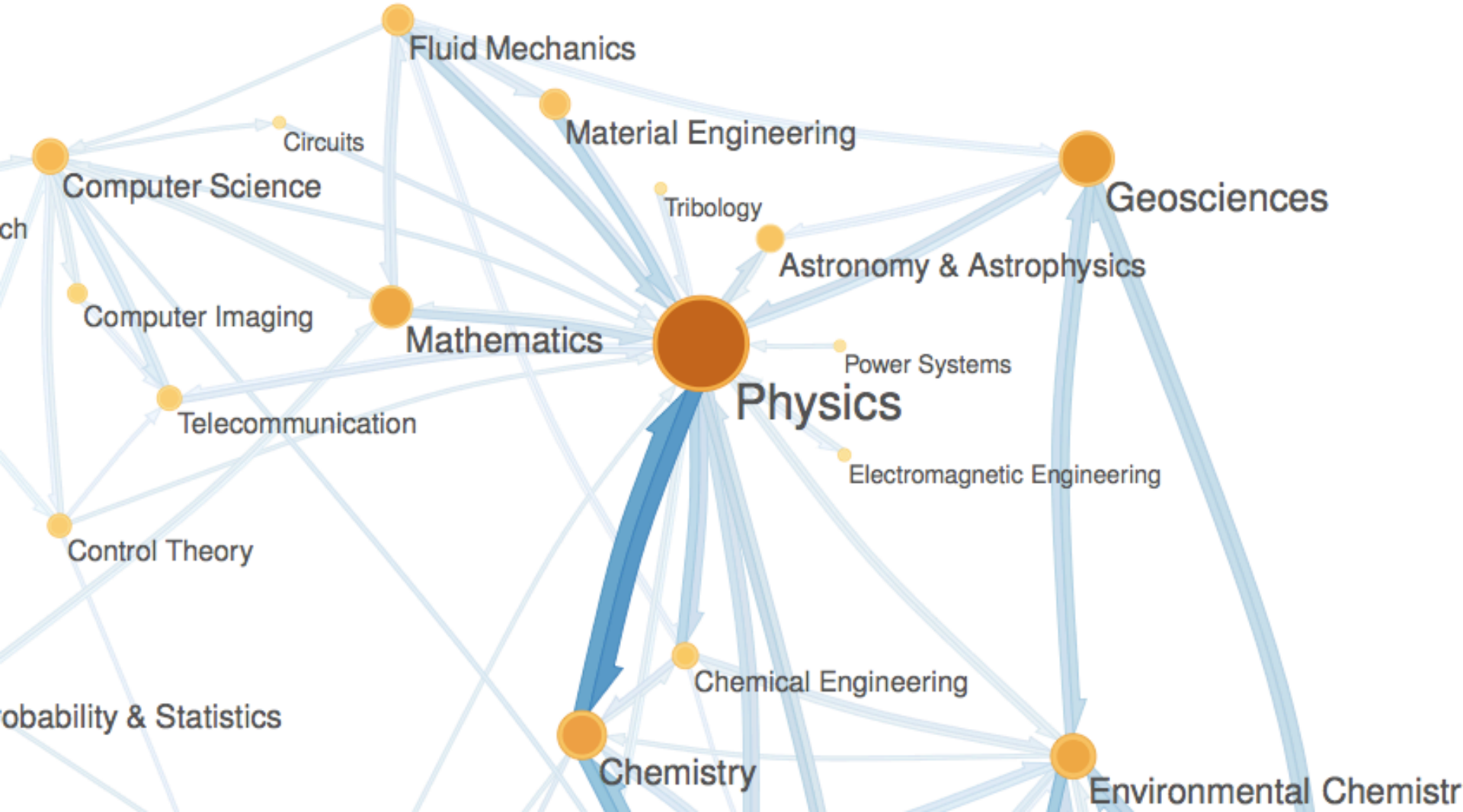
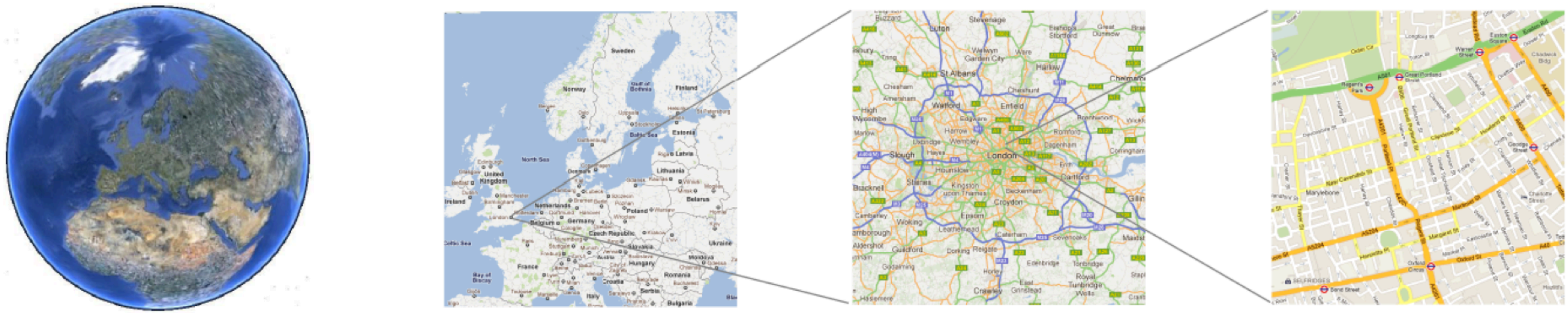
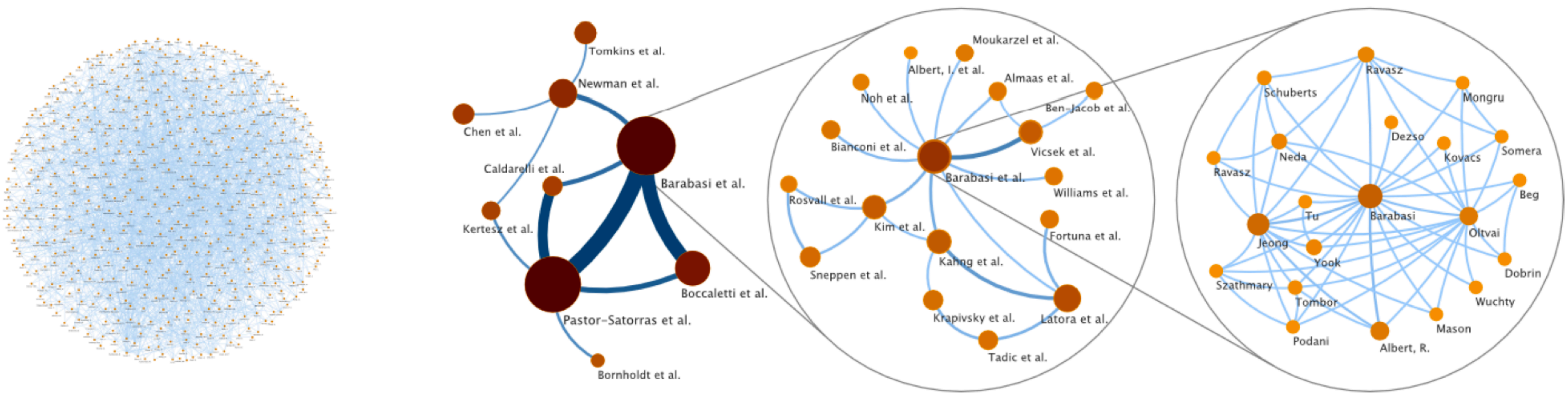
Mapping Knowledge Networks

Jevin West, Information School, University of Washington





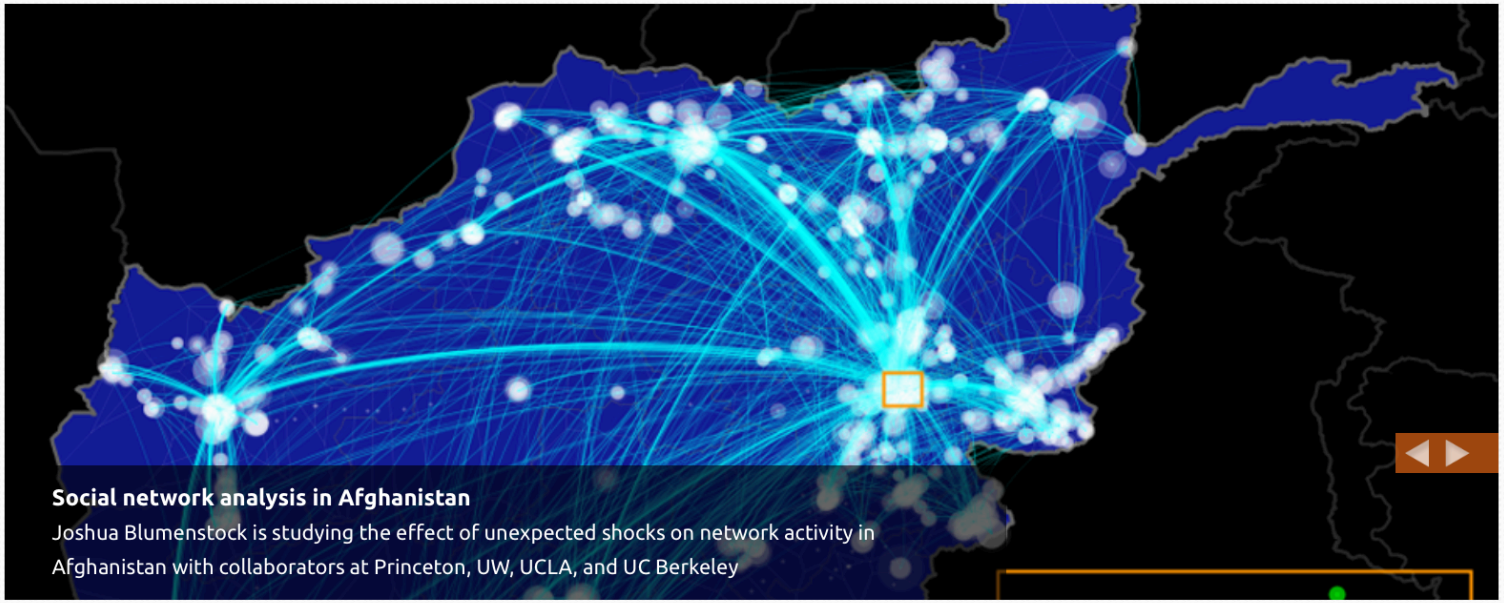

Tomkins et al.
Newman et al.
Chen et al.
Caldarelli et al.
Barabasi et al.
Kertesz et al.
Boccaletti et al.
Pastor-Satorras et al.
Bornholdt et al.
Moukarzel et al.
Albert, I. et al.
Almaas et al.
Noh et al.
Ben-Jacob et al.
Bianconi et al.
Vicsek et al.
Barabasi et al.
Williams et al.
Rosvall et al.
Kim et al.
Fortuna et al.
Kahng et al.
Sneppen et al.
Krapivsky et al.
Latora et al.
Tadic et al.
Ravasz
Schuberts
Mongru
Dezso
Somera
Neda
Kovacs
Ravasz
Beg
Barabasi
Tu
Oltvai
Jeong
Yook
Dobrin
Szathmary
Tombor
Wuchty
Mason
Podani
Albert, R.

