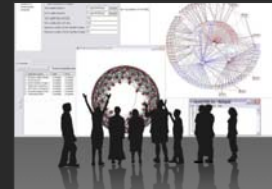
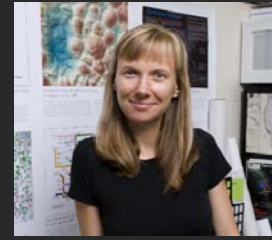


Mapping the Structure and Dynamics of Science

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



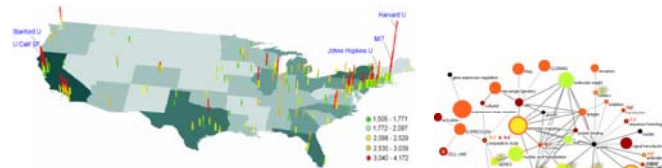
*School of Information, North Campus, Atkins Room #1202, 1075 Beal Avenue
 Video broadcast available at Central Campus, Ebrlicher Room #411
 University of Michigan, Ann Arbor, MI*

Nov 19, 2008



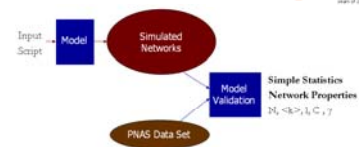
Overview

Computational Scientometrics



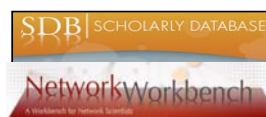
Case Studies:

- Information Diffusion Among Major U.S. Research Institutions
- Identifying Research Topics and Trends
- Modeling the Co-Evolving Author-Paper Networks



Science of Science Cyberinfrastructures

- Scholarly Database
- Network Workbench Tool
- Mapping Science Exhibit

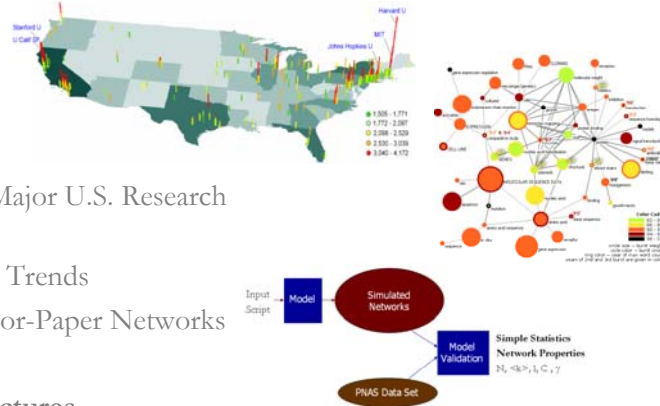


Overview

Computational Scientometrics

Case Studies:

- Information Diffusion Among Major U.S. Research Institutions
- Identifying Research Topics and Trends
- Modeling the Co-Evolving Author-Paper Networks



Science of Science Cyberinfrastructures

- Scholarly Database
- Network Workbench Tool
- Mapping Science Exhibit



Computational Scientometrics: Studying Science by Scientific Means

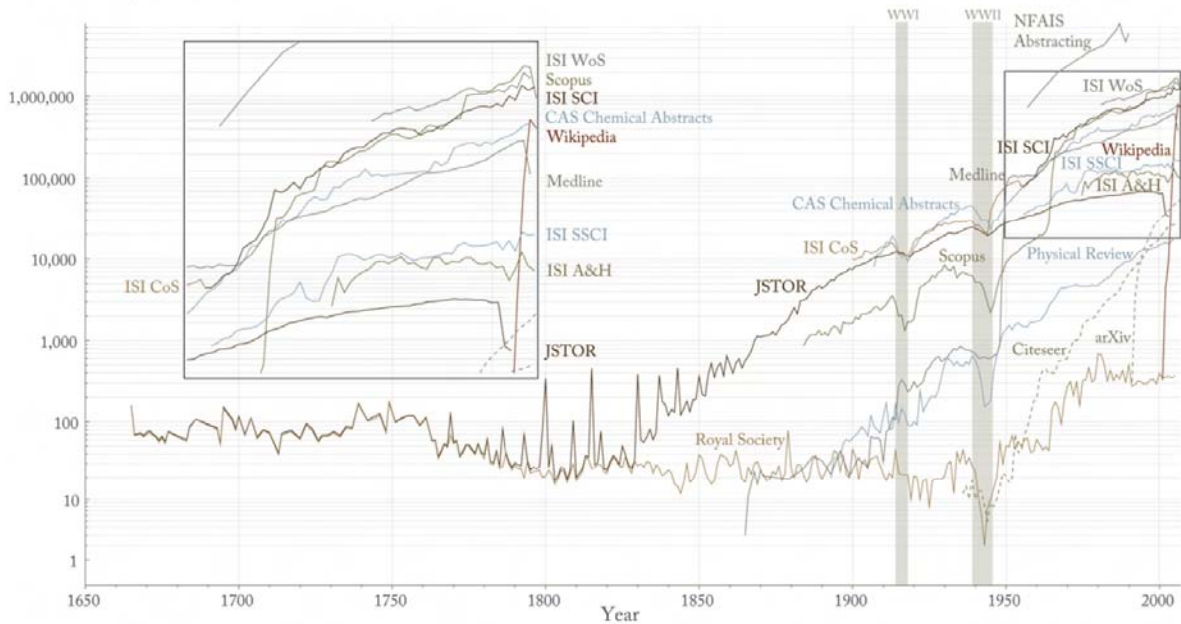
Results are frequently communicated via ‘Science Maps’.



- 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>.

Growth of Scientific Knowledge, 1665 to 2006

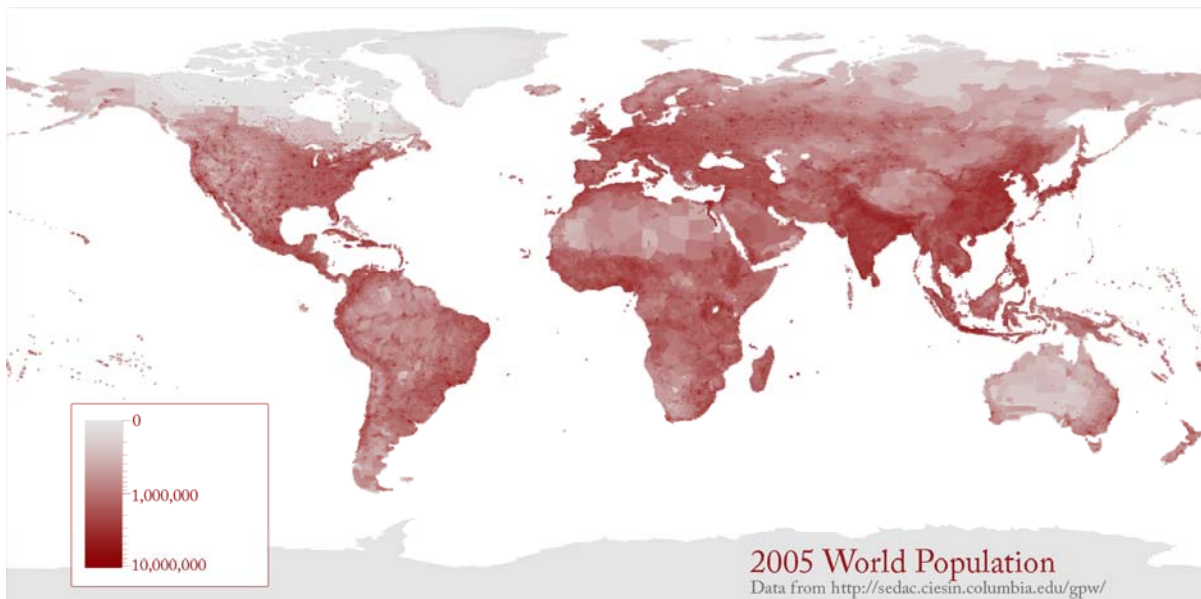
Papers & Wikipedia Entries



Atlas of Science Guiding the Navigation and Management of Scholarly Knowledge. Part I: The Rise of Science and Technology. Chart showing the number of papers/Wikipedia entries for different databases and publication years. Contact Katy Börner <katy@indiana.edu> or Elisha Hardy <ehardy@indiana.edu> for details.

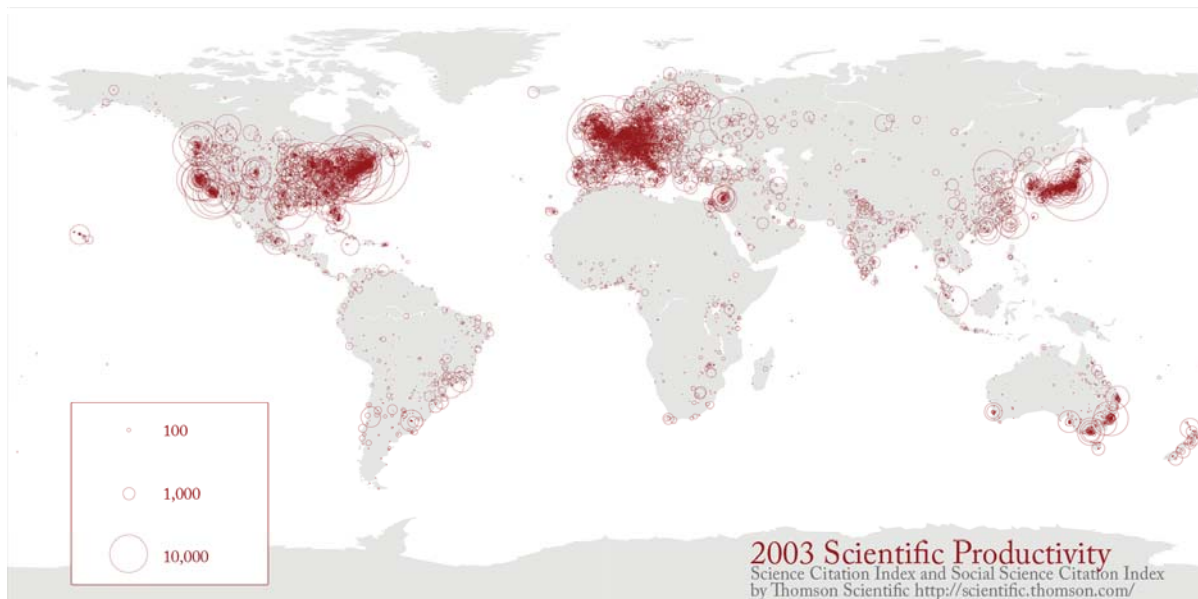
2005 World Population

The population map uses a quarter degree box resolution. Boxes with zero people are given in white. Darker shades of red indicate higher population counts per box using a logarithmic interpolation. The highest density boxes appear in Mumbai, with 11,687,850 people in the quarter degree block, Calcutta (10,816,010), and Shanghai (8,628,088).



