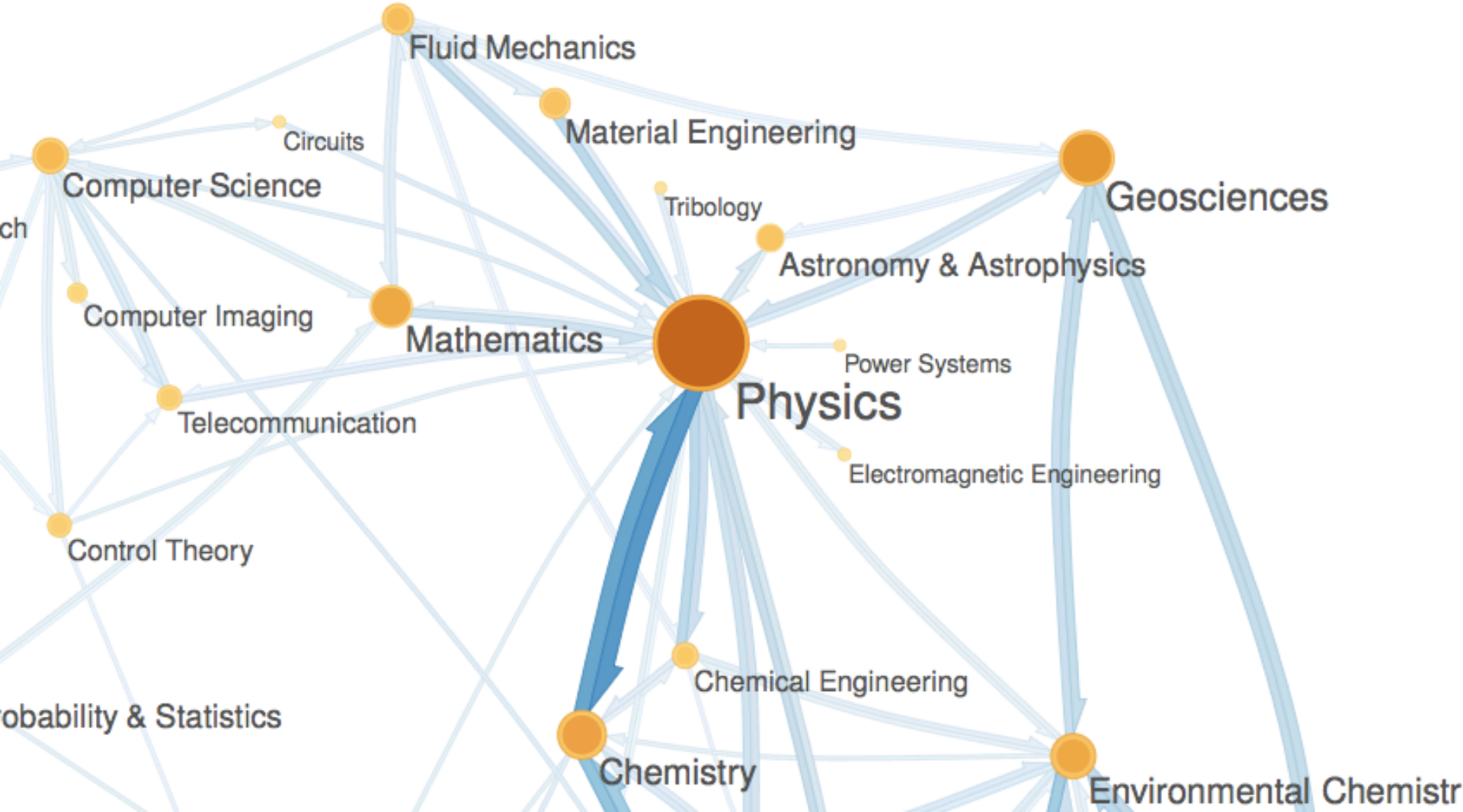
Mapping the emergence of scientific disciplines

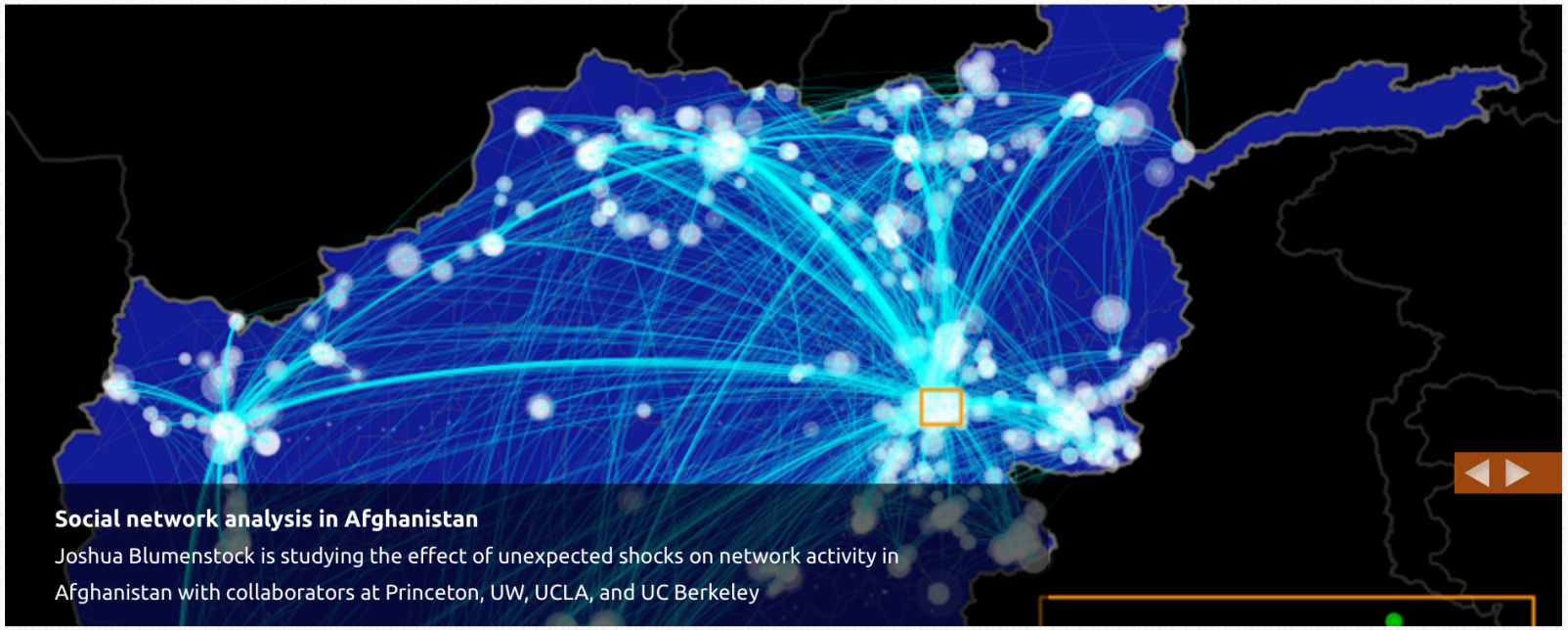
Jevin West, Information School, University of Washington



DataLab

Data Science and Analytics Lab

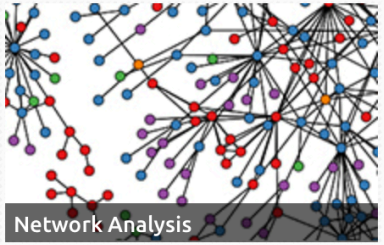
University of Washington Information School



Social network analysis in Afghanistan

Joshua Blumenstock is studying the effect of unexpected shocks on network activity in Afghanistan with collaborators at Princeton, UW, UCLA, and UC Berkeley

Research Focus Areas



News and Updates

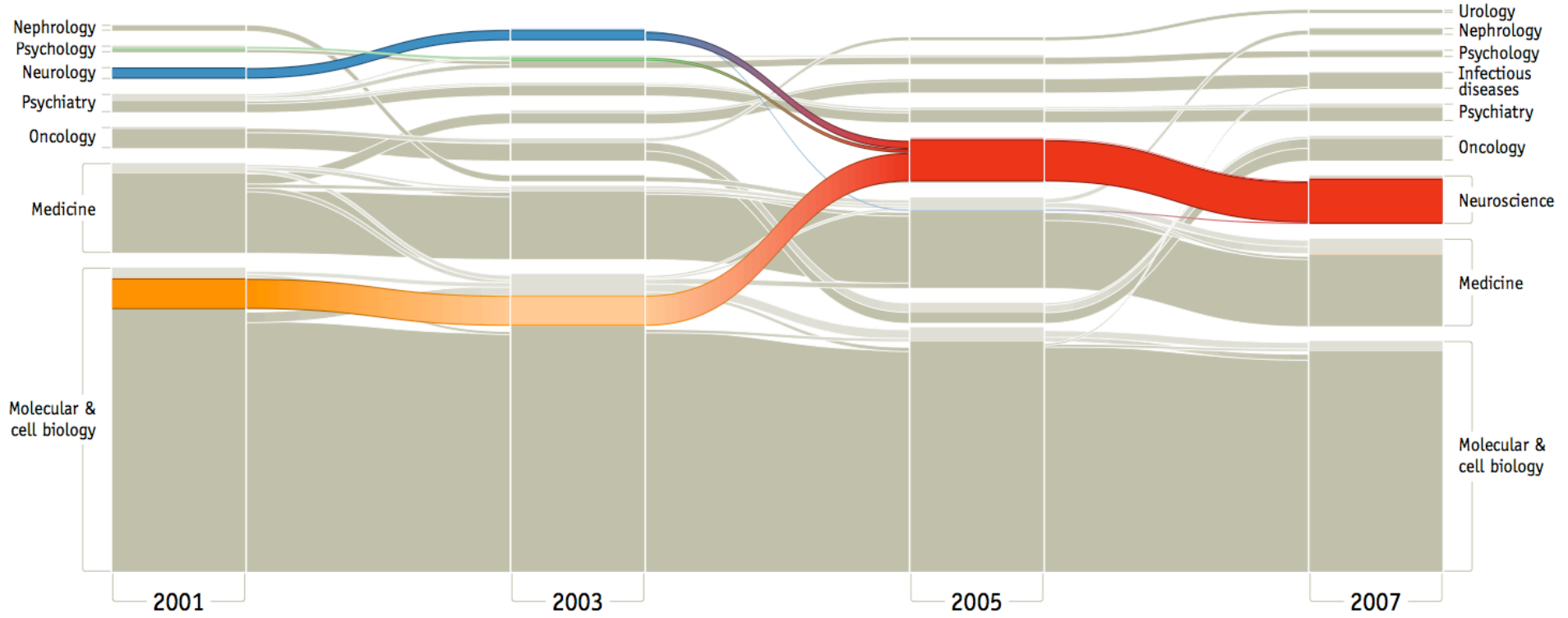
28 **Blumenstock at Population Association of America**

What we do

The DataLab is the nexus for research on Data Science and Analytics at the UW iSchool. We study **large-scale, heterogeneous human data** in an

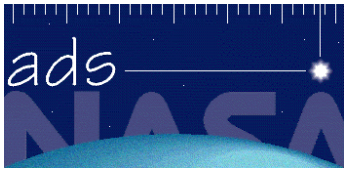
How do we *map* the origins of scientific disciplines?

The Emergence of Neuroscience



Rosvall and Bergstrom (2010) PLoS One

The Scholarly Graph



The Scholarly Graph

Tens of millions articles, patents, books

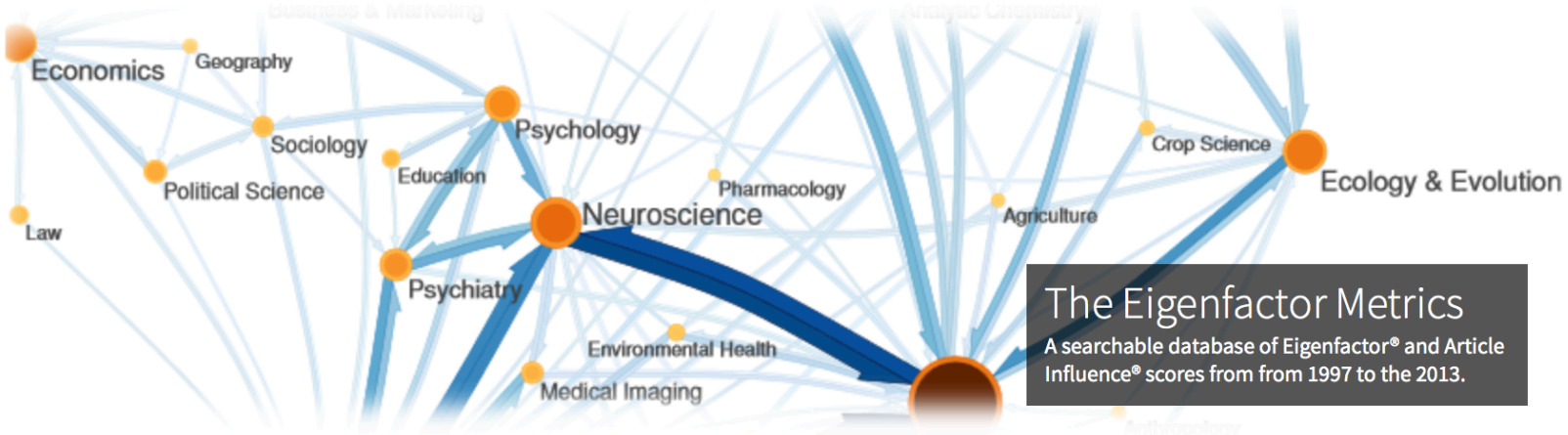
Billions of citation links

Years: 1600 - 2016

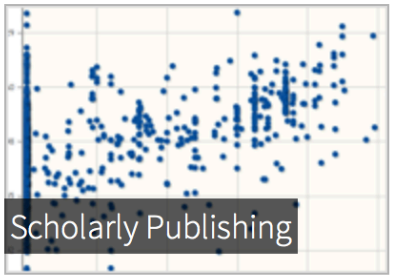
Science of Science

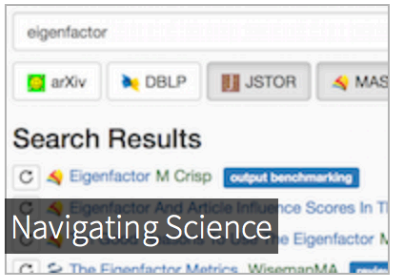
- What ideas, papers and scientists seeded new fields?
- Can we automate hypothesis generation? If so, how would this change science?
- Can we teach the computer to read the literature and understand figures?
- How can we improve scientific 'navigation' in the face of information overload?
- Can we predict new discoveries before they happen?
- How can we improve scientific institutions, reward mechanisms and funding processes?
- How can we facilitate interdisciplinary research and foster innovation rather than just piles of papers?