Data

Compressing ⟷ Finding patterns

If we can find a good code for describing flow on a network, we will have solved the dual problem of finding the important structures with respect to that flow.

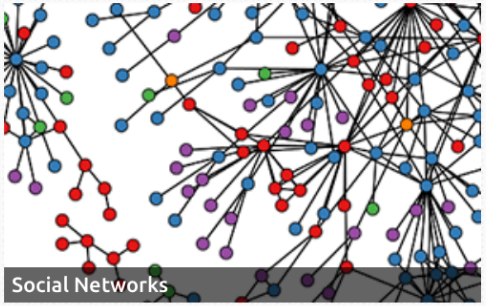
Minimal Description Length (MDL) Statistics

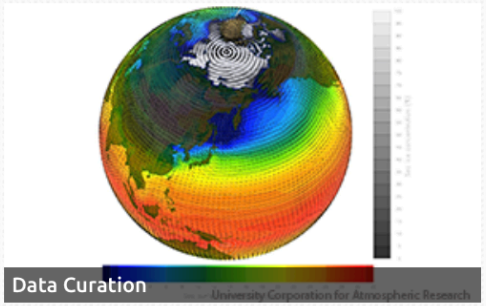


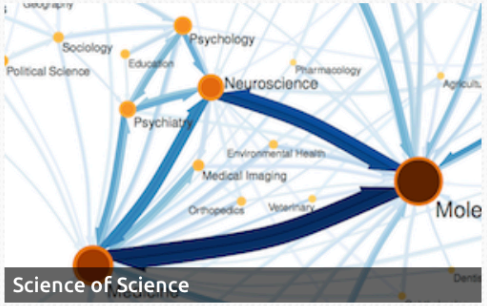
Social network analysis in Afghanistan

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







What We Do

Overview

Over the course of the last decade many disciplines have evolved from recording observations in laboratory notebooks to the use of instruments capable of digitally recording many gigabytes of data in a day. This abundance of data provides unprecedented opportunities for discovery. Tapping its potential requires the application of sophisticated new computational techniques operating on large scale storage, computational and network resources. Since its creation in 2008, the eScience Institute has worked to create the intellectual and physical infrastructure needed to meet this challenge.

At the core of the eScience Institute are individuals who have proven track records in developing and applying advanced computational methods and tools to real world problems. Their task is to seek out and engage researchers across disciplines where eScience approaches are likely to have the greatest impact. To ensure that researchers have access to the necessary physical infrastructure, the Institute has undertaken coordinated planning and support for advanced local and remote computational platforms. This includes developing relationships with commercial and non-commercial service providers as well as the development of shared facilities on campus. This support extends to assistance in the preparation of select proposals where we are able to focus resources, improving their chances for success.

Also in... What We Do

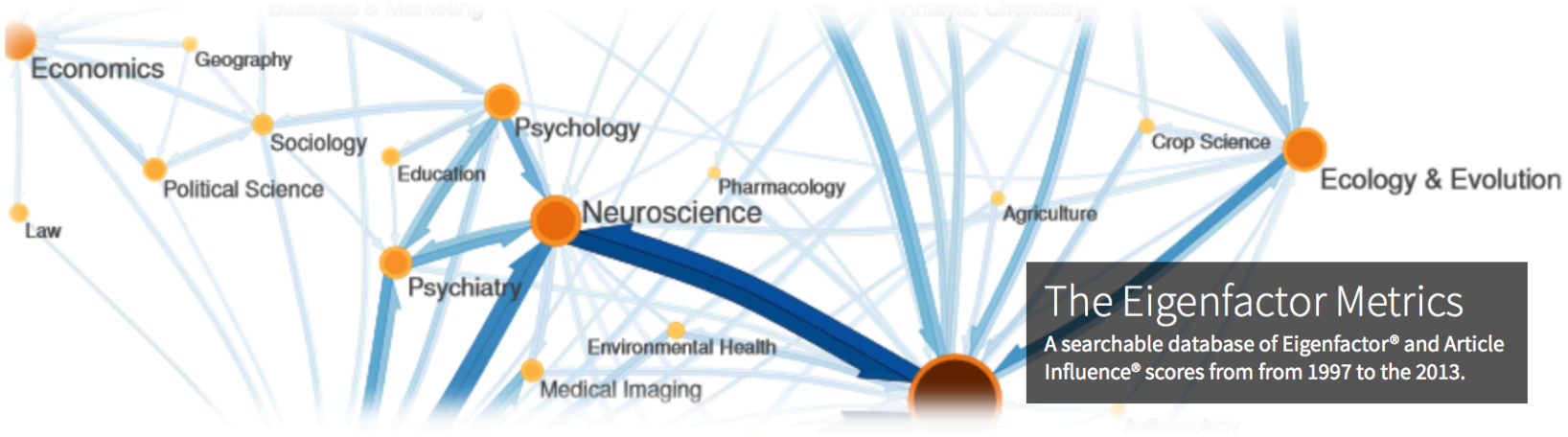
Appliance Gallery
Find and use the eScience Institute's virtual machines equipped with software useful for specific applications.
Campus Compute & Storage
Learn about what UW is doing to support scalable scientific computing on campus
Consulting & Services
From algorithm development to database creation to cloud computing, we can help.
Projects
Explore some of our current collaborations with research scientists.
Relevant Courses
View a list of courses offered in eScience disciplines.
SQLShare Success Stories
Tools
Whether it's database management, visualization, or developer tools, learn about tools we can help you use.

Latest eScience News

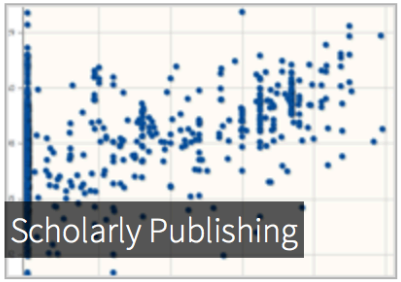
Data Science Incubation Program - Winter 2016
2 hours 4 min ago

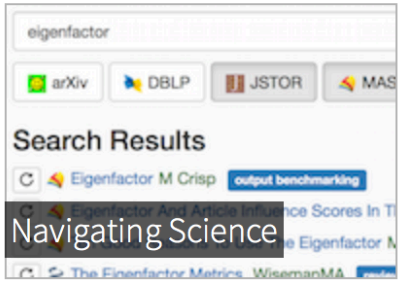
Ben Marwick On How Computers Broke Science

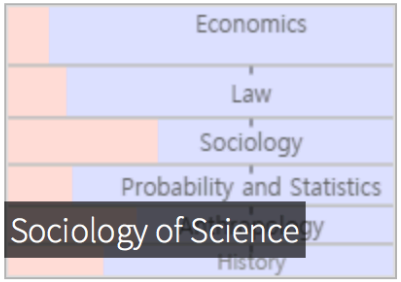
EIGENFACTOR.org



RESEARCH AREAS







NEWS

5 July.

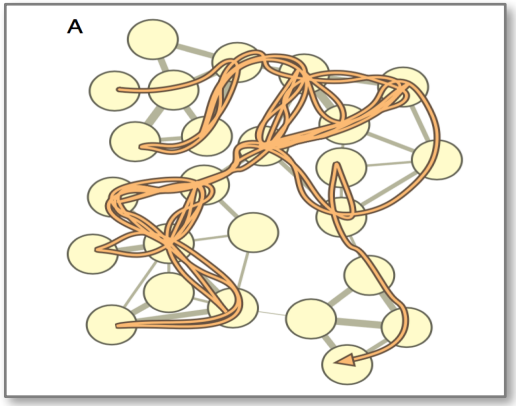
NATURE ON SELF-CITATION

Nature featured our paper on gender differences in self citation.

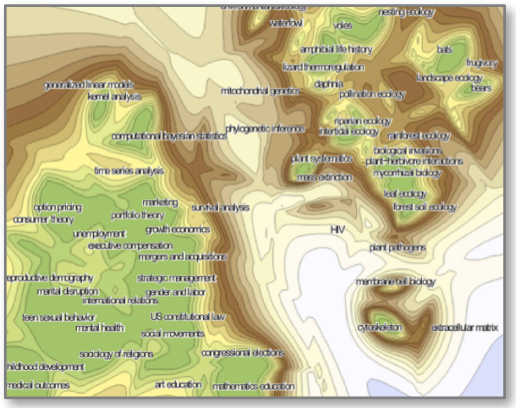
18 June.

THE ECONOMIST ON THE VIZIOMETRICS PROJECT

The Economist published an article describing our Viziometrics.org project.

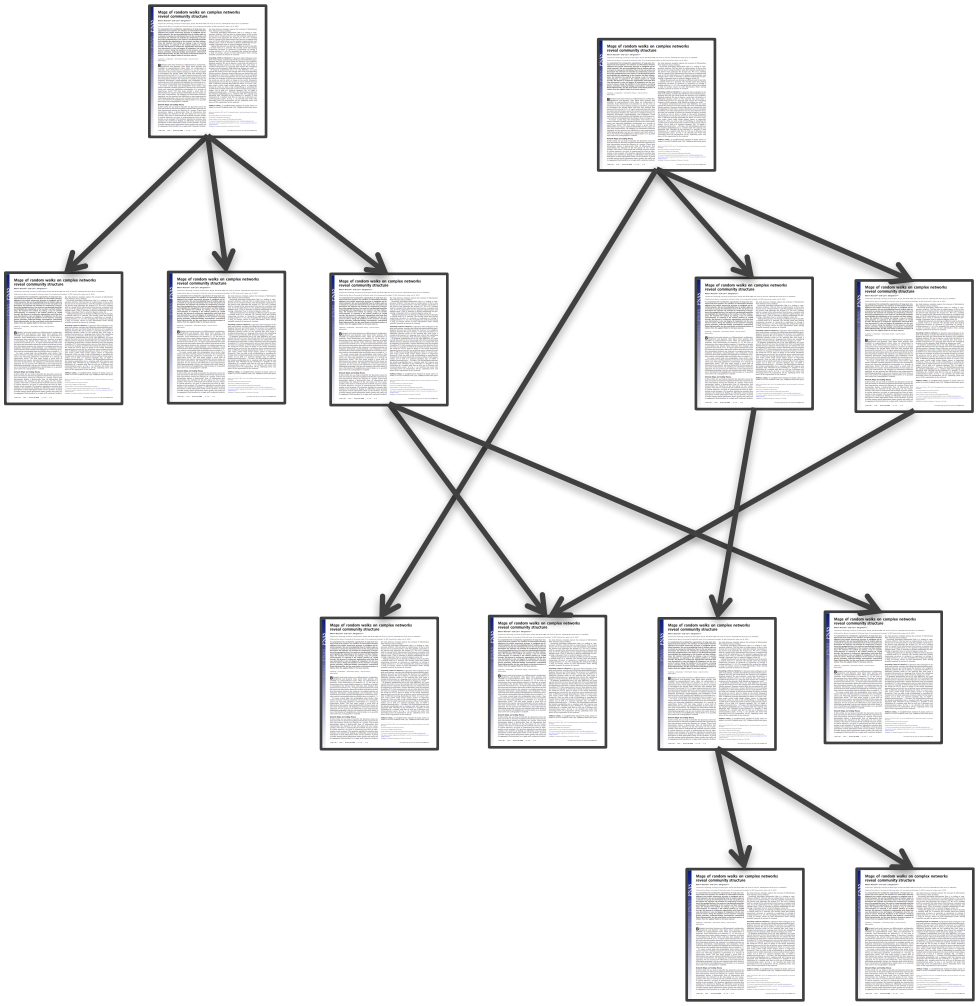


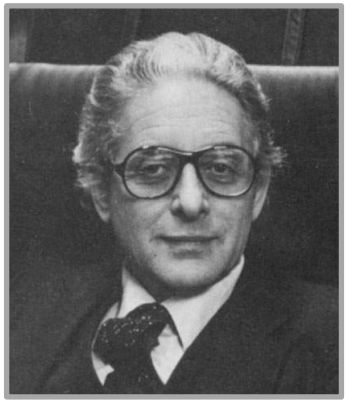
Science of Mapping



Mapping of Science

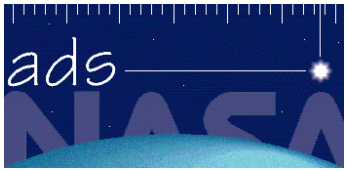
Citations form a vast network





de Solla Price, *Science* (1965)