2003 Scientific Productivity

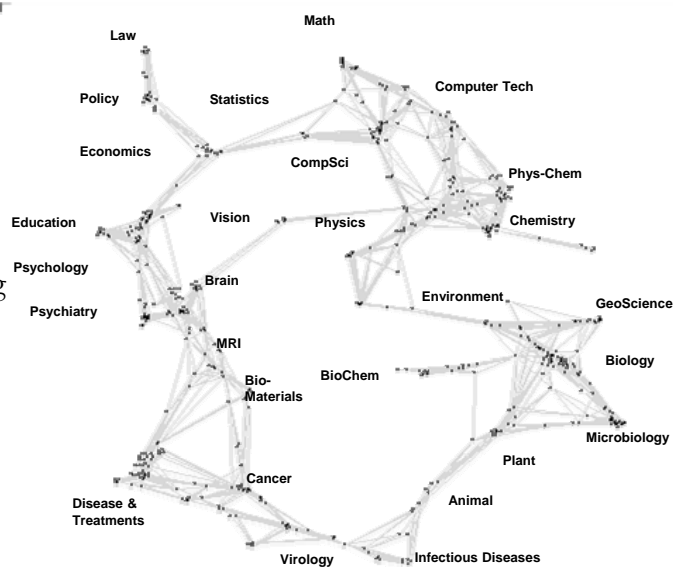
Shown is where science is performed today. Each circle indicates a geographic location at which scholarly papers are published. The larger the circle the more papers are produced. Boston, MA, London, England, and New York, NY are the top three paper production areas. Note the strong resemblance with the Night on Earth and the IP Ownership maps and the striking differences to the world population map.



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.

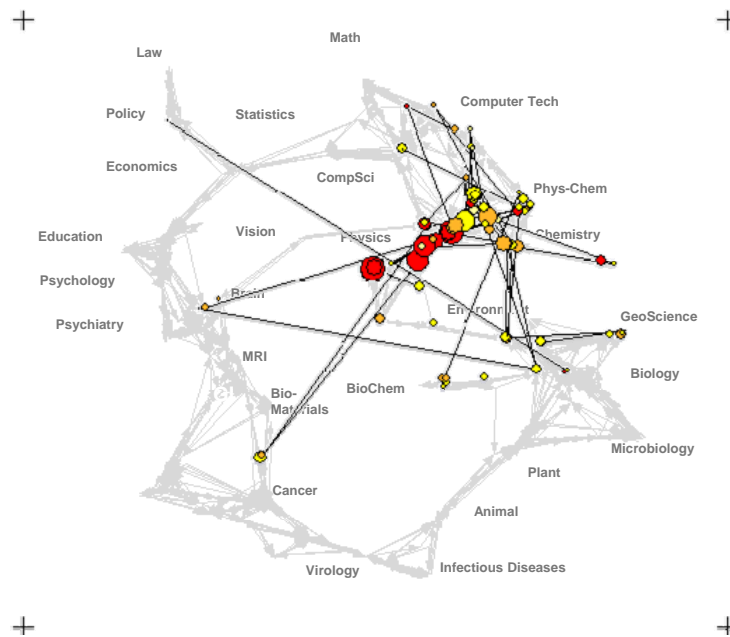
- Uses combined SCI/SSCI from 2002
 - 1.07M papers, 24.5M references, 7,300 journals
 - Bibliographic coupling of papers, aggregated to journals
- Initial ordination and clustering of journals gave 671 clusters
- Coupling counts were reaggregated at the journal cluster level to calculate the
 - (x,y)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

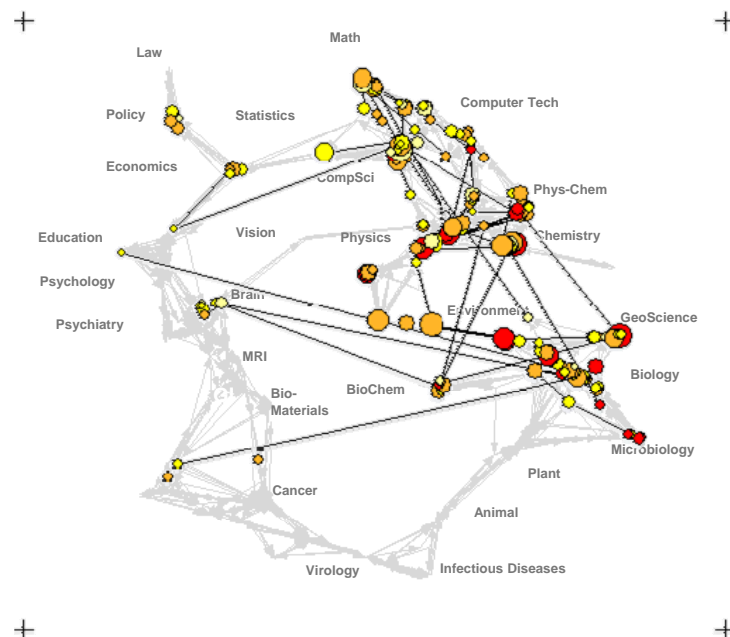
Funding patterns of the US Department of Energy (DOE)



Science map applications: Identifying core competency

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

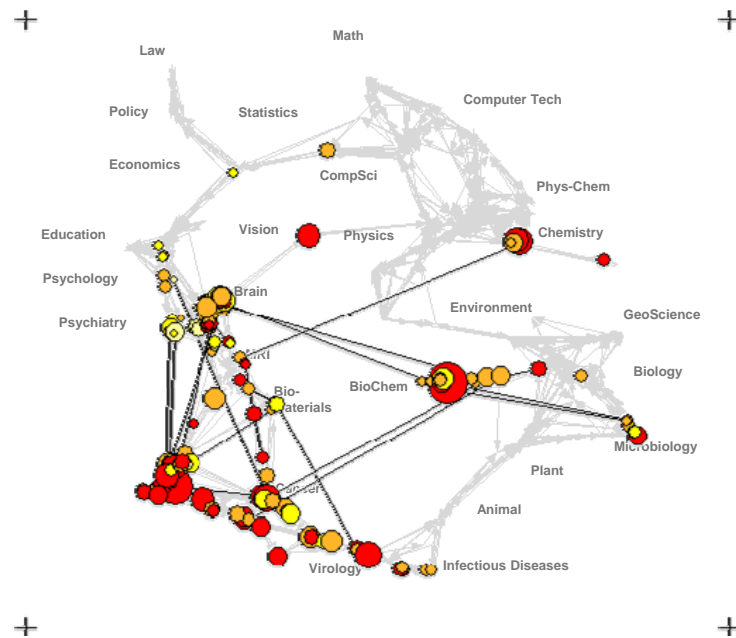
Funding Patterns of the National Science Foundation (NSF)



Science map applications: Identifying core competency

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

Funding Patterns of the National Institutes of Health (NIH)



Opportunities

Advantages for Funding Agencies

- Supports monitoring of (long-term) money flow and research developments, evaluation of funding strategies for different programs, decisions on project durations, funding patterns.
- Staff resources can be used for scientific program development, to identify areas for future development, and the stimulation of new research areas.

Advantages for Researchers

- Easy access to research results, relevant funding programs and their success rates, potential collaborators, competitors, related projects/publications (**research push**).
- More time for research and teaching.

Advantages for Industry

- Fast and easy access to major results, experts, etc.
- Can influence the direction of research by entering information on needed technologies (**industry-pull**).

Advantages for Publishers

- Unique interface to their data.
- Publicly funded development of databases and their interlinkage.

For Society

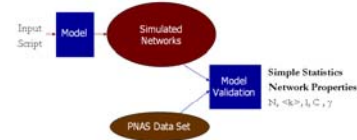
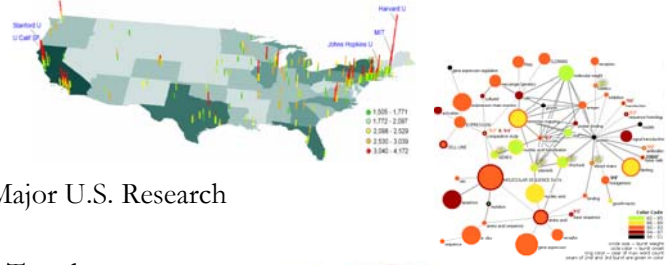
- Dramatically improved access to scientific knowledge and expertise.

Overview

Computational Scientometrics

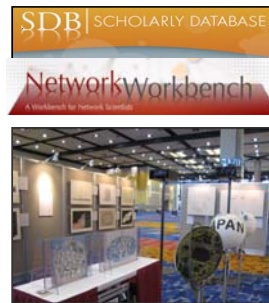
Case Studies:

- Information Diffusion Among Major U.S. Research Institutions
- Identifying Research Topics and Trends
- Modeling the Co-Evolving Author-Paper Networks



Science of Science Cyberinfrastructures

- Scholarly Database
- Network Workbench Tool
- Mapping Science Exhibit



Information Diffusion Among Major U.S. Research Institutions

Börner, Katy, Penumathy, Shashikant, Meiss, Mark & Ke, Weimao. (2006). Mapping the Diffusion of Information among Major U.S. Research Institutions. *Scientometrics*. Vol. 68(3), 415 - 426.

Questions:

1. Does space still matter in the Internet age, i.e., does one still have to study and work at major research institutions in order to have access to high quality data and expertise and to produce high quality research?
2. Does the Internet lead to more global citation patterns, i.e., more citation links between papers produced at geographically distant research institutions?



Contributions:

- Answer to Q1 is YES.
- Answer to Q2 is NO.
- Novel approach to analyzing the dual role of institutions as information producers and consumers and to study and visualize the diffusion of information among them.

GROUP THEORY

What makes a successful team? John Whitfield looks at research that uses massive online databases and network analysis to come up with some rules of thumb for productive collaborations.

By through my recent issue of *Nature*, including this one, and the story is there in black and white almost all original research papers have multiple authors on the title page. In fact, more than 90% of papers published in the past few years, out of a total of some 700,000 papers, are co-authored. The most cited paper in any year was most likely to have been written by a single author than a team, but this pattern reversed in the 1970s. The most cited paper in any year was most likely to have been written by a single author than a team, but this pattern reversed in the 1970s. The most cited paper in any year was most likely to have been written by a single author than a team, but this pattern reversed in the 1970s.

Research in science has always swapped ideas and criticism, but when fields were small, authorship was not such an important mark of achievement. As research expanded by word of mouth, and then through journals and conference proceedings, it became harder for researchers to work in isolation. So people needed to get their ideas and their names on paper, as well as to work with others. This led to the rise of the modern research team.

Research is now done in teams, and this has led to a new way of thinking about science. The results in a clearer picture of science's increasingly collaborative nature, and of the factors that determine a team's success. Funding agencies are not using this work to decide where the money goes — yet. But the researchers behind the work are willing to give tentative tips on what their work reveals. They also think that their work points to ideas of theirs that apply very broadly, whether their papers look like a piece of paper or a piece of software. The first question a reviewer might ask him or herself is: should I collaborate with...

