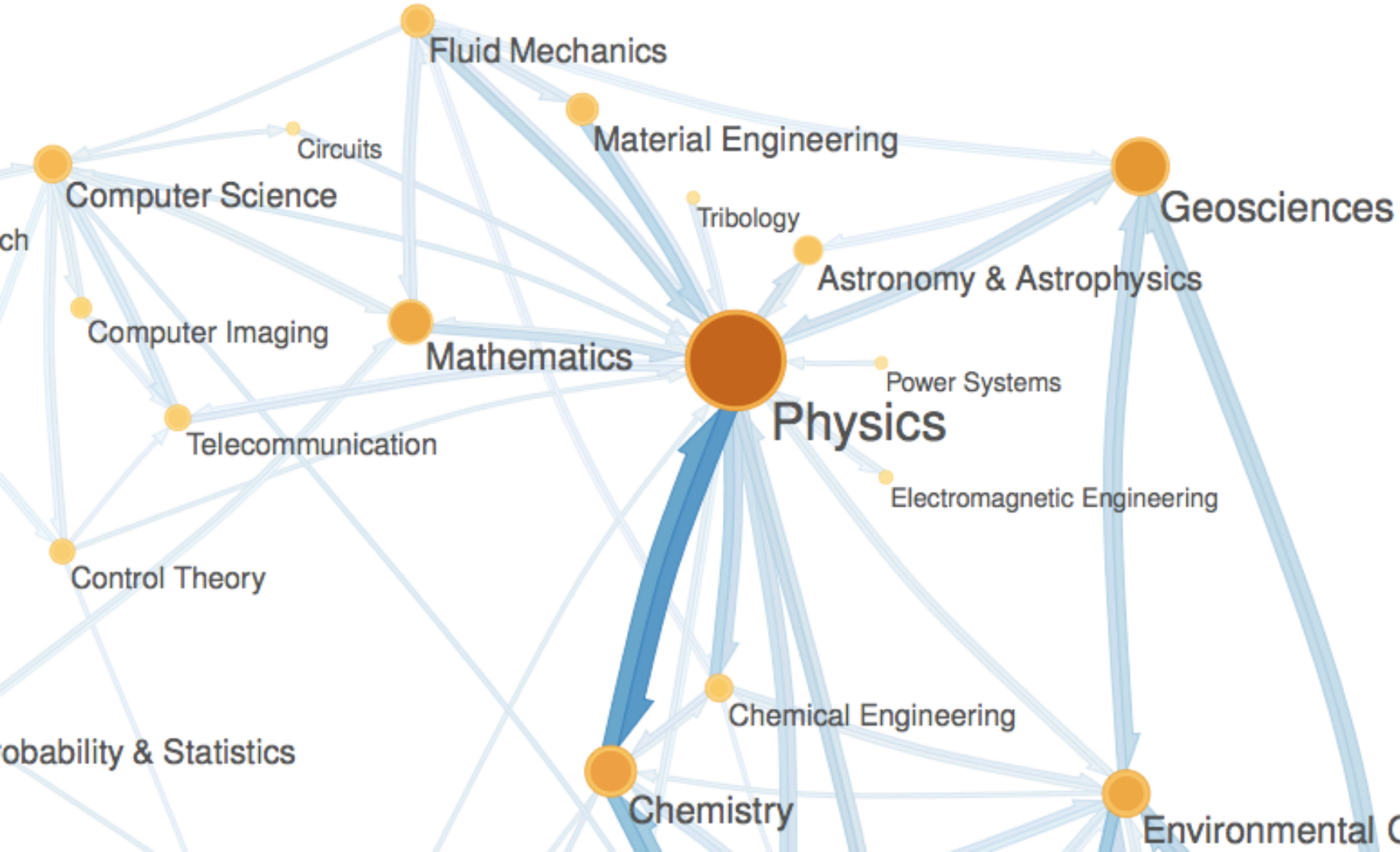
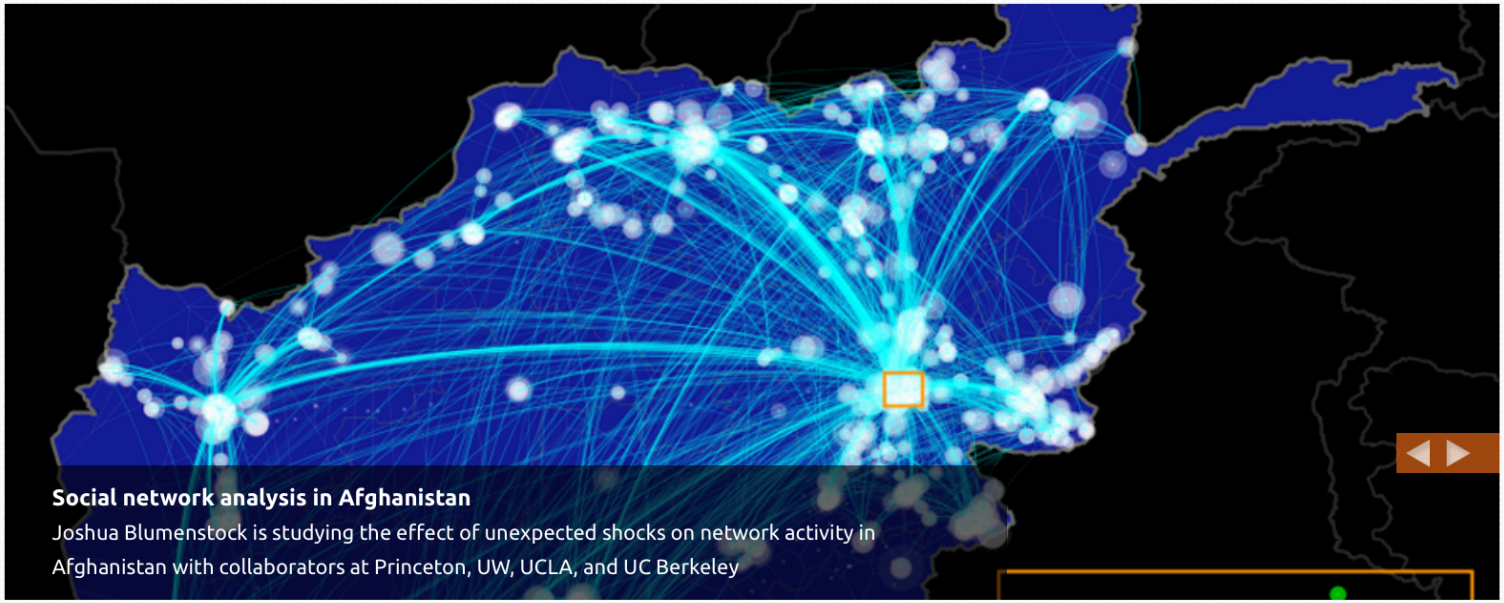
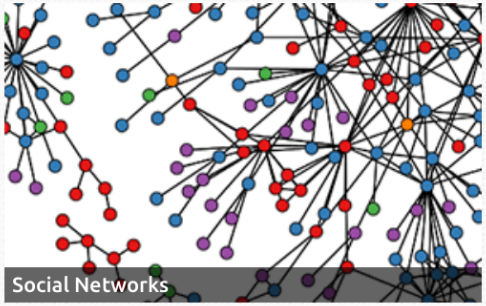
Mapping Knowledge Networks

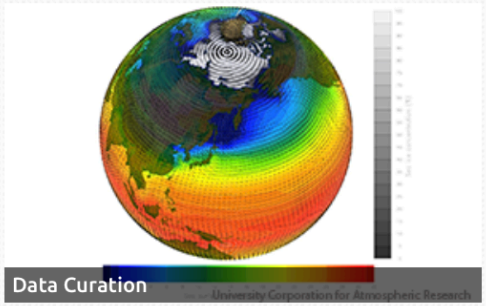
Jevin West, Information School, University of Washington

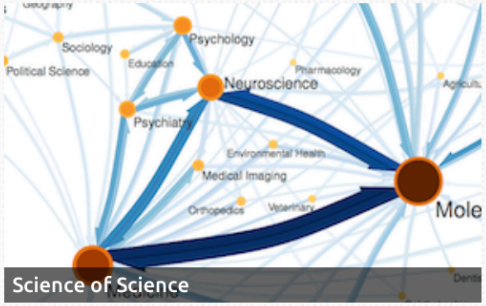




Research Focus Areas







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.

