padhraic Smyth provided significant input for theoretical and practical aspects of this study. We also acknowledge Katy Borner at IU for all of her generous support for this project.

Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006 Visual Browser, 2007, http://scimaps.org/maps/neurovis/


Society for Neuroscience, 2006 - Visual Browser - Netscape
File Edit View Go Bookmarks Tools Window Help
http://scimaps.org/maps/neurovis/
Home
Netscape
Search
Customize...
/~sprao/?q=t...
Netscape
Enter Search Terms
Search
Highlight
Pop-Ups Blocked: 94
Form Fill
Clear Browser History
News
New Tab
Society for Neuroscience, 2006 - Visual Bro...
Cognition and Behavior
Addiction and Drugs of Abuse
Homeostatic and Neuroendocrine Systems
Cognitive, Emotional, and Behavioral State
Cognition and Behavior
Neurogenesis and Gliogenesis
Ischemia
Development
Disorders of the Nervous System
Neurodegenerative Disorders and Movement Dis
Neurotoxicity, Inflammation, and Neuroprotec
Disorders of the Nervous System
Synapses, and Glia
Ligand Gated Ion Channels
POWERED BY Google
Map Data: Society for Neuroscience, 2006 - Terms of Use
Transferring data from www.scimaps.org...

Society for Neuroscience, 2006 Visual Browser

Click to start a bounding box, then click again to end it. A listing of all nodes in the bounding box will be shown. Click once again to clear the bounding box.

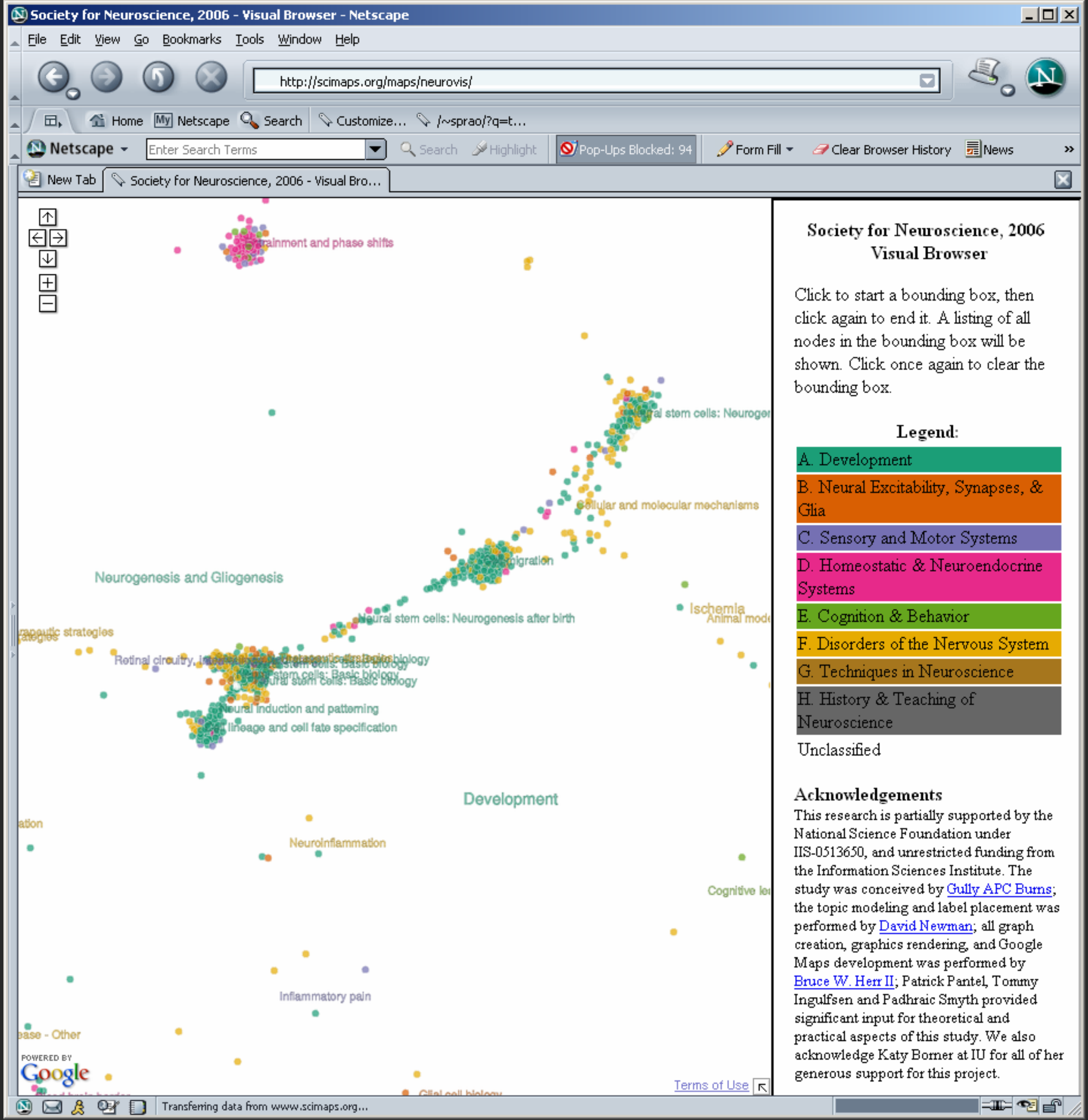
Legend:

- A. Development
- B. Neural Excitability, Synapses, & Glia
- C. Sensory and Motor Systems
- D. Homeostatic & Neuroendocrine Systems
- E. Cognition & Behavior
- F. Disorders of the Nervous System
- G. Techniques in Neuroscience
- H. History & Teaching of Neuroscience
- Unclassified

Acknowledgements

This research is partially supported by the National Science Foundation under IIS-0513650, and unrestricted funding from the Information Sciences Institute. The study was conceived by Gully APC Burns; the topic modeling and label placement was performed by David Newman; all graph creation, graphics rendering, and Google Maps development was performed by Bruce W. Herr II; Patrick Pantel, Tommy Ingulfsen and Padhraic Smyth provided significant input for theoretical and practical aspects of this study. We also acknowledge Katy Borner at IU for all of her generous support for this project.

Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006 Visual Browser, 2007, http://scimaps.org/maps/neurovis/


Society for Neuroscience, 2006 - Visual Browser - Netscape
File Edit View Go Bookmarks Tools Window Help
http://scimaps.org/maps/neurovis/
Home
My Netscape
Search
Customize...
/~sprao/?q=t...
Netscape
Enter Search Terms
Search
Highlight
Pop-Ups Blocked: 94
Form Fill
Clear Browser History
News
New Tab
Society for Neuroscience, 2006 - Visual Bro...
Entrainment and phase shifts
Neural stem cells: Neurogen
Cellular and molecular mechanisms
Migration
Neurogenesis and Gliogenesis
Neural stem cells: Neurogenesis after birth
Ischemia
Animal mode
apeutic strategies
Retinal circuitry
Stem cells: Basic biology
Neural stem cells: Basic biology
Neural induction and patterning
Lineage and cell fate specification
Development
ation
Neuroinflammation
Cognitive le
Inflammatory pain
ease - Other
POWERED BY Google
Glial cell biology
Terms of Use
Transferring data from www.scimaps.org...

Society for Neuroscience, 2006 Visual Browser

Click to start a bounding box, then click again to end it. A listing of all nodes in the bounding box will be shown. Click once again to clear the bounding box.

Legend:

A. Development

B. Neural Excitability, Synapses, & Glia

C. Sensory and Motor Systems

D. Homeostatic & Neuroendocrine Systems

E. Cognition & Behavior

F. Disorders of the Nervous System

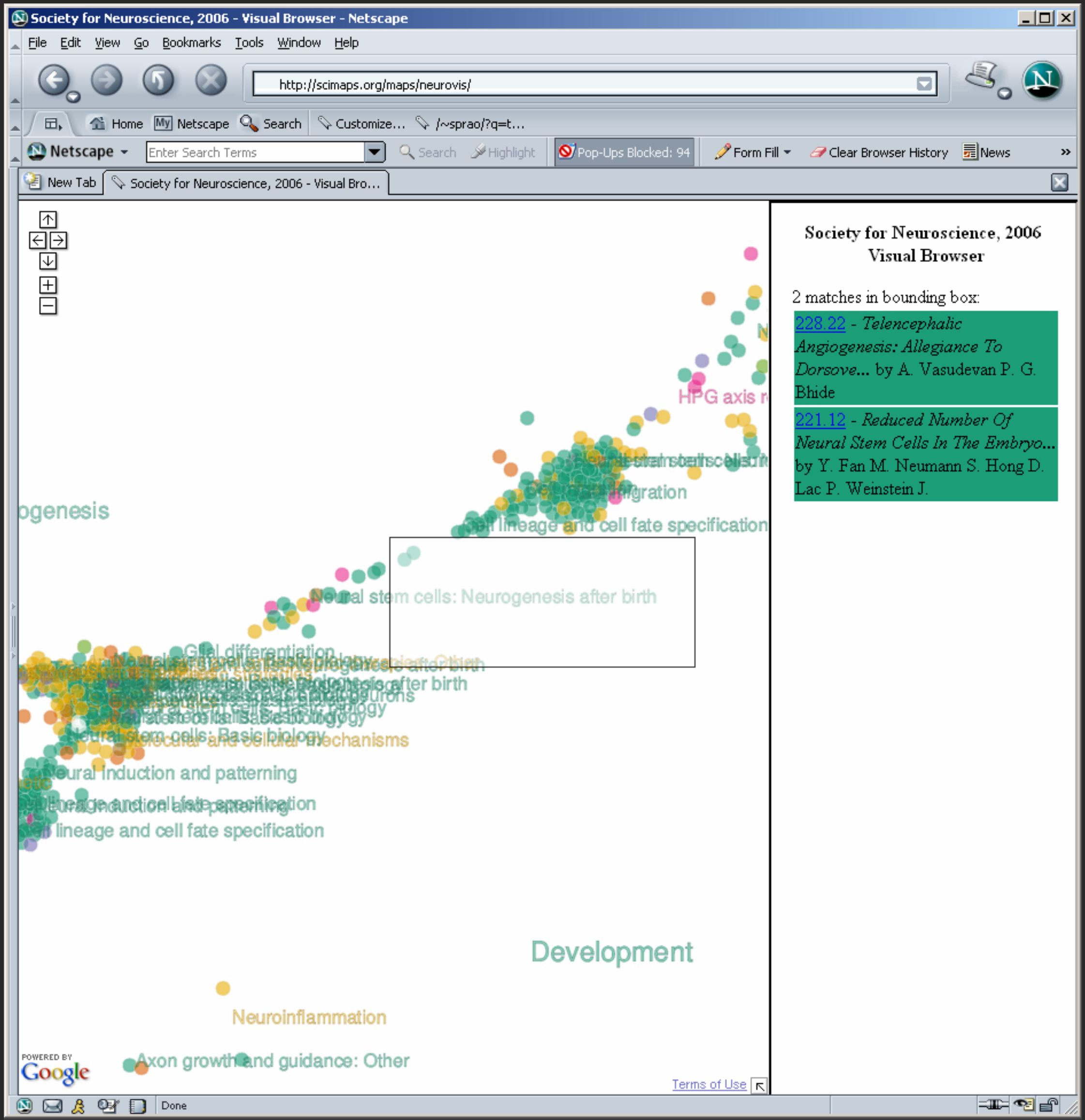
G. Techniques in Neuroscience

H. History & Teaching of Neuroscience

Unclassified

Acknowledgements
This research is partially supported by the National Science Foundation under IIS-0513650, and unrestricted funding from the Information Sciences Institute. The study was conceived by Gully APC Burns; the topic modeling and label placement was performed by David Newman; all graph creation, graphics rendering, and Google Maps development was performed by Bruce W. Herr II; Patrick Pantel, Tommy Ingulfsen and Padhraic Smyth provided significant input for theoretical and practical aspects of this study. We also acknowledge Katy Borner at IU for all of her generous support for this project.

Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006 Visual Browser, 2007, http://scimaps.org/maps/neurovis/


Society for Neuroscience, 2006 - Visual Browser - Netscape
File Edit View Go Bookmarks Tools Window Help
http://scimaps.org/maps/neurovis/
Home
My Netscape
Search
Customize…
/~sprao/?q=t…
Netscape
Enter Search Terms
Search
Highlight
Pop-Ups Blocked: 94
Form Fill
Clear Browser History
News
New Tab
Society for Neuroscience, 2006 - Visual Bro…
HPG axis r
ogenesis
Neuronal migration
Cell lineage and cell fate specification
Neural stem cells: Neurogenesis after birth
Glial differentiation
Neural stem cells: Basic biology
Neural induction and patterning
Cell lineage and cell fate specification
Development
Neuroinflammation
Axon growth and guidance: Other
POWERED BY Google
Terms of Use
Done

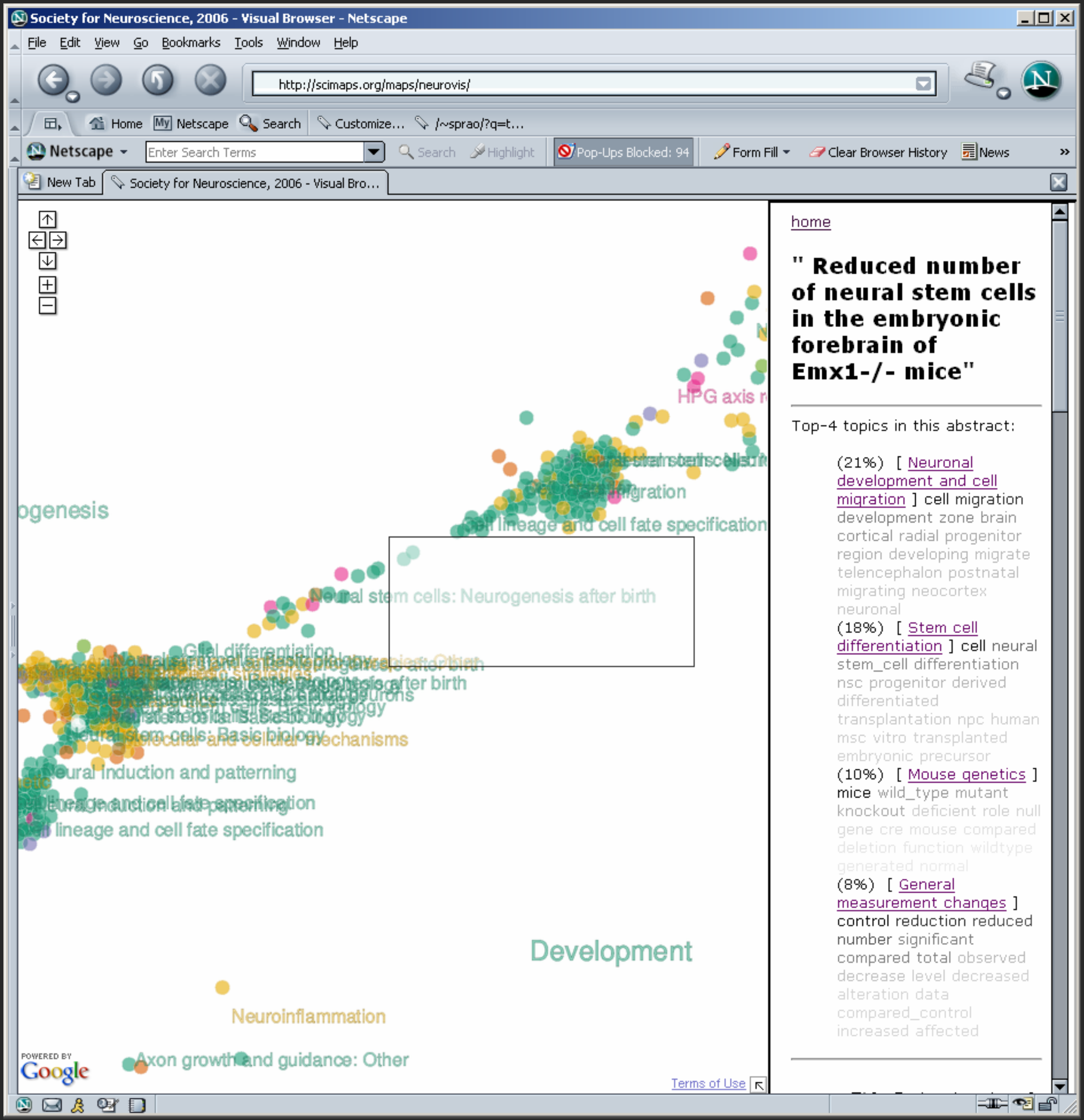
Society for Neuroscience, 2006
Visual Browser

2 matches in bounding box:

228.22 - *Telencephalic Angiogenesis: Allegiance To Dorsove...* by A. Vasudevan P. G. Bhide

221.12 - *Reduced Number Of Neural Stem Cells In The Embryo...* by Y. Fan M. Neumann S. Hong D. Lac P. Weinstein J.

Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006 Visual Browser, 2007, http://scimaps.org/maps/neurovis/


Society for Neuroscience, 2006 - Visual Browser - Netscape
File Edit View Go Bookmarks Tools Window Help
http://scimaps.org/maps/neurovis/
Home
My Netscape
Search
Customize...
/~sprao/?q=t...
Netscape
Enter Search Terms
Search
Highlight
Pop-Ups Blocked: 94
Form Fill
Clear Browser History
News
New Tab
Society for Neuroscience, 2006 - Visual Bro...
HPG axis r
Neuronal migration
Cell lineage and cell fate specification
ogenesis
Neural stem cells: Neurogenesis after birth
Glial differentiation
Neural stem cells: Basic biology
Molecular and cellular mechanisms
Neural induction and patterning
Cell lineage and cell fate specification
Development
Neuroinflammation
Axon growth and guidance: Other
POWERED BY Google
Terms of Use

home

" Reduced number of neural stem cells in the embryonic forebrain of Emx1-/- mice"

Top-4 topics in this abstract:

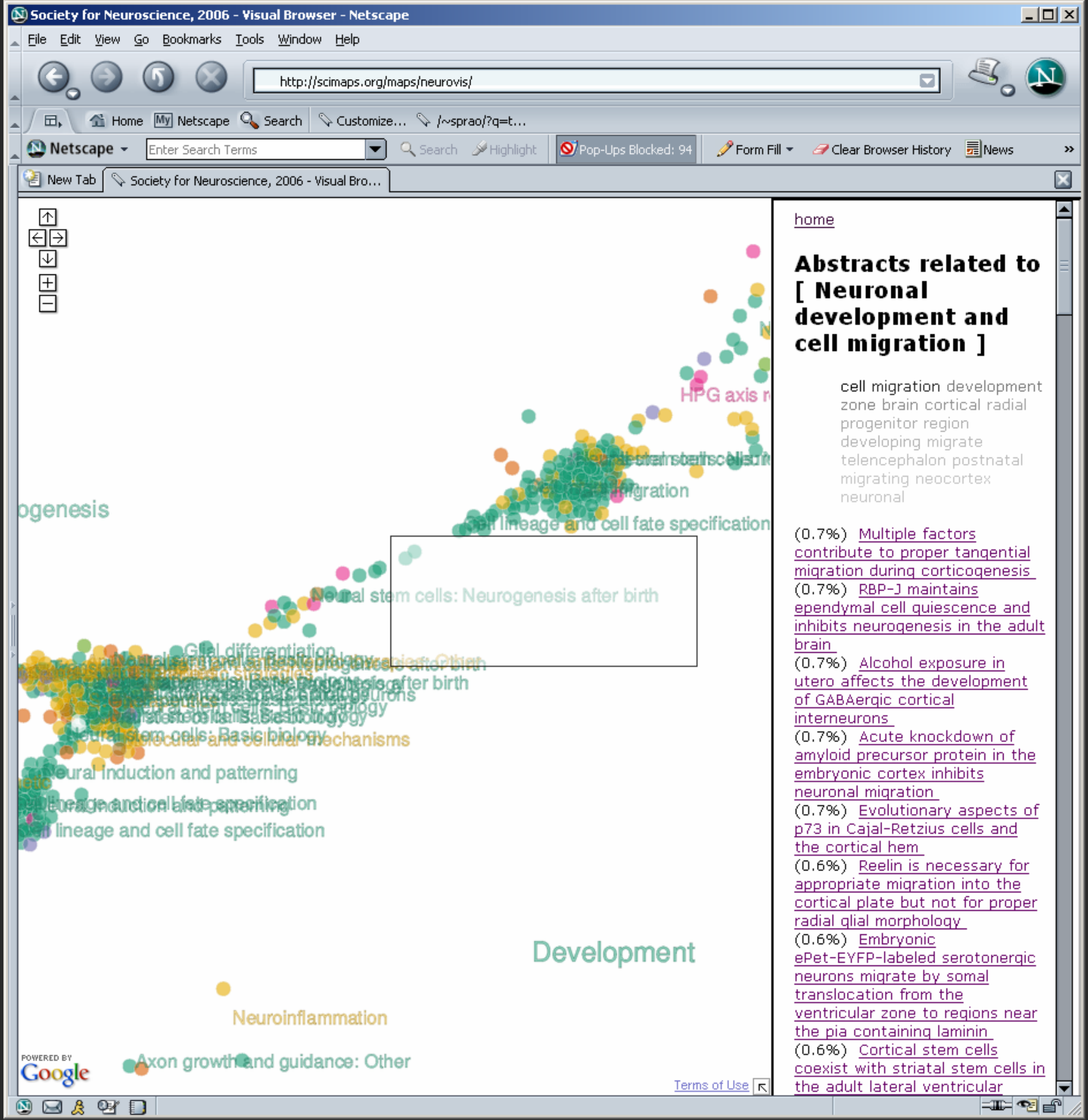
(21%) [Neuronal development and cell migration] cell migration development zone brain cortical radial progenitor region developing migrate telencephalon postnatal migrating neocortex neuronal