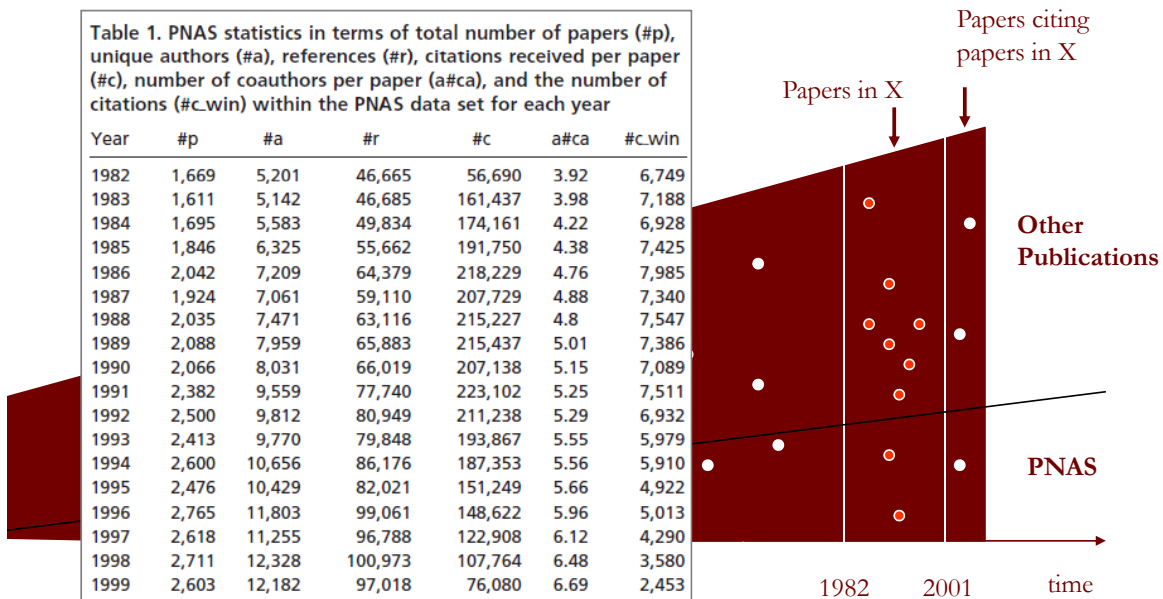
20-Year PNAS Dataset (1982-2001)

45,120 regular articles written by 105,915 unique authors.

114,000 citation references within the set and 472,000 co-author links.

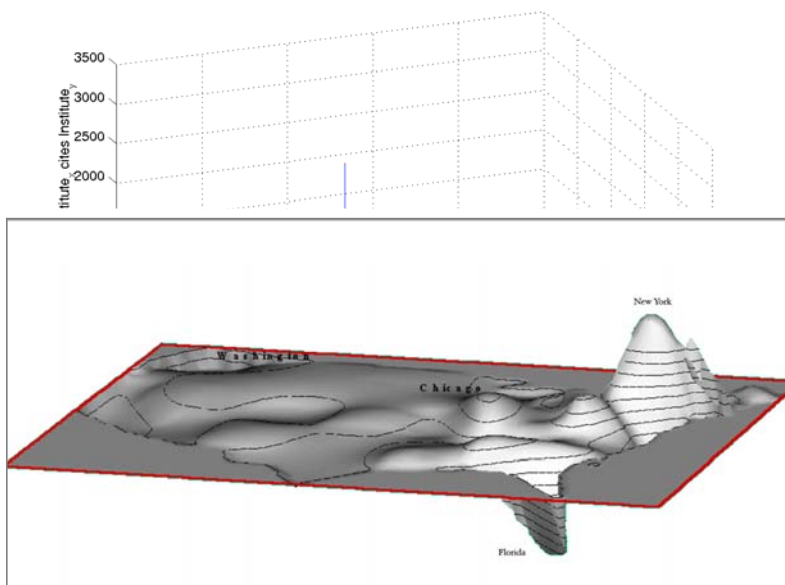
Table 1. PNAS statistics in terms of total number of papers (#p), unique authors (#a), references (#r), citations received per paper (#c), number of coauthors per paper (a#ca), and the number of citations (#c.win) within the PNAS data set for each year

Year	#p	#a	#r	#c	a#ca	#c.win
1982	1,669	5,201	46,665	56,690	3.92	6,749
1983	1,611	5,142	46,685	161,437	3.98	7,188
1984	1,695	5,583	49,834	174,161	4.22	6,928
1985	1,846	6,325	55,662	191,750	4.38	7,425
1986	2,042	7,209	64,379	218,229	4.76	7,985
1987	1,924	7,061	59,110	207,729	4.88	7,340
1988	2,035	7,471	63,116	215,227	4.8	7,547
1989	2,088	7,959	65,883	215,437	5.01	7,386
1990	2,066	8,031	66,019	207,138	5.15	7,089
1991	2,382	9,559	77,740	223,102	5.25	7,511
1992	2,500	9,812	80,949	211,238	5.29	6,932
1993	2,413	9,770	79,848	193,867	5.55	5,979
1994	2,600	10,656	86,176	187,353	5.56	5,910
1995	2,476	10,429	82,021	151,249	5.66	4,922
1996	2,765	11,803	99,061	148,622	5.96	5,013
1997	2,618	11,255	96,788	122,908	6.12	4,290
1998	2,711	12,328	100,973	107,764	6.48	3,580
1999	2,603	12,182	97,018	76,080	6.69	2,453
2000	2,501	12,201	94,181	44,131	7.6	1,354
2001	2,575	13,038	97,450	16,357	8.4	422
Total	45,120		1,509,558	3,230,469		114,003



Citation Matrix

Unsymmetrical direct citation linkage patterns among the top 500 institutions in US. High peak values in the diagonal reflect the high amount of self-citations for all institutions. Medium peak horizontal and vertical lines denote references from and citations to papers written at Harvard University.



Information Sources (Export) and Sinks (Import)

Calculate ratio of the number of references made by an institution divided by the sum of received citations and references made, multiplied by 100.

131 have a value between 0-40% acting mostly as information producers = information sources.

71 have a value between 60-100% and act mostly as information consumers – they reference a large number of papers but the number of citations they receive is comparably low = information sinks.

(Tobler, 1995)

Geographic Location of Received Citations

ESRI's ArcGIS program was used to show the geographic distribution of the top 500 institutions using the Albers equal area projection.

U.S. states are color coded based on the population size in the year 2000. Lighter shades of green represent lower populations.

Overlaid are the top 500 institutions, each represented by a 'citation stick'. The color and height of the stick corresponds to the number of received citations (excluding self citations).

Five institutions produced papers that attracted more than 5,000 citations and are labeled.

Harvard leads with 16,531 citations.



Information Flow Among the Top-5 Consumers and Their Top-10 Producers

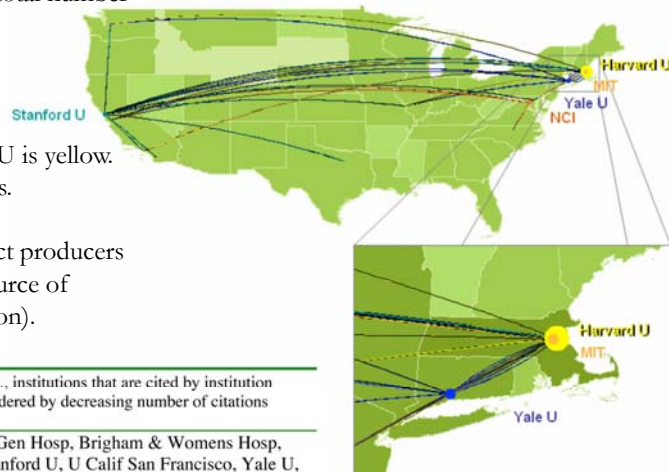
U.S. states are color coded based on the total number of citations received by their institutions (excluding self citations).

Dots indicate the five producers.

Each has a different color, e.g., Harvard U is yellow.

Dot area size depicts number of citations.

Lines represent citations that interconnect producers and consumers shaded from colored (source of information) to white (sink of information).

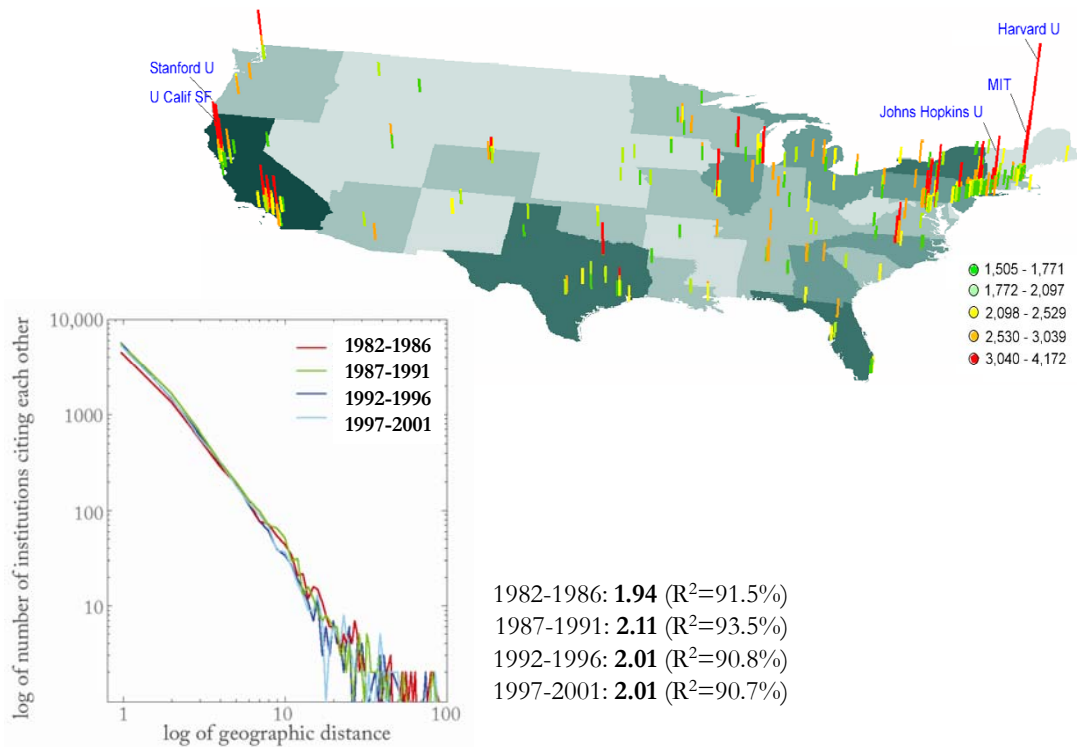


Consumers, i.e., citing institutions	# citations made	Top ten producers, i.e., institutions that are cited by institution listed in first column ordered by decreasing number of citations received.
Harvard U	13,552	MIT, Massachusetts Gen Hosp, Brigham & Womens Hosp, Johns Hopkins U, Stanford U, U Calif San Francisco, Yale U, Rockefeller U, U Washington, Washington U
U Calif SF	4,682	Harvard U, MIT, Stanford U, Johns Hopkins U, U Washington, Washington U, U Calif Berkeley, U Texas, U Calif SD, U Calif LA
MIT	4,655	Harvard U, Whitehead Inst Biomed Res, Johns Hopkins U, Stanford U, U Calif SF, Yale U, Rockefeller U, U Calif LA, Massachusetts Gen Hosp, U Calif Berkeley
NCI (zip: 20814)	4,519	Harvard U, NCI (zip: 20205), NCI (zip: 21701), MIT, Duke U, Johns Hopkins U, NIAID NICHHD, Stanford U, U Calif SF
Yale U	4,464	Harvard U, MIT, Stanford U, Rockefeller U, Johns Hopkins U, Washington U, U Calif SF, U Washington, NCI, Massachusetts Gen Hosp

Paper also shows top-5 producers and their top-10 consumers.

Changes in Citation Behavior Over Time

As time progresses and the amount of produced papers increases, space seems to matter more. Authors are more likely to cite papers generated by authors at close-by institutions.