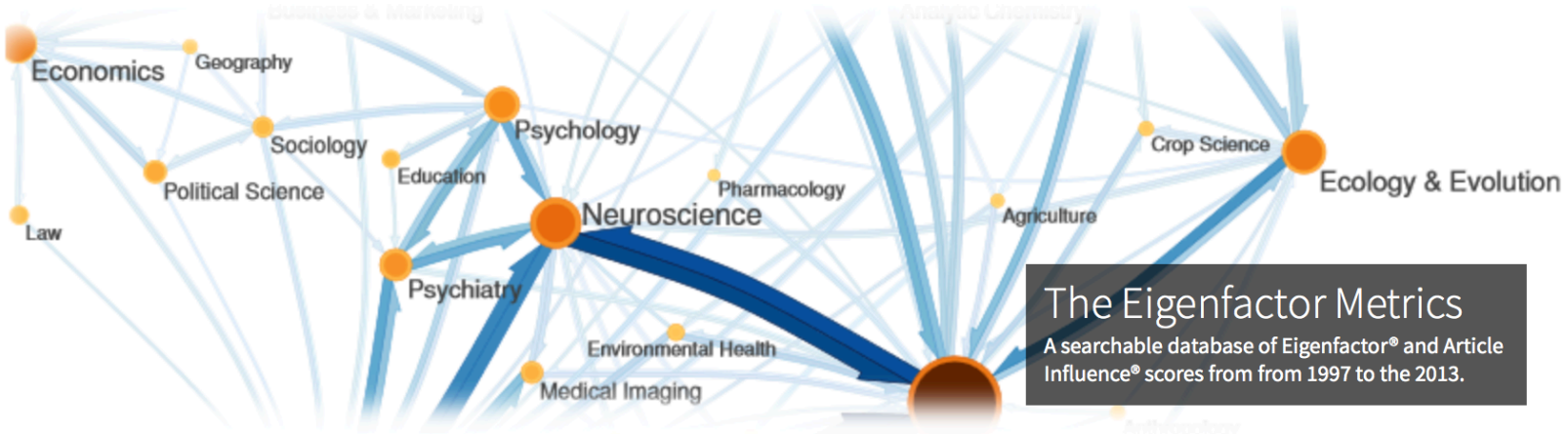
Also in... What We Do

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.

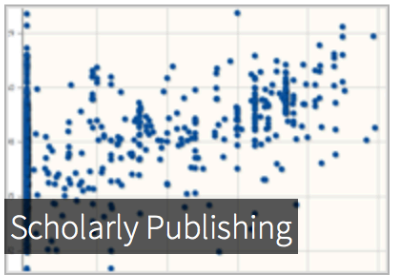
Latest eScience News

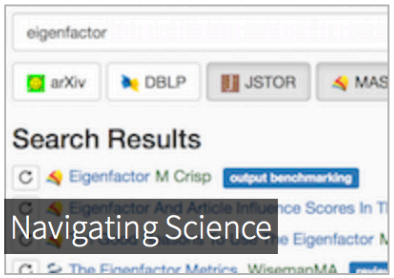
Data Science Incubation Program - Winter 2016
2 hours 4 min ago

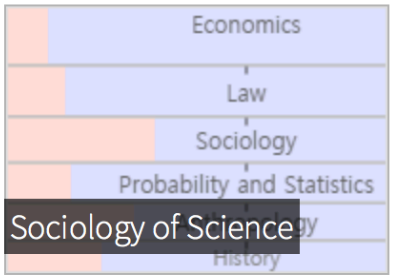
Ben Marwick On How Computers Broke Science



RESEARCH AREAS







NEWS

23 Nov.

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

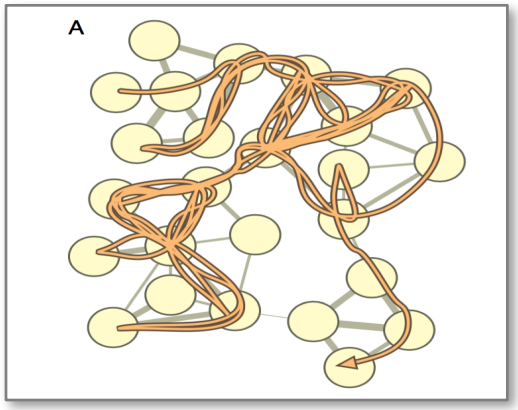
Jevin West discusses the rise of the megajournal and our open access cost effectiveness tool in the *Chronicle of Higher Education*.

---

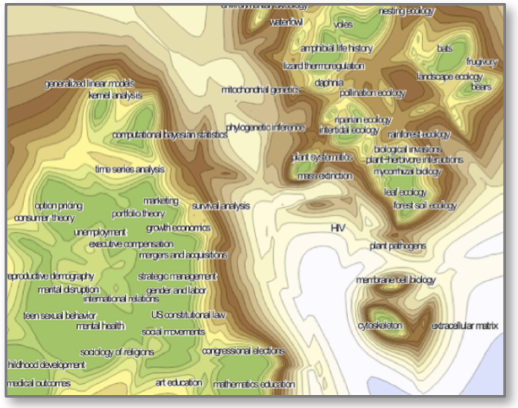
23 Nov.

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

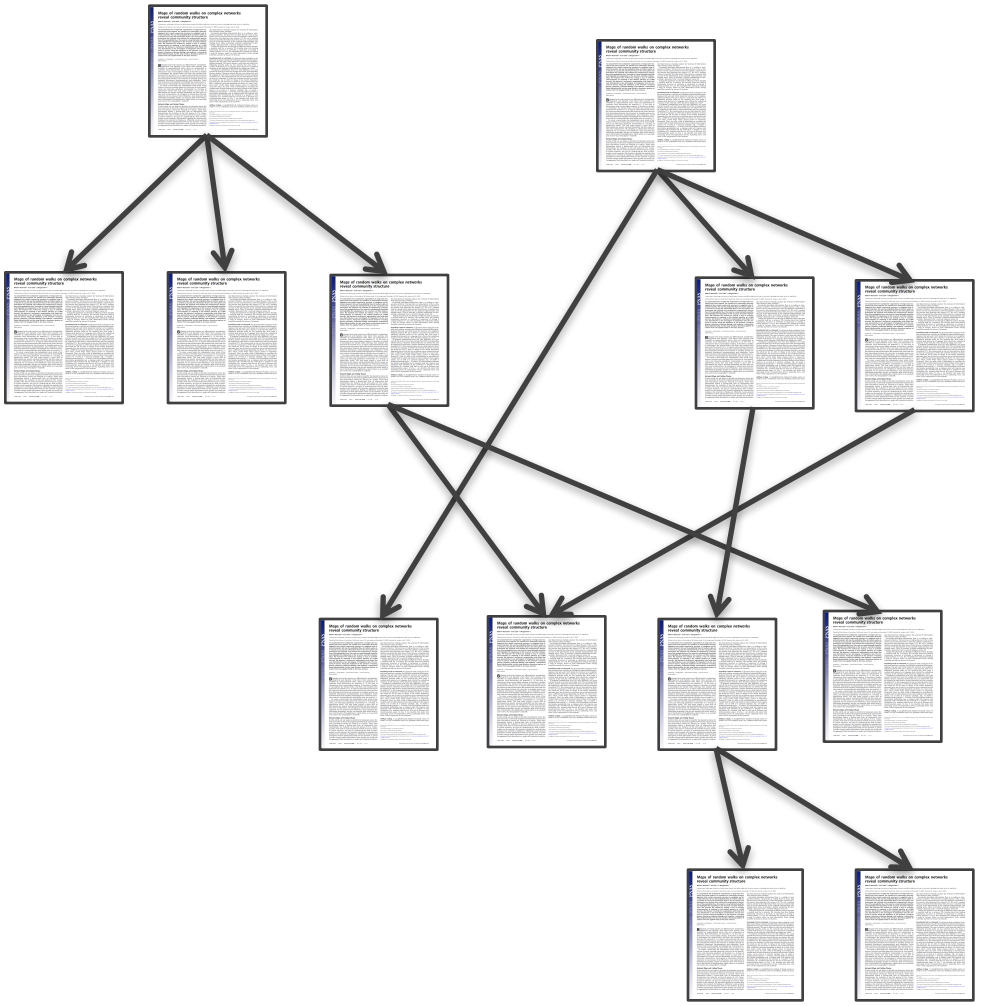


Science of Mapping



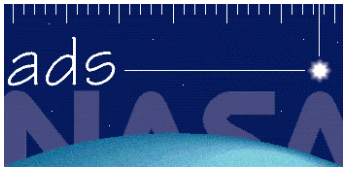
Mapping of Science

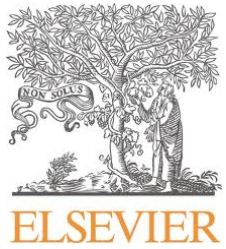
Citations form a vast network



de Solla Price, *Science* (1965)