The Scholarly Graph

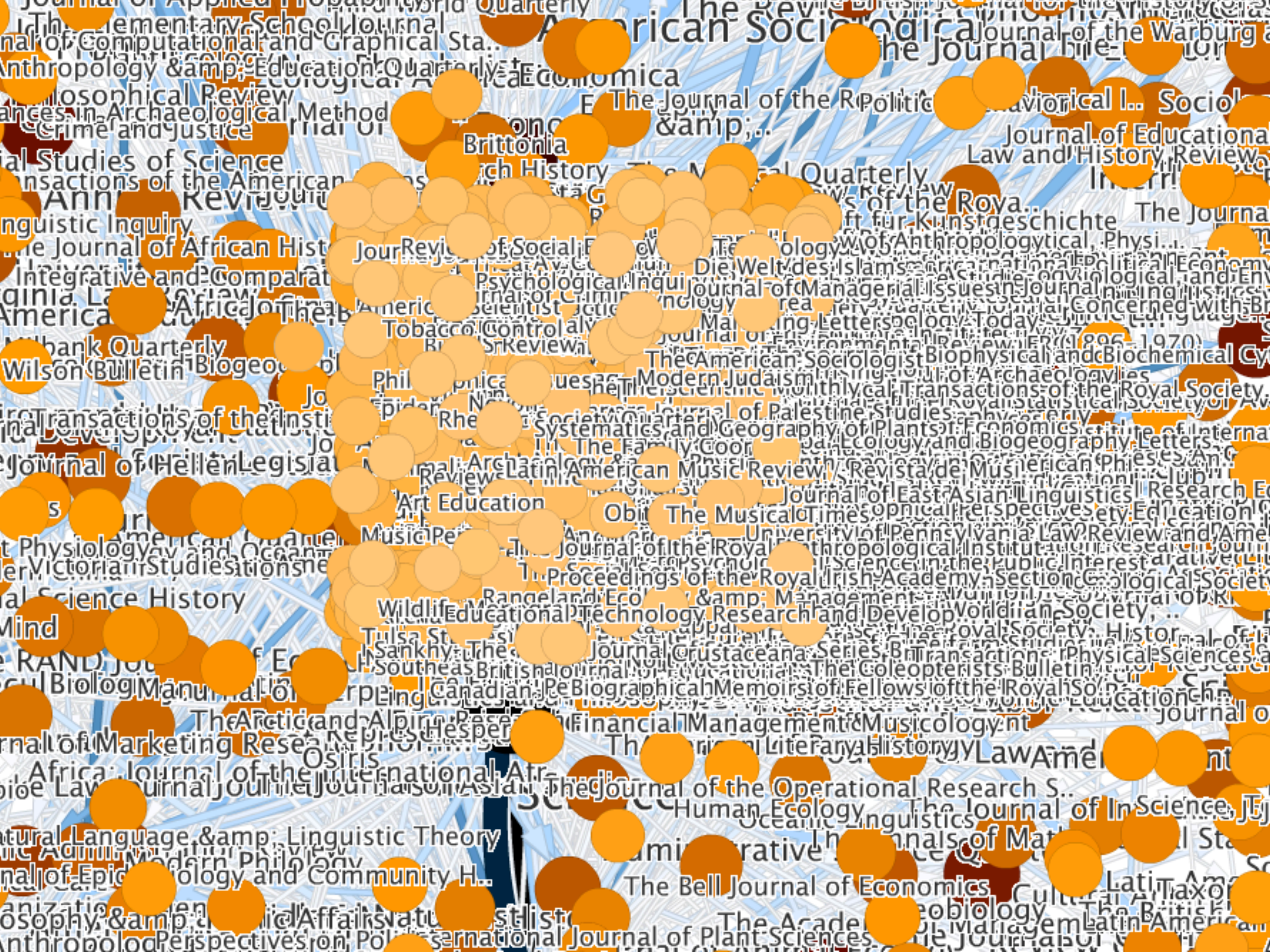


The Scholarly Graph

Tens of millions articles, patents, books

Billions of citation links

Years: 1600 - 2016


The Elementary School Journal
Anthropology & Education Quarterly
Philosophical Review
Crime and Justice
Linguistic Inquiry
Integrative and Comparat
Wilson Bulletin
Journal of the Warburg
Journal of Educationa
Law and History Review
Brittonia
The Journal of the R
Psychological Inqui
Journal of Managerial Issues
Die Welt des Islams
Tobacco Control
The American Sociologist
Biophysical and Biochemical Cy
Modern Judaism
Journal of Palestine Studies
Systematics and Geography of Plants
The Family Coor
Latin American Music Review
Revista de Musi
Art Education
Journal of East Asian Linguistics
The Musical Times
University of Pennsylvania Law Review
Proceedings of the Royal Irish Academy. Section C
Educational Technology Research and Develop
Journal Crustaceana Series
The Coleopterists Bulletin
Biographical Memoirs of Fellows of the Royal So
Financial Management
Musicology
Journal of Marketing Resea
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 Pol
International Journal of Plant Sciences
Victorian Studies
Transactions of the Instit
Mind
Social Science History
Journal of Hellen
Economica
American Sociological

The map equation

frequency of inter-module movements

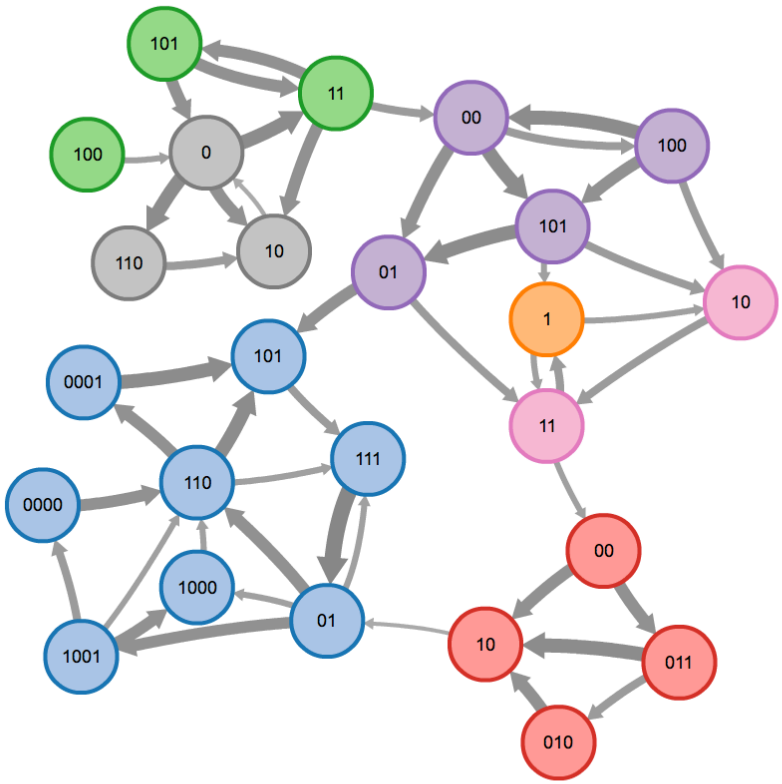
frequency of movements within module *i*

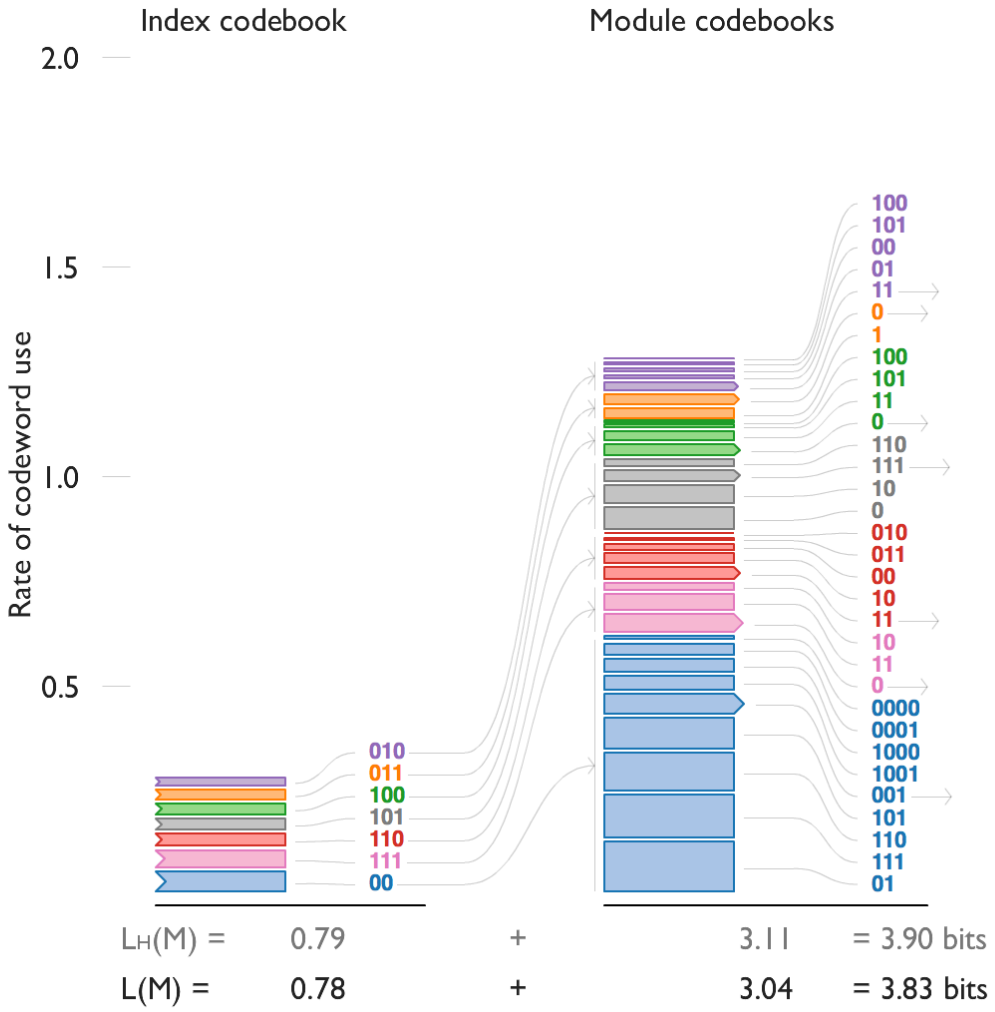
$$L(\mathbf{M}) = q_{\curvearrowright} H(\mathcal{Q}) + \sum_{i=1}^{m} p^{i}_{\circlearrowright} H(\mathcal{P}^{i})$$