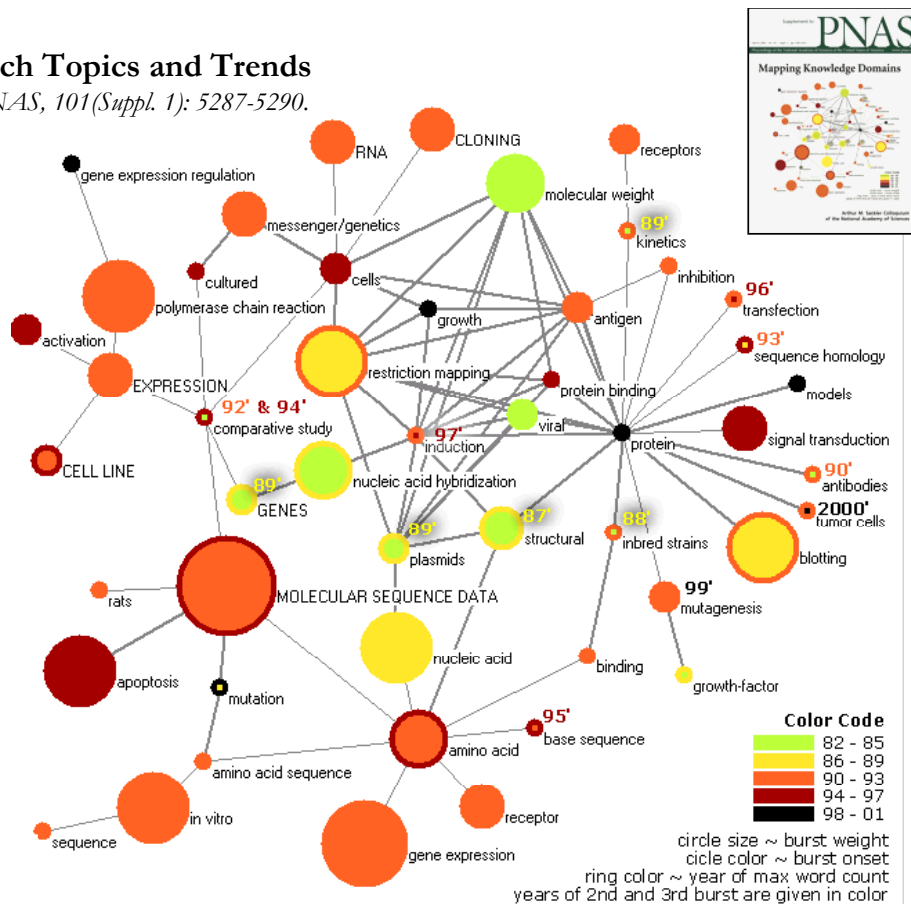
Identifying Research Topics and Trends

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

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

Words burst first before experiencing major usage.

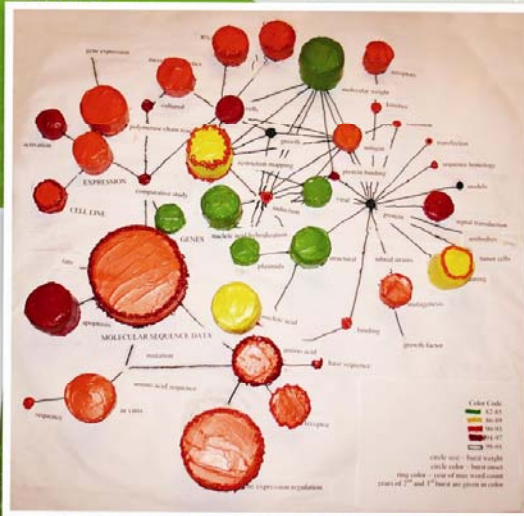
'Protein' and 'model' are among the highly bursty terms in 98-01 and became major research topics since then.



Merry Christmas and Happy New Year! 2008

Jon Burgoyne Katy Börner
Russell J. Duhon

Shravan Rajagopal
Heng (Michael) Zhang
Bruce W. Herr II
Julie M. Smith
Peter A. Hook
Nanli Ma
Chung-Yang (Kenneth) Lee



Kristin E. Reed
Stacy Kowalevzk
Micah Linnemeier
Bryan J. Hook
Nianli Ma
Elisha F. Hardy
Fileve Palmer
Carol Walter
Renpeng Hu
Richard Pinapati
Todd Holloway
Heng (Michael) Zhang
Peter A. Hook
Renpeng Hu
Benjamin Ray Gonzalez Jr
Micah Linnemeier
Peter A. Hook
Benjamin Ray Gonzalez Jr
Micah Linnemeier

Weixia (Bonnie) Huang
Elisha F. Hardy
Fileve Palmer
Bruce W. Herr II
Benjamin Ray Gonzalez Jr
Micah Linnemeier
Peter A. Hook
Benjamin Ray Gonzalez Jr
Micah Linnemeier

Cake created by Kristin Reed and Lydia Nichols. They insisted on having a legend!

<http://ella.slis.indiana.edu/~katy>

<http://scimaps.org>

<http://ivl.slis.indiana.edu>

Modeling the Co-Evolving Author-Paper Networks

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



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

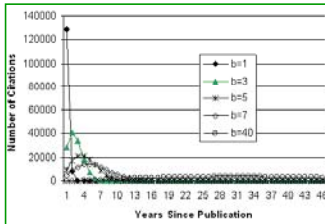
- Co-author and paper-citation networks co-evolve.
- Authors come and go.
- Papers are forever.
- Only authors that are 'alive' are able to co-author.
- All existing (but no future) papers can be cited.
- Information diffusion occurs directly via co-authorships and indirectly via the consumption of other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

```

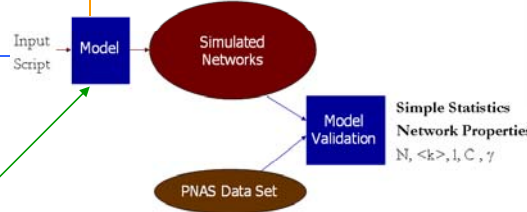
Aging function



```

// 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;
}

```



Model Validation

The properties of the networks generated by this model are validated against the 20-year PNAS data set (1982-2001).

Table 3 Statistics for SIM data

Year	#p	#a	#r	#c	alpha
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	alpha
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	3250469	

Process Model in Pseudocode & Input Parameters

If no topics are considered then the number of topics is one, i.e., all papers and authors have the same topic. If no coauthors are considered then each paper has exactly one author. If the reference path length is 0 then no references are considered for citation.

// 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;
}

```

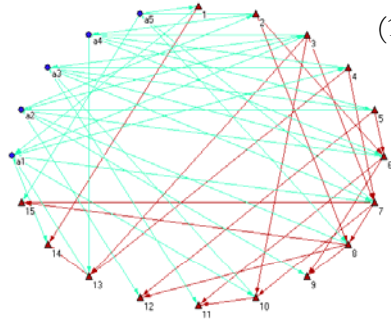
```

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

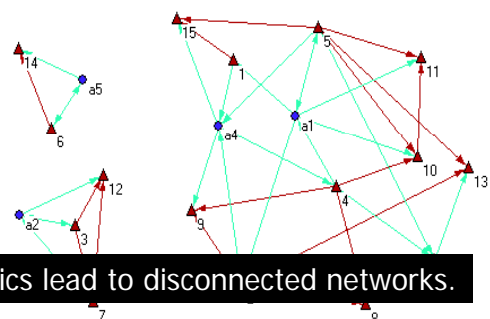
```

The TARL Model: The Effect of Parameters

(0000)

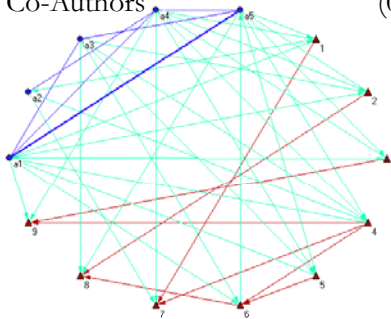


(1000) Topics

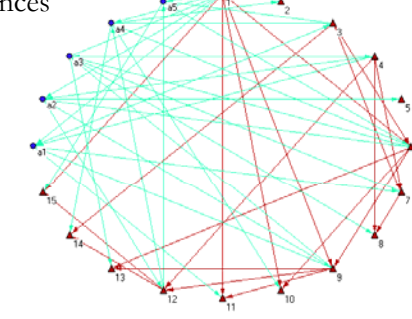


Topics lead to disconnected networks.

(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

Aging function

```

// 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 #_coauthors+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_group of size #_coauthors+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_papers;
        }
      }
    }
  }
  add all new papers to the set of existing papers;
  add new links to author and paper information;
}
    
```

Input Script → **Model** → Simulated Networks

Simulated Networks → Model Validation → Simple Statistics Network Properties (N, <k>, l, C, γ)

PNAS Data Set → Model Validation

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 $\langle k \rangle$, path length l , cluster coefficient C , and power law exponent γ . Source references are given in the left column.

Network	n	$\langle k \rangle$	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	

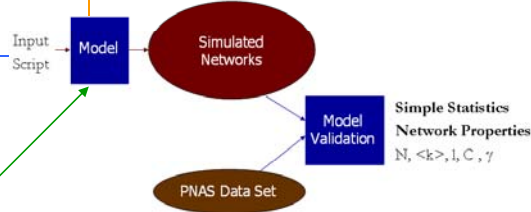
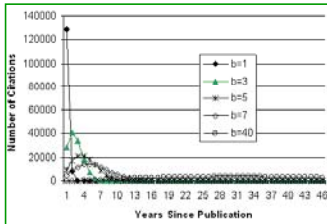
```

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
    
```

```

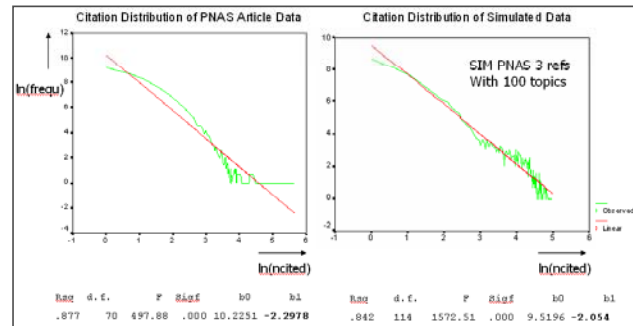
// Initialization
generate #_papers papers and assign a random topic to each paper;
generate #_authors authors and assign a random topic to each author;
randomly assign #_coauthors+1 authors to papers of the same topic;
// Simulation
for each year do {
  add #_new_authors new authors, deactivate authors older than #_author_age;
  for each topic do {
    randomly partition set of authors into author_group of size #_coauthors+1;
    for each author_group do {
      generate new_papers;
      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 cur_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_papers;
      }
    }
  }
  add all new papers to the set of existing papers;
  add new links to author and paper information;
}
    
```

Aging function



Co-Author and Paper-Citation Network Properties

Power Law Distributions



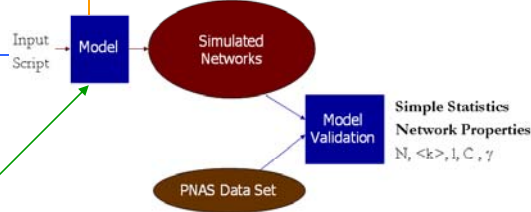
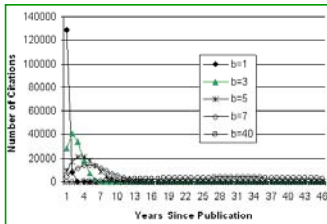
```

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 #_coauthors+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_group of size #_coauthors+1;
    for each author_group do {
      generate new_papers;
      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 cur_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_papers;
      }
    }
  }
  add all new papers to the set of existing papers;
  add new links to author and paper information;
}
    
```

Aging function



Topics: The number of topics is linearly correlated with the clustering coefficient of the resulting network: $C = 0.000073 * \# \text{topics}$. 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.