The Scholarly Graph



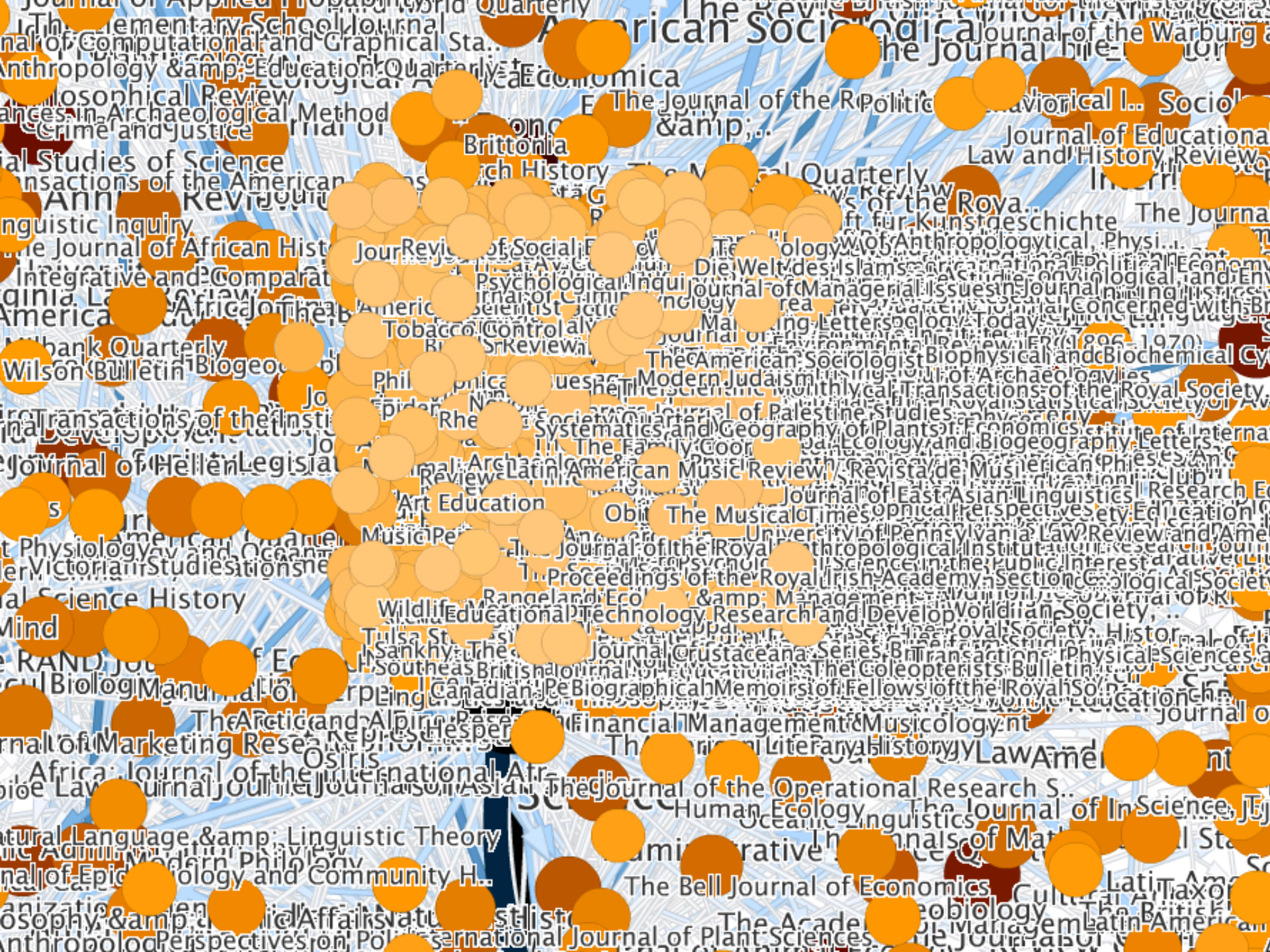


The Scholarly Graph

Tens of millions articles, patents, books

Billions of citation links

Years: 1600 - 2016


The Elementary School Journal
Journal of Computational and Graphical Sta..
Anthropology & Education Quarterly
Ecological Economica
The Journal of the R
Crime and Justice
Britton
Journal of Educationa
Law and History Review
Annual Review
Linguistic Inquiry
The Journal of African Hist
Integrative and Comparat
Psychological Inqui
Journal of Managerial Issues
Tobacco Control
Wilson Bulletin
The American Sociologist
Modern Judaism
Transactions of the Institu
Systematics and Geography of Plants
Journal of Hellen
Latin American Music Review, Revista de Musi
Art Education
The Musical Times
Victorian Studies
Proceedings of the Royal Irish Academy. Section C: Geological Society
Science History
Wildlife Educational Technology Research and Develop
Mind
Journal of Marketing Rese
Osiris
Africa: Journal of the International Af
The Journal of the Operational Research S..
Human Ecology
Natural Language & Linguistic Theory
The Bell Journal of Economics
Perspectives on Po
International Journal of Plant Sciences