code length of module names

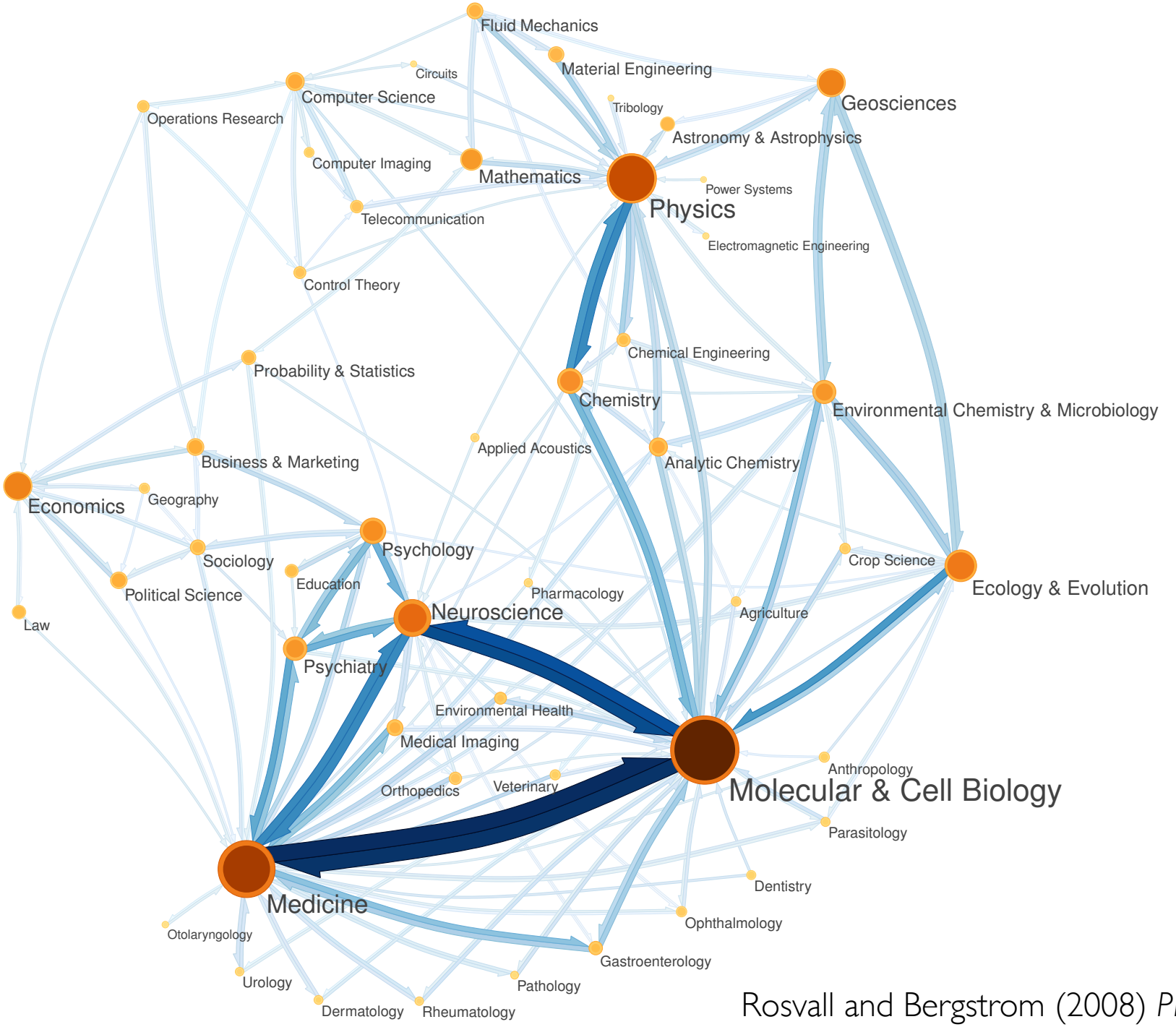
code length of node names in module *i*

Rosvall and Bergstrom (2008) *PNAS*



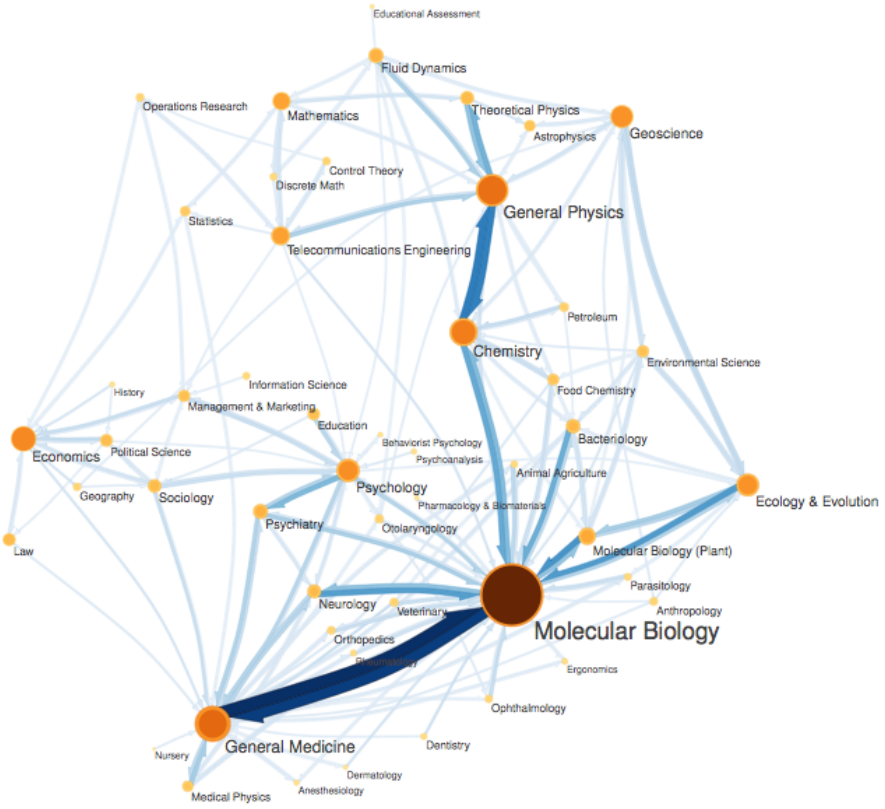


Mapequation.org, Daniel Edler

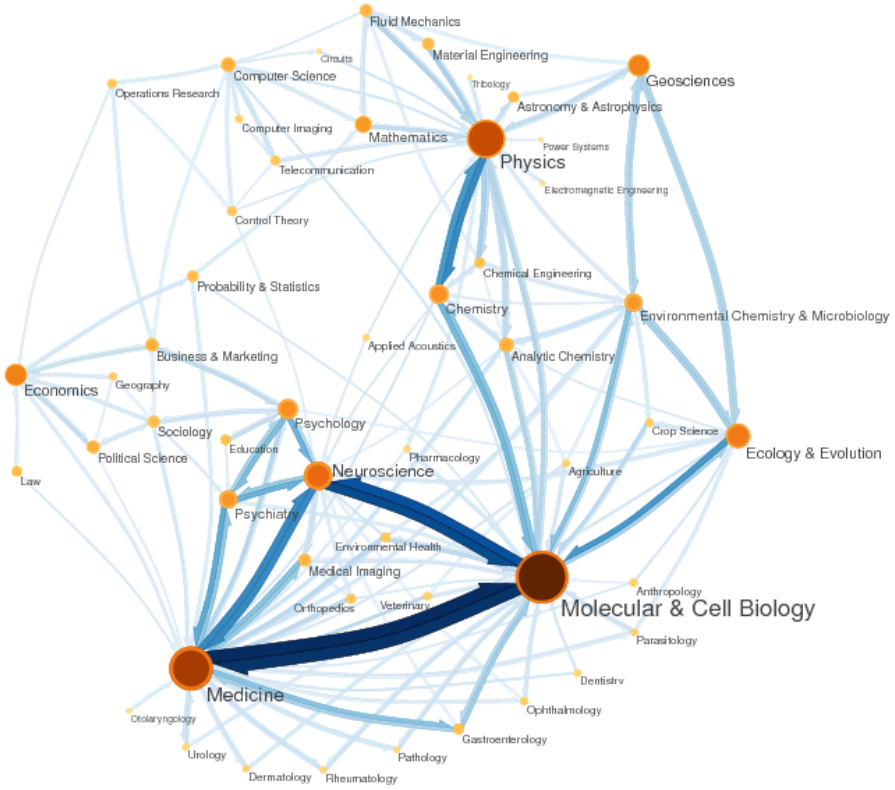


Rosvall and Bergstrom (2008) *PNAS*

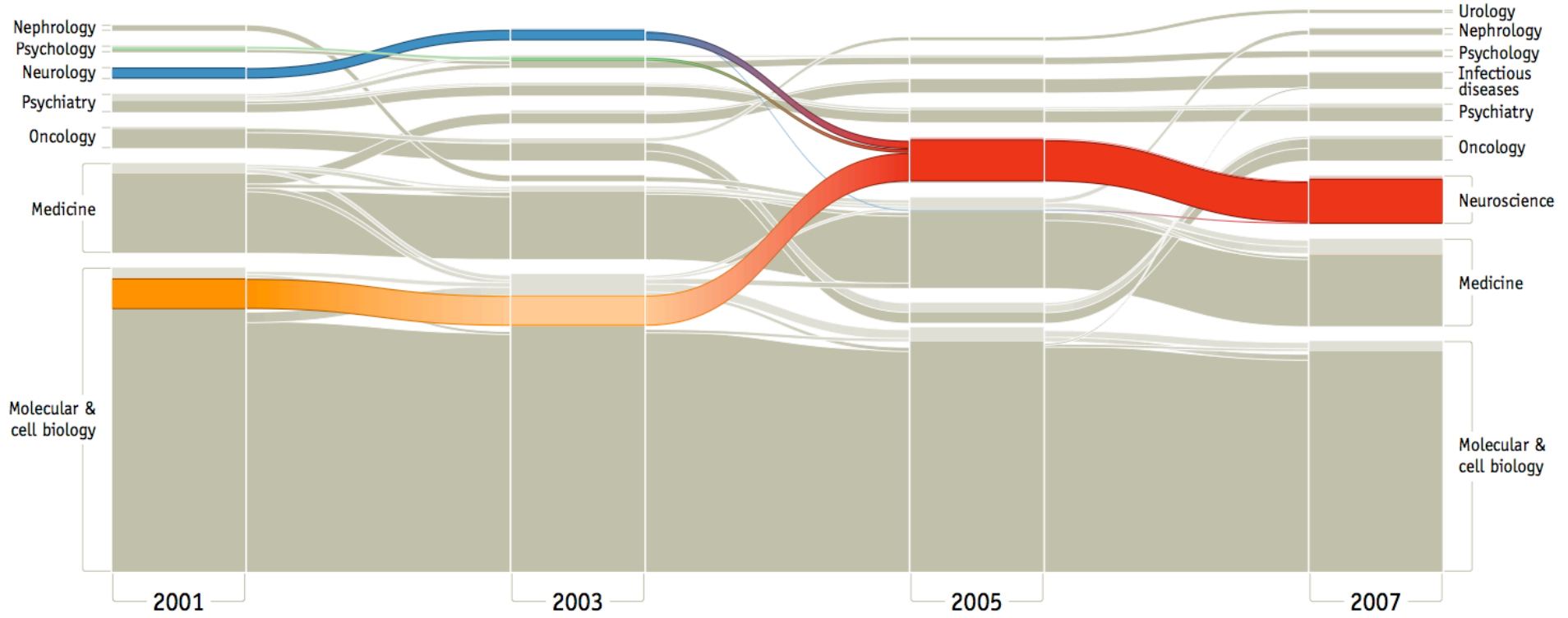
1995


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
Ecology & Evolution
Pharmacology & Biomaterials
Psychiatry
Otolaryngology
Law
Molecular Biology (Plant)
Parasitology
Neurology
Veterinary
Anthropology
Molecular Biology
Orthopedics
Ergonomics
Ophthalmology
General Medicine
Nursery
Dentistry
Dermatology
Medical Physics
Anesthesiology