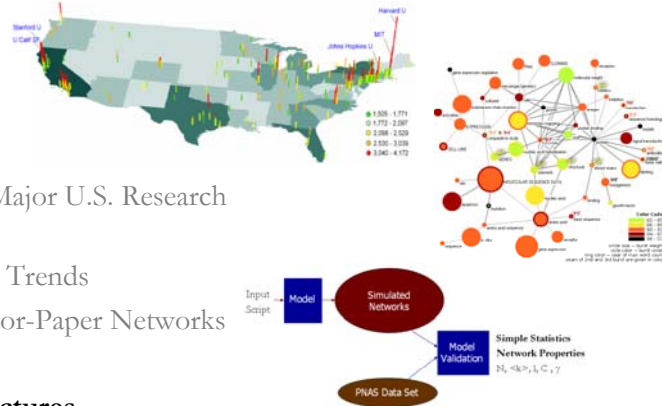
References/Recursive Linking: The length of the chain of paper citation links that is followed to select references for a new paper also influences the clustering coefficient. Temporal clustering is ameliorated by the practice of citing (and hopefully reading!) the papers that were the earlier inspirations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

Overview

Computational Scientometrics

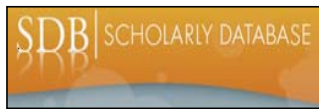
Case Studies:

- Information Diffusion Among Major U.S. Research Institutions
- Identifying Research Topics and Trends
- Modeling the Co-Evolving Author-Paper Networks



Science of Science Cyberinfrastructures

- Scholarly Database
- Network Workbench Tool
- Mapping Science Exhibit



Scholarly Database: Web Interface

Search across publications, patents, grants.

Download records and/or (evolving) co-author, paper-citation networks.

The image shows two screenshots of the Scholarly Database web interface. The left screenshot is the login page, which has two columns: 'IU User' and 'Non-IU User'. The 'IU User' column has a 'Go to IU Login' button. The 'Non-IU User' column has fields for 'Email' and 'Password' with a 'Login' button. Below the login fields are links for 'Not Registered Yet?' and 'In the News'. The right screenshot is the search page, which has a navigation bar with 'Search', 'Edit Profile', 'Admin', 'About', and 'Logout'. The search area includes fields for 'Creators', 'Title', 'Abstract' (with a dropdown for 'PIAS'), 'Full Text', 'First Year' (with a dropdown for '1999'), and 'Last Year' (with a dropdown for '2009'). There are also checkboxes for 'Fulltext (1999 - 2009)', 'Abstract (1999 - 2009)', and 'PIAS (1978 - 2007)'. A 'Search' button is at the bottom. To the right of the search fields is a help text explaining search syntax like 'AND', 'OR', and 'AND/OR'.

Register for free access at <http://sdb.slis.indiana.edu>

Datasets available via the Scholarly Database

Dataset	#Records	Years Coverage	updated	Restricted Access
Medline	16,053,495	1898-2008	Yes	
PhysRev	398,005	1893-2006		Yes
PNAS	16,167	1997-2002		Yes
JCR	59,078	1974,1979,1984,1989,1994-2004		Yes
USPTO	3,710,952	1976-2007	Yes	
NSF	174,835	1985-2003	Yes	
NIH	1,043,804	1972-2002	Yes	
Total	21,456,336	1893-2008	4	3

Aim for comprehensive temporal, geospatial, and topic coverage.



Network Workbench (NWB) Project

Investigators: Katy Börner, Albert-Laszlo Barabasi, Santiago Schnell, Alessandro Vespignani & Stanley Wasserman, Eric Wernert



Software Team: Lead: Micah Linnemeier
 Members: Patrick Phillips, Russell Duhon, Tim Kelley & Ann McCranie
 Previous Developers: Weixia (Bonnie) Huang, Bruce Herr, Heng Zhang, Duygu Balcan, Bryan Hook, Ben Markines, Santo Fortunato, Felix Terkhorn, Ramya Sabbineni, Vivek S. Thakre & Cesar Hidalgo

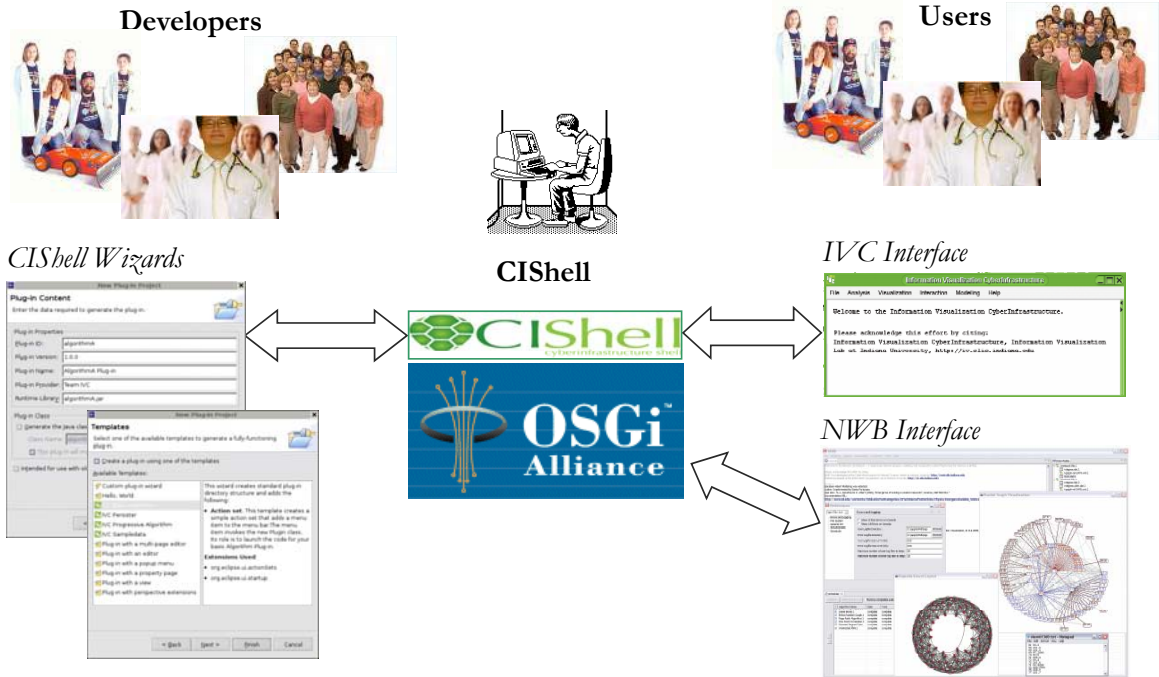


Goal: Develop a large-scale network analysis, modeling and visualization toolkit for physics, biomedical, and social science research.

Amount: \$1,120,926, NSF IIS-0513650 award

Duration: Sept. 2005 - Aug. 2009

Website: <http://nwb.slis.indiana.edu>



Katy Börner: Mapping the Structure and Dynamics of Science 35

Algorithms Currently Available

See <https://nwb.slis.indiana.edu/community> July 1st, 2008

Preprocessing [Edit](#)

- Remove Nodes**
 - [Extract Top Nodes](#)
 - [Extract Nodes Above or Below Val](#)
 - [Delete High Degree Nodes](#)
 - [Delete Random Nodes](#)
 - [Delete Isolates](#)
- Remove Edges**
 - [Extract Top Edges](#)
 - [Extract Edges Above or Below Val](#)
 - [Remove Self Loops](#)
 - [Trim By Degree²](#)
 - [Pathfinder Network Scaling](#)
- Sampling**
 - [Snowball Sampling \(n nodes\)](#)
 - [Node Sampling](#)
 - [Edge Sampling](#)
- Transformations**
 - [Symmetrize](#)
 - [Dichotomize](#)
 - [Multipartite Joining](#)

Modeling [Edit](#)

- General**
 - [Random Graph](#)
 - [Watts-Strogatz Small World](#)
 - [Barabási-Albert Scale-Free](#)
- Structured**
 - [CAN](#)
 - [Chord](#)
- Unstructured**
 - [Hypergrid](#)
 - [PRU](#)
- Other**
 - [TARL](#)
 - [Discrete Network Dynamics](#)

Analysis [Edit](#)


- General Purpose**
 - [Network Analysis Toolkit²](#)
- Unweighted & Undirected**
 - Based on degree/**
 - [Node Degree](#)
 - [Node Distribution](#)
 - Based on clustering**
 - [k-Nearest Neighbor](#)
 - [Watts Strogatz Clustering Coefficient](#)
 - [Watts Strogatz Clustering Coefficient](#)
 - Based on path**
 - [Diameter](#)
 - [Average Shortest Path](#)
 - [Shortest Path Distribution](#)
 - [Node Betweenness Centrality](#)
 - Based on components**
 - [Connected Components](#)
 - [Weak Component Clustering](#)
 - K-Core**
 - [Extract K-Core²](#)
 - [Annotate K-Core²](#)
- Unweighted & Directed**
 - Based on degree**
 - [Node Indegree](#)
 - [Node Outdegree](#)
 - [Indegree Distribution](#)
 - [Outdegree Distribution](#)
 - Based on local graph structure**
 - [k-Nearest Neighbor](#)
 - [Single Node In-Out Degree Correl²](#)
 - Unnamed Category?**
 - [Page Rank](#)
 - Based on local graph structure**
 - [Dyad Reciprocity²](#)
 - [Arc Reciprocity²](#)
 - [Adjacency Transitivity²](#)
 - Based on components**
 - [Weak Component Clustering](#)
 - [Extract Attractors²](#)

Visualization [Edit](#)

- Tools**
 - [GUESS](#)
 - [GnuPlot²](#)
- Predefined Positions Layout**
 - [DrL \(VxOrd\)](#)
 - [Pre-defined Positions \(prefuse beta\)²](#)
- Move**
 - [Circular](#)
- Tree Layouts**
 - [Radial Tree \(prefuse alpha\)](#)
 - [Radial Tree with Annotations \(prefuse beta\)²](#)
 - [Tree Map](#)
 - [Tree View](#)
 - [Balloon Graph \(prefuse alpha\)²](#)
- Network Layouts**
 - [Force Directed with Annotation \(prefuse beta\)](#)
 - [Kamada-Kawai \(JUNG\)](#)
 - [Fruchterman-Reingold \(JUNG\)](#)
 - [Fruchterman-Reingold with Annotation \(prefuse beta\)](#)
 - [Spring \(JUNG\)](#)
 - [Small World \(prefuse alpha\)](#)
- Other Layouts**
 - [Parallel Coordinates \(demo\)²](#)
 - [LaNet \(k-Core Decomposition\)](#)

Scientometrics [Edit](#)

- Extract Network From Table**
 - [Extract Co-Authorship Network](#)
 - [Extract Co-Occurrence Network From Table²](#)
 - [Extract Directed Network From Table²](#)
- Extract Network From Another Network**
 - [Extract Bibliographic Coupling Similarity Network](#)
 - [Extract Co-Citation Similarity Network²](#)
- Cleaning**
 - [Remove ISI Duplicate Records](#)
 - [Detect Duplicate Nodes](#)
 - [Remove Rows With Multitudinous Fields²](#)



INDIANA UNIVERSITY

IU News Room

Sunday, May 4, 2008

IU News from all eight campuses

Browse by Campus [v] Services & Resources [v]

Search [input type="text"/> GO

For Journalists | Archives | Site Index | Contact Us | Public Affairs

Newsroom Home > Indiana University Media Relations > News Release

Last modified: Tuesday, April 8, 2008

\$1.2 million NIH project will help track and predict epidemics

E-mail this page Print this page

FOR IMMEDIATE RELEASE
April 8, 2008

BLOOMINGTON, Ind. -- The National Institutes of Health has given \$1.2 million to Indiana University researchers to build the ultimate international epidemic research tool.