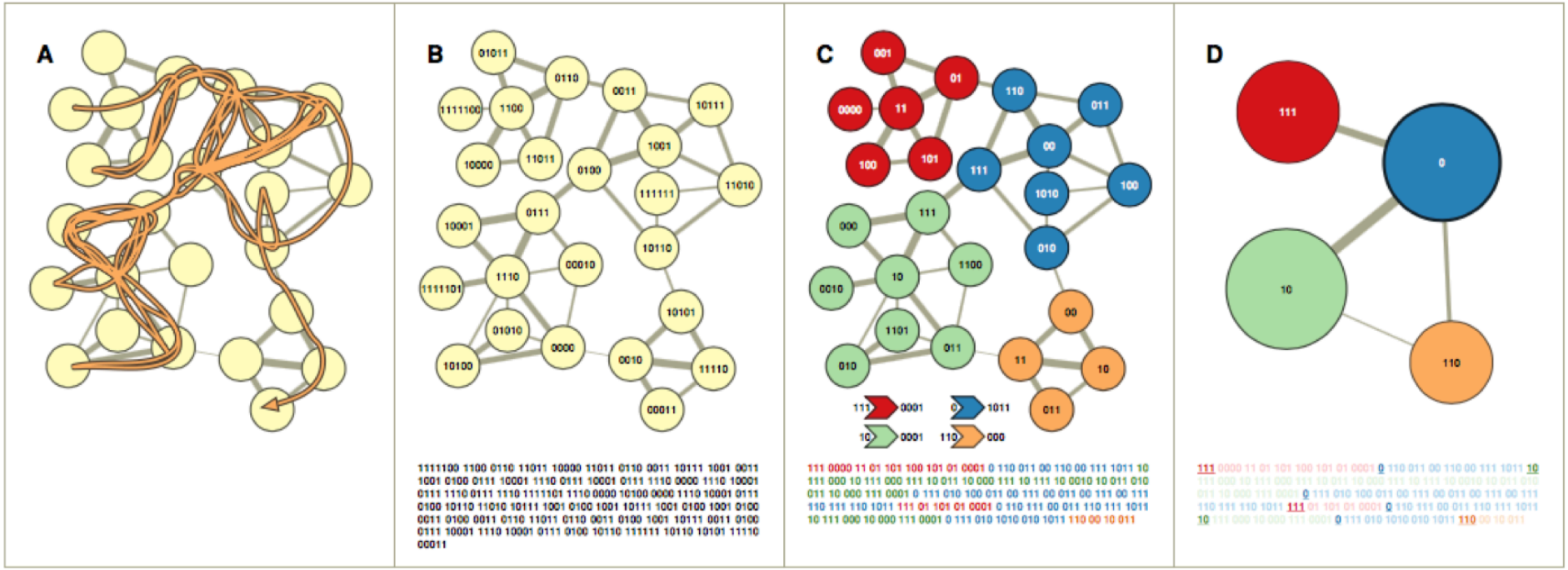
Data

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

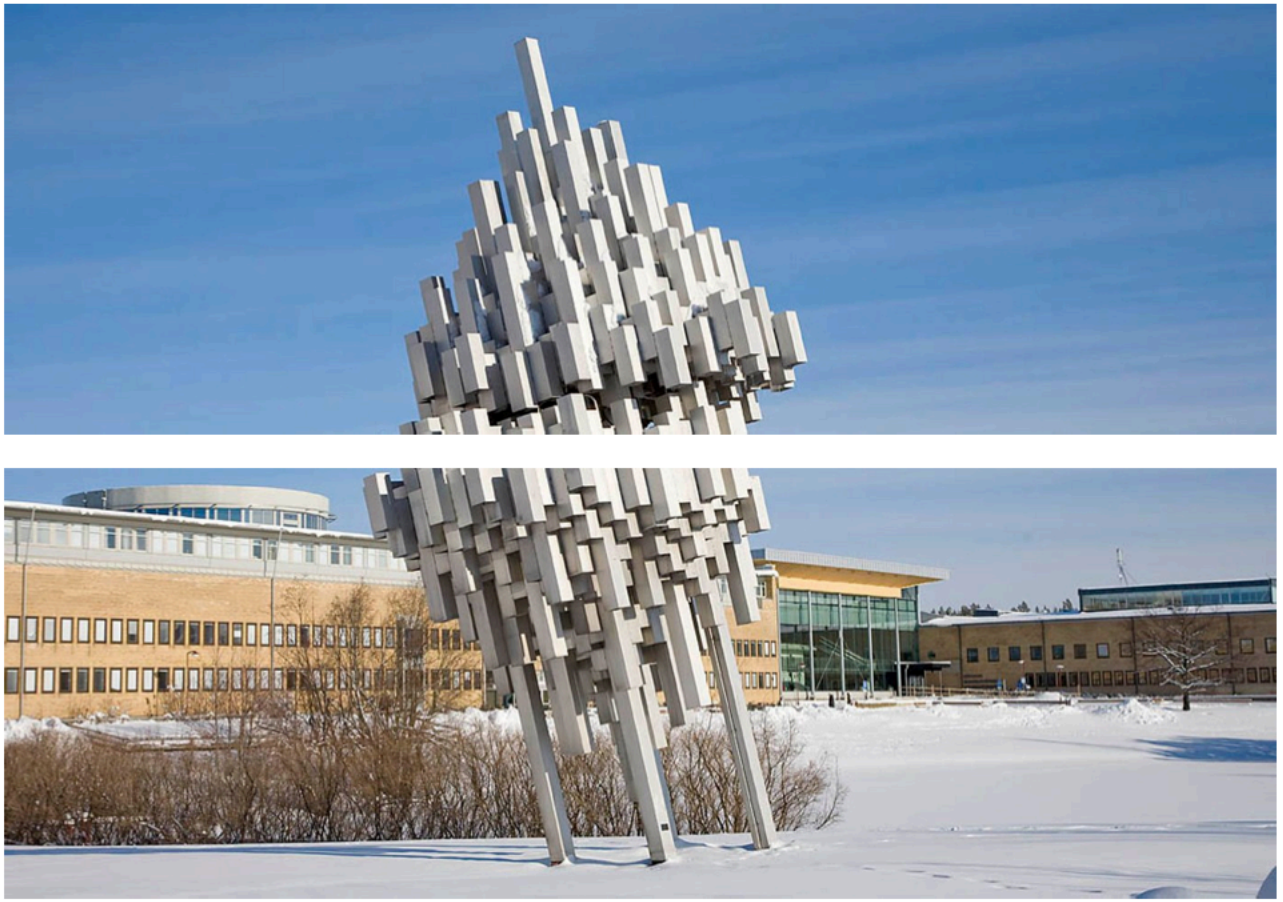
Finding regularities in the dynamics on networks



Rosvall and Bergstrom (2008) PNAS

compressing ⟷ finding patterns

5.8 MB (TIFF) ⟶ 0.9 MB (TIFF + LZW)



5.8 MB (TIFF) ⟶ 2.8 MB (TIFF + LZW)

The map equation

frequency of inter-module movements

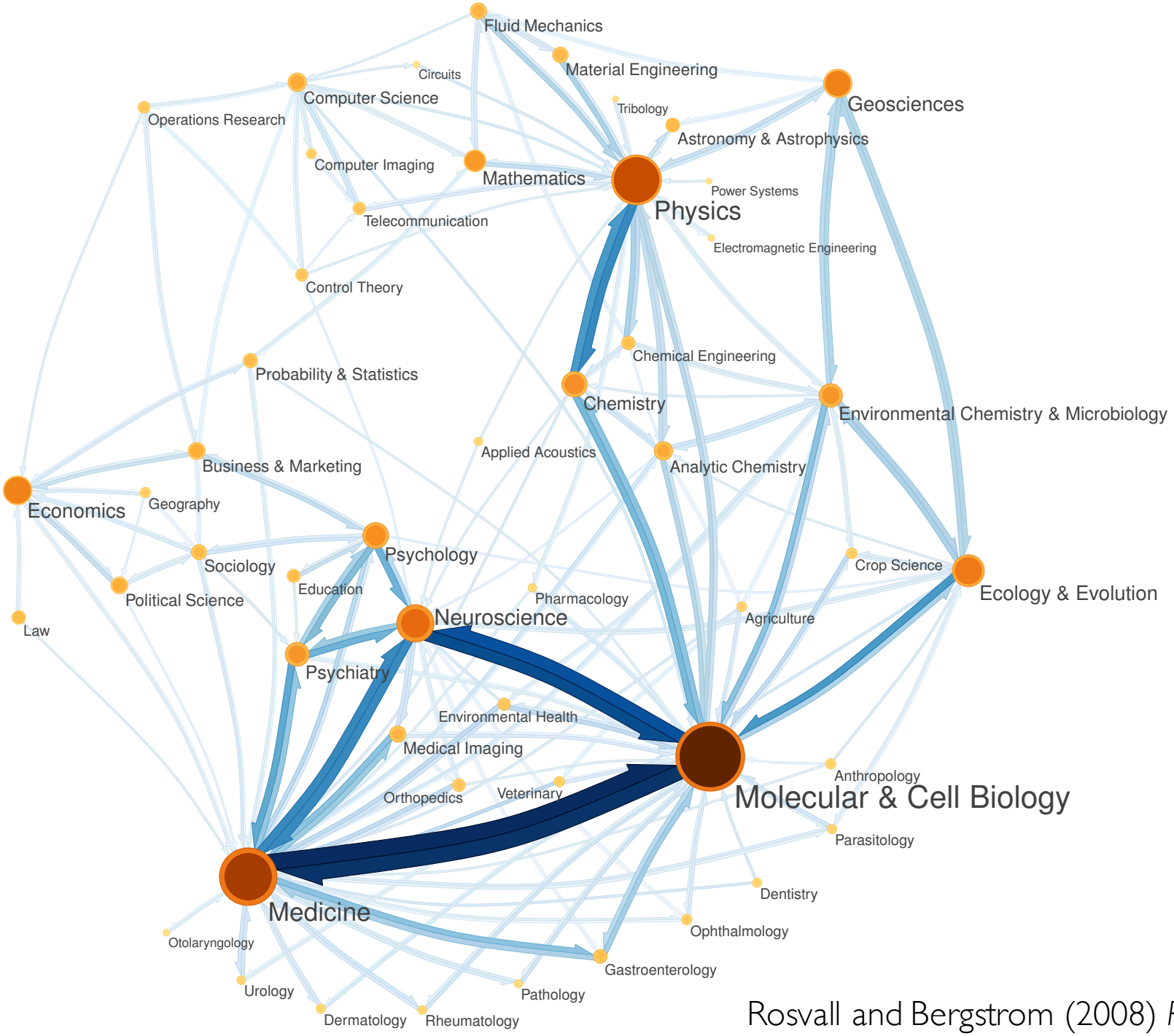
frequency of movements within module *i*