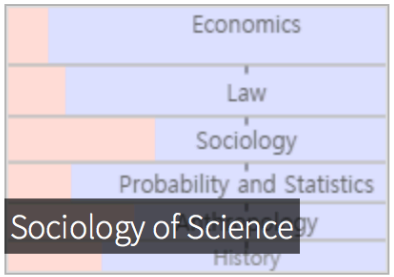
EIGENFACTOR.org



RESEARCH AREAS







NEWS

23 Nov.

JEVIN WEST ON MEGAJOURNALS IN THE *CHRONICLE OF HIGHER EDUCATION*

Jevin West discusses the rise of the megajournal and our open access cost effectiveness tool in the *Chronicle of Higher Education*.

---

23 Nov.

EIGENFACTOR TEAM PLACES SECOND IN MICROSOFT RESEARCH'S WSDM CUP

The WSDM Cup Challenge asked teams to use 30GB of data from the Microsoft Academic Graph to rank the

Babel Home About API

babel.eigenfactor.org

oren etzioni

Ian Wesley-Smith

DBLP JSTOR MAS PLOS PubMed

s

al Methods For Analyzing Speedup Learning Experiments O Etzioni satisfaction programs

-Face And Computer-Mediated Communities Amitai Etzioni, Oren Etzioni 1998 resources sustained

cument Clustering O Zamir document clustering

Communities: Virtual Vs. Real A Etzioni 1996 implications internet

al Methods For Analyzing Speedup Learning Experiments. O Etzioni 1993 scheduling problems

al Methods For Analyzing Speedup Learning Experiments O Etzioni 1993 generating abstractions

Get Related Web Document Clustering: A Feasibility Demonstration O Zamir 1997 document clustering

Get Related Web Document Clustering: A Feasibility Demonstration. O Zamir 1997 browsing large

Get Related Sound And Efficient Closed- World Reasoning O Etzioni proving problem

Get Related Appears In Comm. OfACM O Etzioni scalable comparison-shopping

« Previous 1 2 3 4 5 6 7 8 9 10 Next »

Papers related to Statistical Methods For Analyzing Speedup Learning Experiments O Etzioni satisfaction programs

Get Related Automatically Configuring Constraint Satisfaction Programs: A Case Study S Minton 1995 satisfaction programs

Get Related Abstraction Via Approximate Symmetry T Ellman 1992 satisfaction programs

Get Related Integrating Heuristics For Constraint Satisfaction Problems: A Case Study S Minton 1992 satisfaction programs

Get Related An Analytic Learning System For Specializing Heuristics S Minton 1992 satisfaction programs

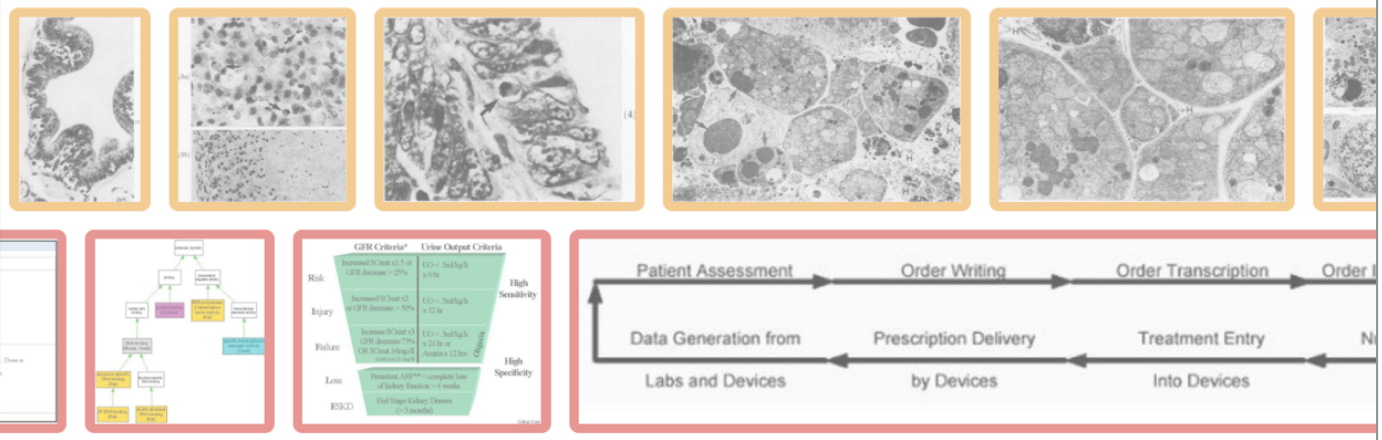
Get Related Automated Synthesis Of Constrained Generators W Braudaway 1988 satisfaction programs

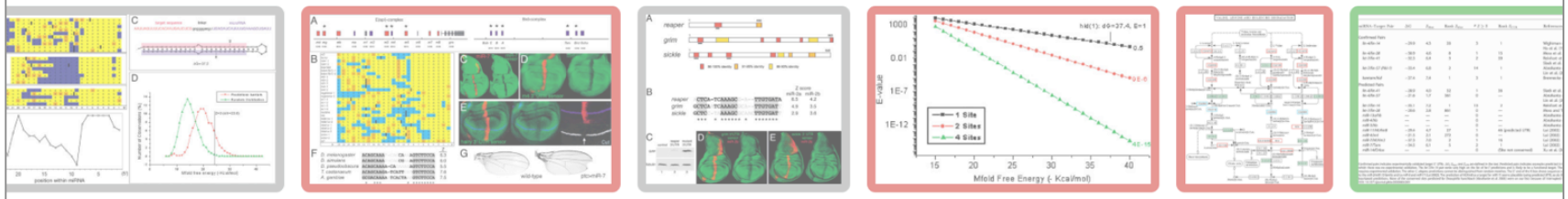
VizioMetrix About Search Crowdsourcing viziometrics.org

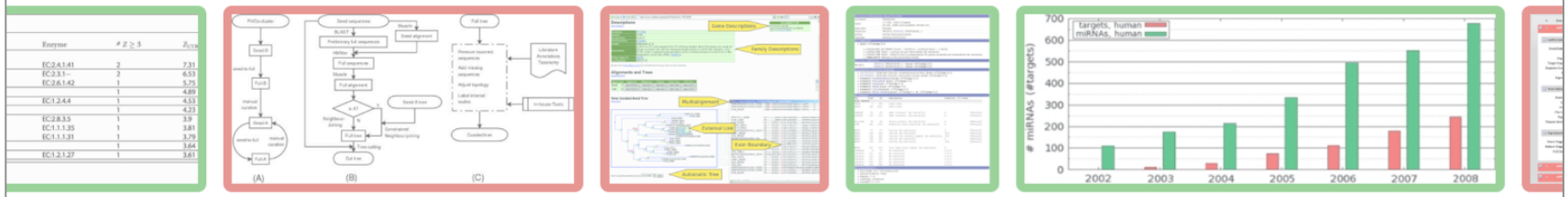
Poshen Lee

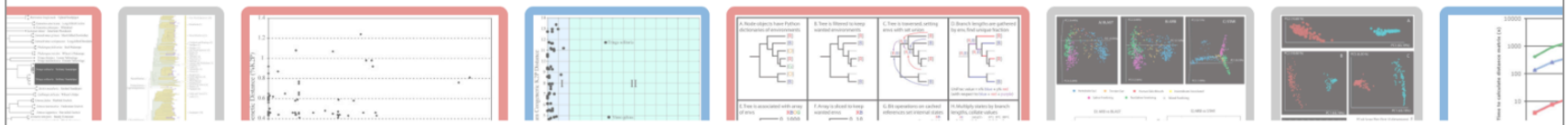
Impact | Keywords or Cluster, Result Ordered by Impact | Search

Composite Equation Diagram Photo Plot Table



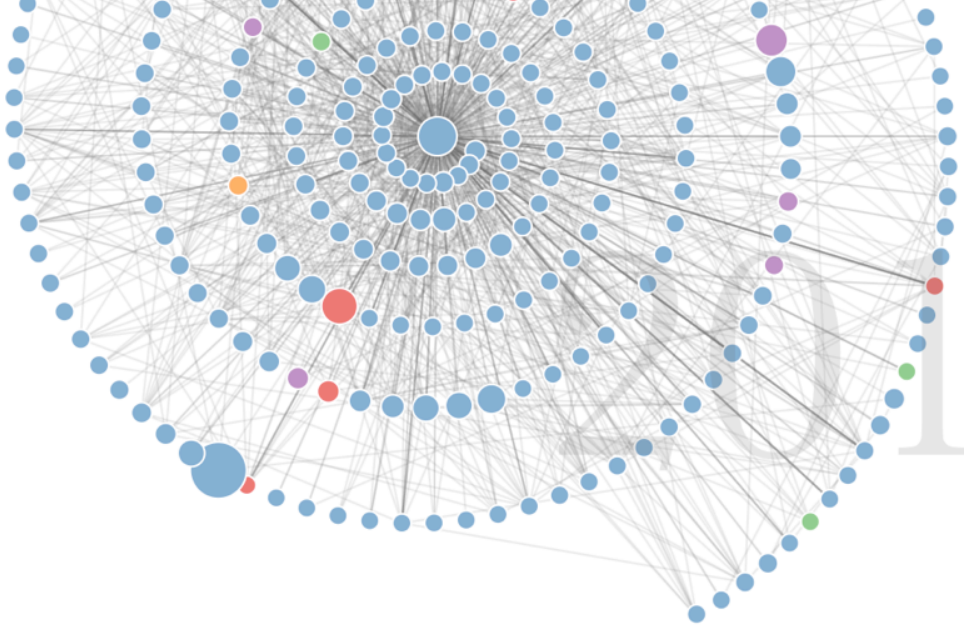


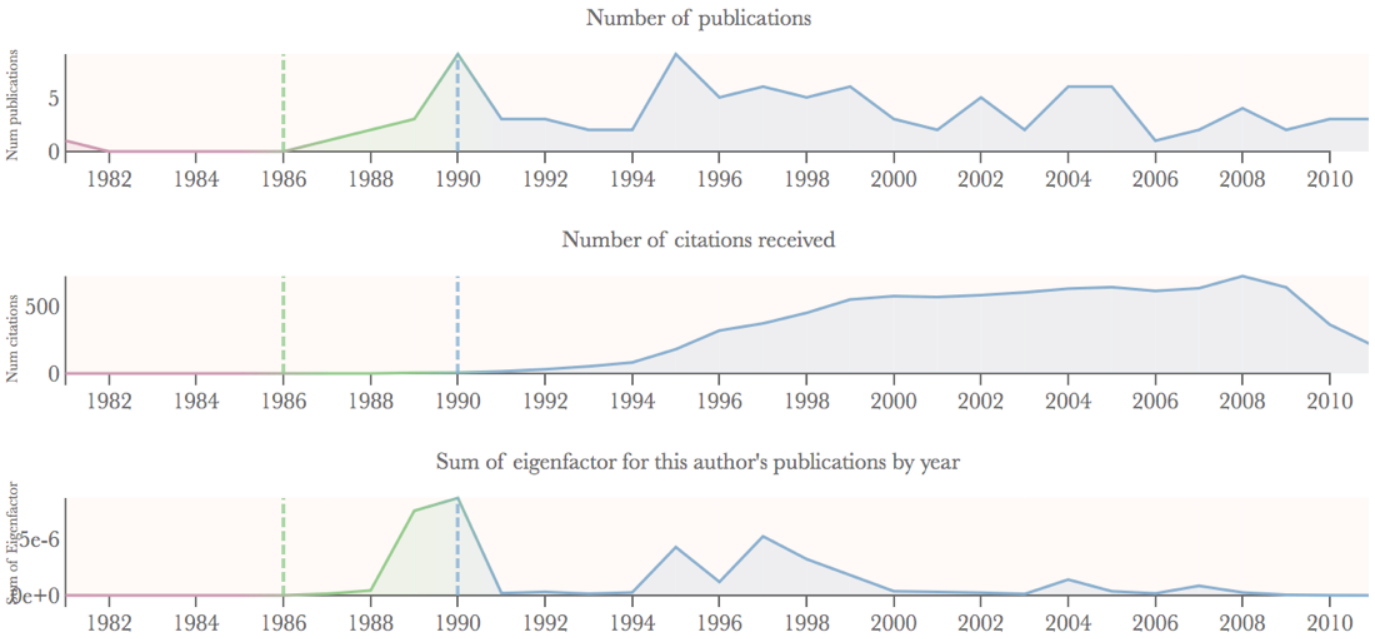




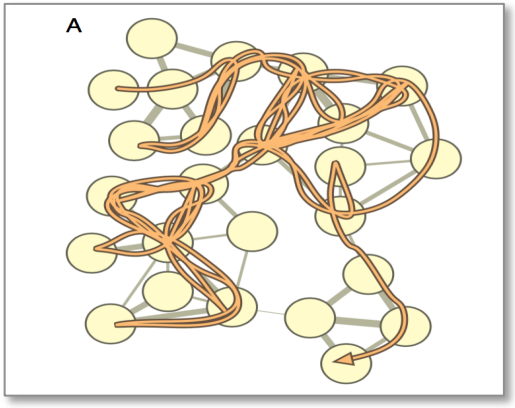
A project of the eScience Institue at the University of Washington