Media Contacts

- Neal Moore
ngmoore@indiana.edu
317-278-9208
- David Bricker
brickerd@indiana.edu
812-856-9035

News by Topic

- General News
- Graduate Studies
- Life Sciences
- Science
- Technology

More Topics >>

News by Category

Katy Börner: Mapping the Structure and Dynamics of Science 37

Mapping Science Exhibit – 10 Iterations in 10 years

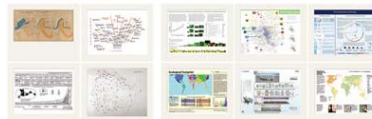
<http://scimaps.org/>



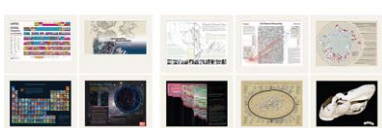
The Power of Maps (2005)



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)

The Power of Forecasts (2007)



How to Lie with Science Maps (2014)

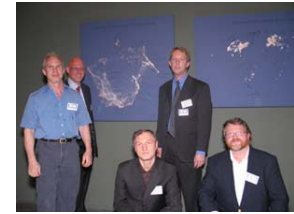


Exhibit has been shown in 49 venues on four continents. Also at

- NSF, 10th Floor, 4201 Wilson Boulevard, Arlington, VA.
- Chinese Academy of Sciences, China, May 17-Nov. 15, 2008.
- University of Alberta, Edmonton, Canada, Nov 10-Jan 31, 2009
- Center of Advanced European Studies and Research, Bonn, Germany, Dec. 11-19, 2008.

Illuminated Diagram Display

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



Large-scale, high resolution prints illuminated via projector or screen.

Questions:

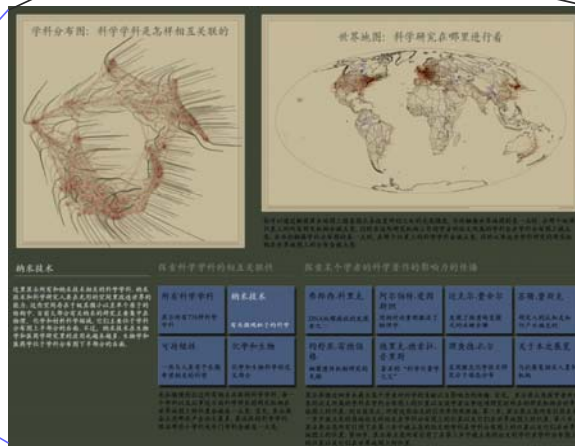
- Who is doing research on what topic and where?
- What is the 'footprint' of interdisciplinary research fields?
- What impact have scientists?

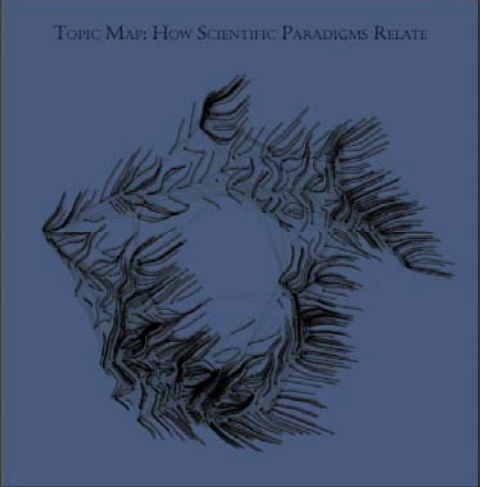


Interactive touch panel.


Contributions:

- Interactive, high resolution interface to access and make sense of data about scholarly activity.





TOPIC MAP: HOW SCIENTIFIC PARADIGMS RELATE



GEOGRAPHIC MAP: WHERE SCIENCE GETS DONE

You may run your finger over each of these maps to control the lighting on the other: touching a place on the world map will light up topics studied in that place; touching a paradigm on the topic map will light up the places that study that topic.

Nanotechnology

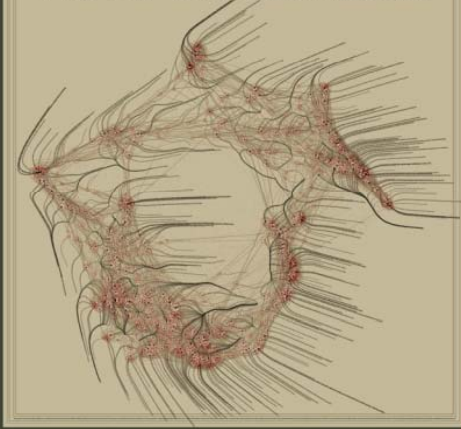
This overlay shows the distribution of nanotechnology within the paradigms of science. The majority of current work in nanotechnology takes places in physics, chemistry, and materials science, at the upper right portion of the map. However, an increasing amount of nanotechnology is being applied in the biological and medical sciences, at the lower right.

All Topics <i>Sweep through all 776 scientific paradigms</i>	Nanotechnology <i>Science on the tiny scale of molecules</i>	Francis H. C. CRICK <i>Co-discovered DNA's double helix</i>	Albert EINSTEIN <i>Revitalized physics with Relativity theories</i>	Michael E. FISHER <i>Models critical phase transitions of matter</i>	Susan T. FISKE <i>Connects perception and stereotypes</i>
Sustainability <i>The science behind our long-term hopes</i>	Biology & Chemistry <i>The interface between these two vital fields</i>	Joshua LEDERBERG <i>Pioneer in bacterial genetic mechanisms</i>	Derek J. de Solla PRICE <i>Known as the "Father of Scientometrics"</i>	Richard N. ZARE <i>Uses laser chemistry in molecular dynamics</i>	About this display <i>People & organizations that helped create it</i>

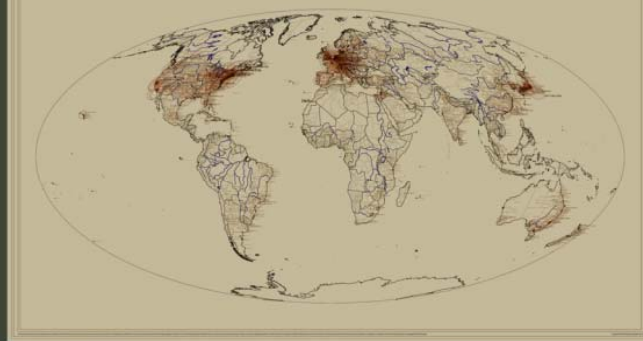
We sweep slowly through adjoining related topics, lighting up the places in the world that study each topic. You may select a subset of the topics that deal with these three interesting subjects by touching it.

A single person's spreading influence is shown as a series of four snapshots. First, we light only topics and places relating to that person's papers—papers that are still highly cited today. The second lights everything that cites that original work. Note that this first-generation impact extends to far more topics than did the original work. The third snapshot lights science that cites the second, and the fourth lights science that cites the third.

学科分布图：科学学科是怎样相互关联的



世界地图：科学研究在哪里进行着



你可以通过触摸屏在地图上随意指点来改变所到之处的光亮强度。当你触摸世界地图的某一点时，在那个地理位置上的所有研究机构会被点亮。同时在这些研究机构工作的学者的论文所属的学科会在学科分布图上被点亮。而当你触摸学科分布图的某一点时，在那个位置上的科学学科会被点亮，同时从事这些学科研究的研究机构在世界地图上的分布会被点亮。

纳米技术

这里显示所有和纳米技术相关的科学学科。纳米技术和科学研究人在无形的空间里改造世界的的能力。这些空间存在于其微小以至单个原子的结构中。目前大部分有关纳米的研究主要集中在物理、化学和材料科学领域。它们主要位于学科分布图上半部分的右面。不过，纳米技术在生物学和医药学研究里的应用也越来越多。生物学和医药学位于学科分布图下半部分的右面。



探索科学学科的相互关联性

所有科学学科 显示所有776种科学学科	纳米技术 有关微观粒子的科学
可持续性 一些与人类寄予长期希望相关的科学	化学和生物 化学和生物科学的交叉部分

光标缓慢的扫过所有相互关联的科学学科，每一个学科以及从事这方面科学研究的研究机构在世界地图上的位置会被逐一点亮。首先，显示屏会点亮那些产出论文最多、最活跃的科学学科，然后那些小学科或冷门学科会被逐一点亮。

探索某个学者的科学著作的影响力的传播

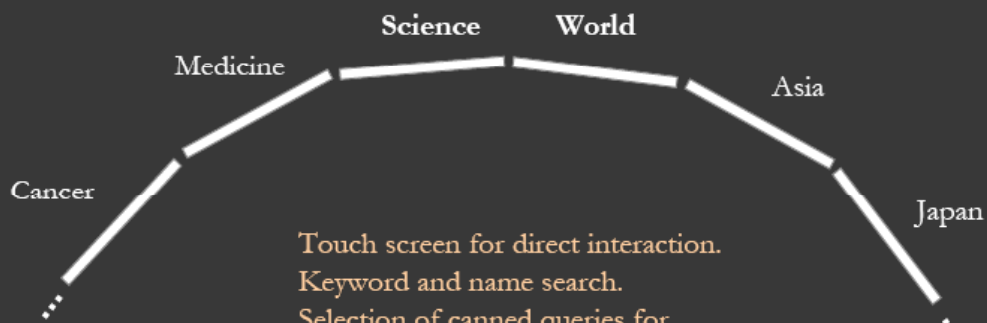
弗郎西·科里克 DNA双螺旋结构的发现者之一	阿尔伯特·爱因斯坦 用相对论重新激活了物理学	迈克尔·费舍尔 发现了物质转变模式的关键步骤	苏珊·费斯克 研究人的认知是如何产生偏见的
约舒亚·雷德伯格 细菌遗传机制研究先驱	德里克·德索拉·普里斯 著名的“科学计量学之父”	理查德·扎尔 采用激光化学技术研究分子动态分布	关于本次展览 与此展览相关人员和机构

显示屏通过四步来展示某个学者对科学的贡献以及影响力的传播。首先，显示屏点亮该学者所发表的论文所属的学科在学科分布图上的位置以及该学者从事这项研究时所在的研究机构在世界地图上的位置。到目前为止，所有这些论文的引用率仍然很高。第二步，显示屏点亮所有引用在第一步中被点亮的原始论文的论文在学科分布图上的位置以及它们在世界地图上的位置。第三步，显示屏点亮所有引用了在第二步中被点亮的论文的论文在学科分布图上的位置以及它们在世界地图上的位置。第四步，显示屏点亮所有引用了在第三步中被点亮的论文的论文在学科分布图上的位置以及它们在世界地图上的位置。

Re-implementation of Illuminated Diagram Software

by Advanced Visualization Lab, Indiana University

Drives unlimited number of ID screens.