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

code length of module names

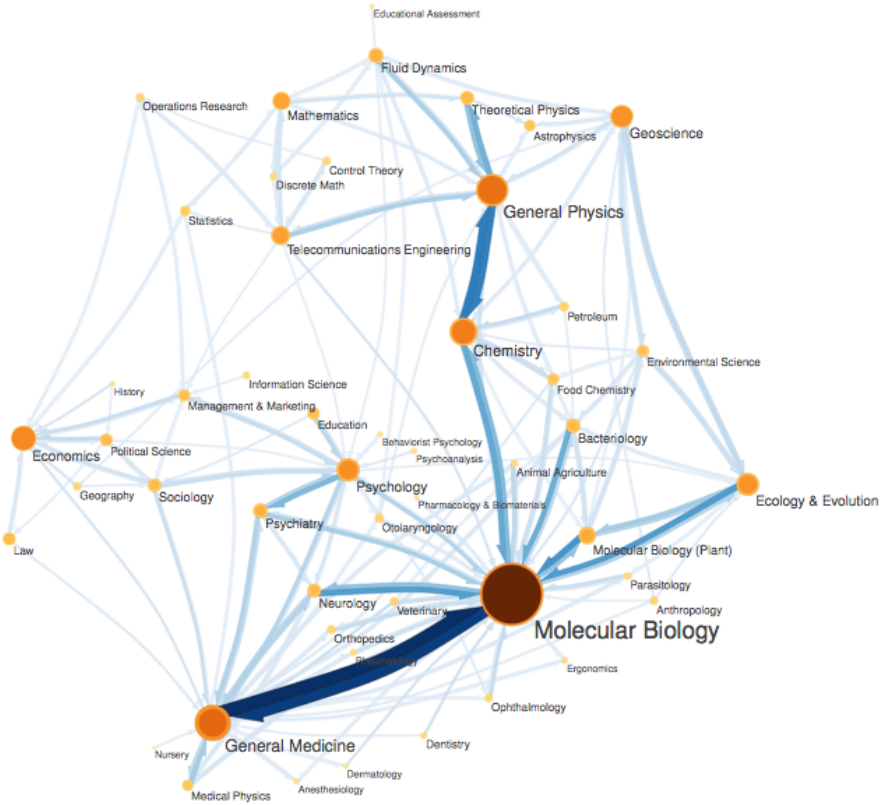
code length of node names in module *i*

Rosvall and Bergstrom (2008) *PNAS*

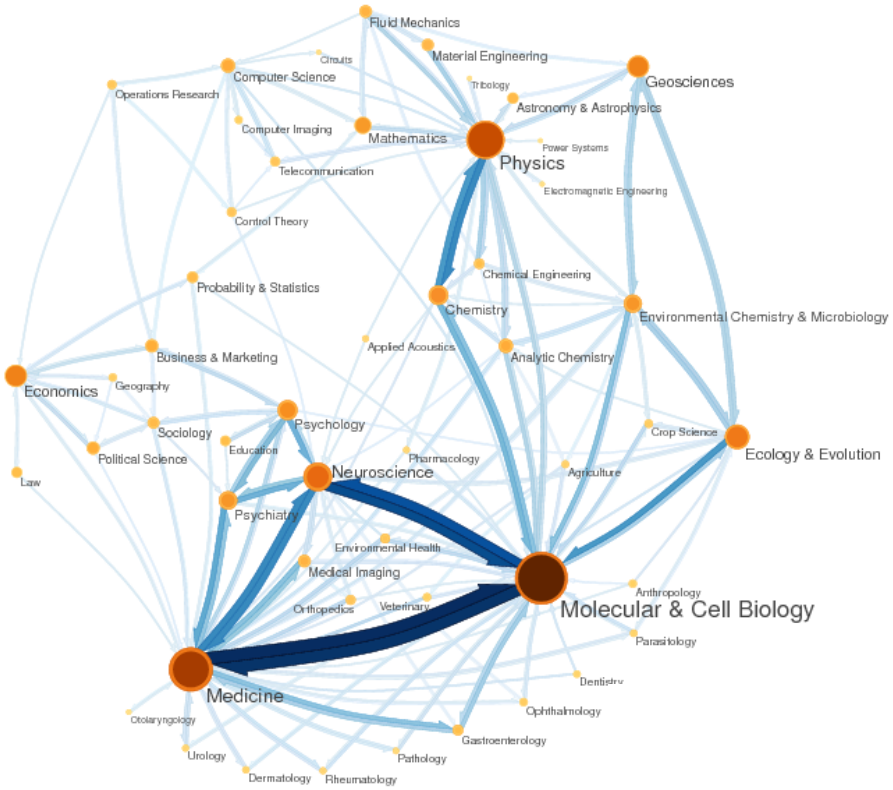


Rosvall and Bergstrom (2008) *PNAS*

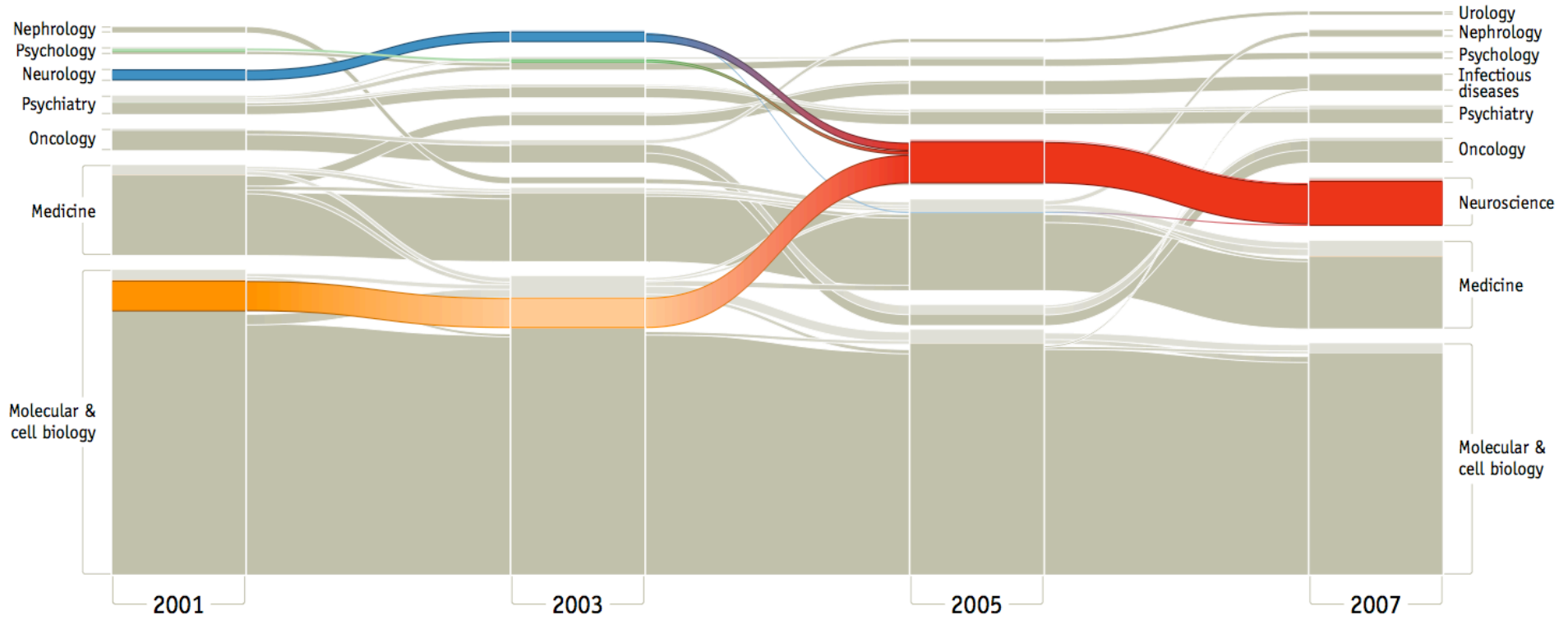
1995


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
Animal Agriculture
Geography
Sociology
Psychology
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