Jason Portenoy

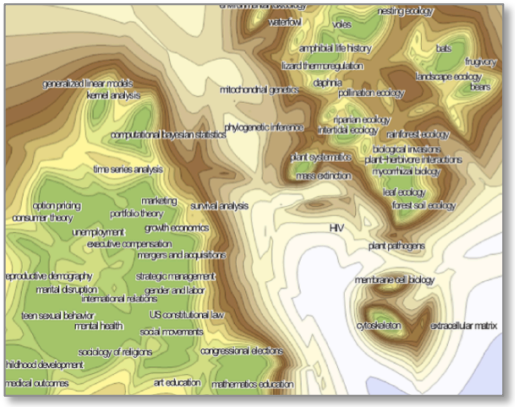




scholar.eigenfactor.org

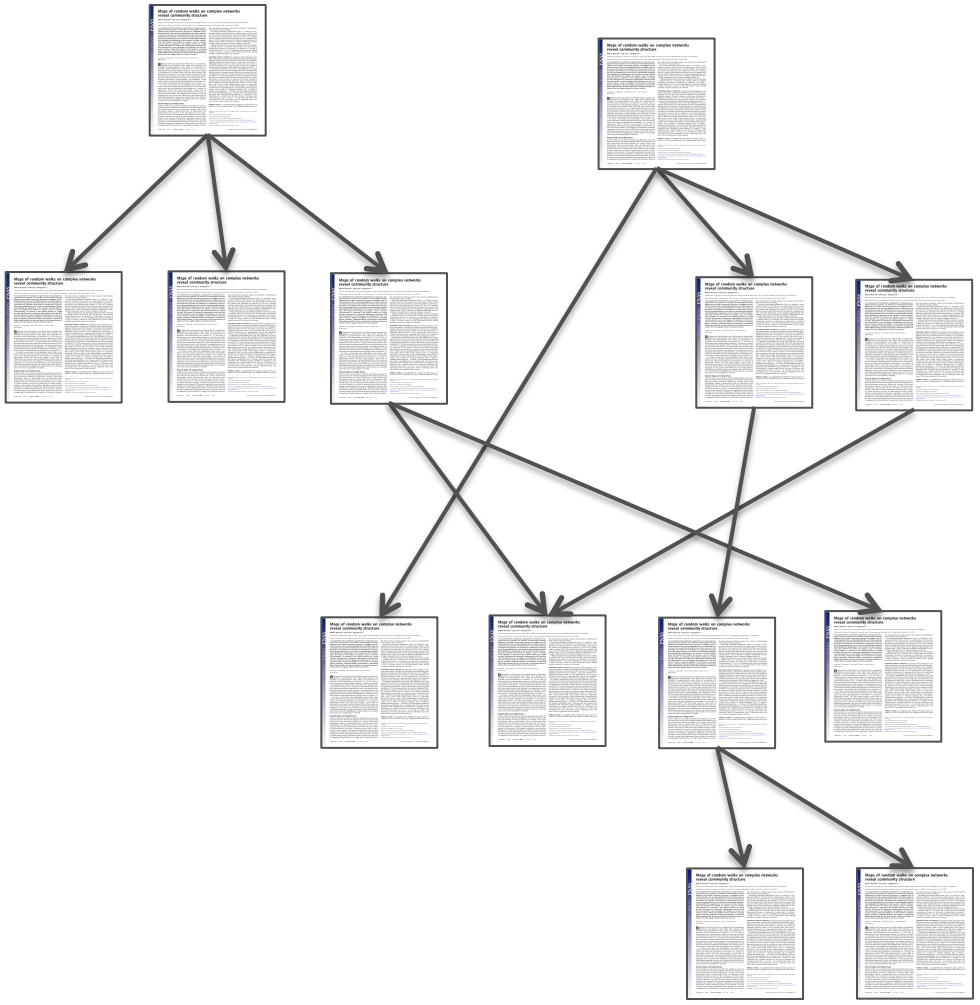


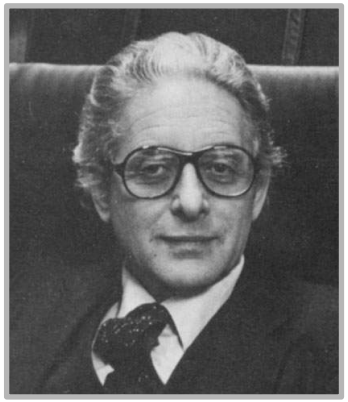
Science of Mapping



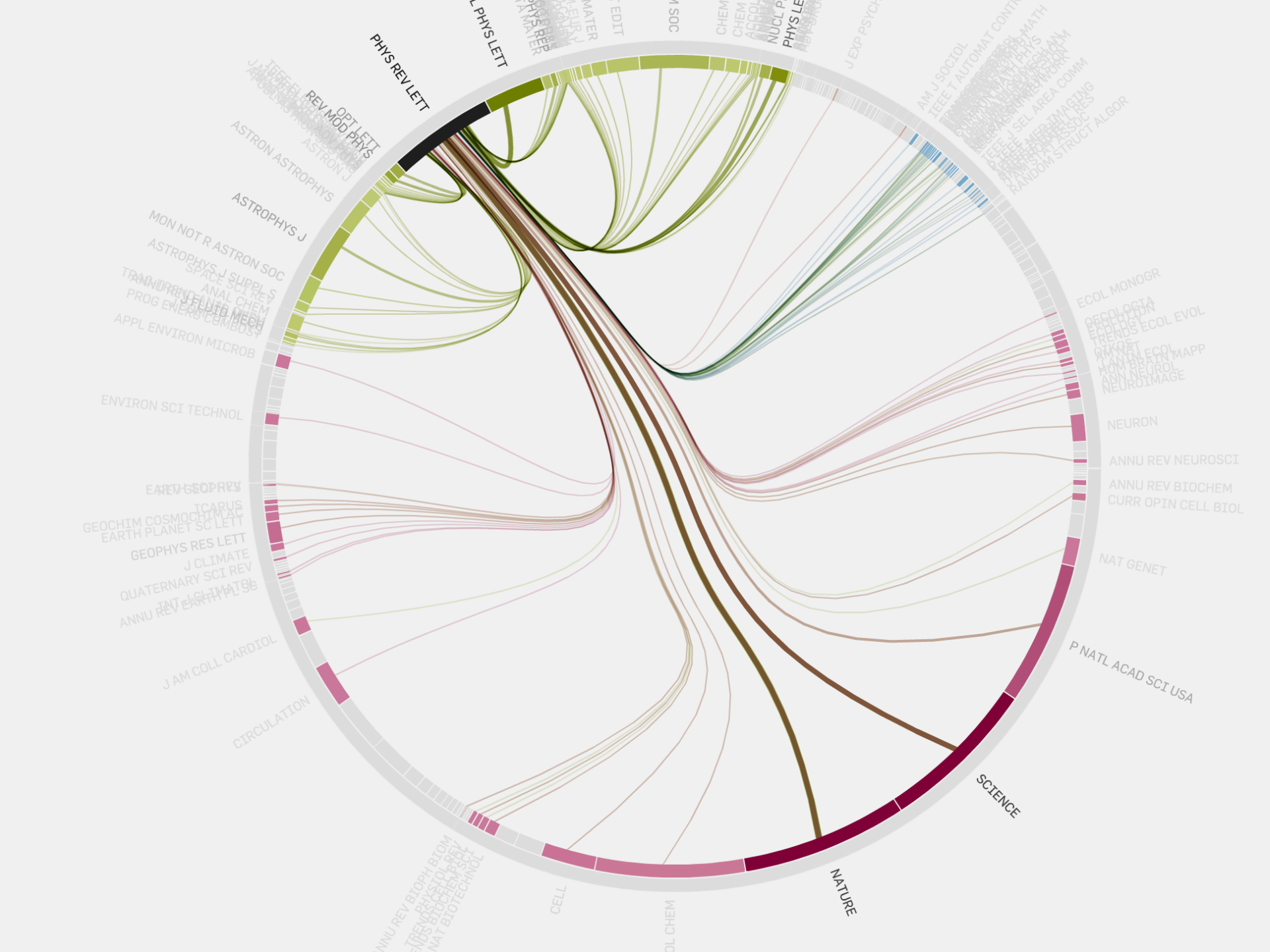
Mapping of Science

Citations form a vast network

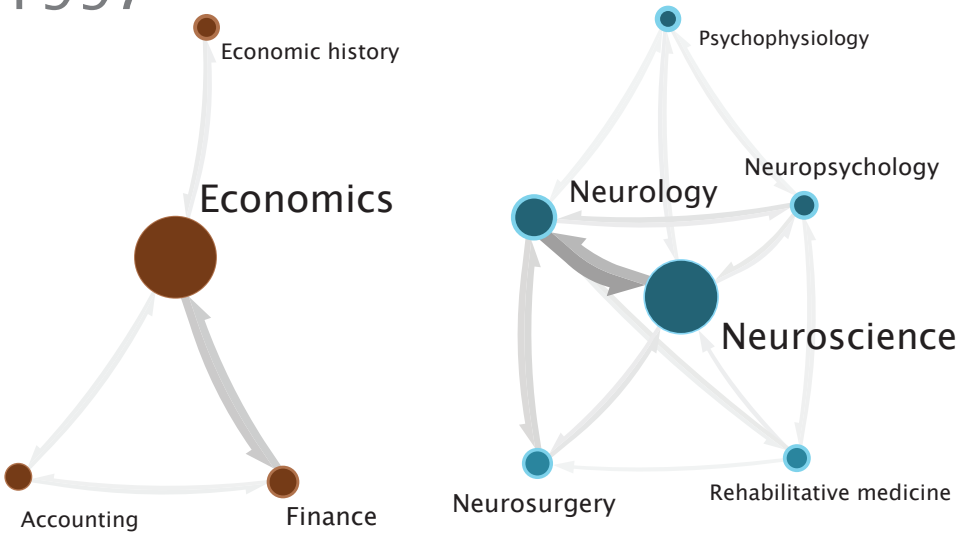




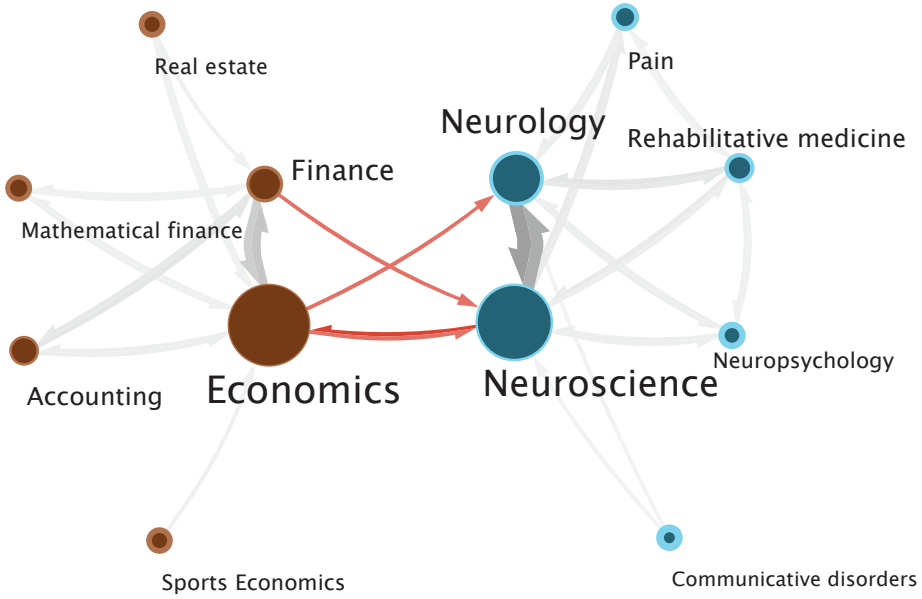
de Solla Price, *Science* (1965)


PHYS REV LETT
ASTROPHYS J
NATURE
SCIENCE
P NATL ACAD SCI USA
NAT GENET
CURR OPIN CELL BIOL
ANNU REV BIOCHEM
ANNU REV NEUROSCI
NEURON
NEUROIMAGE
ECOL MONOGR
ENVIRON SCI TECHNOL
GEOPHYS RES LETT
J CLIMATE
J AM COLL CARDIOL
CIRCULATION
CELL
ASTRON ASTROPHYS
REV MOD PHYS
MON NOT R ASTRON SOC
APPL ENVIRON MICROB
ICARUS

1997



2010



NEUROECONOMICS

Recent common ancestor

Of the 14 million pairs of papers, 7.1M of them have a common ancestor—a hit rate of about 50%. Among these 7.1M LCAs, here are the ten most frequent papers (and their frequencies):