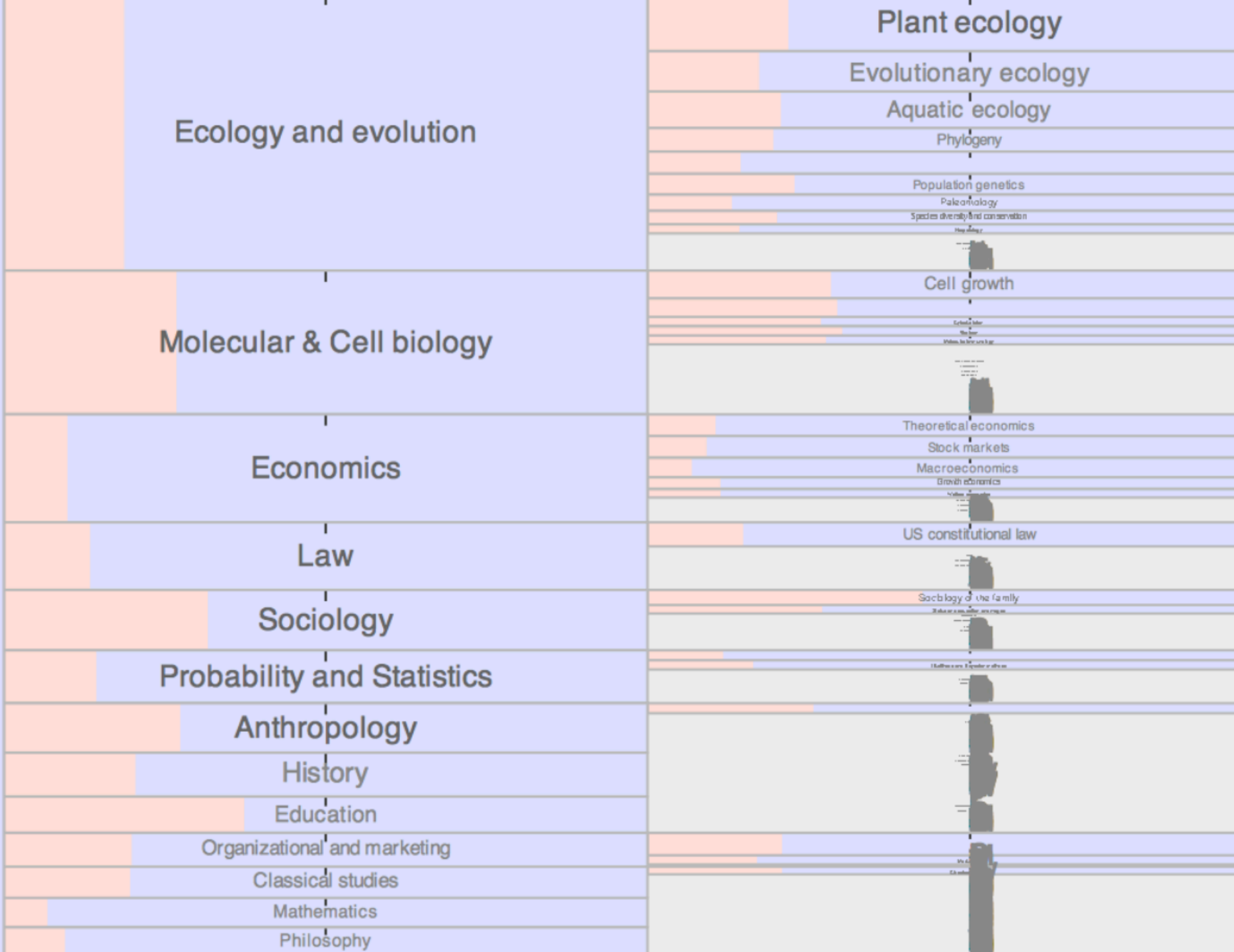
2004


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

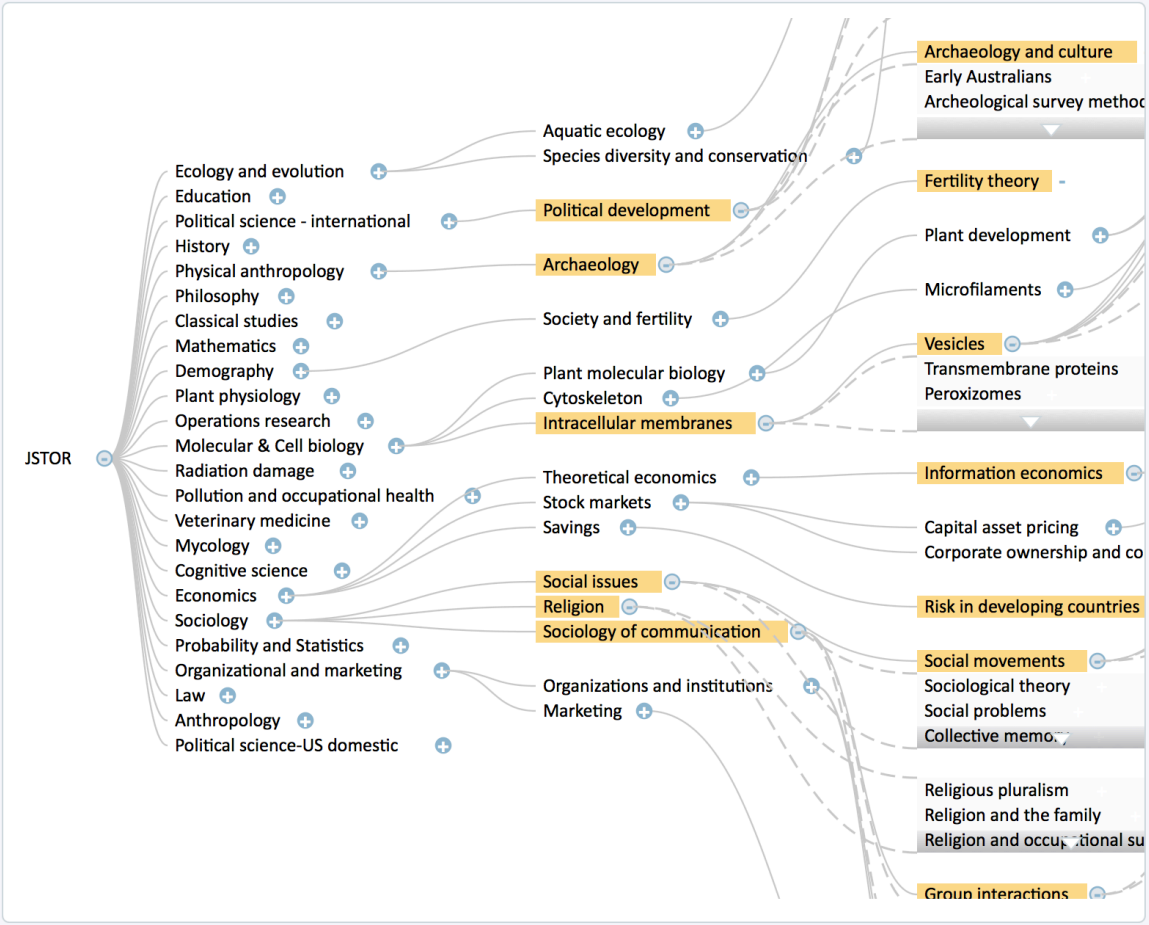
The Emergence of Neuroscience



Rosvall and Bergstrom (2010) PLoS One


Ecology and evolution
Plant ecology
Evolutionary ecology
Aquatic ecology
Phylogeny
Population genetics
Paleontology
Species diversity and conservation
Molecular & Cell biology
Cell growth
Economics
Theoretical economics
Stock markets
Macroeconomics
Growth economics
Law
US constitutional law
Sociology
Sociology of the family
Probability and Statistics
Anthropology
History
Education
Organizational and marketing
Classical studies
Mathematics
Philosophy

"Network"



Find Papers

by **title** network

by **field**

by **author**

by **journal**

Active Queries: clear all

keyword: **network**

Top Papers

Sort by Year (newest)

Using Siting Algorithms in the Design of Marine Reserve Networks
Heather Leslie - *Ecological Applications* (2003)

Mechanism of Filopodia Initiation by Reorganization of a Dendritic Network
Tatyana Svitkina - *The Journal of Cell Biology* (2003)

Network Structure and Knowledge Transfer: The Effects of Cohesion and Range
Ray Reagans - *Administrative Science Quarterly* (2003)

A General Model for Designing Networks of Marine Reserves
Enric Sala - *Science* (2002)