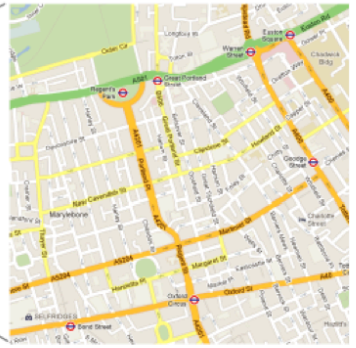
2004

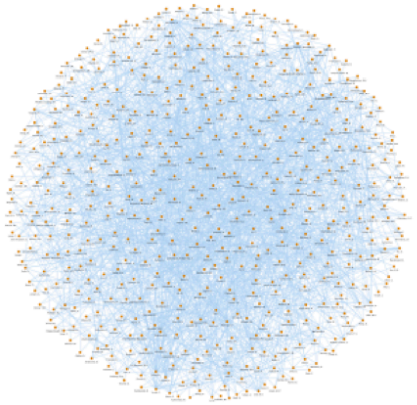

Fluid Mechanics
Material Engineering
Circuits
Computer Science
Geosciences
Tribology
Operations Research
Astronomy & Astrophysics
Computer Imaging
Mathematics
Power Systems
Physics
Telecommunication
Electromagnetic Engineering
Control Theory
Chemical Engineering
Probability & Statistics
Chemistry
Environmental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