- Touch screen for direct interaction.
 Keyword and name search.
 Selection of canned queries for
- interdisciplinary research areas
 - famous people
 - activity patterns, e.g., bursts, trends, etc.





Science of Science Cyberinfrastructure
— P O R T A L —

Provided by the [Cyberinfrastructure for Network Science Center](#) at Indiana University.

Introduction
E. O. Wilson writes in *Consilience: The Unity of Knowledge* (1998): "Features that distinguish science from pseudoscience are repeatability, economy, mensuration, heuristics, and consilience." Please see Börner's [recent presentation](#) at the *A Deeper Look at the Visualization of Scientific Discovery* NSF Workshop for a general introduction of the needs and the resources provided here.

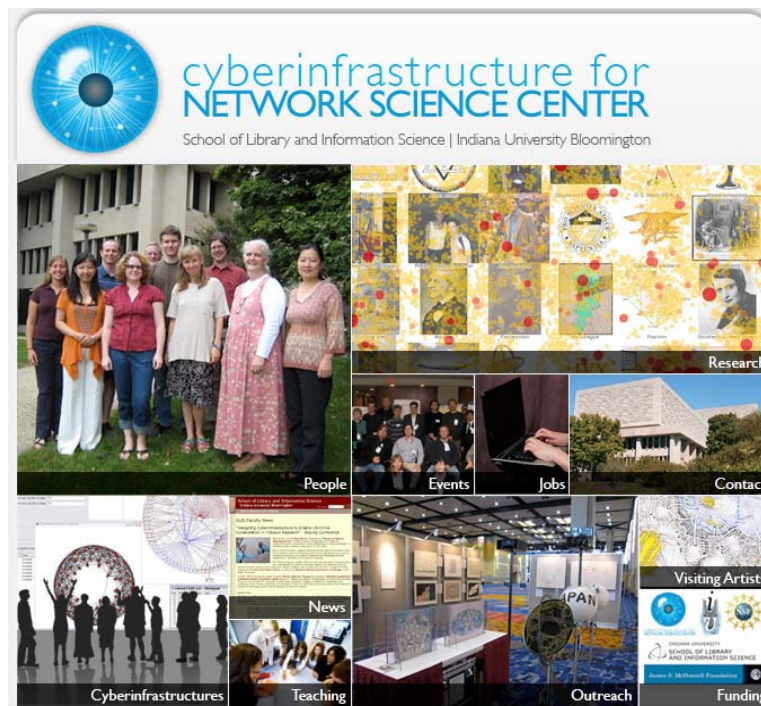
Needs Analysis
As part of the "TLIS: Towards a Macroscopic for Science Policy Decision Making" NSF SBE-0738111 award, interviews with science policy makers are conducted to identify what science of science research results and tools might be most desirable and effective. So far, 20 formal, one-hour interviews have been conducted with science policy makers at university campus level, program officer level, and division director level for governmental, state, and private foundations. Data compilation will start in October 2008 and resulting report can be ordered by sending a request to Mark Price (maaprice@indiana.edu).

Conceptualization of Science
A 'science of science' requires a theoretically grounded and practically useful conceptualization of the structure and evolution of science. A special journal issue entitled "[Science of Science: Conceptualizations and Models of Science](#)" edited by [Katy Börner](#), Indiana University & [Andrea Scharnhorst](#), Royal Netherlands Academy of Arts and Sciences invites contributions on this topic. It will be published in the *Journal of Informetrics* 3(1) in January 2009.

Scholarly Database
The [Scholarly Database \(SDB\)](#) at Indiana University aims to serve researchers and practitioners interested in the analysis, modeling, and visualization of large-scale scholarly datasets. The database currently provides access to over 20 million papers, patents and grants. Resulting datasets can be downloaded in bulk. Register for free access at <https://sdb.slis.indiana.edu/>.

Cyberinfrastructures
The Scientometrics filling of the [Network Workbench \(NWB\) Tool](#) provides a unique distributed, shared resources environment for large-scale network analysis, modeling, and visualization. Thomson Scientific/ISI, Scopus and Google Scholar data, EndNote and Bibtext files, or NSF awards can be read and diverse networks can be extracted and studied. Download [User Manual with focus on Scientometrics](#).

<http://sci.slis.indiana.edu>



cyberinfrastructure for
NETWORK SCIENCE CENTER
School of Library and Information Science | Indiana University Bloomington

People

Research

Events

Jobs

Contact

News

Cyberinfrastructures

Teaching

Outreach

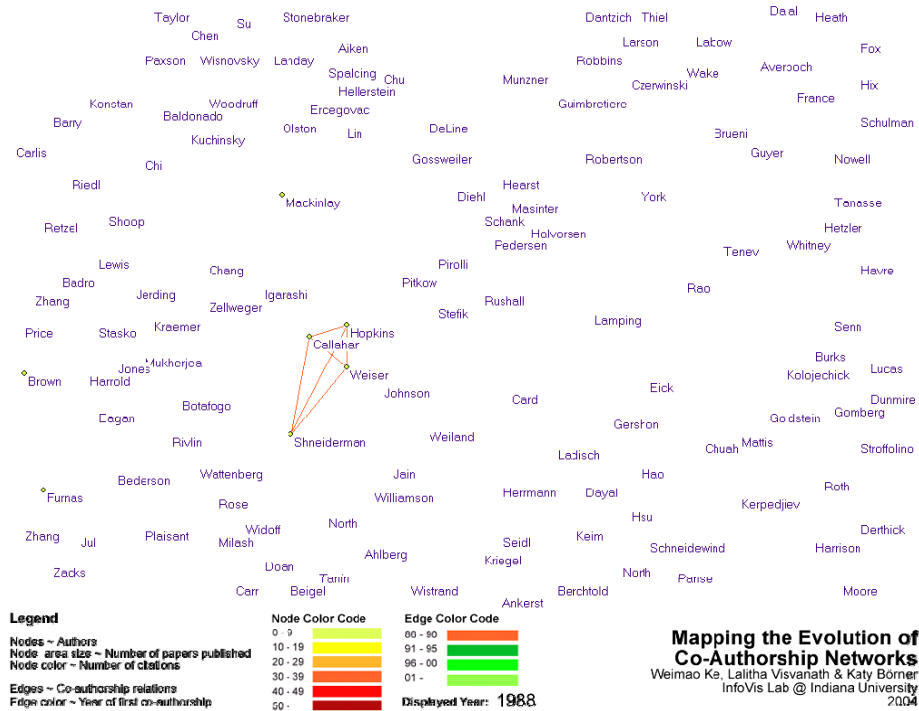
Visiting Artists

Funding

<http://cns.slis.indiana.edu>

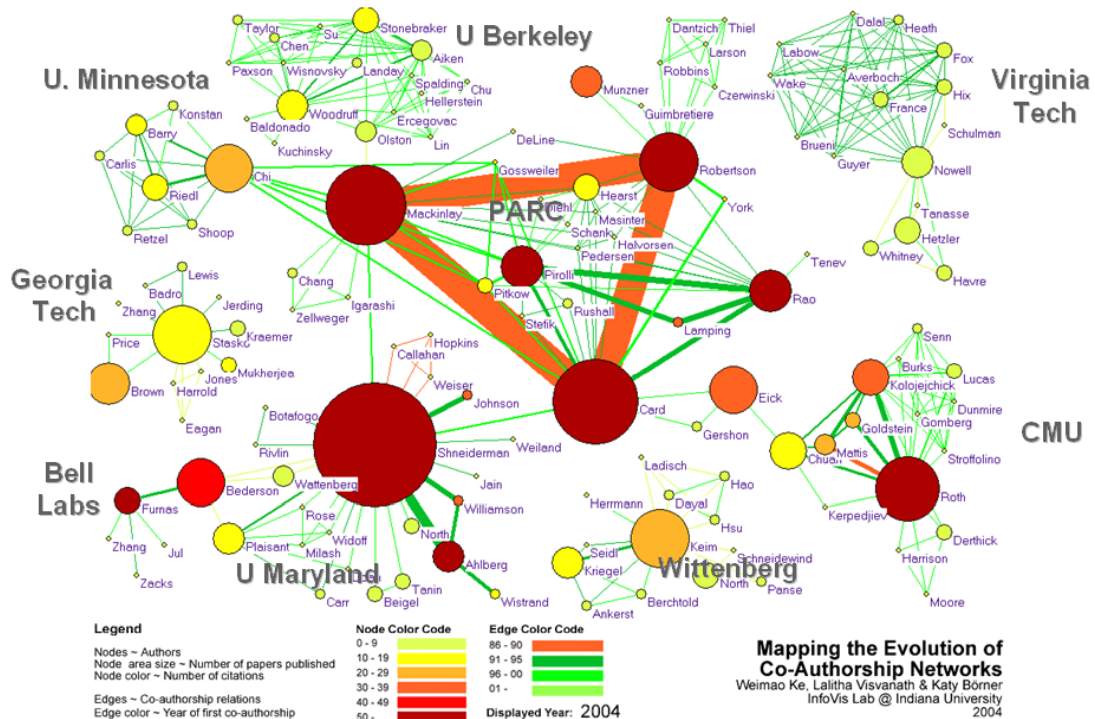
Mapping the Evolution of Co-Authorship Networks

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.



Studying the Emerging Global Brain: Analyzing and Visualizing the Impact of Co-Authorship Teams

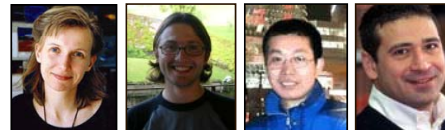
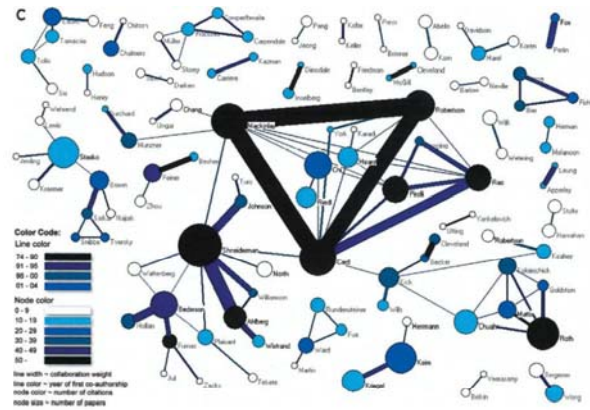
Börner, Dall'Asta, Ke & Vespignani (2005) *Complexity*, 10(4):58-67.

Research question:

- Is science driven by prolific single experts or by high-impact co-authorship teams?

Contributions:

- New approach to allocate citational credit.
- Novel weighted graph representation.
- Visualization of the growth of weighted co-author network.
- Centrality measures to identify author impact.
- Global statistical analysis of paper production and citations in correlation with co-authorship team size over time.
- Local, author-centered entropy measure.