1. (47,129) *Some Methods for Strengthening the Common $\chi 2$ Tests* (Cochran, 1954)
2. (35,585) *The Evolution of Reciprocal Altruism* (Trivers, 1971)
3. (34,195) *On the Mathematical Foundations of Theoretical Statistics* (Fisher, 1922)
4. (34,093) *The Tragedy of the Commons* (Hardin, 1968)
5. (32,067) *Some Difficulties of the Determination Problem* (Harrison, 1933)
6. (29,458) *Diverse Doctrines of Evolution, Their Relation to the Practice of Science and of Life* (Jennings, 1927)
7. (28,149) *An Analysis of Transformations* (Box, 1964)
8. (26,000) *Fitting the Negative Binomial Distribution to Biological Data* (Bliss, 1953)
9. (25,410) *A Method for Cluster Analysis* (Edwards, 1965)
10. (24,611) *A Theory of the Allocation of Time* (Becker, 1965)

http://blog.halper.in/posts/2014/10/09/moore-sloan-fun-with-academic-lineage/

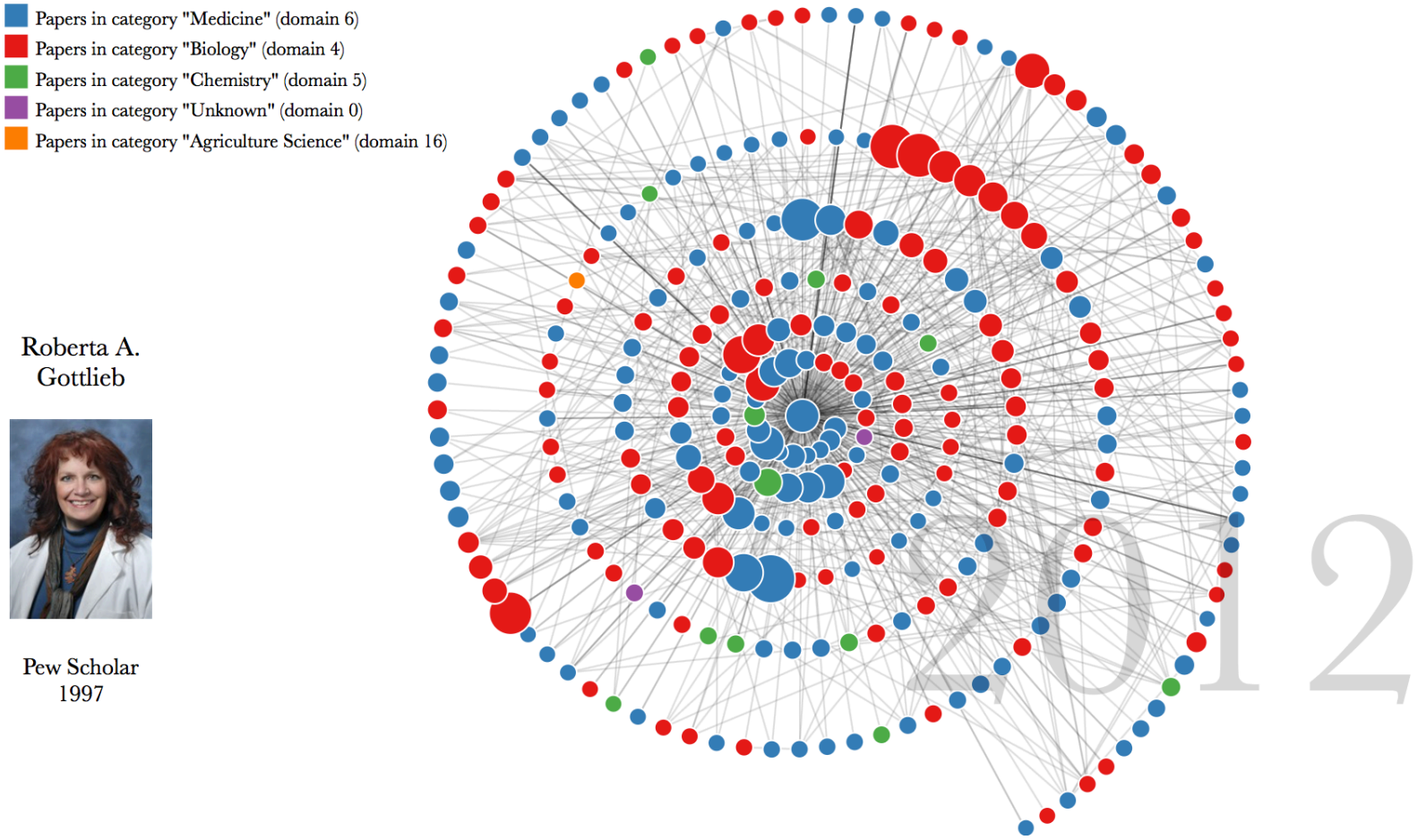
scholar.eigenfactor.org

Visualizing Scholarly Influence Over Time

Influence of Pew Scholars

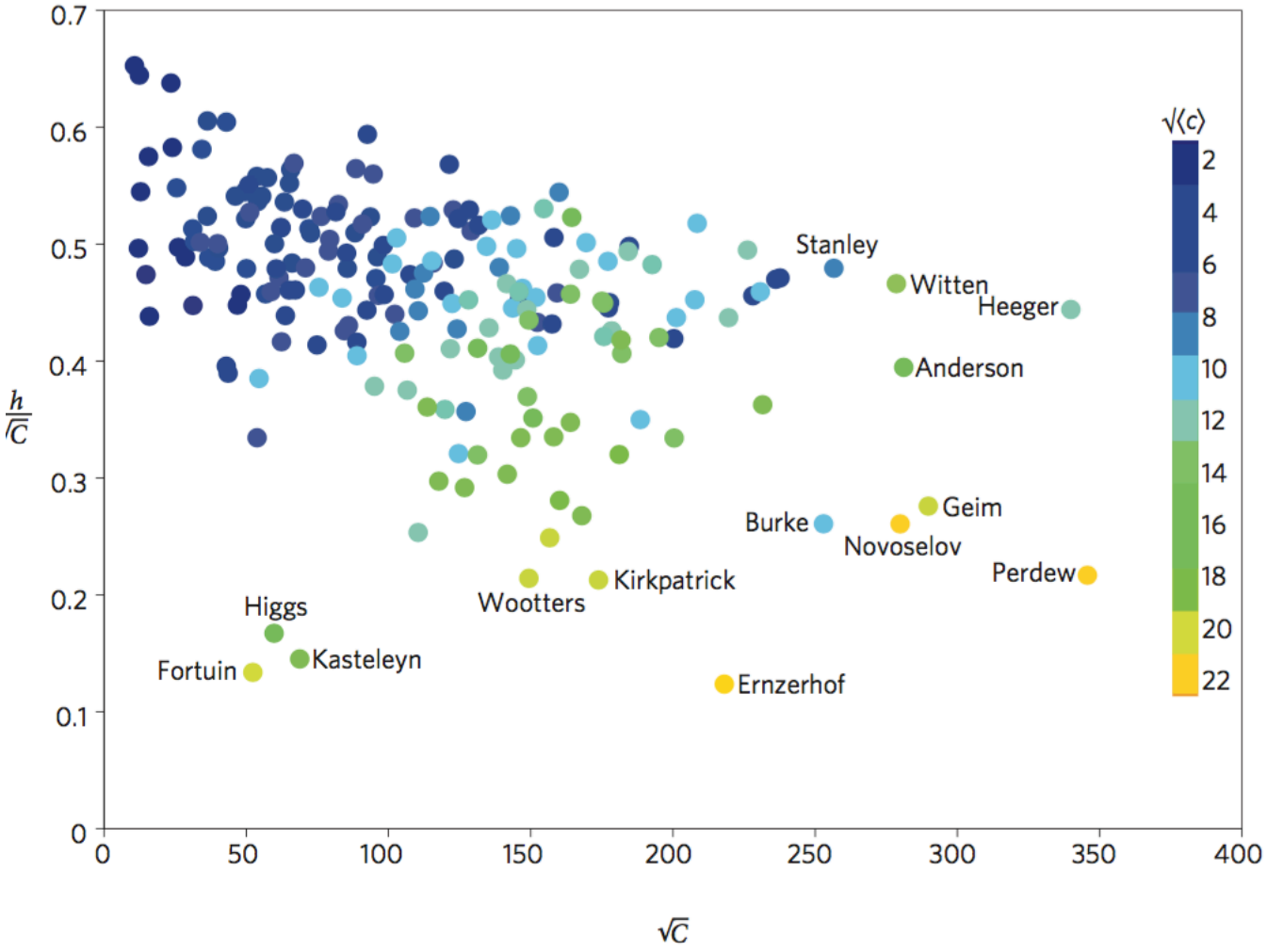
Roberta A. Gottlieb

Learn More



Portenoy et al. (2016) Leveraging Citation Networks to Visualize Scholarly Influence Over Time. *in review*

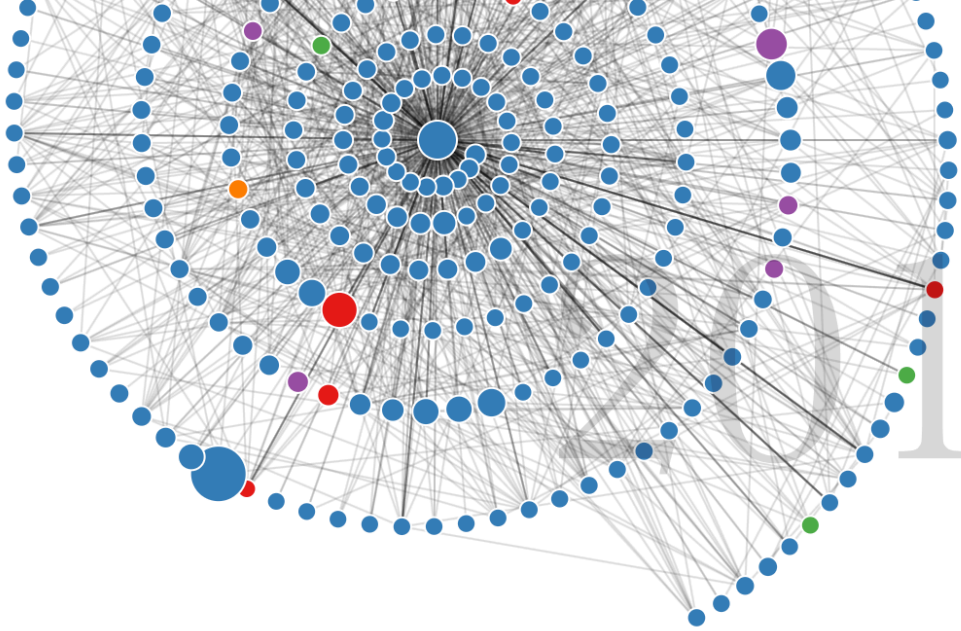
An evisceration of the H-index…



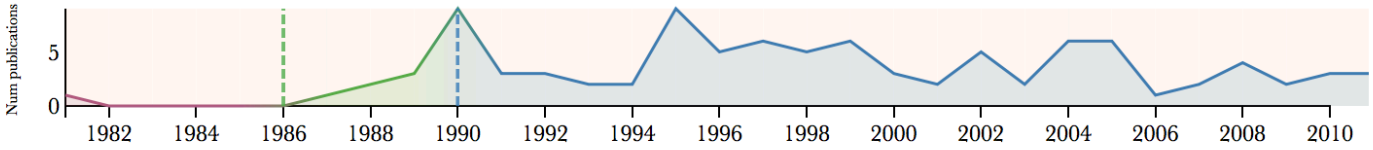
S. N. Dorogovtsev and J. F. F. Mendes (2015) Nature Physics