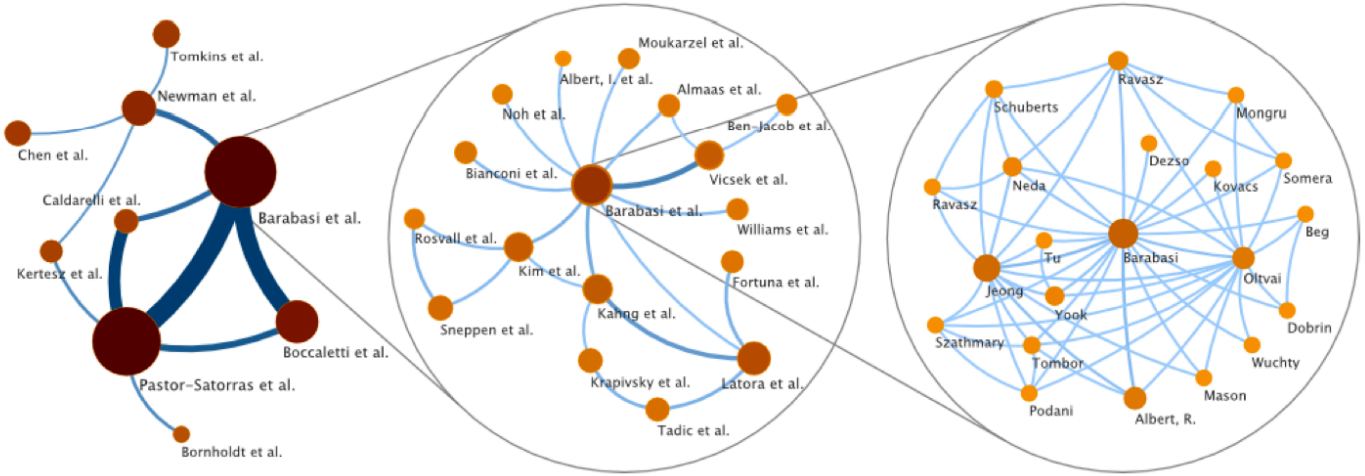
The Emergence of Neuroscience

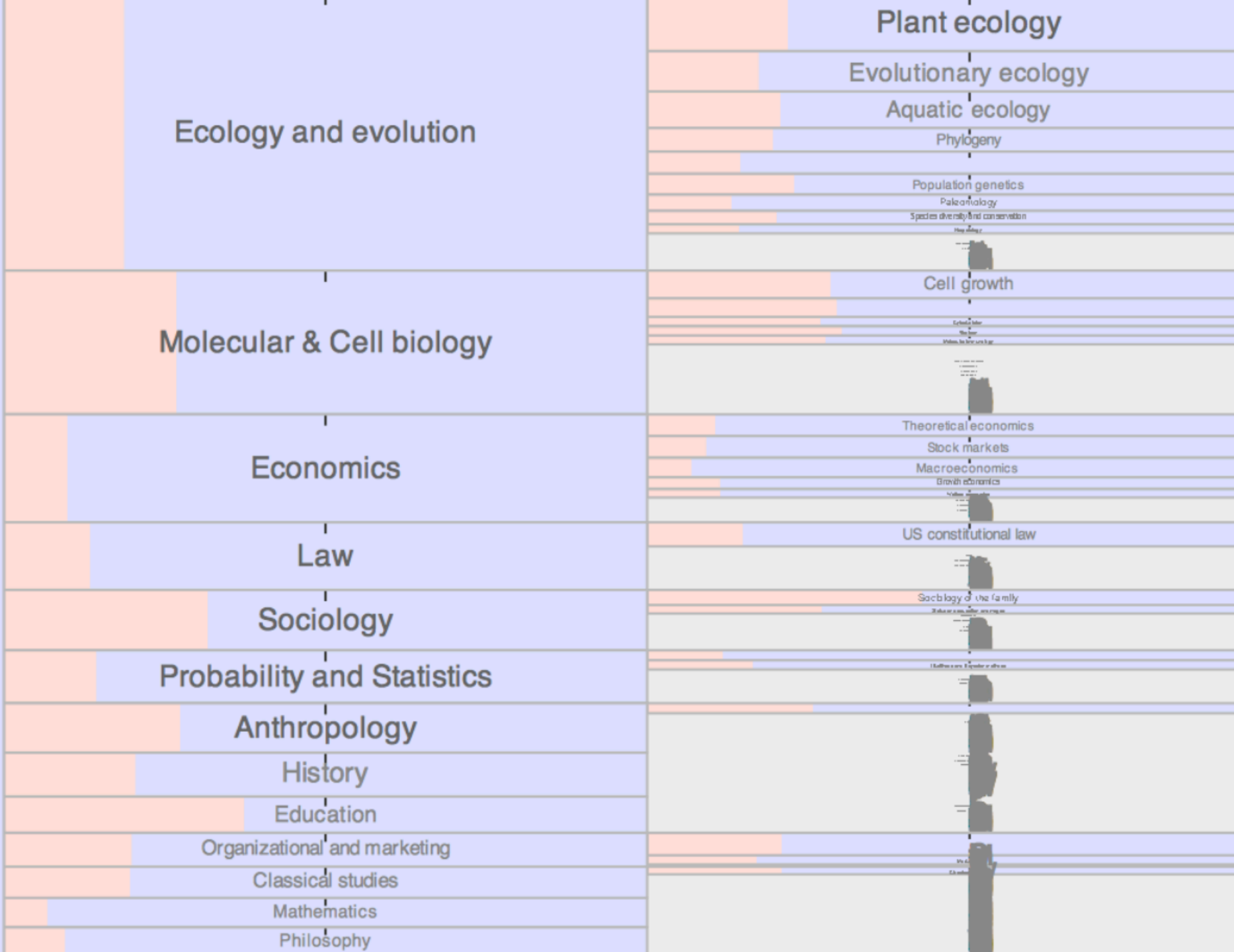


Rosvall and Bergstrom (2010) PLoS One

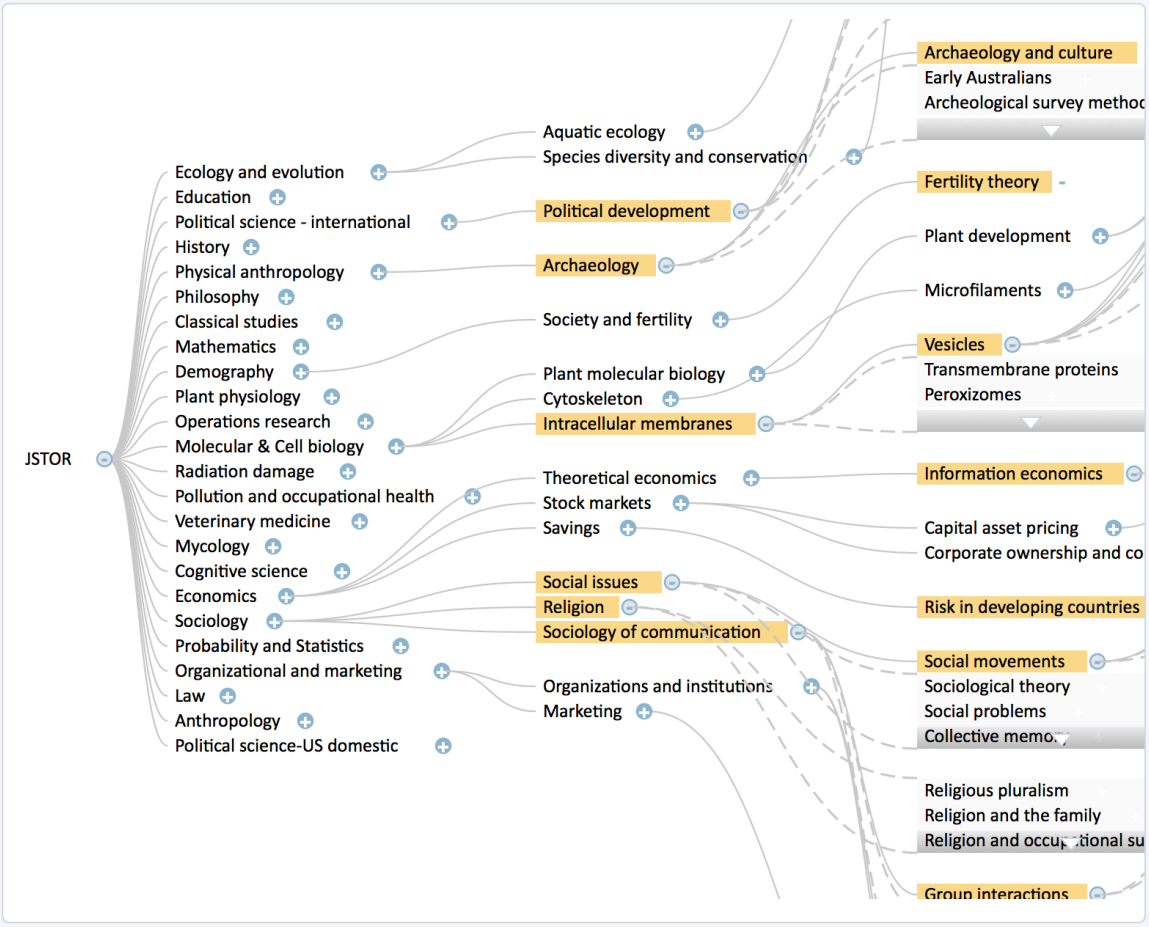





Tomkins et al.
Newman et al.
Chen et al.
Caldarelli et al.
Barabasi et al.
Kertesz et al.
Boccaletti et al.
Pastor-Satorras et al.
Bornholdt et al.
Moukarzel et al.
Albert, I. et al.
Almaas et al.
Noh et al.
Ben-Jacob et al.
Bianconi et al.
Vicsek et al.
Barabasi et al.
Williams et al.
Rosvall et al.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


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)

Next Steps

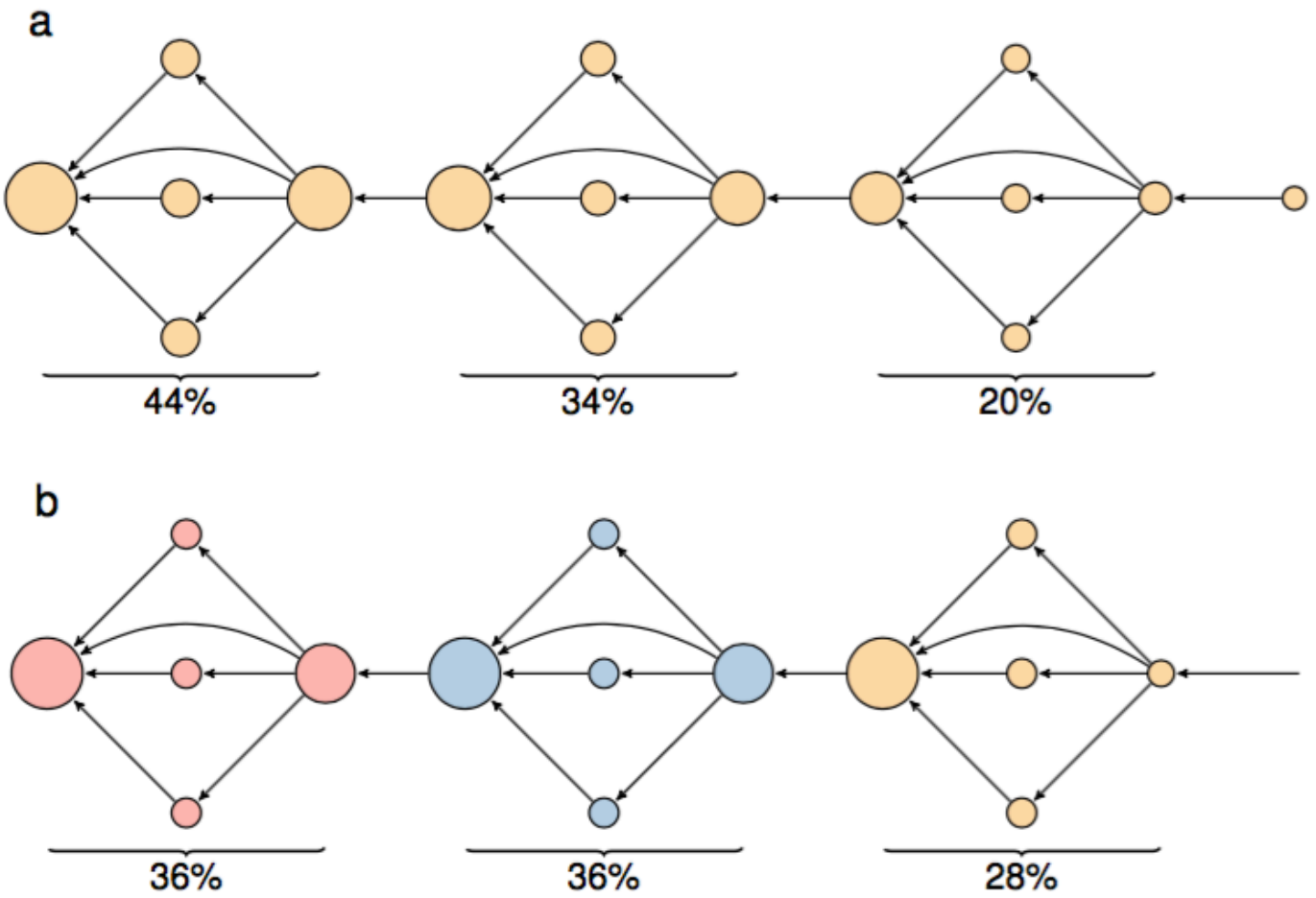
Higher Resolution Maps

Article-level Ranking

Author-centric Maps

Interactive Visualizations

Article-level Ranking and Mapping



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

WSDM CUP CHALLENGE

SIGN-UPS FOR THE WSDM CUP CHALLENGE ARE NOW CLOSED

The Graph

The Microsoft Academic Graph is a heterogeneous graph containing scientific publication records, citation relationships between publications, as well as authors, institutions, journal and conference "venues," and fields of study.