(18%) [Stem cell differentiation] cell neural stem_cell differentiation nsc progenitor derived differentiated transplantation npc human msc vitro transplanted embryonic precursor

(10%) [Mouse genetics] mice wild_type mutant knockout deficient role null gene cre mouse compared deletion function wildtype generated normal

(8%) [General measurement changes] control reduction reduced number significant compared total observed decrease level decreased alteration data compared_control increased affected

Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006 Visual Browser, 2007, http://scimaps.org/maps/neurovis/


Society for Neuroscience, 2006 - Visual Browser - Netscape
File Edit View Go Bookmarks Tools Window Help
http://scimaps.org/maps/neurovis/
Home
My Netscape
Search
Customize...
/~sprao/?q=t...
Netscape
Enter Search Terms
Search
Highlight
Pop-Ups Blocked: 94
Form Fill
Clear Browser History
News
New Tab
Society for Neuroscience, 2006 - Visual Bro...
HPG axis
ogenesis
Cell migration
Cell lineage and cell fate specification
Neural stem cells: Neurogenesis after birth
Glial differentiation
Neural stem cells: Basic biology
Cellular and molecular mechanisms
Neural induction and patterning
Cell lineage and cell fate specification
Development
Neuroinflammation
Axon growth and guidance: Other
POWERED BY Google
Terms of Use

home

Abstracts related to [Neuronal development and cell migration]

cell migration development zone brain cortical radial progenitor region developing migrate telencephalon postnatal migrating neocortex neuronal

(0.7%) Multiple factors contribute to proper tangential migration during corticogenesis
(0.7%) RBP-J maintains ependymal cell quiescence and inhibits neurogenesis in the adult brain
(0.7%) Alcohol exposure in utero affects the development of GABAergic cortical interneurons
(0.7%) Acute knockdown of amyloid precursor protein in the embryonic cortex inhibits neuronal migration
(0.7%) Evolutionary aspects of p73 in Cajal-Retzius cells and the cortical hem
(0.6%) Reelin is necessary for appropriate migration into the cortical plate but not for proper radial glial morphology
(0.6%) Embryonic ePet-EYFP-labeled serotonergic neurons migrate by somal translocation from the ventricular zone to regions near the pia containing laminin
(0.6%) Cortical stem cells coexist with striatal stem cells in the adult lateral ventricular

Bruce W. Herr II, Gully Burns (USC), David Newman (UCI), Society for Neuroscience, 2006 Visual Browser, 2007, http://scimaps.org/maps/neurovis/

Wikipedian Activity

Studying large scale social networks such as Wikipedia

Vizzards 2007 Entry

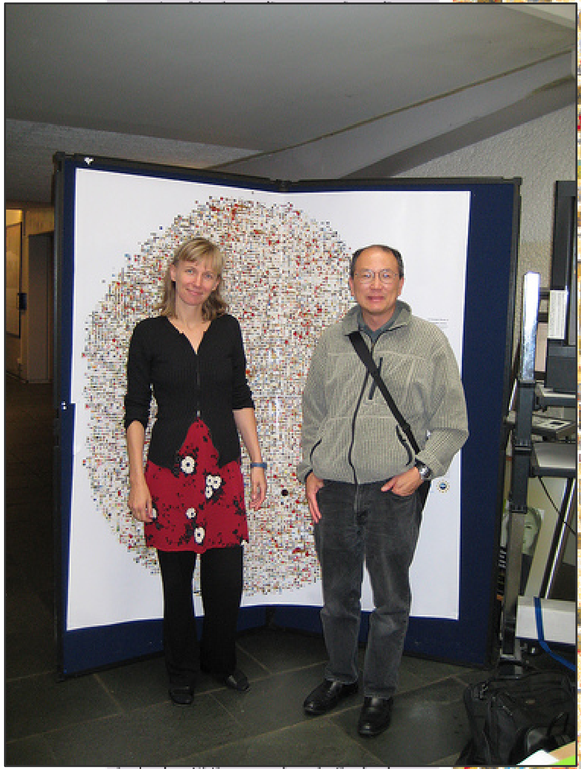
Second Sight: An Emergent Mosaic of Wikipedian Activity, The NewScientist, May 19, 2007

Second sight

Image: Bruce W. Herr and Todd M. Holloway

Power struggle

How do you keep track of the bubbling mass of information that is Wikipedia? This chaotic-looking mosaic is one attempt to show which topics are



locked until the mood cools (locked pages at the time of writing include entries on Sheffield Wednesday football club, Mikhail Gorbachev and pigs).

The mosaic has been commended in a competition for images that visualise network dynamics, coinciding with this week's International Workshop and Conference on Network Science in Bloomington.

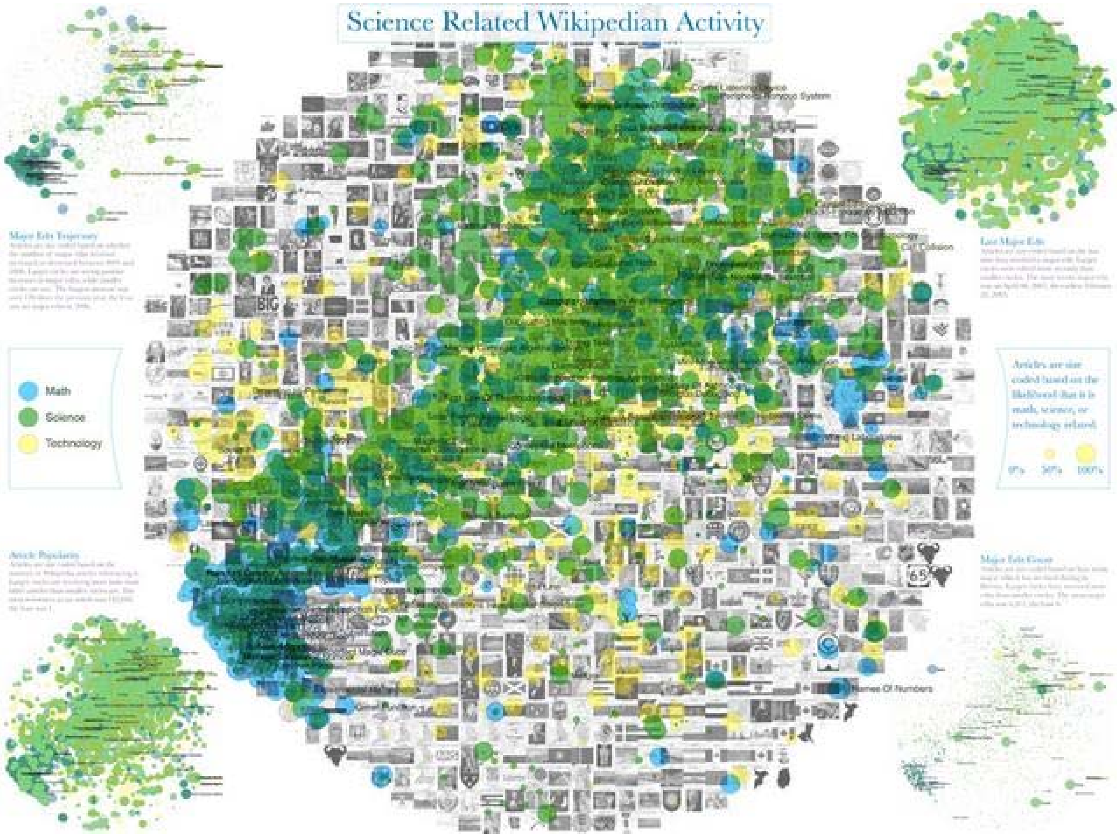
Science Related Wikipedian Activity

http://scimaps.org/dev/map_detail.php?map_id=165

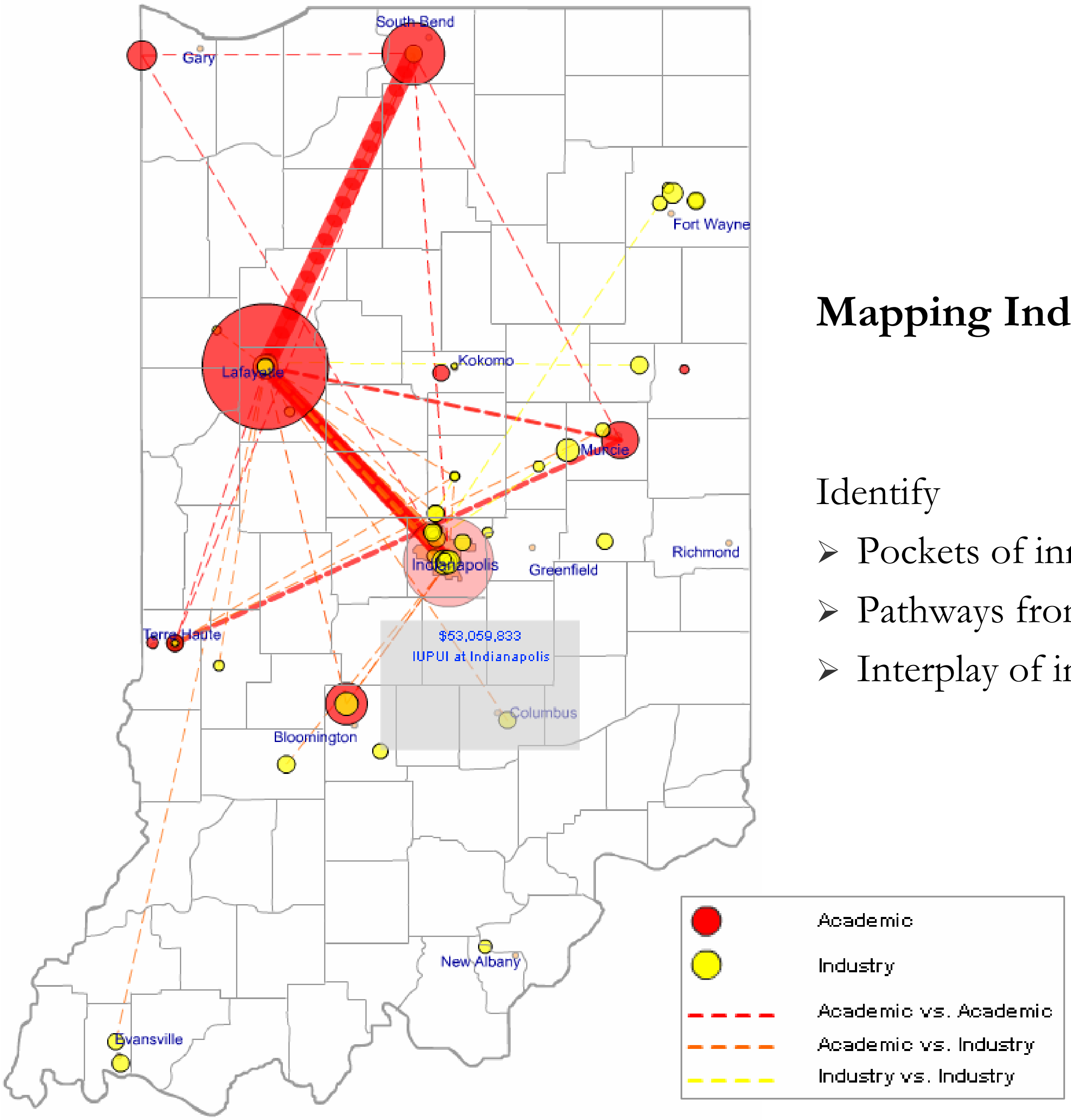
Same base map.