Philip A.
Hieter

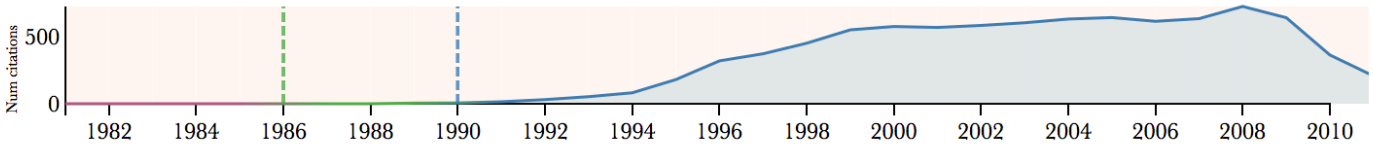
Pew Scholar
1986



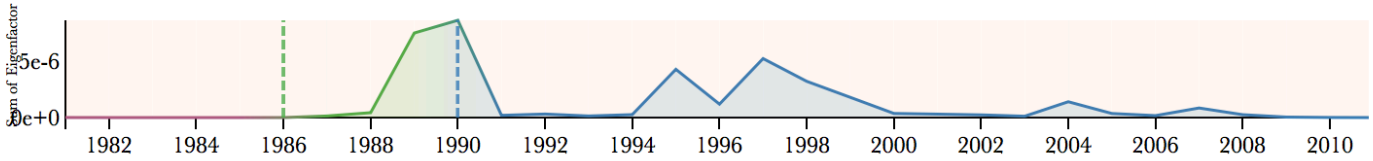
Number of publications



Number of citations received



Sum of eigenfactor for this author's publications by year



Visualizing Scholarly Influence Over Time

Influence of Pew Scholars

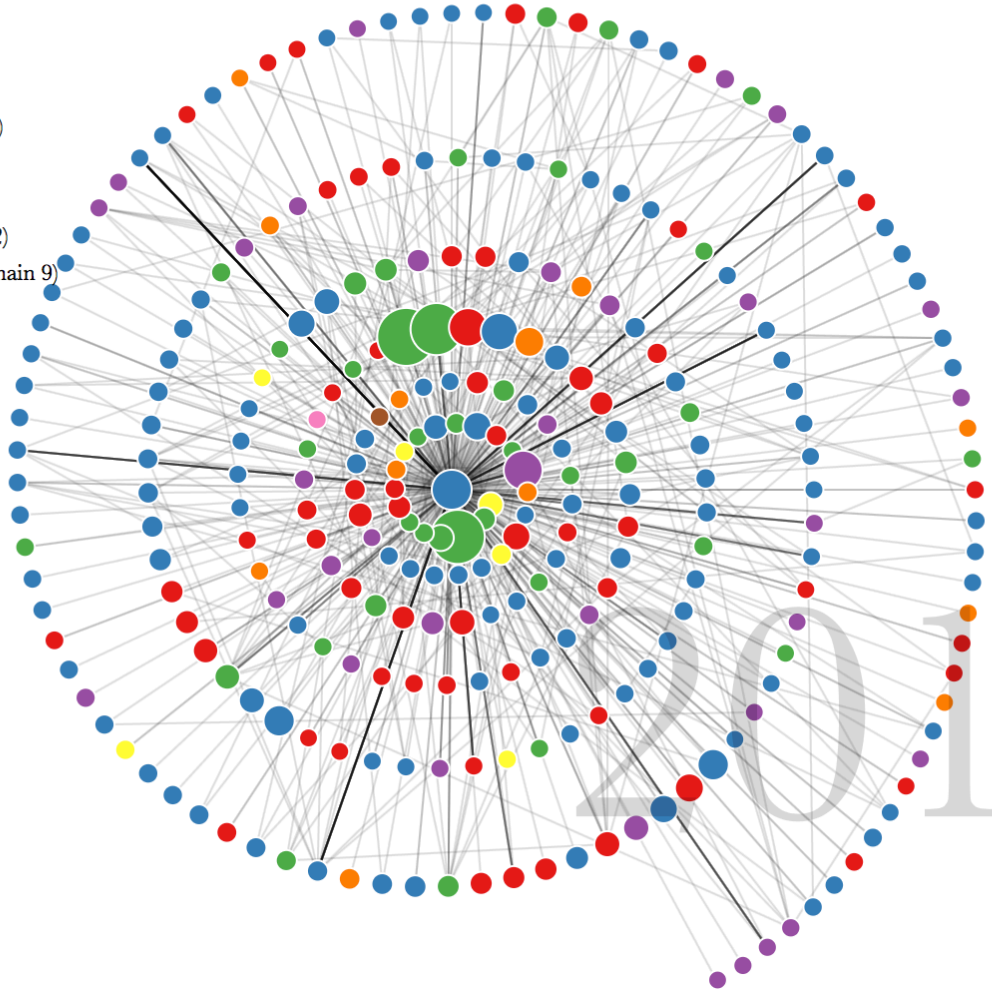
Mark W. Grinstaff

Learn More

- Papers in category "Chemistry" (domain 5)
- Papers in category "Medicine" (domain 6)
- Papers in category "Biology" (domain 4)
- Papers in category "Material Science" (domain 12)
- Papers in category "Engineering" (domain 8)
- Papers in category "Physics" (domain 19)
- Papers in category "Computer Science" (domain 2)
- Papers in category "Environmental Sciences" (domain 9)

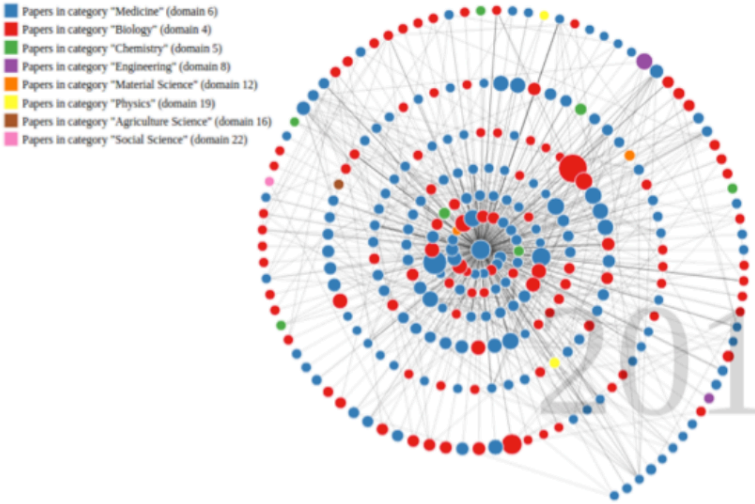
Mark W.
Grinstaff

Pew Scholar
1999



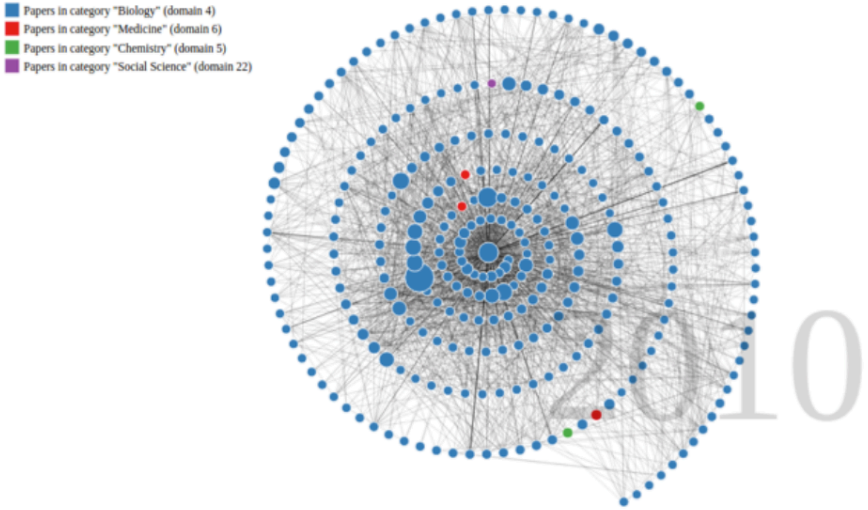
2012

Comparing Authors



A more sparse network indicates fewer citations between papers shown in the network. This could be a result of the central scholar having impact across a wider set of academic communities.

A denser network means that the papers that cite the central author also tend to cite each other.



The map equation

frequency of inter-module movements

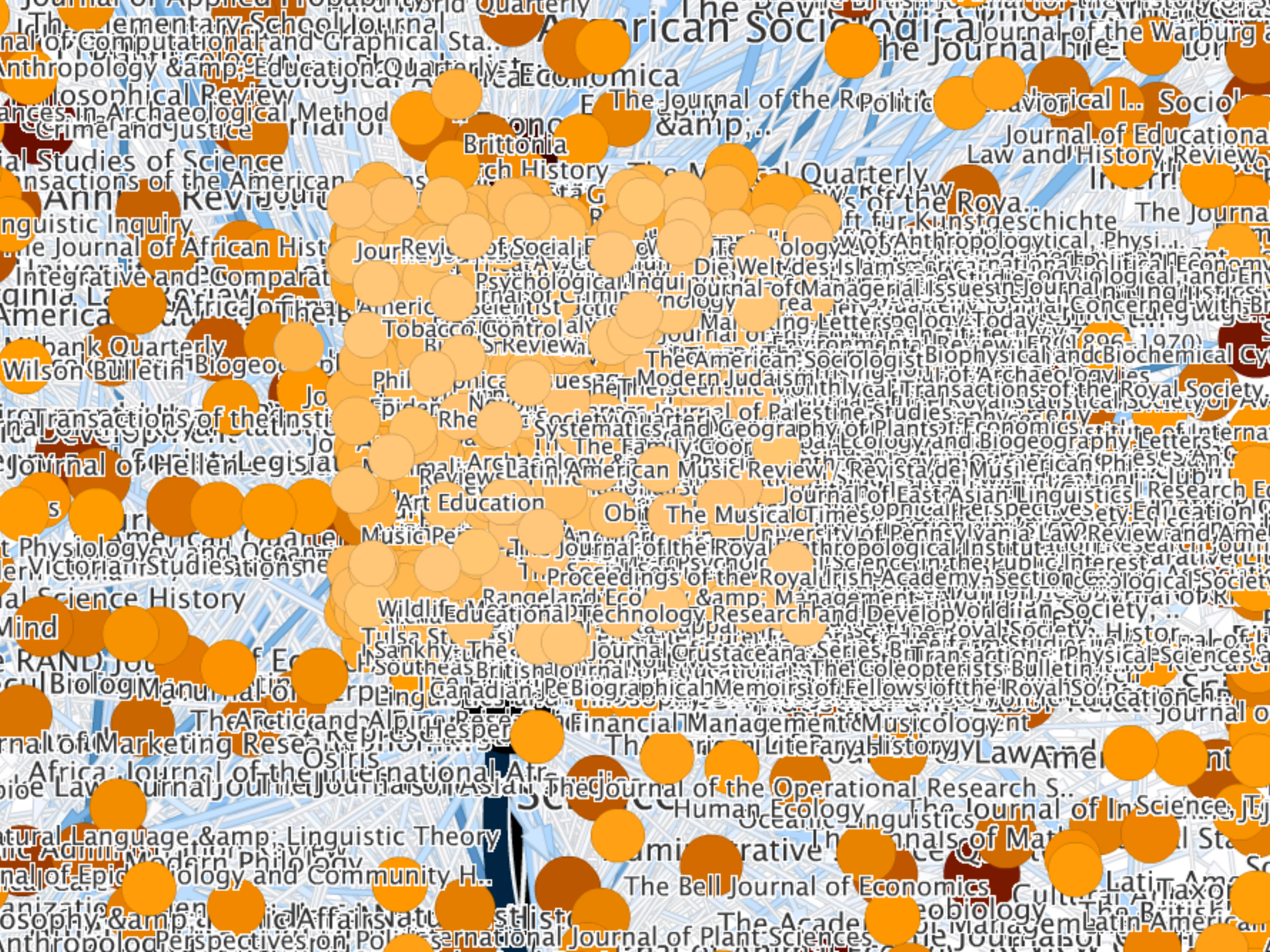
frequency of movements within module *i*

$$L(\mathbf{M}) = q_{\curvearrowright} H(\mathcal{Q}) + \sum_{i=1}^{m} p^{i}_{\circlearrowright} H(\mathcal{P}^{i})$$