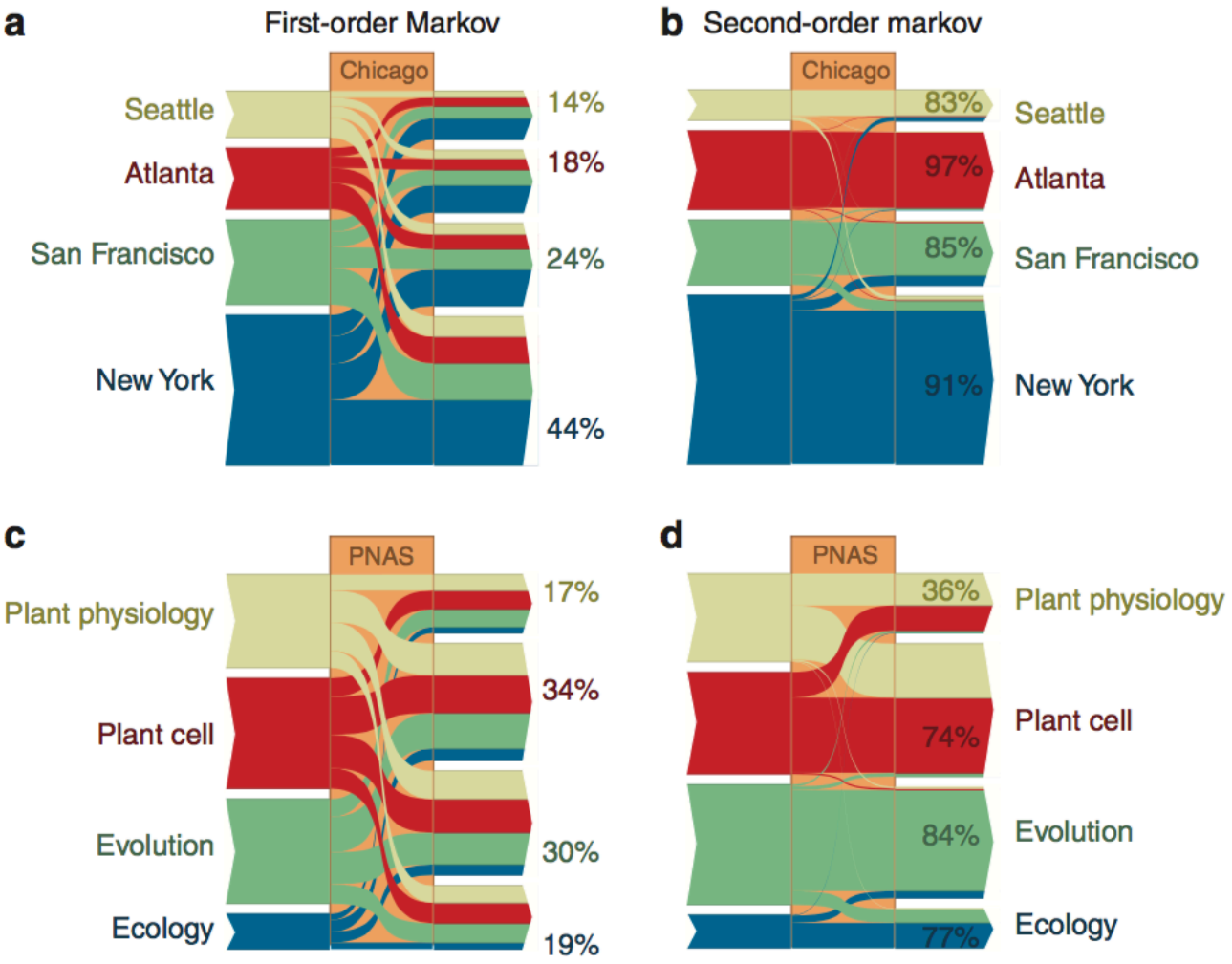
The Density of Social Networks and Fertility Decisions: Evidence from South Nyanza District, Kenya
Hans-Peter Kohler - *Demography* (2001)

A New Dynamin-Like Protein, ADL6, Is Involved in Trafficking from the trans-Golgi Network to the Central Vacuole in Arabidopsis
Jing Bo Jin - *The Plant Cell* (2001)

Comparing Sequenced Segments of the Tomato and Arabidopsis Genomes: Large-Scale Duplication Followed by Selective Gene Loss Creates a Network of Synteny
Hsin-Mei Ku - *Proceedings of the National Academy of Sciences of the United States of America* (2000)

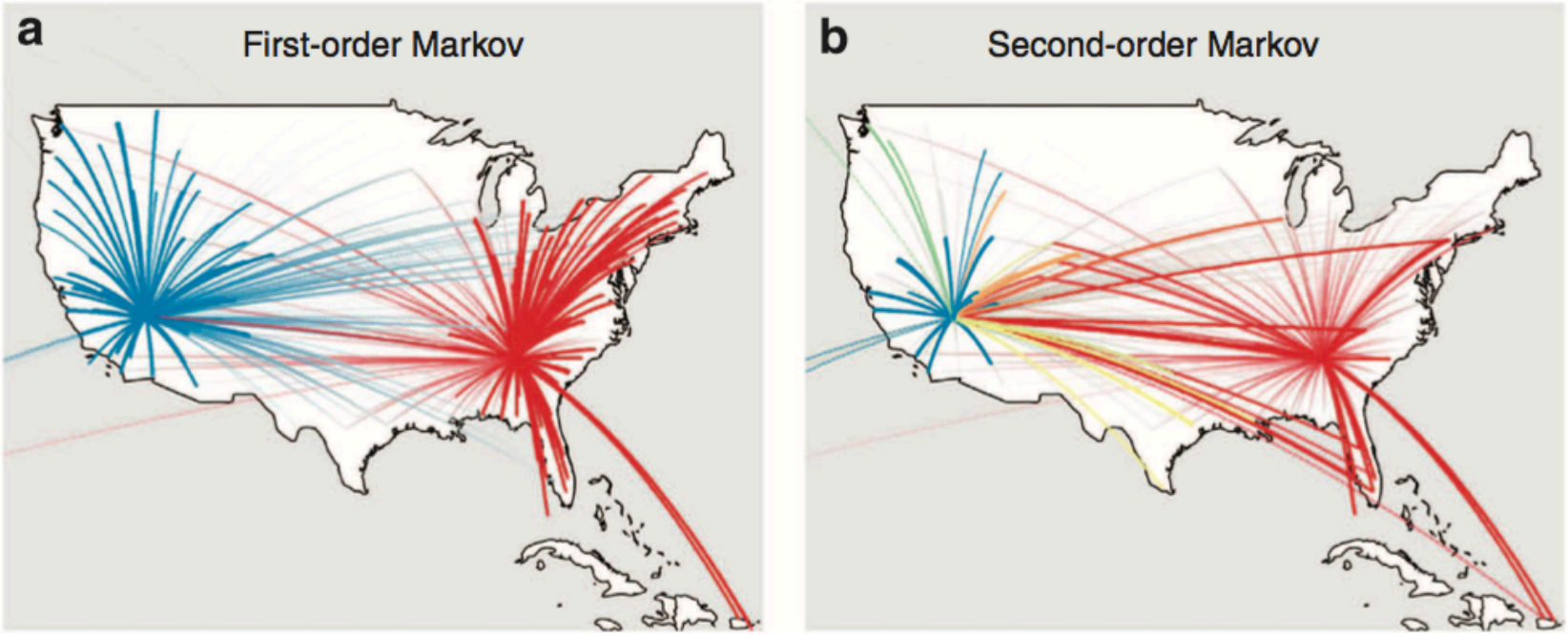
A Noncooperative Model of Network Formation
Venkatesh Bala - *Econometrica* (2000)

Higher Order Dynamics



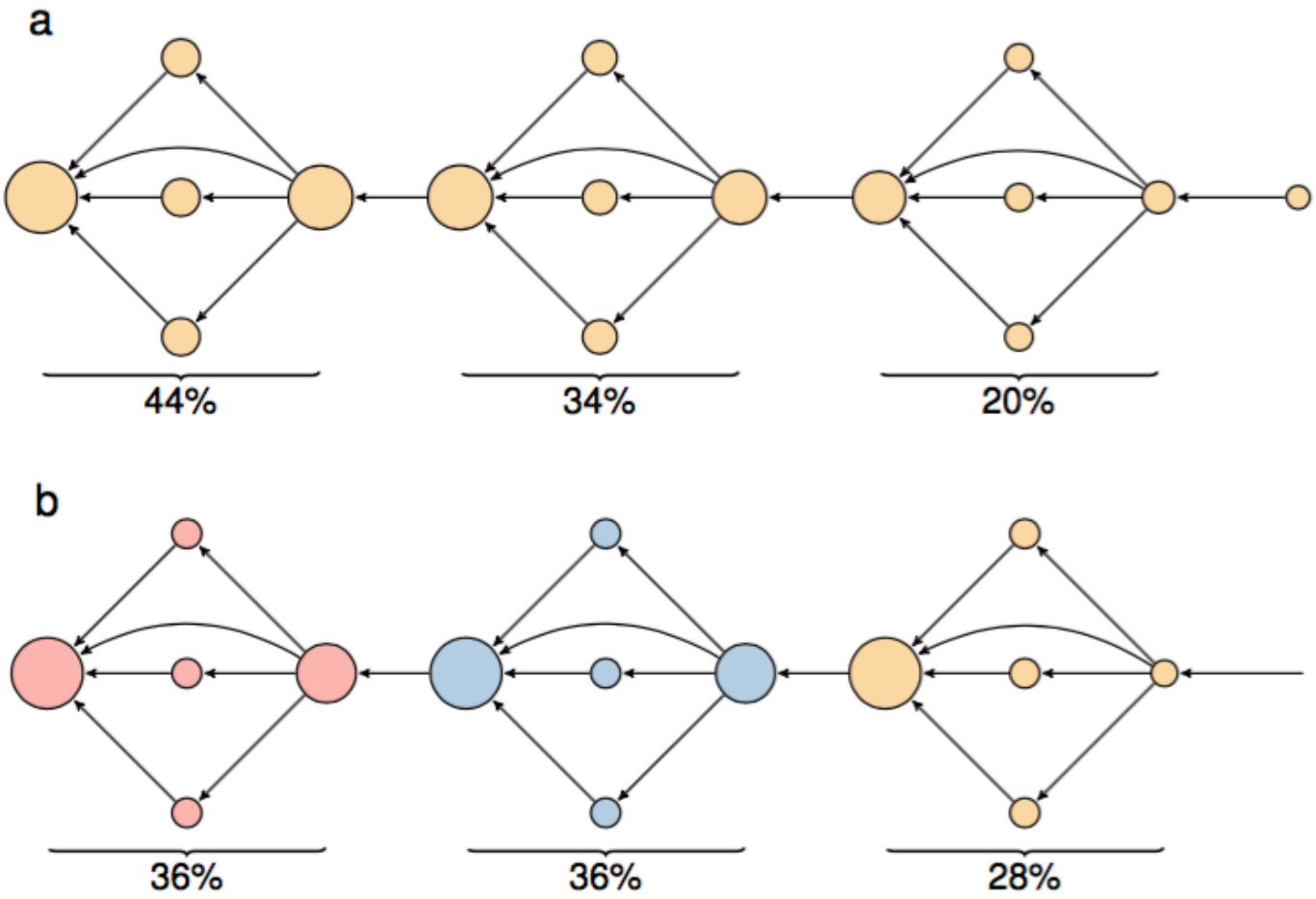
Rosvall et al. (2014) Memory in network flows and its effects on spreading dynamics and community detection. *Nature Communications*