Overlaid are 3,599 math (blue), 6,474 science (green), and 3,164 technology relevant articles (yellow).
All other articles are given in grey.

Corners show articles size coded according to
-article edit activity (top left),
- number of major edits (top right),
- number of bursts in edit activity (bottom, right)
- indegree (bottom left).



Mapping Indiana's Intellectual Space



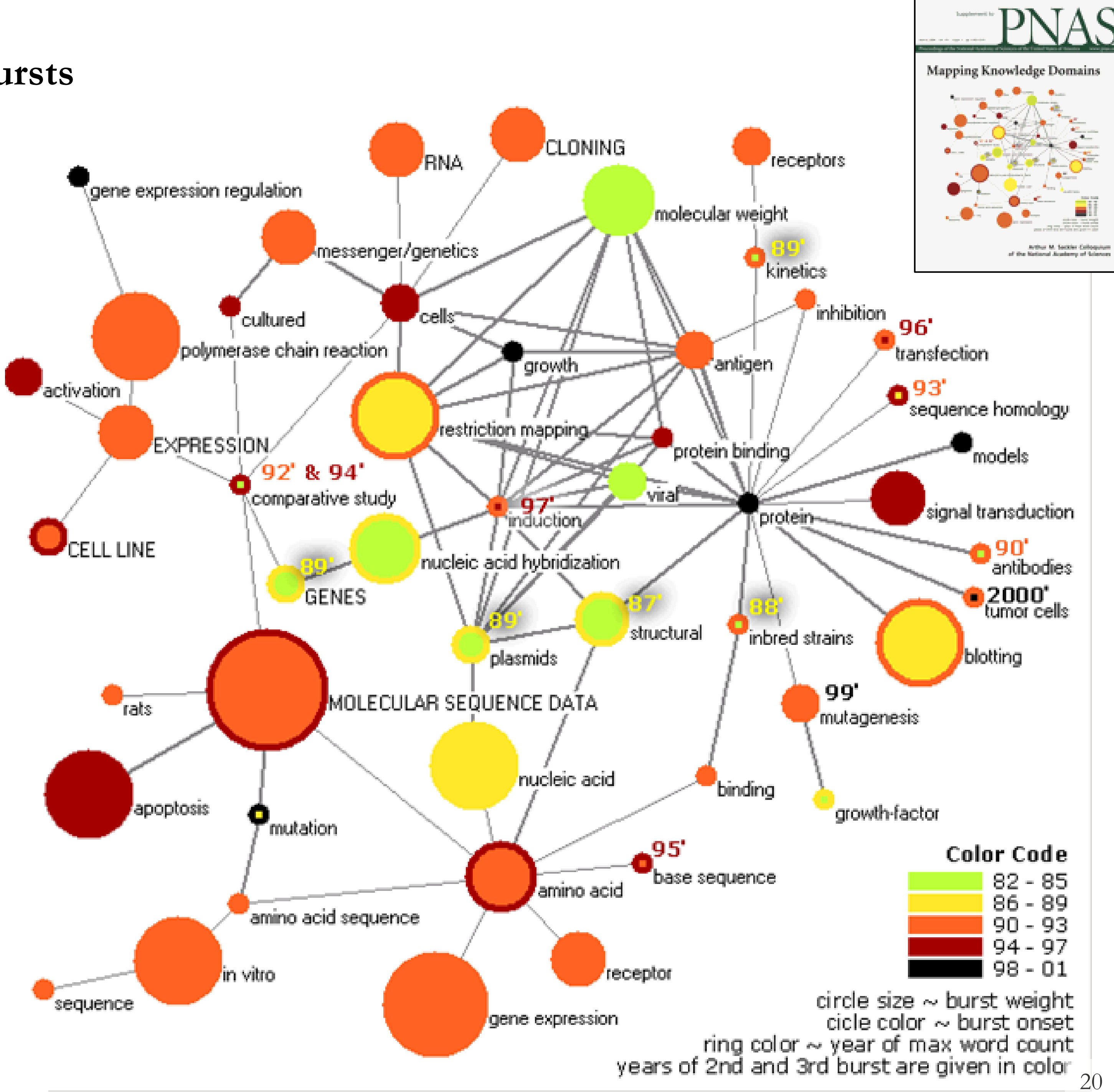
Identify

- Pockets of innovation
- Pathways from ideas to products
- Interplay of industry and academia

Mapping Topic Bursts

Co-word space of the top 50 highly frequent and bursty words used in the top 10% most highly cited PNAS publications in 1982-2001.

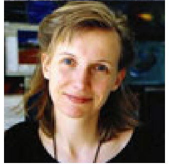
Mane & Börner. (2004) PNAS, 101(Suppl. 1): 5287-5290.

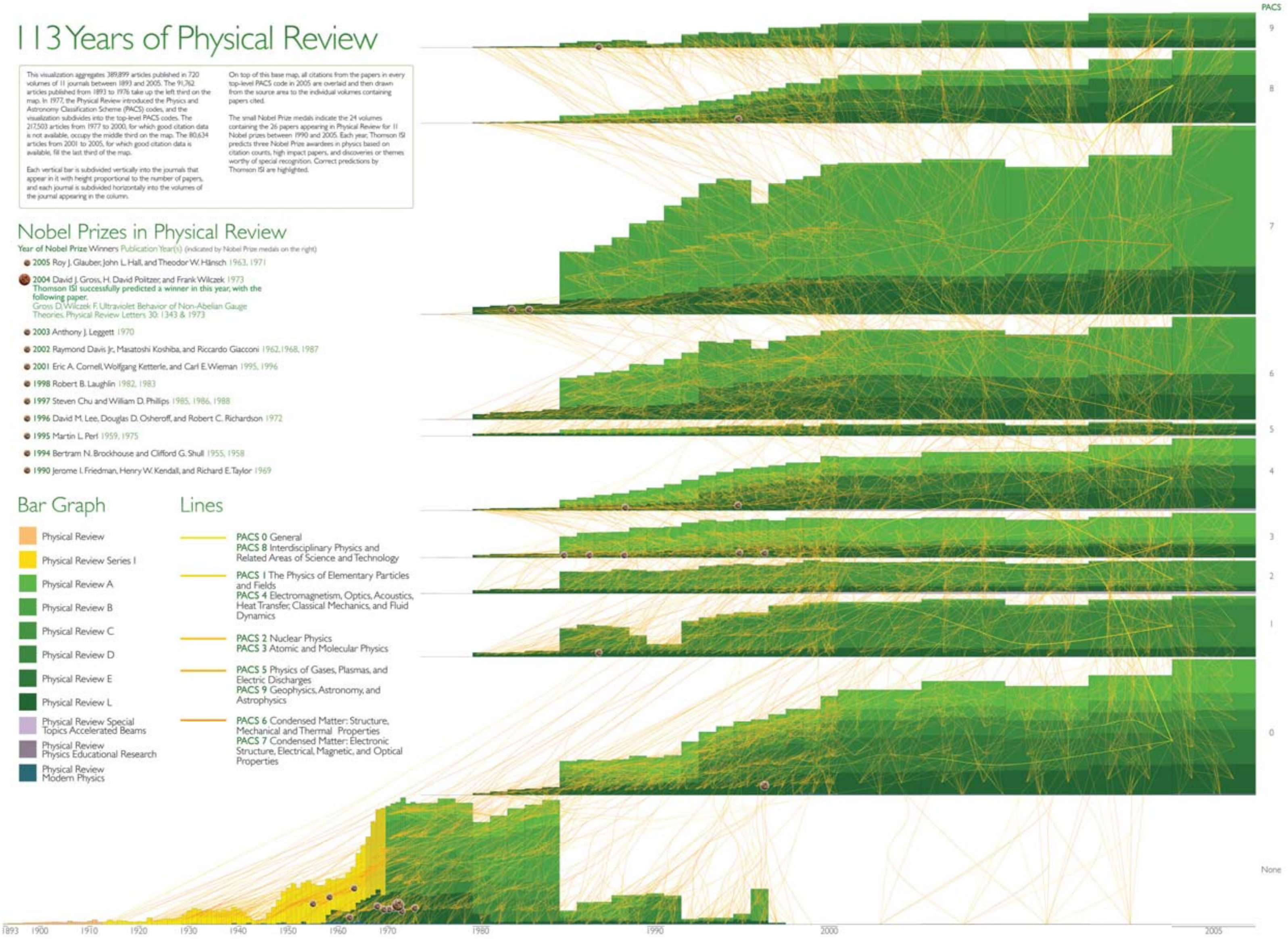


circle size ~ burst weight
cicle color ~ burst onset
ring color ~ year of max word count
years of 2nd and 3rd burst are given in color

113 Years of Physical Review

Bruce W. Herr II and Russell Duhon (Data Mining & Visualization), Elisha F. Hardy (Graphic Design), Shashikant Penumarthy (Data Preparation) and Katy Börner (Concept)





113 Years of Physical Review

This visualization aggregates 389,899 articles published in 720 volumes of 11 journals between 1893 and 2005. The 91,762 articles published from 1893 to 1976 take up the left third on the map. In 1977, the Physical Review introduced the Physics and Astronomy Classification Scheme (PACS) codes, and the visualization subdivides into the top-level PACS codes. The 217,503 articles from 1977 to 2000, for which good citation data is not available, occupy the middle third on the map. The 80,634 articles from 2001 to 2005, for which good citation data is available, fill the last third of the map.