code length of module names

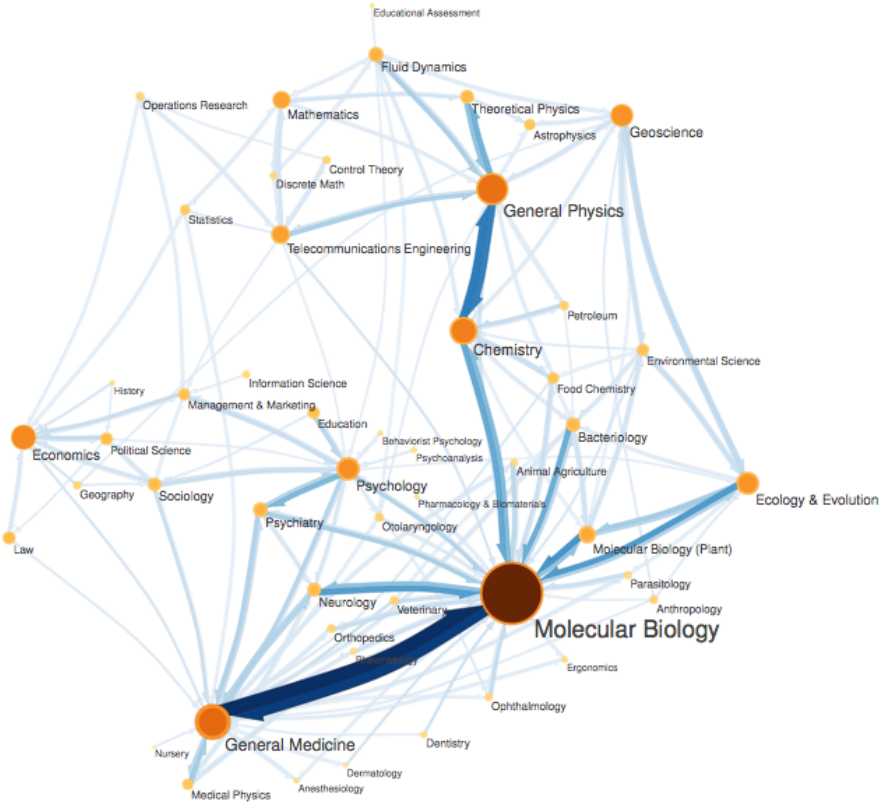
code length of node names in module *i*

Rosvall and Bergstrom (2008) *PNAS*

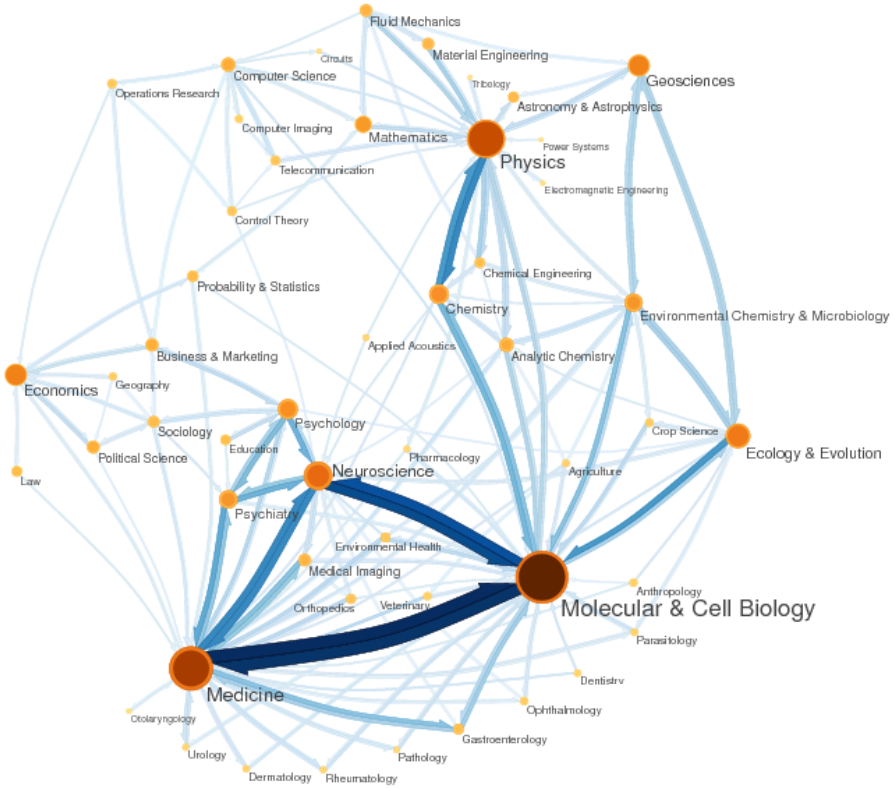


The Elementary School Journal
Ecological
Economica
Philosophical Review
Crime and Justice
Brittonia
Law and History Review
Linguistic Inquiry
Integrative and Comparative
Die Welt des Islams
Psychological Inquiry
Journal of Managerial Issues
Tobacco Control
The American Sociologist
Biophysical and Biochemical
Wilson Bulletin
Modern Judaism
Transactions of the Institute
Journal of Palestine Studies
Systematics and Geography of Plants
The Family Coordinator
Latin American Music Review
Revista de Música
Journal of East Asian Linguistics
Art Education
The Musical Times
University of Pennsylvania Law Review
Victorian Studies
Proceedings of the Royal Irish Academy. Section C
Social Science History
Educational Technology Research and Development
Mind
Journal Crustaceana Series B
Transactions: Physical Sciences
The Coleopterists Bulletin
Biographical Memoirs of Fellows of the Royal Society
The Arctic and Alpine Research
Financial Management
Musicology
Osiris
The Journal of the Operational Research S..
Human Ecology
Natural Language & Linguistic Theory
Modern Philology
Journal of Epidemiology and Community H..
The Bell Journal of Economics
Perspectives on Politics
International Journal of Plant Sciences

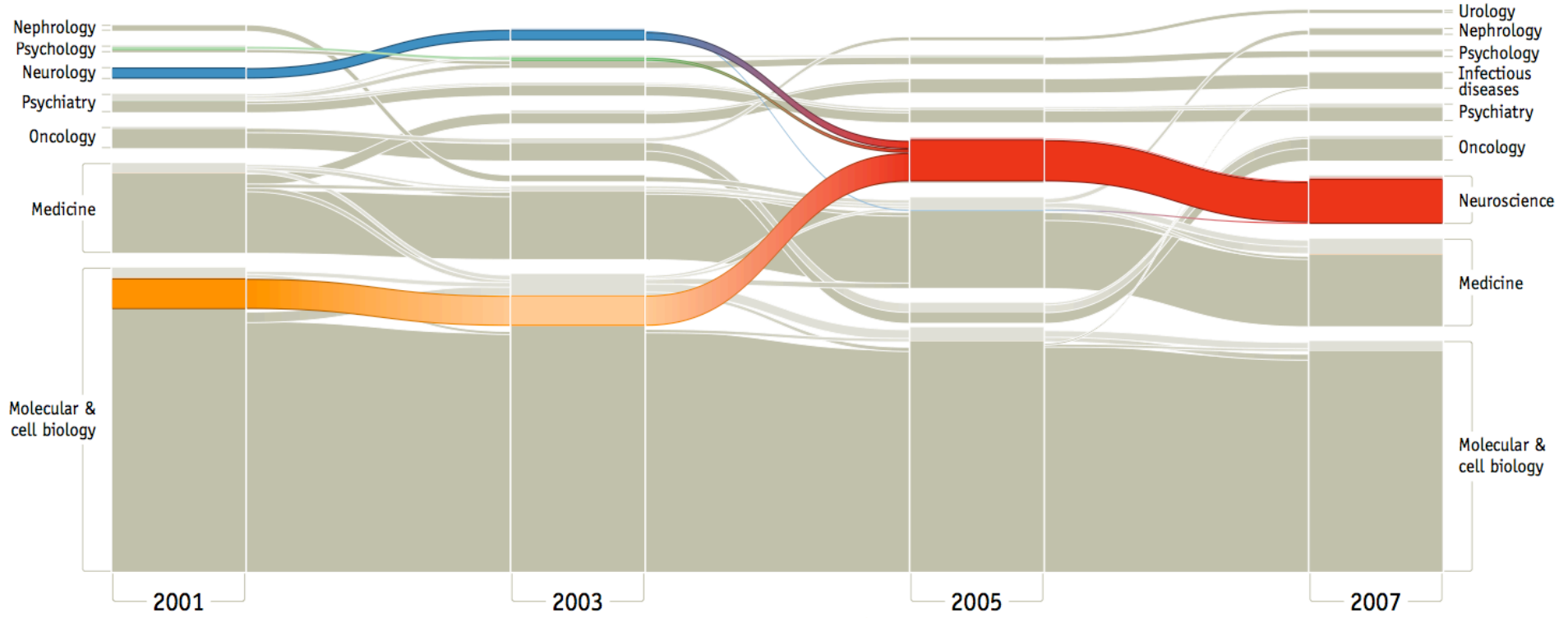
1995


Educational Assessment
Fluid Dynamics
Operations Research
Mathematics
Theoretical Physics
Astrophysics
Geoscience
Control Theory
Discrete Math
Statistics
General Physics
Telecommunications Engineering
Petroleum
Chemistry
Environmental Science
Information Science
History
Management & Marketing
Food Chemistry
Education
Bacteriology
Economics
Political Science
Behaviorist Psychology
Psychoanalysis
Psychology
Animal Agriculture
Geography
Sociology
Pharmacology & Biomaterials
Ecology & Evolution
Psychiatry
Otolaryngology
Law
Molecular Biology (Plant)
Parasitology
Neurology
Veterinary
Anthropology
Molecular Biology
Orthopedics
Ergonomics
Ophthalmology
General Medicine
Nursery
Dentistry
Dermatology
Medical Physics
Anesthesiology

2004

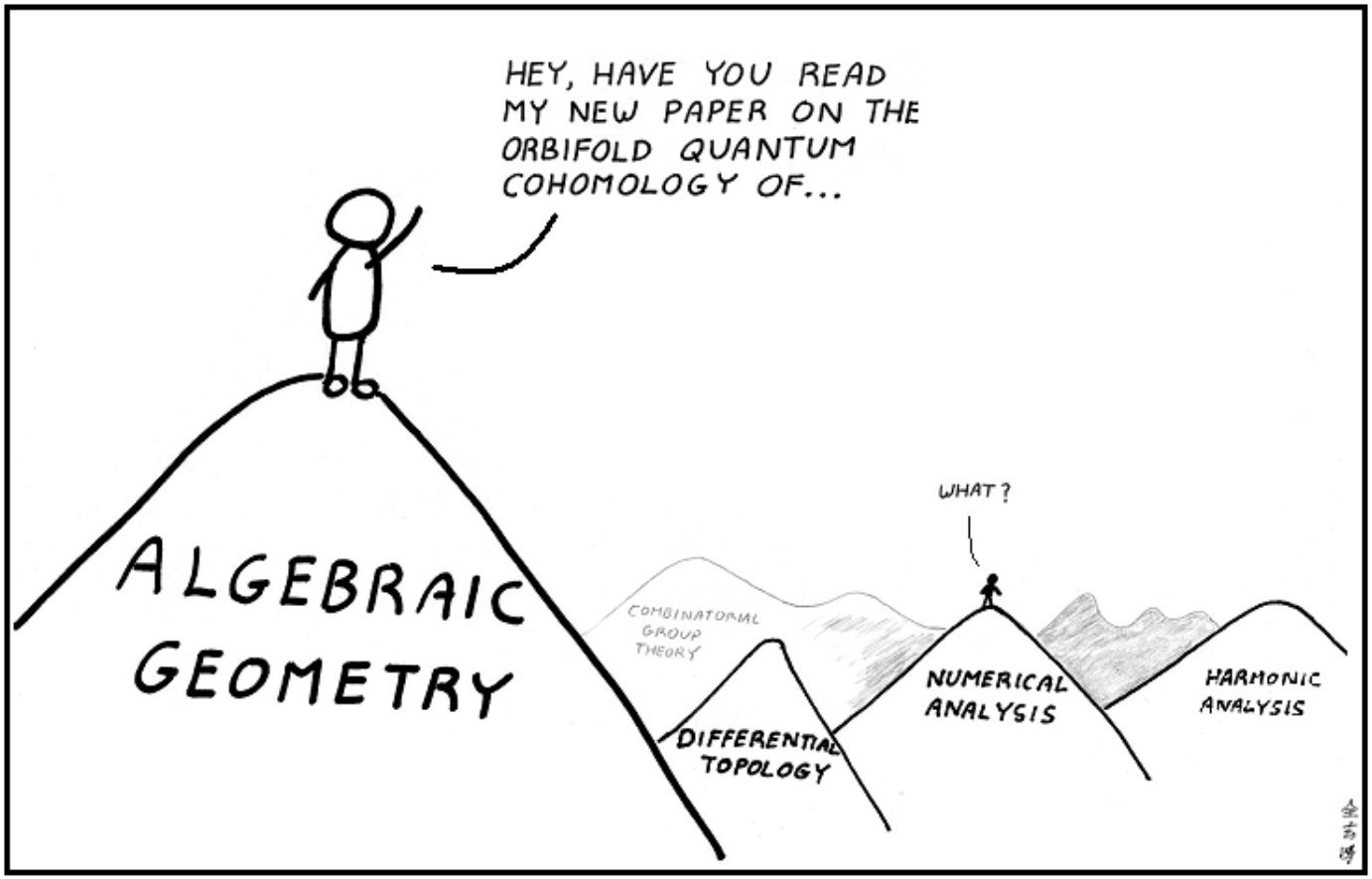

Fluid Mechanics
Material Engineering
Circuits
Computer Science
Geosciences
Tribology
Operations Research
Astronomy & Astrophysics
Computer Imaging
Mathematics
Power Systems
Physics
Telecommunication
Electromagnetic Engineering
Control Theory
Chemical Engineering
Probability & Statistics
Chemistry
Environmental Chemistry & Microbiology
Applied Acoustics
Analytic Chemistry
Business & Marketing
Economics
Geography
Psychology
Sociology
Crop Science
Education
Political Science
Pharmacology
Ecology & Evolution
Neuroscience
Agriculture
Law
Psychiatry
Environmental Health
Medical Imaging
Anthropology
Orthopedics
Veterinary
Molecular & Cell Biology
Parasitology
Dentistry
Medicine
Ophthalmology
Otolaryngology
Gastroenterology
Urology
Pathology
Dermatology
Rheumatology

The Emergence of Neuroscience

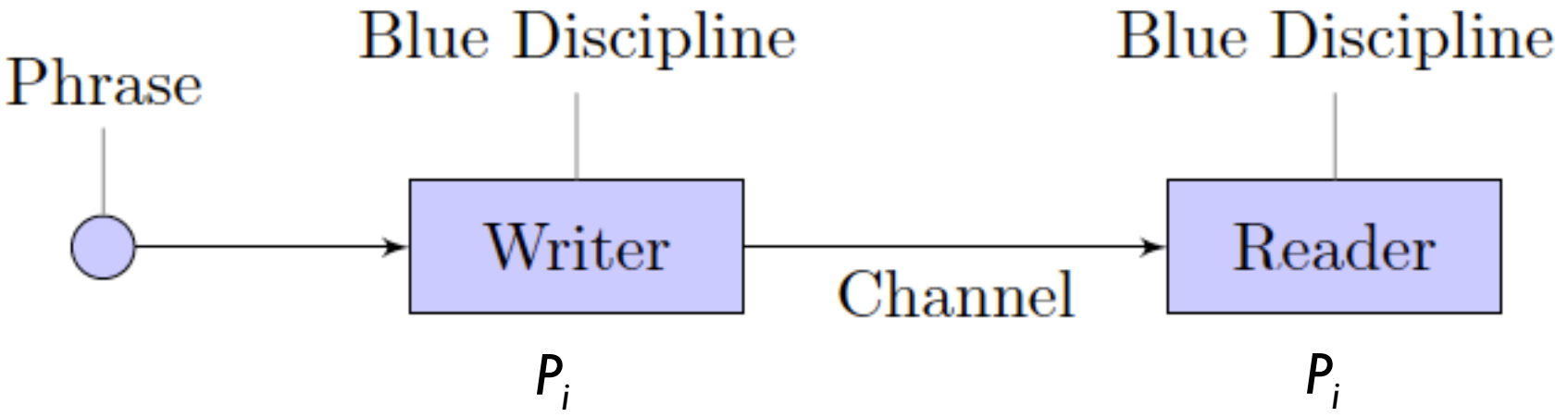


Rosvall and Bergstrom (2010) PLoS One

The jargon barriers of science



The Landscape of Modern Mathematics



X ~ space of all phrases

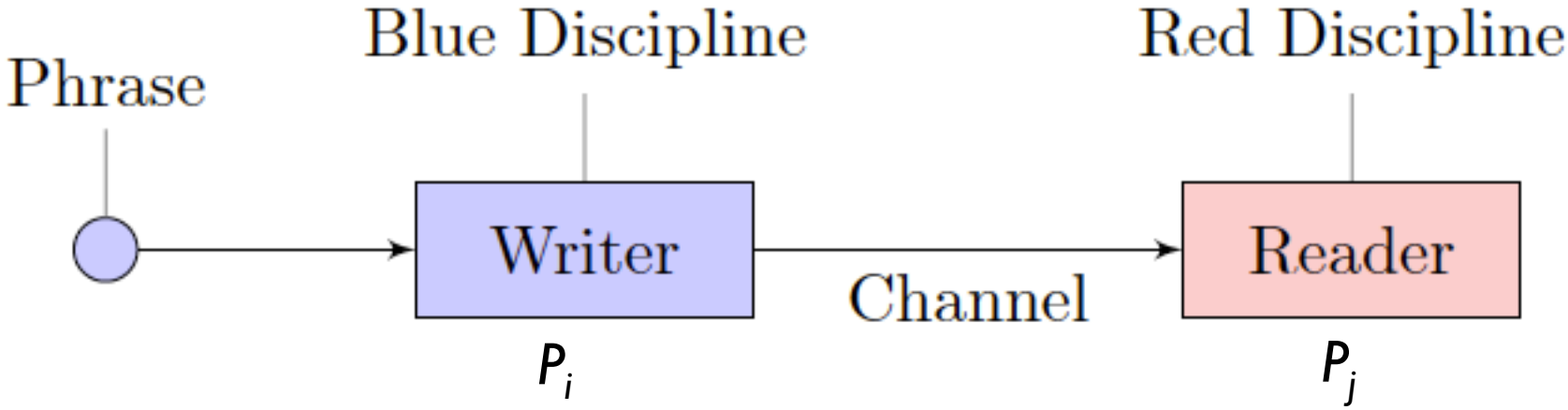
P_i ~ probability distribution over χ_i with values $x \in$ ***X***