Higher Resolution Maps

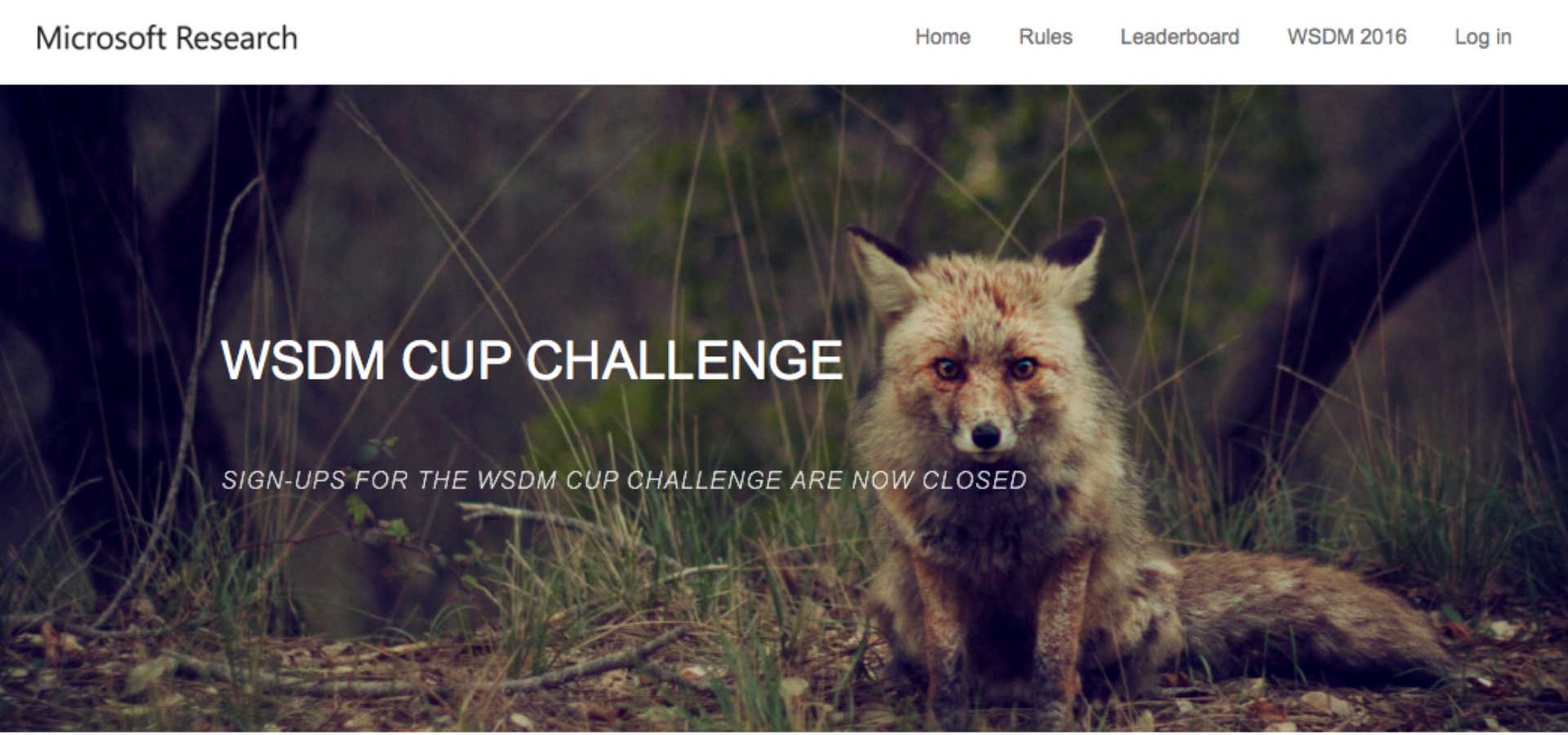


Rosvall et al. (2014) Memory in network flows and its effects on spreading dynamics and community detection. *Nature Communications*

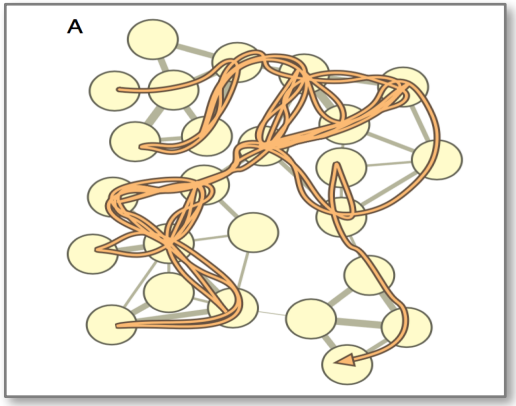
Article-level Ranking and Mapping



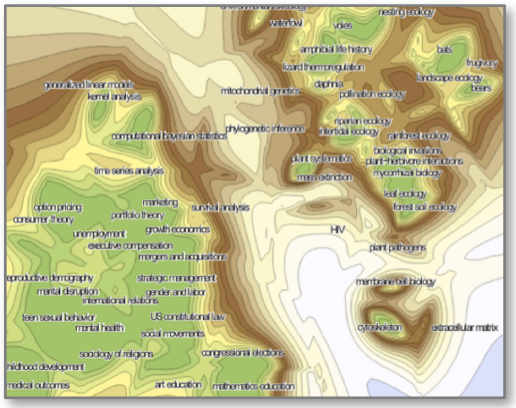
West et al. (2016) Ranking and mapping article-level citation networks. *in prep.*



Wesley-Smith et al. (2016) Static Ranking of Scholarly Papers using Article-Level Eigenfactor (ALEF). *arxiv:1606.08534*



Science of Mapping

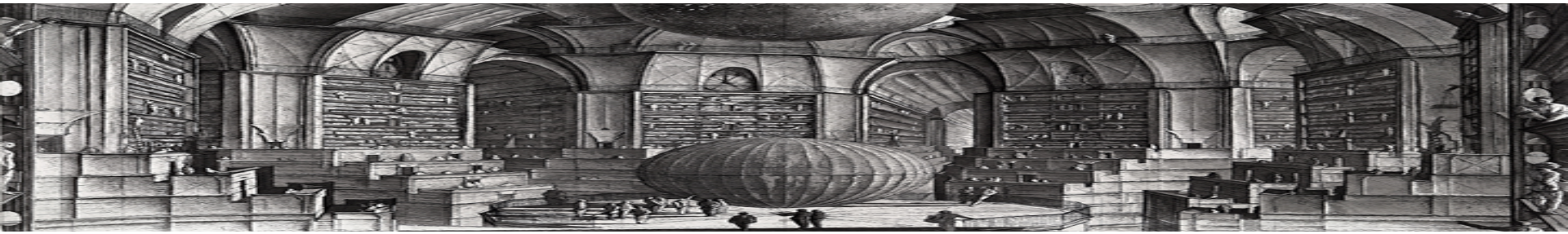


Mapping of Science

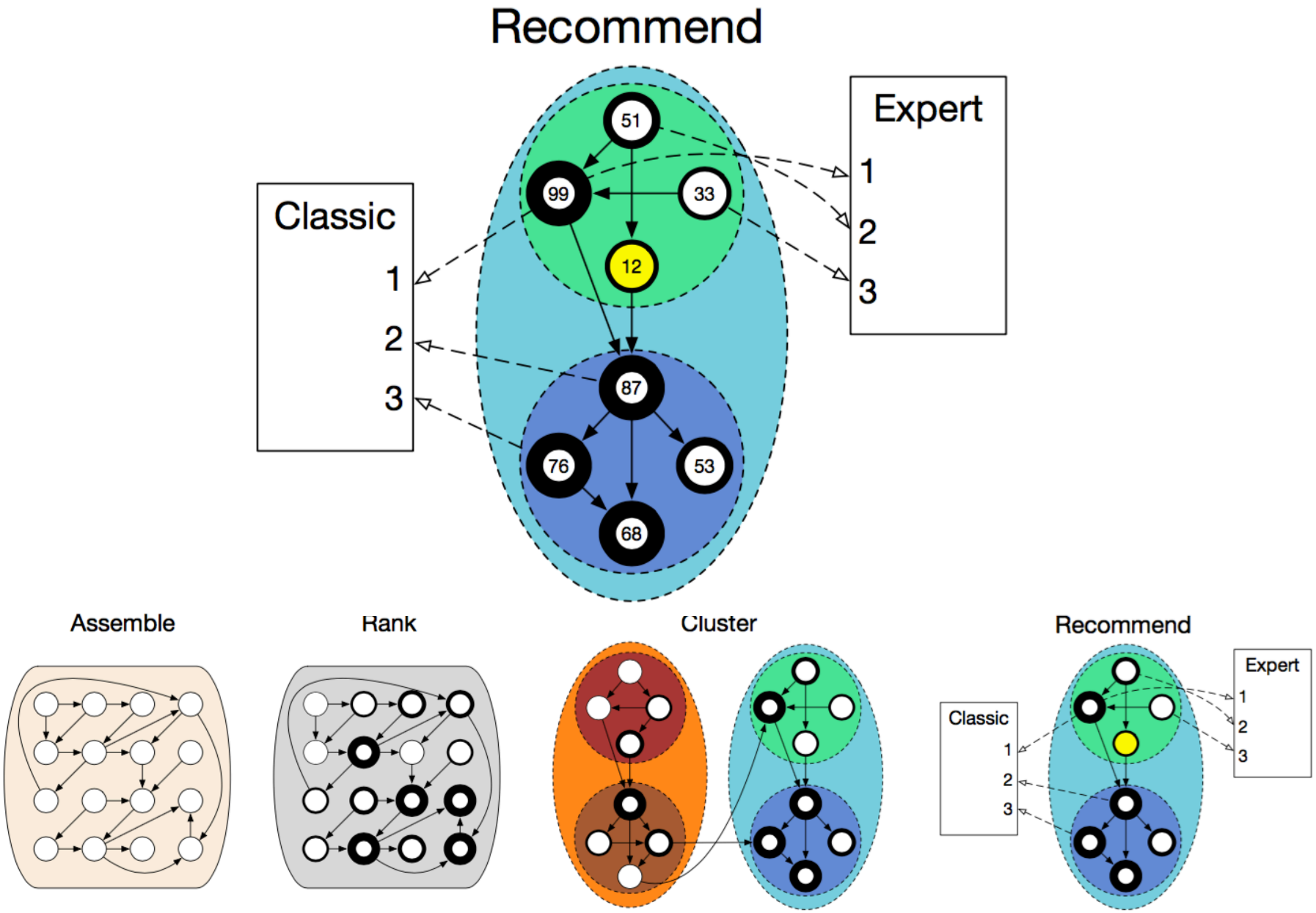
JW

Explore the recommendations

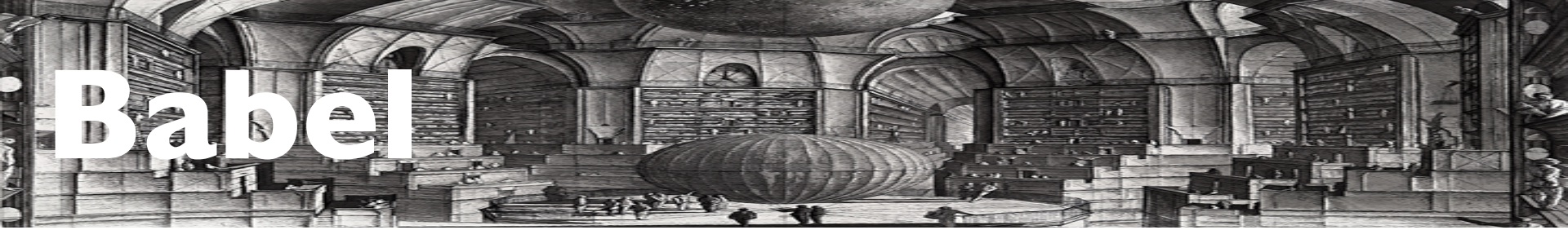
babel.eigenfactor.org



jevinw@uw.edu



West, Wesley-Smith, Bergstrom (2016) A recommendation system based on hierarchical clustering of an article-level citation network. *IEEE Transactions on Big Data*



Babel

- Facilitate research and *implementation* of recommendations
- Bibliographic data at scale
- Freely available and open source
- Evaluation of recommendations
- Audience: publishers, researchers, developers
- API Standardization & Endpoint Discovery

babel.eigenfactor.org

JSTOR Sustainability BETA

Help us make this site better!

Send Comments Login

Search

CONSERVATION BIOLOGY OVERVIEW

Conservation is the scientific study of the nature and of Earth's biodiversity with the aim of protecting species, their habitats, and ecosystems from excessive rates of extinction and the erosion of biotic interactions. It is an interdisciplinary subject drawing on natural and social sciences, and the practice of natural resource management.The conservation ethic is based on the findings of conservation biology.