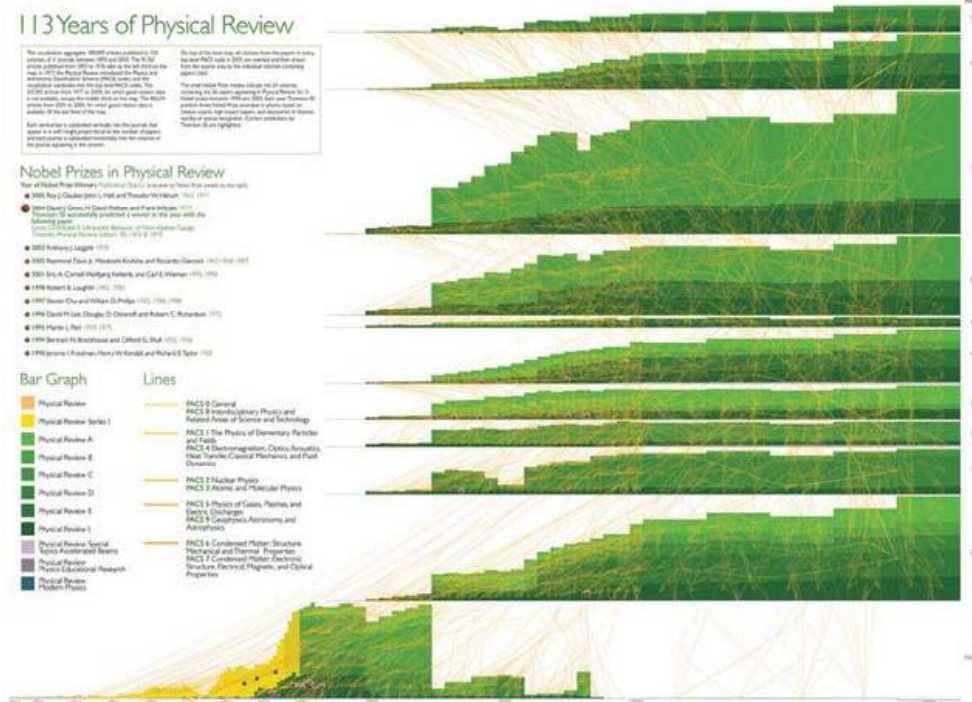


47

113 Years of Physical Review

http://scimaps.org/dev/map_detail.php?map_id=171

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

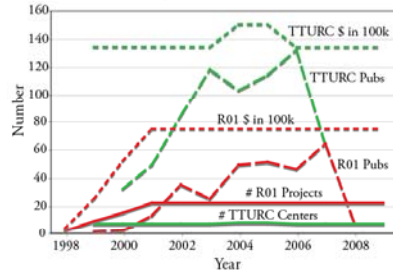


Mapping Transdisciplinary Tobacco Use Research Centers Publications

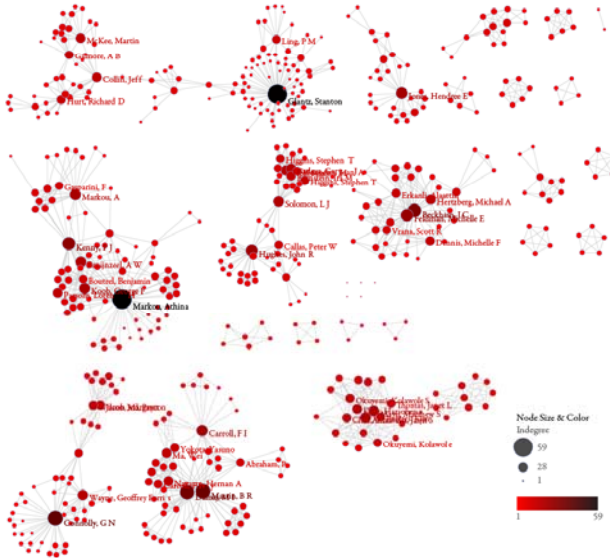
Compare R01 investigator based funding with TTURC Center awards in terms of number of publications and evolving co-author networks.

Zoss & Börner, forthcoming.

R01 & TTURC Project Information



R01 Co-Author Network

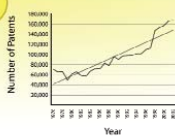


TTURC Co-Author Network



Examining the Evolution and Distribution of Patent Classifications

1 Patents Granted Over the Last 20 Years



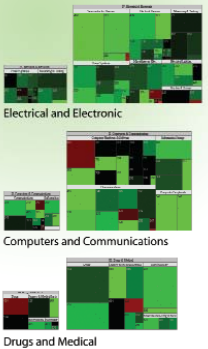
Class	Title	Patents
2801	Chemistry of Carbon Compounds	13,348
4307	English Writing and Body Training Compositions	10,527
2414	Compositions	9,148
6306	Vegetables	7,064
6708	Stocks, Bonds, or Miscellaneous Articles	6,971
751	Manufacturing and Testing	6,162
232	Internal-Combustion Engines	4,942
340	Communications/Chemical	4,283
2344	Electrical Components and Circuits/Power Systems	3,789
27	Metal Working	2,684
Total		35,019

Class	Title	Patents
311	Drug, Bio/Writing and Body Training Compositions	28,041
434	Chemical/Carbon Compounds	13,376
431	Chemistry, Multiple-Stage/Inhibitor/Aldehyde	12,674
434	Drug, Bio/Writing and Body Training Compositions	12,027
678	Stocks, Bonds, or Miscellaneous Articles	13,314
257	Audio, Solid-State Device/Asic, Transistors, Colloidal Crystals	12,024
395	Information Processing System Organization	8,095
393	Computer Graphics Processing, Operator Interface, Processing and Selective Visual Display Systems	6,519
237	Optical Systems and Devices	5,131
263	Static Information Storage and Retrieval	4,822
Total		140,000

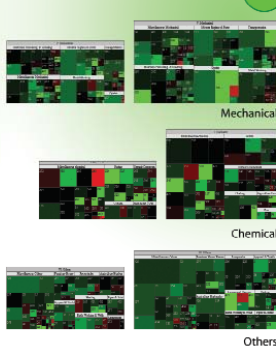
In the United States, each patent gets assigned to one out of more than 450 classes covering broad application domains. An examination of the size and growth of patent classes provides insight about patenting trends.

Treemaps, a space-filling technique developed in the HCI Lab at the University of Maryland, are used to communicate major results. Treemaps represent a tree structure as nested rectangles with each rectangle representing a node. A rectangular area is first allocated to hold the representation of the tree, and this area is then subdivided into a set of rectangles that represent the top level of the tree. This process continues recursively on the resulting rectangles to represent each lower level of the tree. The parent-child relationship is indicated by enclosing the child rectangle by its parent rectangle. Typically, the size of each rectangle corresponds to the size of the node. Additional information about a node, e.g., its age or value, can be represented by the color of the respective rectangle.

2 Fast Growth Domains 1983 - 1987 / 1998 - 2002

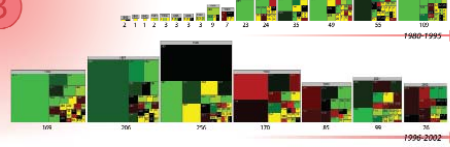


2 Slow Growth Domains 1983 - 1987 / 1998 - 2002



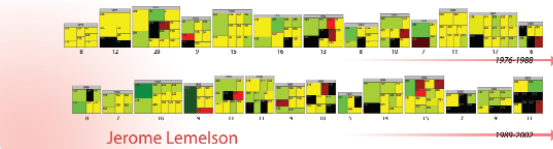
Shown is a comparison of the patent class space for 1983 to 1987 and 1998 to 2002. There is a predominance of growth in the 1998 to 2002 patent space, which correlates to the increase in patent grants during this period. By comparing the growth in categories, one can distinguish between domains that have been receiving a larger amount of patent grants.

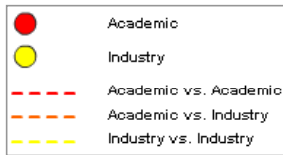
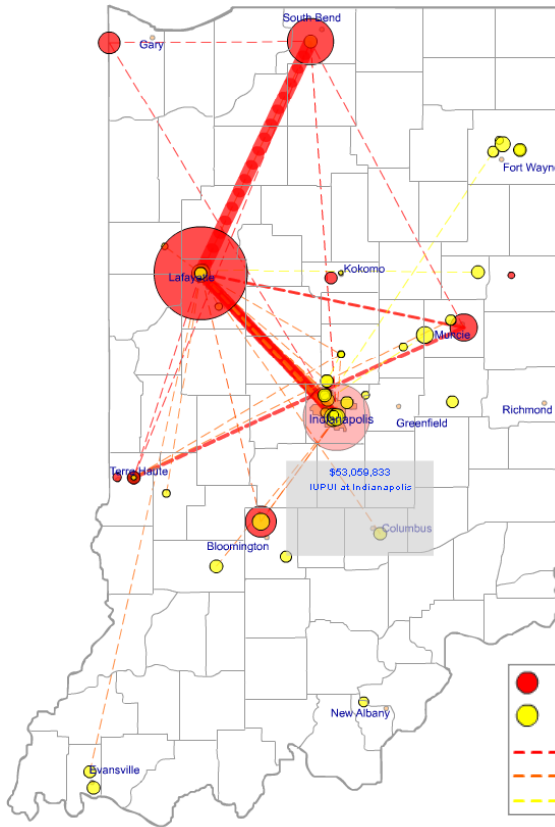
3 Apple Computer



Depicted above is how Apple Computers' portfolio has changed in yearly increments from 1980 to 2002.

Legend:
Green = Increase in number of patent grants in particular class.
Red = Decrease in number of patent grants in particular class.
Yellow = No patents granted in that class in the past five years.
Size = Number of patent grants in a particular class.





Mapping Indiana's Intellectual Space

Identify

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

Wikipedian Activity

Studying large scale social networks such as Wikipedia

Vizzards 2007 Entry

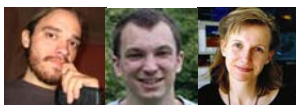
Second Sight: An Emergent Mosaic of Wikipedian Activity, *The NewScientist*, May 19, 2007

Rendered as Google Map:

<http://scimaps.org/maps/wikipedia>

Jan 8th, 2008 Data Version on Gigapan:

<http://gigapan.org/viewGigapan.php?id=5042>



Second sight

Image: Bruce W. Herr and Todd M. Holloway

Power struggle



To spot where arguments are taking place, Herr suggests. If rival contributors are repeatedly changing each other's entries, for example, a page could be 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.

www.newscientist.com

19 May 2007 | NewScientist | 95

Science Related Wikipedian Activity

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

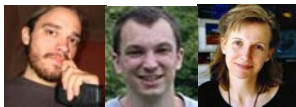
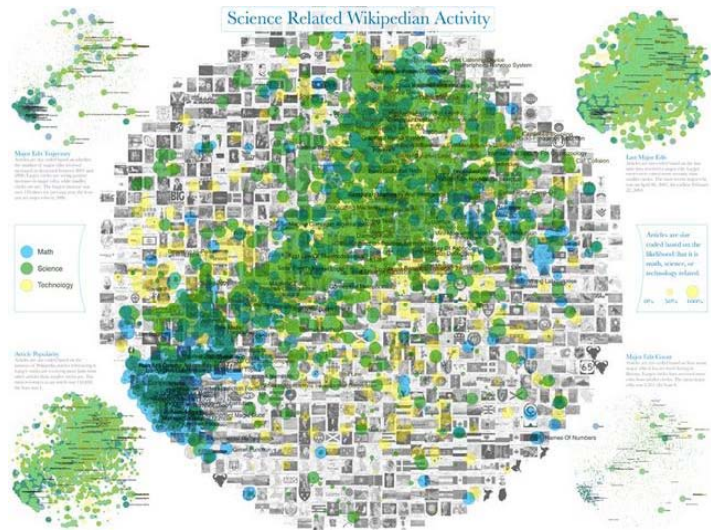
Same base map.

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).



GEOGRAPHIC MAP: WHERE SCIENCE GETS DONE