Each vertical bar is subdivided vertically into the journals that appear in it with height proportional to the number of papers, and each journal is subdivided horizontally into the volumes of the journal appearing in the column.

On top of this base map, all citations from the papers in every top-level PACS code in 2005 are overlaid and then drawn from the source area to the individual volumes containing papers cited.

The small Nobel Prize medals indicate the 24 volumes containing the 26 papers appearing in Physical Review for 11 Nobel prizes between 1990 and 2005. Each year, Thomson ISI predicts three Nobel Prize awardees in physics based on citation counts, high impact papers, and discoveries or themes worthy of special recognition. Correct predictions by Thomson ISI are highlighted.

Nobel Prizes in Physical Review

Year of Nobel Prize Winners Publication Year(s) (indicated by Nobel Prize medals on the right)

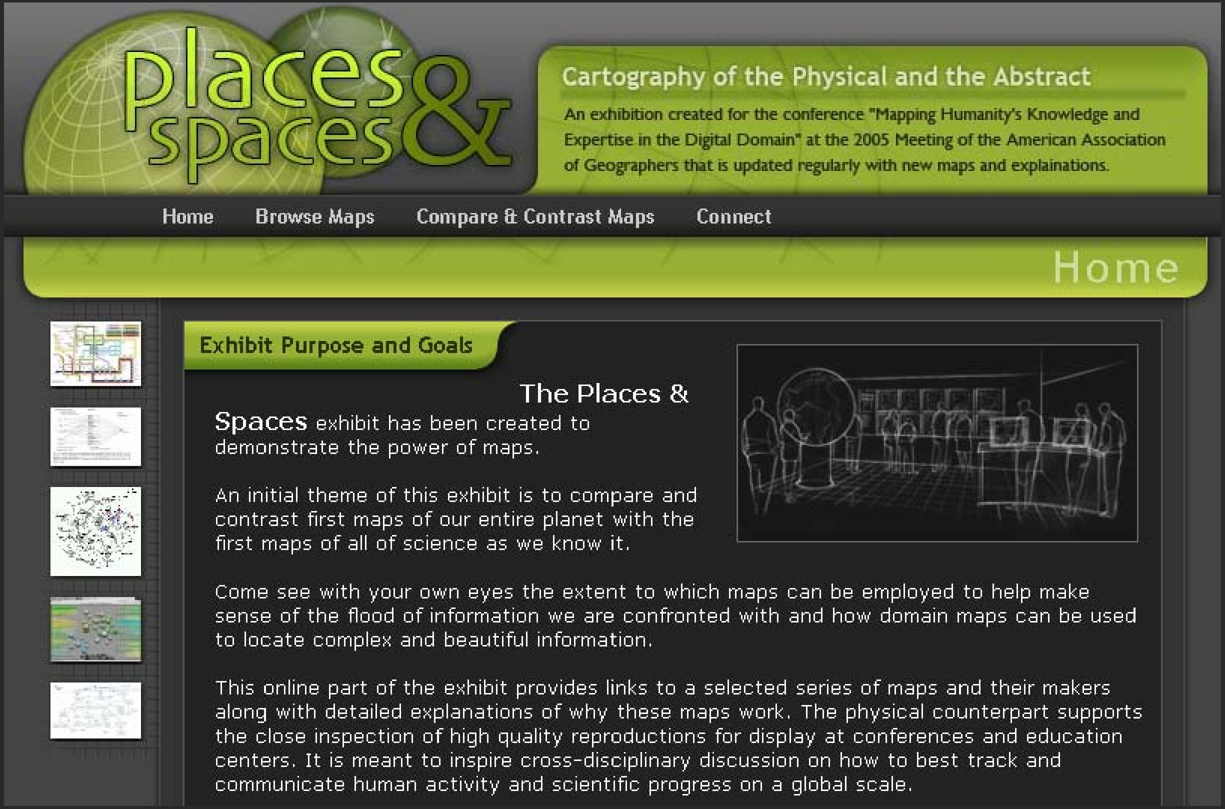
- **2005** Roy J. Glauber, John L. Hall, and Theodor W. Hänsch 1963, 1971
- **2004** David J. Gross, H. David Politzer, and Frank Wilczek 1973
  Thomson ISI successfully predicted a winner in this year, with the following paper.
  Gross D, Wilczek F. Ultraviolet Behavior of Non-Abelian Gauge Theories. Physical Review Letters 30: 1343 & 1973
- **2003** Anthony J. Leggett 1970
- **2002** Raymond Davis Jr., Masatoshi Koshiba, and Riccardo Giacconi 1962, 1968, 1987
- **2001** Eric A. Cornell, Wolfgang Ketterle, and Carl E. Wieman 1995, 1996
- **1998** Robert B. Laughlin 1982, 1983
- **1997** Steven Chu and William D. Phillips 1985, 1986, 1988
- **1996** David M. Lee, Douglas D. Osheroff, and Robert C. Richardson 1972
- **1995** Martin L. Perl 1959, 1975
- **1994** Bertram N. Brockhouse and Clifford G. Shull 1955, 1958
- **1990** Jerome I. Friedman, Henry W. Kendall, and Richard E. Taylor 1969

Bar Graph

- 🟧 Physical Review
- 🟨 Physical Review Series I
- 🟩 Physical Review A
- 🟩 Physical Review B
- 🟩 Physical Review C
- 🟩 Physical Review D
- 🟩 Physical Review E
- 🟩 Physical Review L
- 🟪 Physical Review Special Topics Accelerated Beams
- 🟪 Physical Review Physics Educational Research
- 🟦 Physical Review Modern Physics

Lines

- — **PACS 0** General
  PACS 8 Interdisciplinary Physics and Related Areas of Science and Technology
- — **PACS 1** The Physics of Elementary Particles and Fields
  PACS 4 Electromagnetism, Optics, Acoustics, Heat Transfer, Classical Mechanics, and Fluid Dynamics
- — **PACS 2** Nuclear Physics
  PACS 3 Atomic and Molecular Physics
- — **PACS 5** Physics of Gases, Plasmas, and Electric Discharges
  PACS 9 Geophysics, Astronomy, and Astrophysics
- — **PACS 6** Condensed Matter: Structure, Mechanical and Thermal Properties
  PACS 7 Condensed Matter: Electronic Structure, Electrical, Magnetic, and Optical Properties



Cartography of the Physical and the Abstract

An exhibition created for the conference "Mapping Humanity's Knowledge and Expertise in the Digital Domain" at the 2005 Meeting of the American Association of Geographers that is updated regularly with new maps and explainations.

Exhibit Purpose and Goals

The Places & Spaces exhibit has been created to demonstrate the power of maps.

An initial theme of this exhibit is to compare and contrast first maps of our entire planet with the first maps of all of science as we know it.

Come see with your own eyes the extent to which maps can be employed to help make sense of the flood of information we are confronted with and how domain maps can be used to locate complex and beautiful information.

This online part of the exhibit provides links to a selected series of maps and their makers along with detailed explanations of why these maps work. The physical counterpart supports the close inspection of high quality reproductions for display at conferences and education centers. It is meant to inspire cross-disciplinary discussion on how to best track and communicate human activity and scientific progress on a global scale.

Places & Spaces: Mapping Science

a science exhibit that introduces people to maps of sciences, their makers and users.

http://scimaps.org.

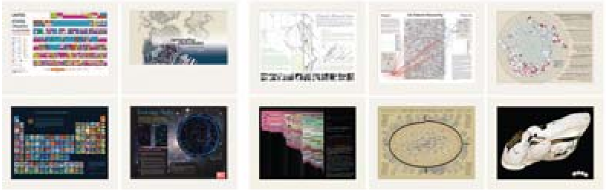
Exhibit Curators: Dr. Katy Börner & Elisha Hardy

Mapping Science Exhibit – 10 Iterations in 10 years

The Power of Maps (2005)

Science Maps for Economic Decision Makers (2008)

The Power of Reference Systems (2006)



Science Maps for Science Policy Makers (2009)
Science Maps for Scholars (2010)
Science Maps as Visual Interfaces to Digital Libraries (2011)
Science Maps for Kids (2012)
Science Forecasts (2013)

How to Lie with Science Maps (2014)

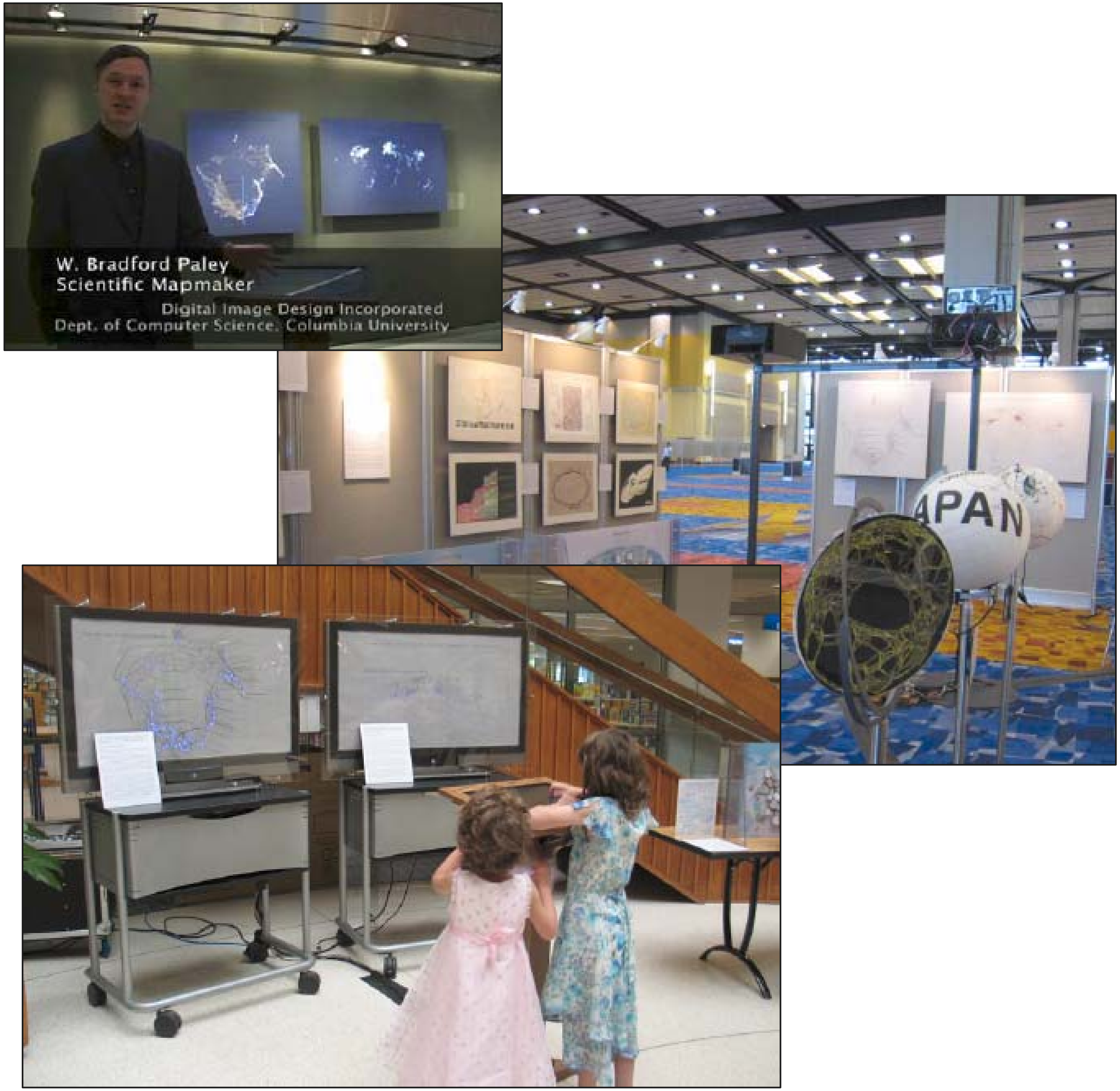
The Power of Forecasts (2007)

scimaps.org

Illuminated Diagram Display

W. Bradford Paley, Kevin W. Boyack, Richard Kalvans, and Katy Börner (2007) Mapping, Illuminating, and Interacting with Science. SIGGRAPH 2007, San Diego, CA.