Source: Wikipedia

Source: Wikimedia Commons

Influential Articles

1960s 1970s 1980s 1990s 2000s 2010s 2020

- The Canonical Distribution of Commonness and ...
- An Equilibrium Theory of Insular Zoogeography
- Turnover Rates in Insular Biogeography: ...
- The Statistics and Biology of the ...

Related Topics

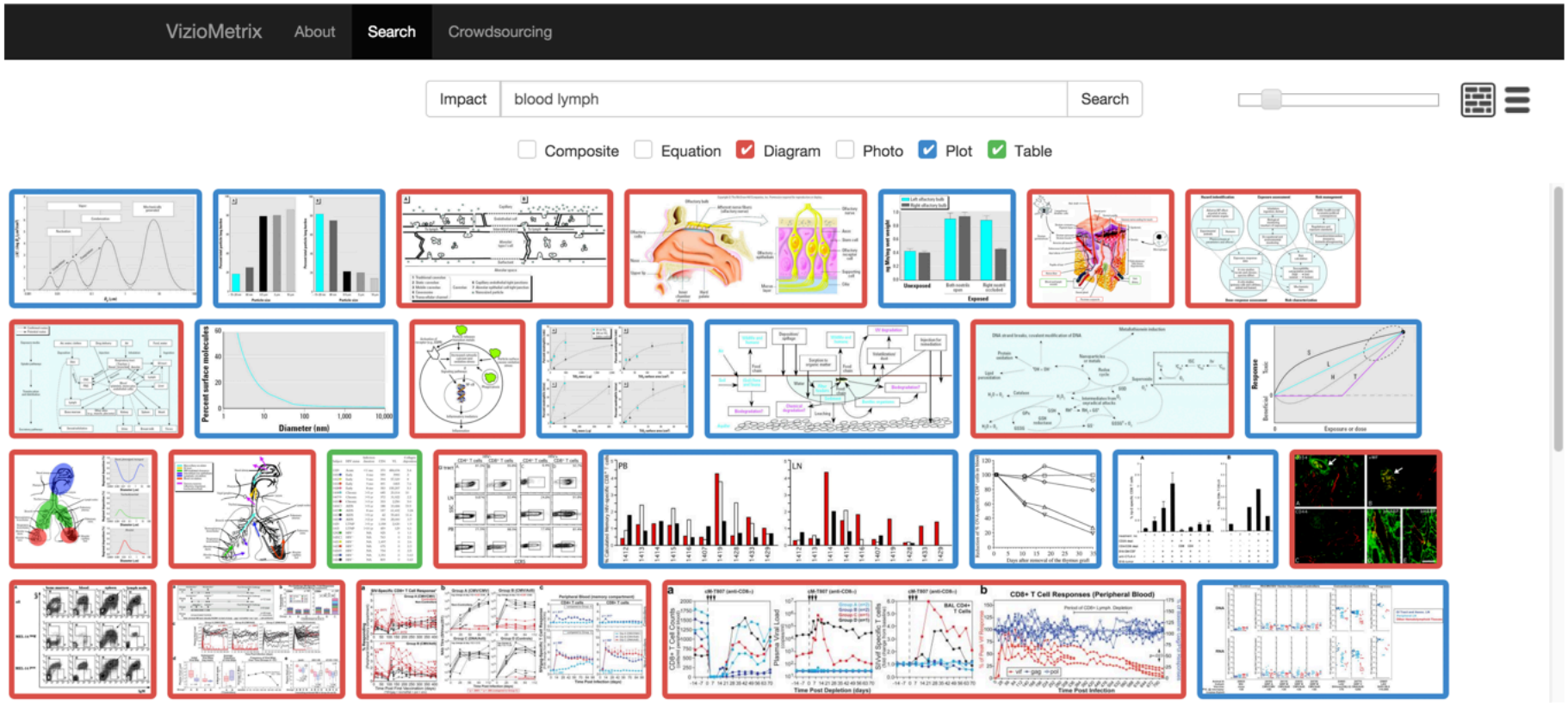
Species

Habitat conservation

http://labs.jstor.org/sustainability/topic/Biodiversity

Figure-Centric Search Engine

viziometrics.org



Questions

- How do patterns of encoding visual information in the literature vary across disciplines?
- How have patterns of encoding visual information in the literature evolved over time?
- Is there any link between patterns of encoding visual information and scientific impact?

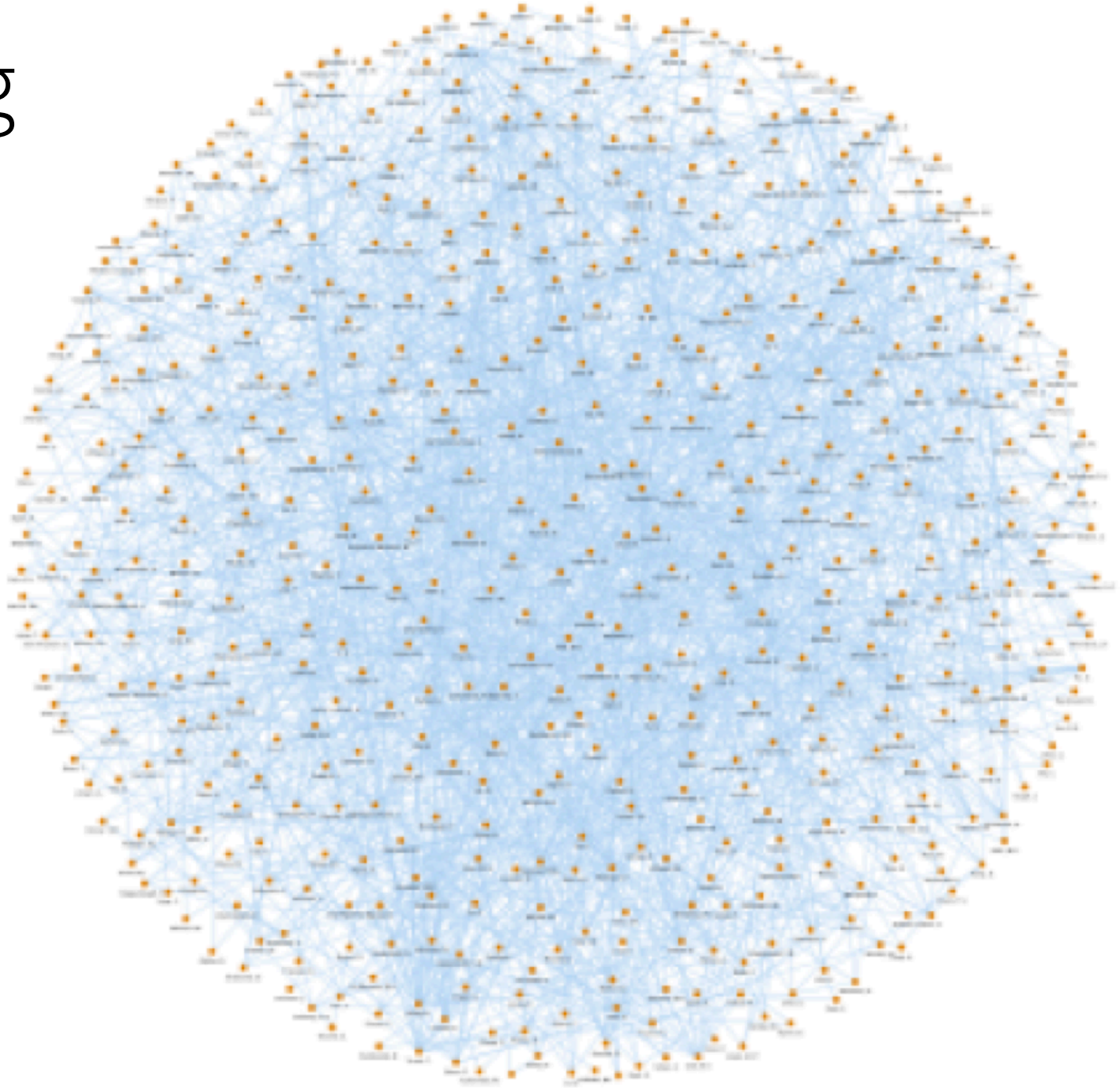
How can we better utilize visual information in the search and navigation process?

Challenges

1. scaling

2. mechanism

Scaling



RelaxMap

A

Scalable and Efficient Flow-Based Community Detection for Large-Scale Graph Analysis

SEUNG-HEE BAE, DANIEL HALPERIN, and JEVIN WEST, University of Washington
MARTIN ROSVALL, Umeå University
BILL HOWE, University of Washington

Community detection is a powerful approach to uncover important structures in large networks. For real networks that often describe the flow of some entity, flow-based community detection methods are particularly important. Infomap is a flow-based community detection algorithm that optimizes the objective function known as the map equation. Third-party benchmarks have found that Infomap is the most effective algorithm for identifying clusters in large graphs. Unfortunately, though Infomap works well, it is an inherently serial algorithm and thus cannot take advantage of multi-core processing in modern computers, limiting its use for analyzing large graphs quickly.

In this paper, we propose a novel algorithm to optimize the map equation called RelaxMap. RelaxMap provides two important improvements over Infomap: parallelization, so that the map equation can be optimized over much larger graphs, and prioritization, so that the most important work occurs first, iterations take less time, and the algorithm converges faster. We implement these techniques using OpenMP on shared-memory multicore systems, and evaluate our approach on a variety of graphs from standard graph clustering benchmarks as well as real graph datasets. Our evaluation shows that both techniques are effective: RelaxMap achieves 70% parallel efficiency on 8 cores, and prioritization improves algorithm performance by an additional 20%–50% in average, depending on the graph properties. Additionally, RelaxMap converges in the similar number of iterations and provides solutions of equivalent quality as the serial Infomap implementation.

Categories and Subject Descriptors: I.5.3 [**Clustering**]: Algorithms; F.1.2 [**Modes of Computation**]: Parallelism and Concurrency

General Terms: Design, Algorithms, Performance

Additional Key Words and Phrases: Community Detection, Graph Analysis, Parallelization, Prioritization

ACM Reference Format:
Seung-Hee Bae, Daniel Halperin, Jevin West, Martin Rosvall, and Bill Howe, 2014. Scalable and efficient flow-based community detection for large-scale graph analysis. *ACM Trans. Knowl. Discov. Data.* V, N, Article A (January 2014), 29 pages.
DOI:http://dx.doi.org/10.1145/0000000.0000000

1. INTRODUCTION

Community detection in large graphs is emerging as a first-class technique in a number of applications: functional similarity in biological networks [Gavin et al. 2006; Guimera and Amaral 2005], collaboration communities in research networks [Girvan and Newman 2002], and the macro-structure

Summary

- Study the *Science of Science*
- Assemble scholarly knowledge graph into machine readable formats
- Ask questions about the origin and evolution of ideas and fields, interdisciplinarity, impact assessment and sociology of science
- Develop clustering algorithms for automatically detecting scholarly communities in the literature
- Building statistical and visualization tools that improve navigation, make relevant connections and facilitate knowledge discovery
- Challenges: scaling, mechanism
- Eigenfactor.org, Viziometrics.org, Babel.eigenfactor.org

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

jevinw@uw.edu

jevinwest.org

Eigenfactor.org

@jevinwest