- writer chooses phrases with probability $p_i(x)$
- optimal codeword has length $-\log_2 p_i(x)$

expected message length $H(X_i) = -\sum_{x \in \mathcal{X}} p_i(x) \log_2 p_i(x)$

assumption: language of each scientific field is *optimized* based on frequency of phrases

Vilhena (2014) Sociological Science



cross entropy

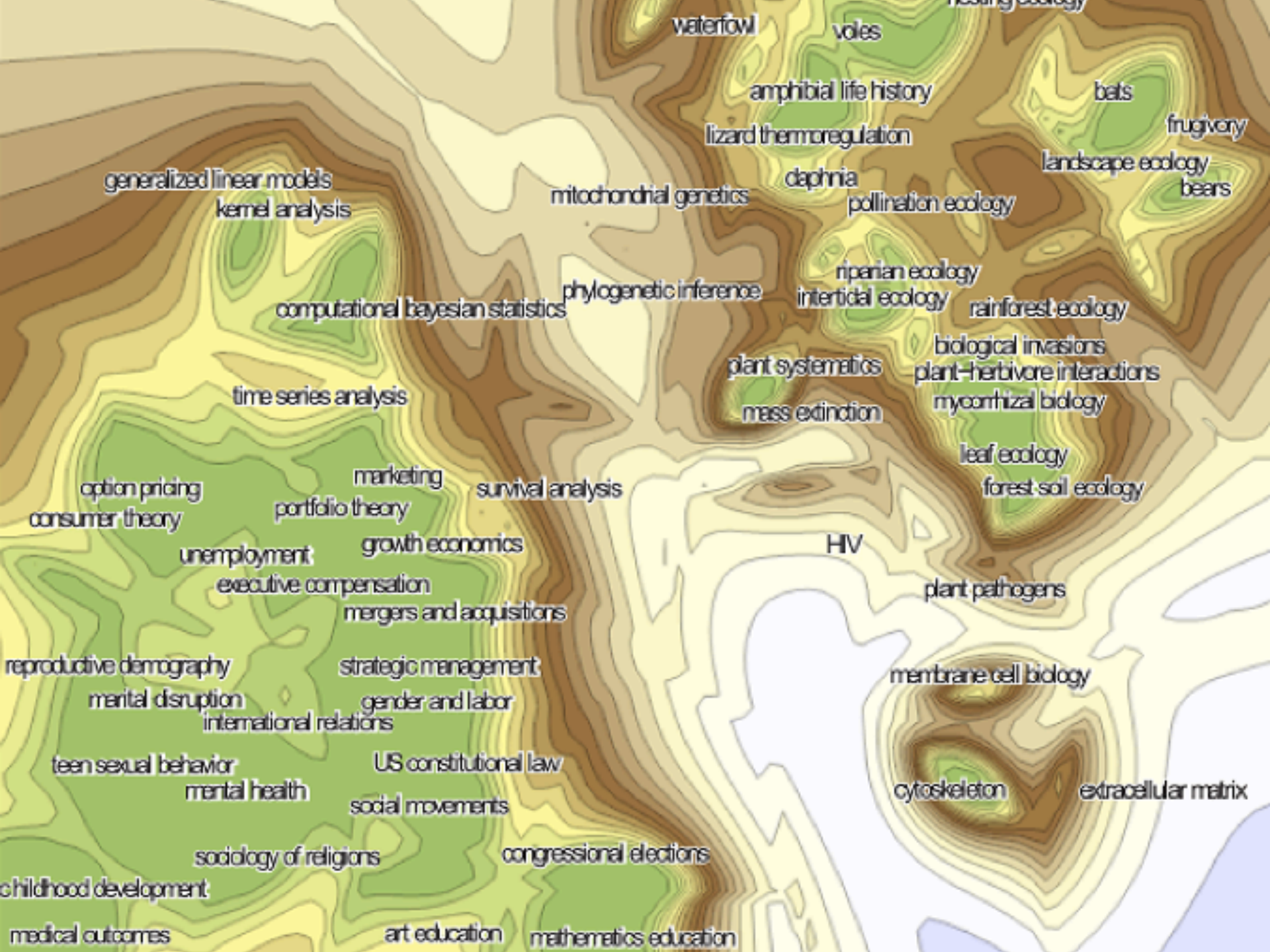
expected message length: $Q(p_i||p_j) = -\sum_{x \in \mathcal{X}} p_i(x) \log_2 p_j(x)$

efficiency of communication

$$E_{ij} = \frac{H(X_i)}{Q(p_i||p_j)} = \frac{-\sum_{x \in \mathcal{X}} p_i(x) \log_2 p_i(x)}{-\sum_{x \in \mathcal{X}} p_i(x) \log_2 p_j(x)}$$

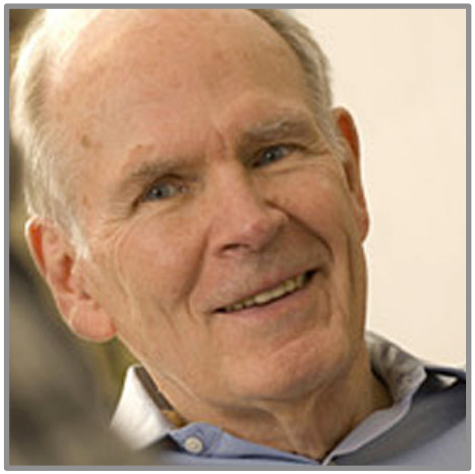
$$C_{ij} = 1 - E_{ij}$$

cultural hole


waterfowl
voles
amphibial life history
bats
frugivory
lizard thermoregulation
landscape ecology
generalized linear models
daphnia
bears
kernel analysis
mitochondrial genetics
pollination ecology
riparian ecology
phylogenetic inference
intertidal ecology
computational bayesian statistics
rainforest ecology
biological invasions
plant systematics
plant–herbivore interactions
time series analysis
mycorrhizal biology
mass extinction
leaf ecology
marketing
option pricing
survival analysis
forest soil ecology
portfolio theory
consumer theory
growth economics
HIV
unemployment
executive compensation
plant pathogens
mergers and acquisitions
reproductive demography
strategic management
membrane cell biology
marital disruption
gender and labor
international relations
teen sexual behavior
US constitutional law
cytoskeleton
extracellular matrix
mental health
social movements
sociology of religions
congressional elections
childhood development
medical outcomes
art education
mathematics education

Translational Lag

Expression of Polypeptides in Yeast



A Life History of Innovation

inventors

Nature, PNAS, Cell → 1980

UW licenses to WRF →

lawyers 1987

Patent Application →

examiners

Patent Granted → 1995

managers

$240 million in revenue →

$31 million WRF -> UW → 2014

Washington Research FOUNDATION

What is the time from  to patent?

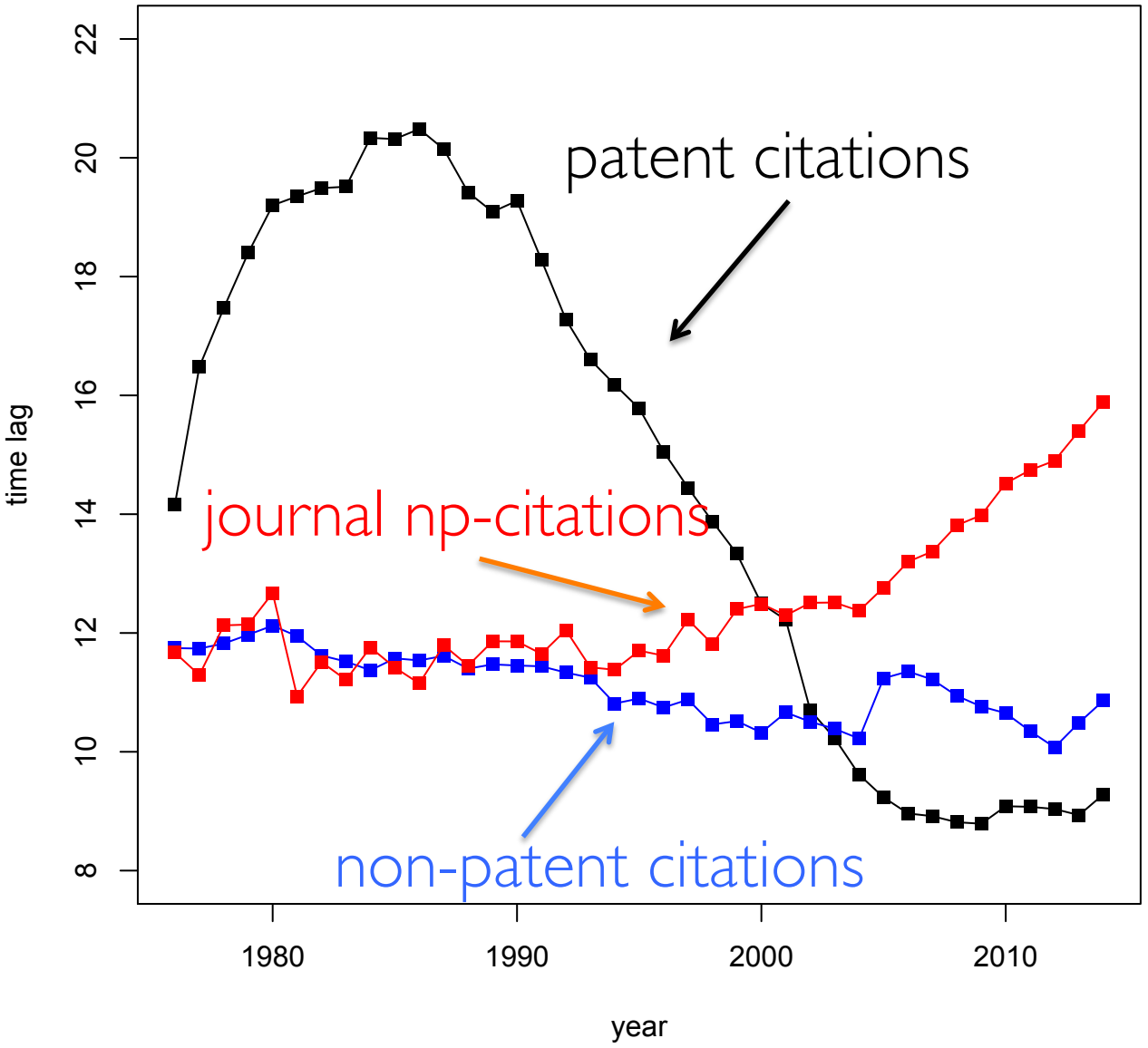
Methods

- Extracted 111,038,761 citations (1976 – 2014)
  - 92,680,292 patent citations
  - 18,358,469 non-patent citations
  - 761,759 non-patent, journal citations
  - 20,470,405 examiner citations (2001 – 2014)
- Extracted years from patents and non-patents

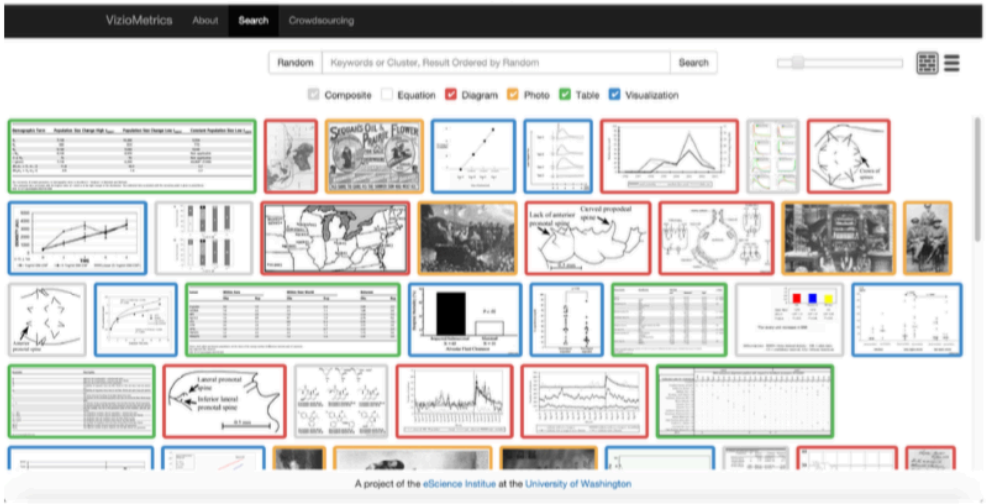
  Svensson, K. et al. Evolution of subspecies of . Journal of Bacteriology, Jun. 2005, vol. 187, No. 11, pp.3903-3908