YouTube Video: http://www.youtube.com/watch?v=bXABcOABG4E


W. Bradford Paley
Scientific Mapmaker
Digital Image Design Incorporated
Dept. of Computer Science, Columbia University
APAN

The TARL Model (Topics, Aging, and Recursive Linking)

Börner, Katy, Maru, Jeegar & Goldstone, Robert. (2004). The Simultaneous Evolution of Author and Paper Networks. PNAS. Vol. 101(Suppl. 1), 5266-5273.

The TARL model incorporates

- A partitioning of authors and papers into topics,
- Aging, i.e., a bias for authors to cite recent papers, and
- A tendency for authors to cite papers cited by papers that they have read resulting in a rich get richer effect.

The model attempts to capture the roles of authors and papers in the production, storage, and dissemination of knowledge.

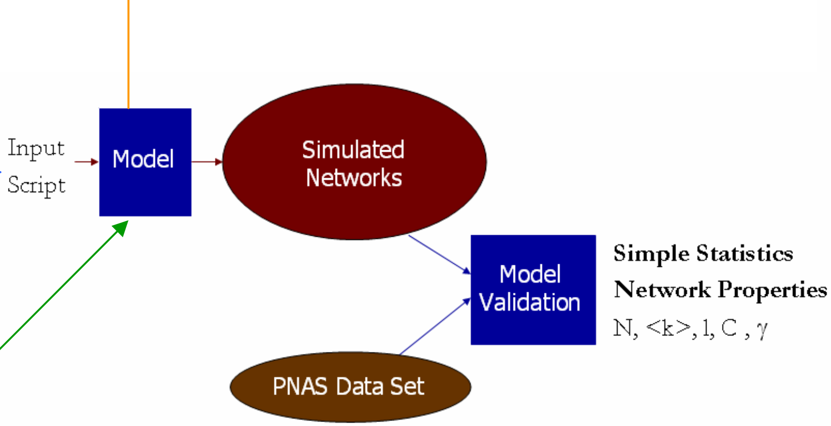
Model Assumptions

- Co-author and paper-citation networks co-evolve.
- Authors come and go.
- Papers are forever.
- Only authors that are 'alive' are able to co-author.
- All existing (but no future) papers can be cited.
- Information diffusion occurs directly via co-authorships and indirectly via the consumption of other authors' papers.

- Preferential attachment is modeled as an emergent property of the elementary, local networking activity of authors reading and citing papers, but also the references listed in papers.

```
---------------------------------------------
Model Parameters (0=without, 1=with)
---------------------------------------------
0/1  Topics
0/1  Co-Authors
0/1  Consider References
0    Aging Function
---------------------------------------------
Model Initialization Values
---------------------------------------------
2  # Years
5  # Authors in Start Year
5  # Papers in Start Year
2  # Papers Consumed (Referenced) per Paper
1  # Papers Produced per Author each Year
5  # Topics
1  # Co-Author(s) per Author
1   # Levels References are Considered
```

```
// Initialization
generate #_papers  papers and assign a random topic to each paper;
generate #_authors  authors and assign a random topic to each author;
randomly assign #_co-authors+1 authors to papers of the same topic;
// Simulation
for each year  do {
      add #_new_authors new authors, deactivate authors older than #_author_age;
      for each topic do {
            randomly partition set of authors into author_groups of size #_co-authors+1;
            for each author_group do {
                  for each new_paper to be produced, do {
                        generate new_paper;
                        randomly select #_read_ papers  from existing papers;
                        get all references of read_ papers  up to #_reference_path_length;
                        for each new_paper_reference do {
                              select a time_slice  from (start year to curr_year-1) with probability given in aging_function;
                              randomly select a paper published or cited in this time_slice; as a new_paper_reference;
                              add the new_paper_reference to new_paper;
                        }
                  }
            }
      }
      add all new papers to the set of existing papers;
      add new links to author and paper information;
}
```

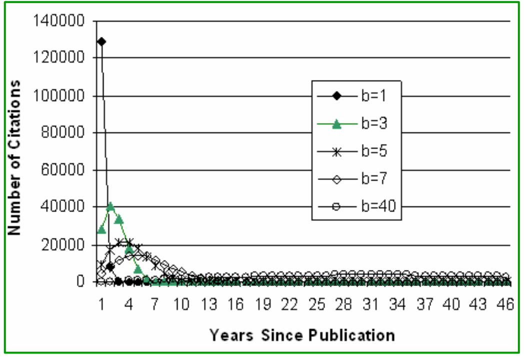


Simple Statistics

Network Properties

N, <k>, l, C , γ

Aging function



Model Validation

The properties of the networks generated by this model are validated against a 20-year data set (1982-2001) of documents of type article published in the Proceedings of the National Academy of Science (PNAS) – about 106,000 unique authors, 472,000 co-author links, 45,120 papers cited within the set, and 114,000 citation references within the set.

Table 3 Statistics for SIM data

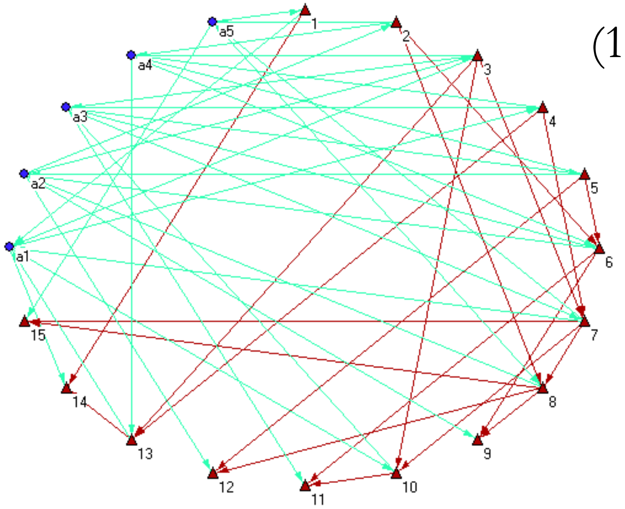
Year	#p	#a	#r	#c	a#ca
1981	1624	3953	0	756	8.21
1982	1040	5200	31200	112161	4
1983	1118	5590	33540	21397	4
1984	1197	5985	35910	10224	4
1985	1275	6375	38250	6184	4
1986	1353	6765	40590	4687	4
1987	1432	7160	42960	3573	4
1988	1510	7550	45300	2816	4
1989	1589	7945	47670	2219	4
1990	1667	8335	50010	1853	4
1991	1745	8725	52350	1634	4
1992	1824	9120	54720	1431	4
1993	1902	9510	57060	1167	4
1994	1981	9905	59430	1040	4
1995	2059	10295	61770	767	4
1996	2137	10685	64110	632	4
1997	2216	11080	66480	522	4
1998	2294	11470	68820	400	4
1999	2373	11865	71190	265	4
2000	2451	12255	73530	125	4
2001	2529	12645	75870	0	4
Total	**37316**		**1070760**	**173853**	

Table 2. PNAS Statistics

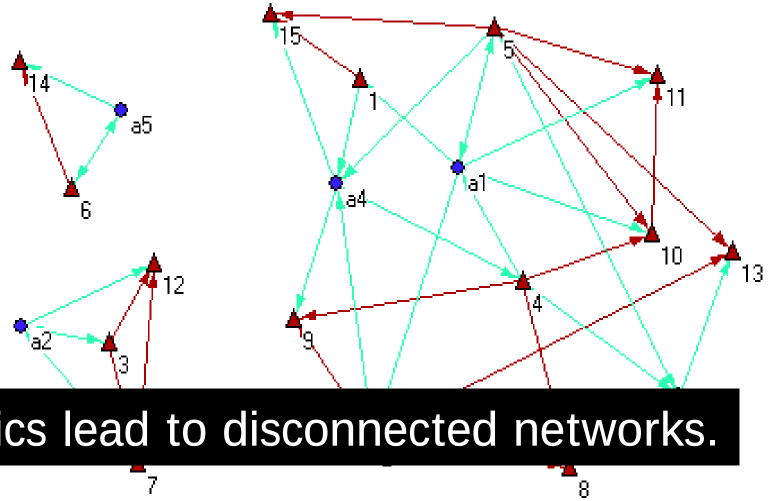
Year	#p	#a	#r	#c	a#ca
1982	1669	5201	46665	156690	3.92
1983	1611	5142	46685	161437	3.98
1984	1695	5583	49834	174161	4.22
1985	1846	6325	55662	191750	4.38
1986	2042	7209	64379	218229	4.76
1987	1924	7061	59110	207729	4.88
1988	2035	7471	63116	215227	4.8
1989	2088	7959	65883	215437	5.01
1990	2066	8031	66019	207138	5.15
1991	2382	9559	77740	223102	5.25
1992	2500	9812	80949	211238	5.29
1993	2413	9770	79848	193867	5.55
1994	2600	10656	86176	187353	5.56
1995	2476	10429	82021	151249	5.66
1996	2765	11803	99061	148622	5.96
1997	2618	11255	96788	122908	6.12
1998	2711	12328	100973	107764	6.48
1999	2603	12182	97018	76080	6.69
2000	2501	12201	94181	44131	7.6
2001	2575	13038	97450	16357	8.4
Total	**45120**		**1509558**	**3230469**	