- Only use first, unique instances of citing patent

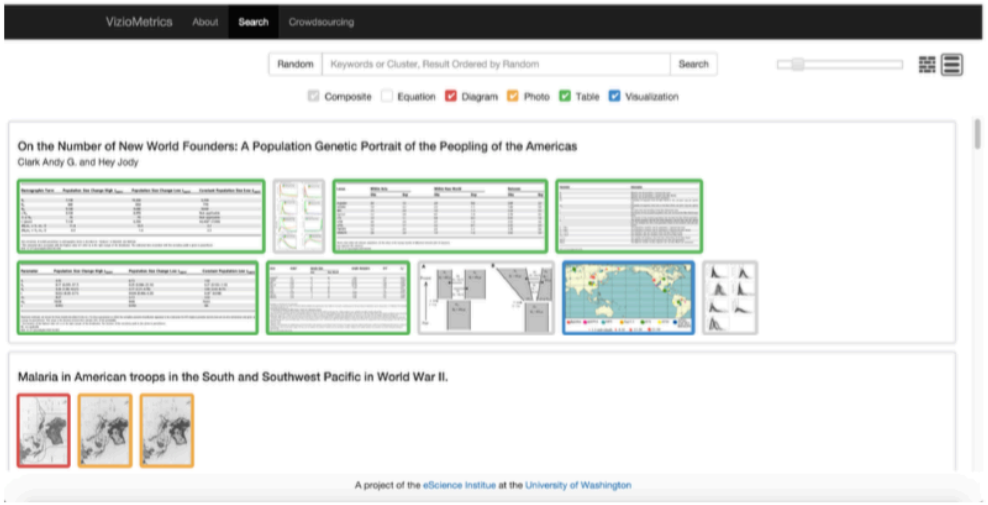
Are patents becoming more myopic?



viziometrics.org

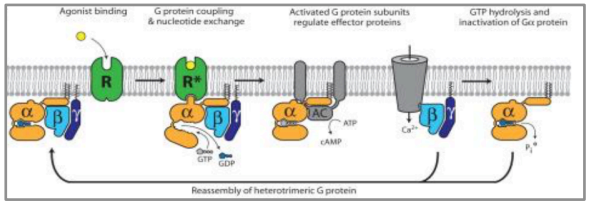


(b)

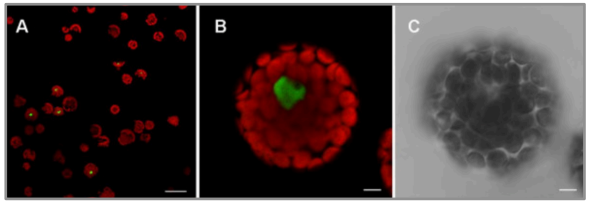


$$w_i = \sum_j^n (Z_{ij} + Z_{ij}^T)$$

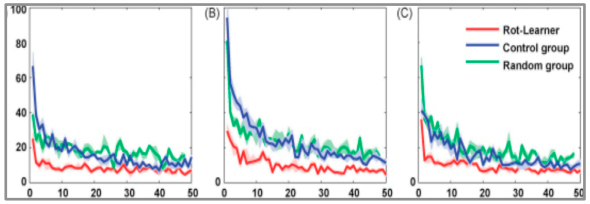
Equations (394)



Schematics (769)



Photos (782)

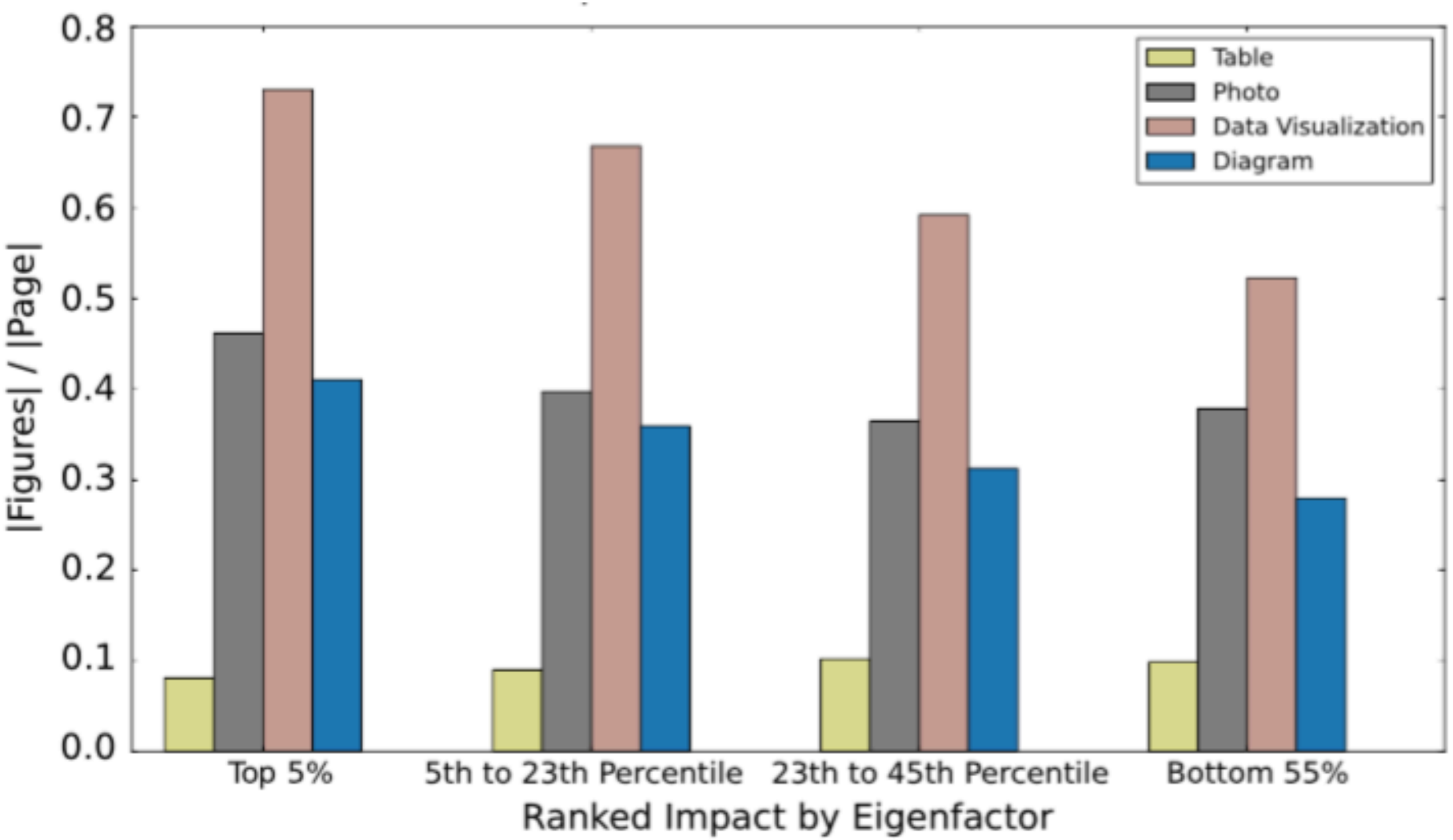


Plots (890)

	PW reading	PW reading	PW reading	W reading	W reading
	W RT	*PW RT*	*CTL*	*W RT*	*PW RT*
MOG → LOT	0.28	0.18	***0.58***	***−0.70***	***−0.50***
MOG → LP	−0.22	***−0.52***	−0.04	0.27	−0.03
LOT → LP	0	0.10	0.24	***−0.56***	***−0.60***
LOT → IFG	0.38	0.17	0.40	0.43	0.13
LP → IFG	0.26	0.05	0.31	0.03	−0.03

Tables (436)

Impact versus Figure Density



Summary

- Assembled scientific knowledge into a machine readable form
- Interrogating science from the meta-view
- Trying to better understand the origins of scientific ideas
- Building tools that facilitate discovery at the interdisciplinary boundaries of science

The Scholarly Graph

What else should we ask of this data?

Acknowledgements

Carl Bergstrom, Department of Biology, University of Washington

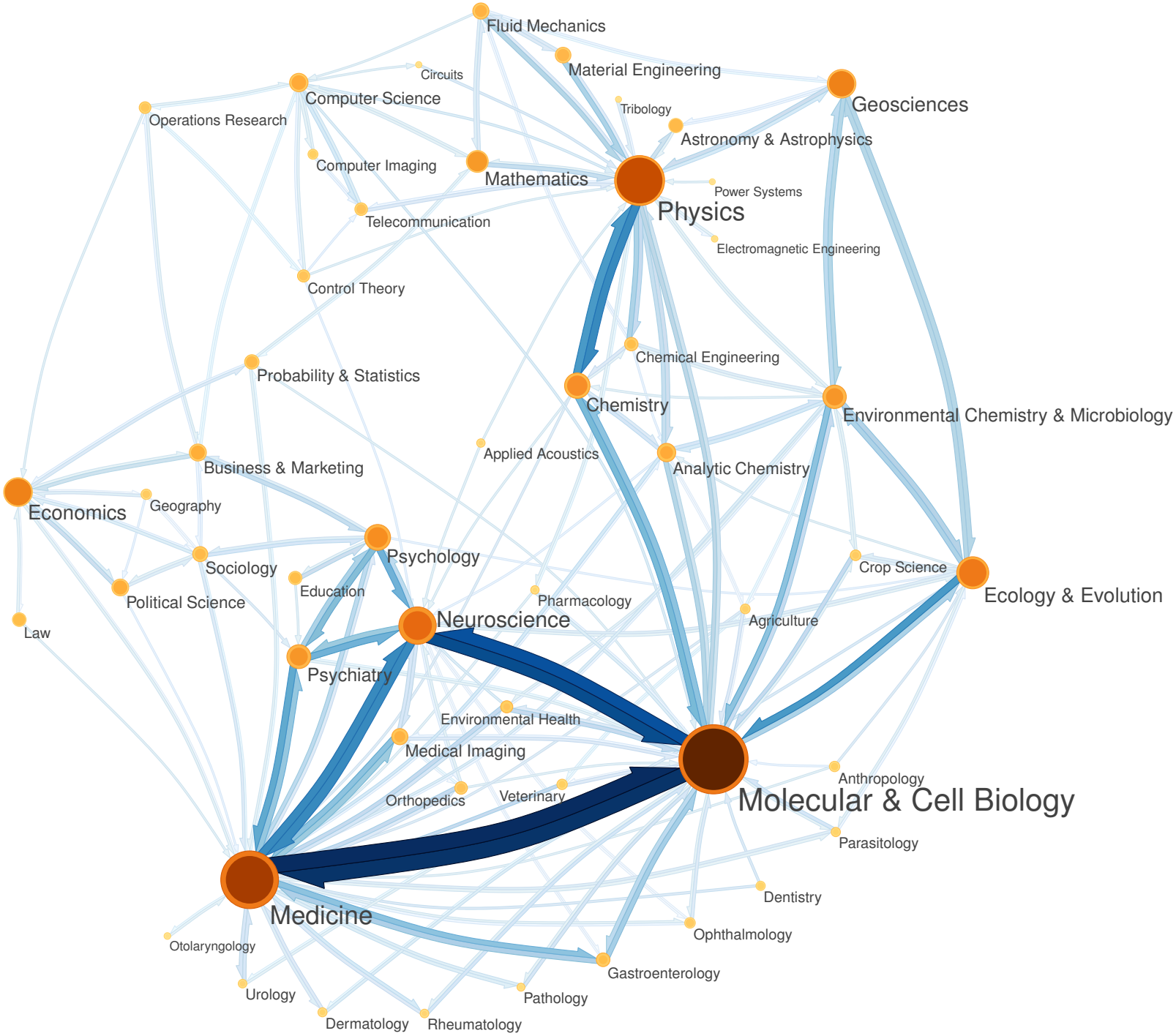
Martin Rosvall, Department of Physics, Umea University

Ian Wesley-Smith, Information School, University of Washington

Jason Portenoy, Information School, University of Washington

Bill Howe, eScience, CSE, University of Washington

Poshen Lee, CSE, University of Washington


Fluid Mechanics
Material Engineering
Circuits
Computer Science
Geosciences
Operations Research
Tribology
Astronomy & Astrophysics
Computer Imaging
Mathematics
Power Systems
Physics
Telecommunication
Electromagnetic Engineering
Control Theory
Chemical Engineering
Probability & Statistics
Chemistry
Environmental Chemistry & Microbiology
Applied Acoustics
Business & Marketing
Analytic Chemistry
Economics
Geography
Psychology
Crop Science
Sociology
Education
Political Science
Pharmacology
Ecology & Evolution
Law
Neuroscience
Agriculture
Psychiatry
Environmental Health
Medical Imaging
Anthropology
Orthopedics
Veterinary
Molecular & Cell Biology
Parasitology
Dentistry
Medicine
Ophthalmology
Otolaryngology
Gastroenterology
Urology
Pathology
Dermatology
Rheumatology