The TARL Model: The Effect of Parameters

(0000)

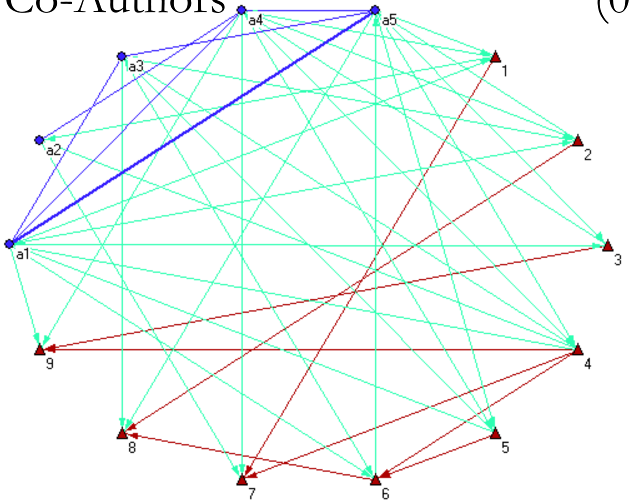


(1000) Topics

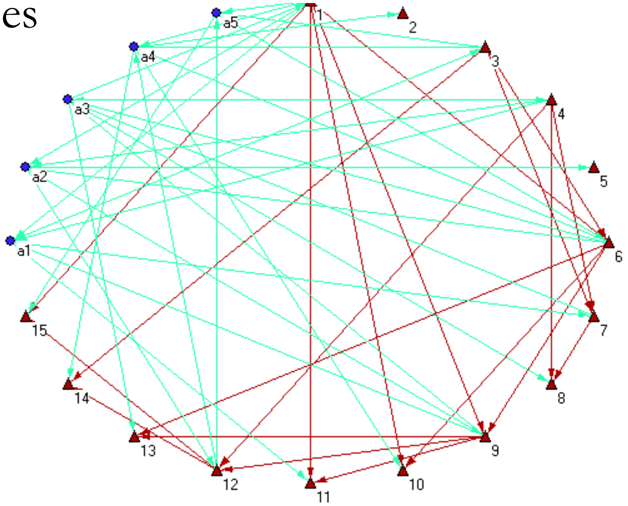


Topics lead to disconnected networks.

(0100) Co-Authors



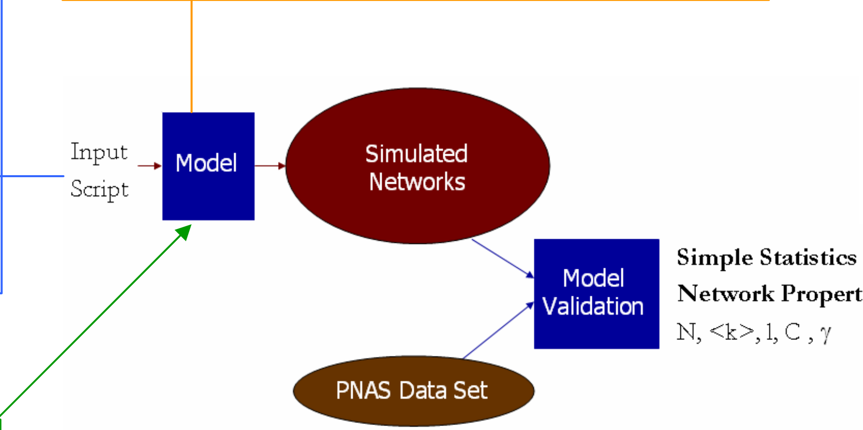
(0010) References



Co-authoring leads to fewer papers.

```
--------------------------------------------
Model Parameters (0=without, 1=with)
--------------------------------------------
0/1  Topics
0/1  Co-Authors
0/1  Consider References
0    Aging Function
--------------------------------------------
Model Initialization Values
--------------------------------------------
2  # Years
5  # Authors in Start Year
5  # Papers in Start Year
2  # Papers Consumed (Referenced) per Paper
1  # Papers Produced per Author each Year
5  # Topics
1  # Co-Author(s) per Author
1   # Levels References are Considered
```

```
// Initialization
generate #_papers papers and assign a random topic to each paper;
generate #_authors authors and assign a random topic to each author;
randomly assign #_co-authors+1 authors to papers of the same topic;
// Simulation
for each year do {
    add #_new_authors new authors, deactivate authors older than #_author_age;
    for each topic do {
        randomly partition set of authors into author_groups of size #_co-authors+1;
        for each author_group do {
            for each new_paper to be produced, do {
                generate new_paper;
                randomly select #_read_ papers from existing papers;
                get all references of read_ papers up to #_reference_path_length;
                for each new_paper_reference do {
                    select a time_slice from (start year to curr_year-1) with probability given in aging_function;
                    randomly select a paper published or cited in this time_slice; as a new_paper_reference;
                    add the new_paper_reference to new_paper;
                }
            }
        }
    }
    add all new papers to the set of existing papers;
    add new links to author and paper information;
}
```

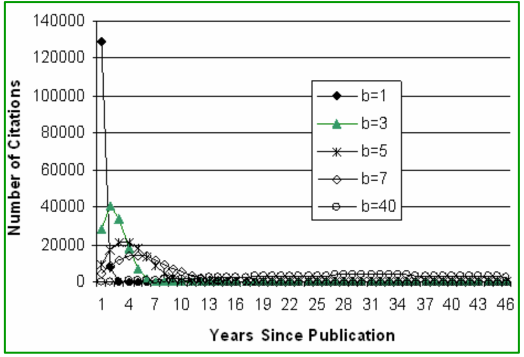


Simple Statistics

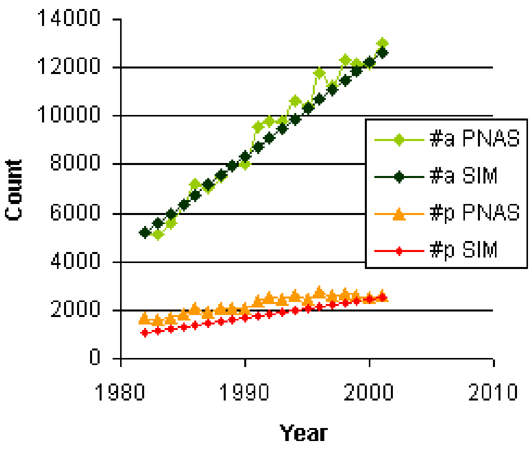
Network Properties

N, <k>, l, C , γ

Aging function



Counts for Papers and Authors



Counts for Citations

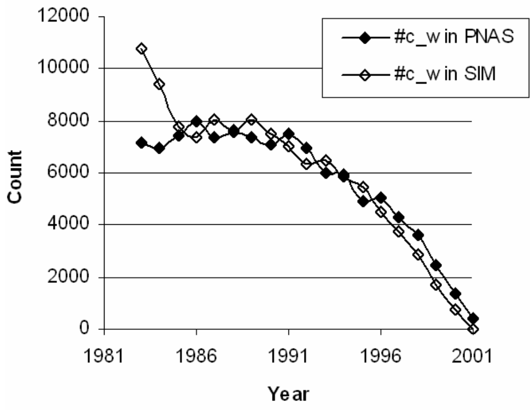
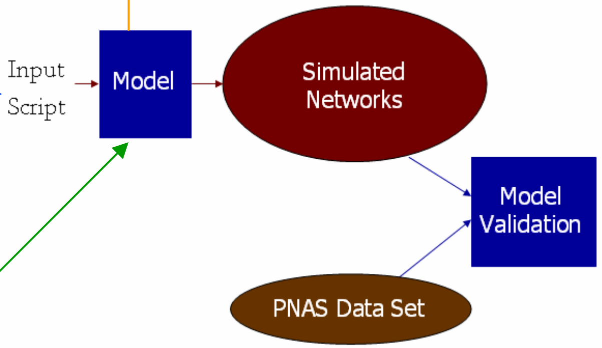


Table 2. Properties of co-author & paper citation networks comprising number of nodes n, average node degree <k>, path length l, cluster coefficient C, and power law exponent γ. Source references are given in the left column.

Network	n	<k>	l	C	γ	Reference
Co-authorship networks						
LANL	52,909	9.7	5.9	0.43	--	Newman, (2001a; 2001b; 2001c)
MEDLINE	1,520,251	18.1	4.6	0.066	--	
SPIRES	56,627	1.73	4.0	0.726	1.2	
NCSTRL	11,994	3.59	9.7	0.496	--	
Math.	70,975	3.9	9.5	0.59	2.5	Barabasi et al., (2002)
Neurosci.	209,293	11.5	6	0.76	2.1	
PNAS	105,915	8.97	5.89	0.399	2.54	
Paper-citation networks						
ISI	783,339	8.57	--	--	3	Redner, (1998)
PhysRev	24,296	14.5	--	--	3	
PNAS	45,120	3.53	--	0.081	2.29	
SIM	37,114	2.13	--	0.074	2.05	

```
// Initialization
generate #_papers  papers and assign a random topic
generate #_authors  authors and assign a random top
randomly assign #_co-authors+1 authors to papers o
// Simulation
for each year  do {
    add #_new_authors new authors, deactivate aut
    for each topic do {
        randomly partition set of authors into autho
        for each author_group do {
            for each new_paper to be produced,
                generate new_paper;
                randomly select #_read_ paper
                get all references of read_ pap
                for each new_paper_reference
                    select a time_slice  from (
                    randomly select a paper p
                    add the new_paper_refere
                }
            }
        }
    }
    add all new papers to the set of existing papers;
    add new links to author and paper information;
}
```

```
-----------------------------------------
Model Parameters (0=without, 1=with)
-----------------------------------------
0/1  Topics
0/1  Co-Authors
0/1  Consider References
0    Aging Function
-----------------------------------------
Model Initialization Values
-----------------------------------------
2  # Years
5  # Authors in Start Year
5  # Papers in Start Year
2  # Papers Consumed (Referenced) per Paper
1  # Papers Produced per Author each Year
5  # Topics
1  # Co-Author(s) per Author
1   # Levels References are Considered
```



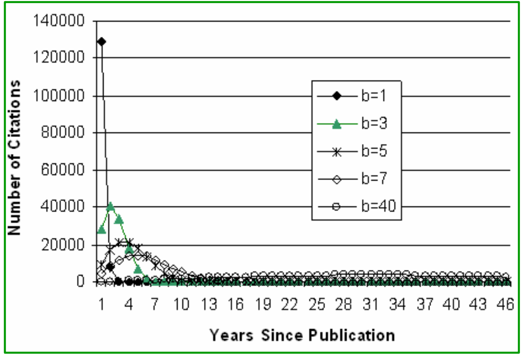
Co-Author and Paper-Citation Network Properties

Simple Statistics

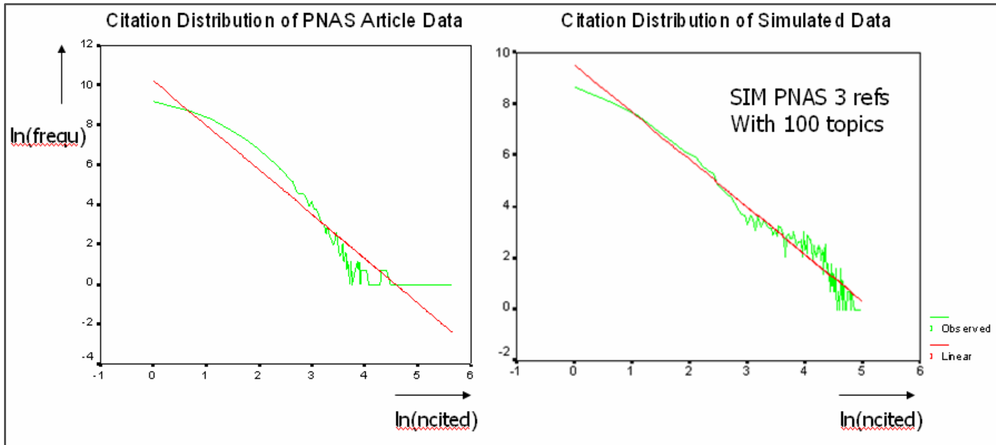
Network Properties

N, <k>, l, C , γ

Aging function



Power Law Distributions

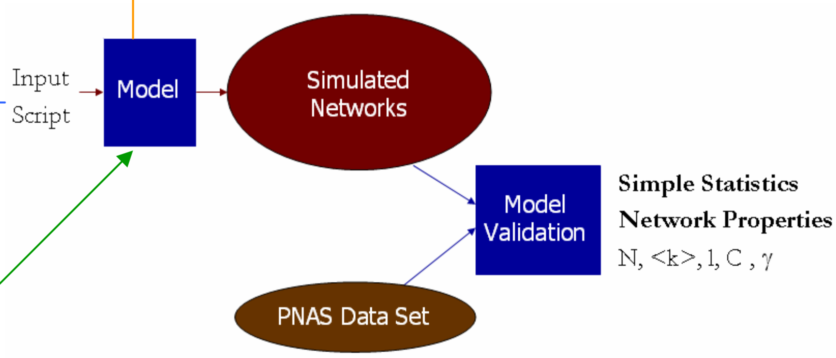


```
Rsq   d.f.       F  Sigf      b0       b1
.877    70  497.88  .000 10.2251 -2.2978
```

```
Rsq    d.f.       F  Sigf      b0      b1
.842   114  1572.51  .000  9.5196 -2.054
```

```
-------------------------------------------
Model Parameters (0=without, 1=with)
-------------------------------------------
0/1  Topics
0/1  Co-Authors
0/1  Consider References
0    Aging Function
-------------------------------------------
Model Initialization Values
-------------------------------------------
2  # Years
5  # Authors in Start Year
5  # Papers in Start Year
2  # Papers Consumed (Referenced) per Paper
1  # Papers Produced per Author each Year
5  # Topics
1  # Co-Author(s) per Author
1   # Levels References are Considered
```

```
// Initialization
generate #_papers papers and assign a random topic to each paper;
generate #_authors authors and assign a random topic to each author;
randomly assign #_co-authors+1 authors to papers of the same topic;
// Simulation
for each year do {
    add #_new_authors new authors, deactivate authors older than #_author_age;
    for each topic do {
        randomly partition set of authors into author_groups of size #_co-authors+1;
        for each author_group do {
            for each new_paper to be produced, do {
                generate new_paper;
                randomly select #_read_ papers from existing papers;
                get all references of read_ papers up to #_reference_path_length;
                for each new_paper_reference do {
                    select a time_slice from (start year to curr_year-1) with probability given in aging_function;
                    randomly select a paper published or cited in this time_slice; as a new_paper_reference;
                    add the new_paper_reference to new_paper;
                }
            }
        }
    }
    add all new papers to the set of existing papers;
    add new links to author and paper information;
}
```

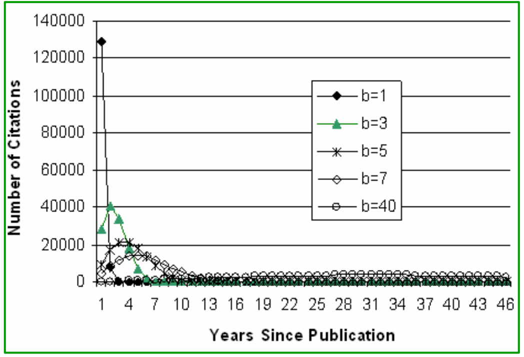


Simple Statistics

Network Properties

N, <k>, l, C , γ

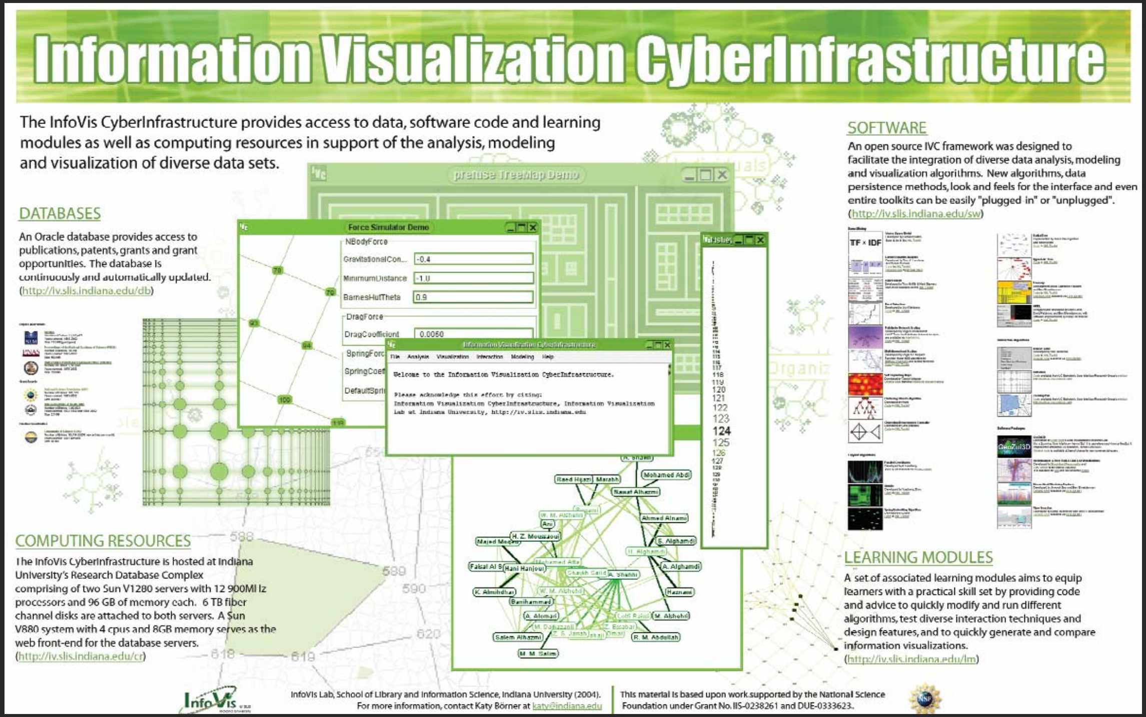
Aging function



Topics: The number of topics is linearly correlated with the clustering coefficient of the resulting network: C= 0.000073 * #topics. Increasing the number of topics increases the power law exponent as authors are now restricted to cite papers in their own topics area.

Aging: With increasing b, and hence increasing the number of older papers cited as references, the clustering coefficient decreases. Papers are not only clustered by topic, but also in time, and as a community becomes increasingly nearsighted in terms of their citation practices, the degree of temporal clustering increases.

References/Recursive Linking: The length of the chain of paper citation links that is followed to select references for a new paper also influences the clustering coefficient. Temporal clustering is ameliorated by the practice of citing (and hopefully reading!) the papers that were the earlier inspirations for read papers.



Information Visualization CyberInfrastructure

The InfoVis CyberInfrastructure provides access to data, software code and learning modules as well as computing resources in support of the analysis, modeling and visualization of diverse data sets.

DATABASES

An Oracle database provides access to publications, patents, grants and grant opportunities. The database is continuously and automatically updated. (http://iv.slis.indiana.edu/db)

SOFTWARE

An open source IVC framework was designed to facilitate the integration of diverse data analysis, modeling and visualization algorithms. New algorithms, data persistence methods, look and feels for the interface and even entire toolkits can be easily "plugged in" or "unplugged". (http://iv.slis.indiana.edu/sw)

COMPUTING RESOURCES

The InfoVis CyberInfrastructure is hosted at Indiana University's Research Database Complex comprising of two Sun V1280 servers with 12 900MHz processors and 96 GB of memory each. 6 TB fiber channel disks are attached to both servers. A Sun V880 system with 4 cpus and 8GB memory serves as the web front-end for the database servers. (http://iv.slis.indiana.edu/cr)

LEARNING MODULES

A set of associated learning modules aims to equip learners with a practical skill set by providing code and advice to quickly modify and run different algorithms, test diverse interaction techniques and design features, and to quickly generate and compare information visualizations. (http://iv.slis.indiana.edu/lm)

Welcome to the Information Visualization CyberInfrastructure.

Please acknowledge this effort by citing:
Information Visualization CyberInfrastructure, Information Visualization Lab at Indiana University, http://iv.slis.indiana.edu

InfoVis Lab, School of Library and Information Science, Indiana University (2004).
For more information, contact Katy Börner at katy@indiana.edu

This material is based upon work supported by the National Science Foundation under Grant No. IIS-0238261 and DUE-0333623.

CAREER: Visualizing Knowledge Domains. NSF IIS-0238261 award (Katy Börner, $451,000) Sept. 03-Aug. 08.
http://iv.slis.indiana.edu/

CIShell
cyberinfrastructure shell

SDB | SCHOLARLY DATABASE

Scholarly Database
http://sdb.slis.indiana.edu

NetworkWorkbench
A Workbench for Network Scientists

NSF SBC Sun microsystems
HPNAP High Performance Network Applications Program
INDIANA 21st Century Research & Technology Fund

SEI: Network Workbench: A Large-Scale Network Analysis, Modeling and Visualization Toolkit for Biomedical, Social Science and Physics Research. NSF IIS-0513650 award (Katy Börner, Albert-Laszlo Barabasi, Santiago Schnell, Alessandro Vespignani & Stanley Wasserman, Eric Wernert (Senior Personnel), $1,120,926) Sept. 05 - Aug. 08. http://nwb.slis.indiana.edu

The End.