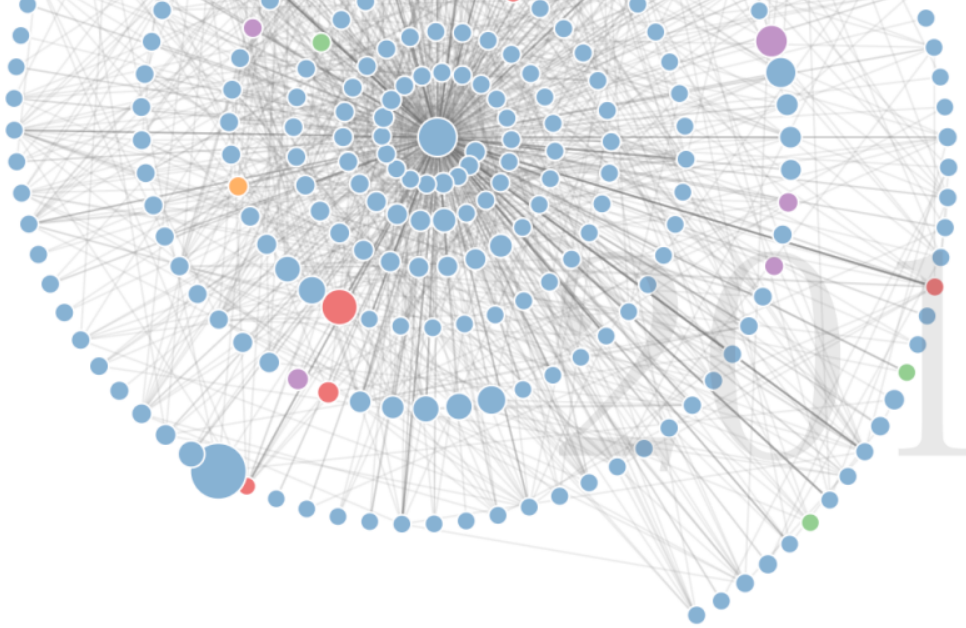
The Data

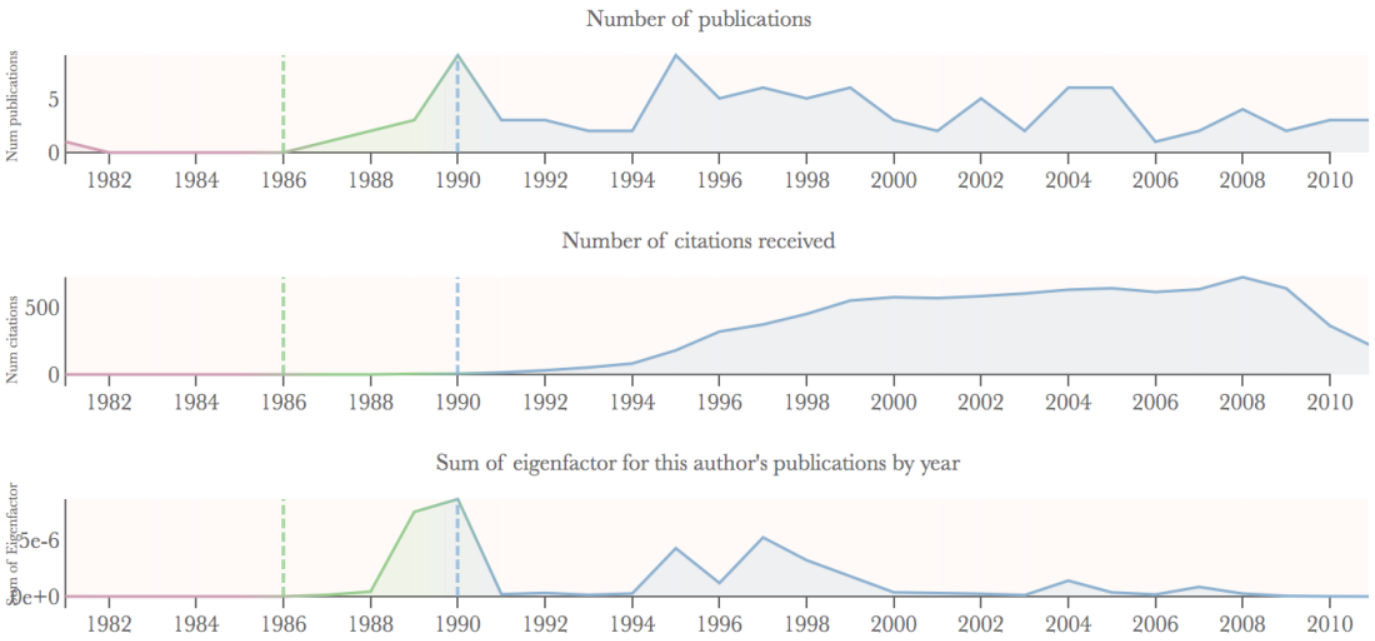
This data is available as a set of zipped text files stored in Microsoft Azure blob storage and available via HTTP. The file size (zipped) is ~30GB and may be downloaded here.

The Challenge

The goal of the Ranker Challenge is to assess the query-independent importance of scholarly articles, using data from the Microsoft Academic Graph.

Jason Portenoy





scholar.eigenfactor.org

scholar.eigenfactor.org

Visualizing Scholarly Influence Over Time

Influence of Pew Scholars

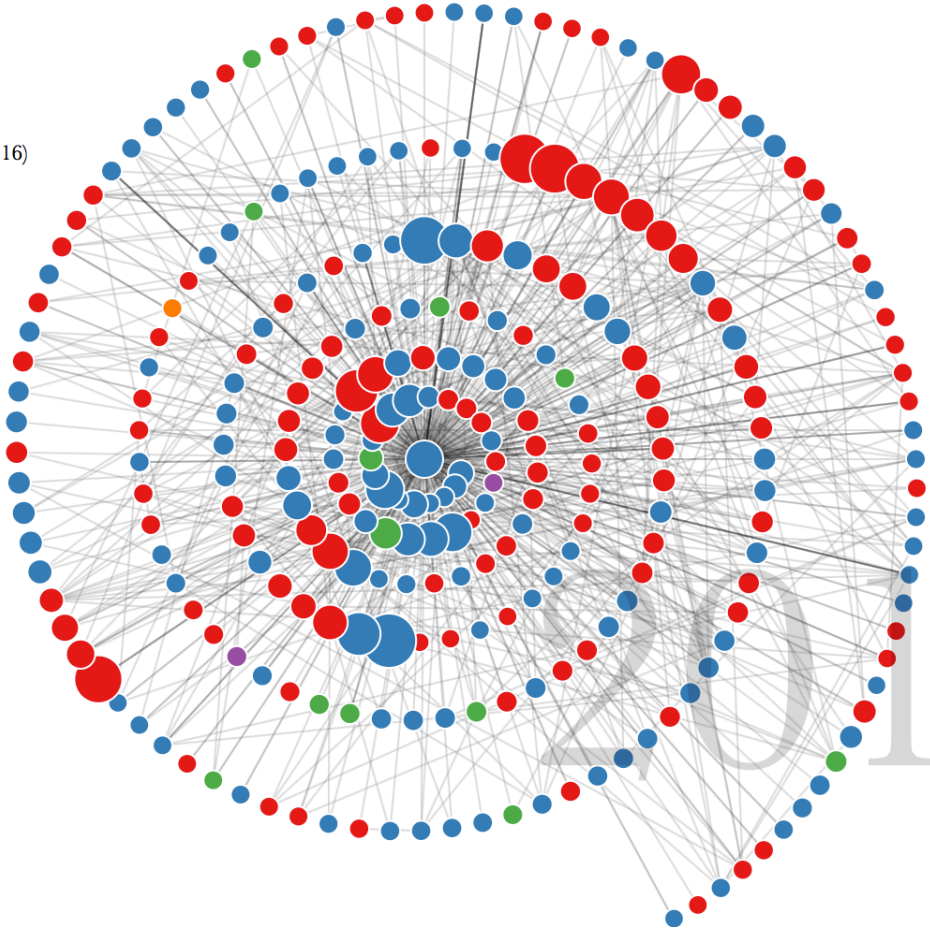
Roberta A. Gottlieb

Learn More

Papers in category "Medicine" (domain 6)
Papers in category "Biology" (domain 4)
Papers in category "Chemistry" (domain 5)
Papers in category "Unknown" (domain 0)
Papers in category "Agriculture Science" (domain 16)

Roberta A. Gottlieb

Pew Scholar
1997



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

Visualizing Scholarly Influence Over Time

Influence of Pew Scholars

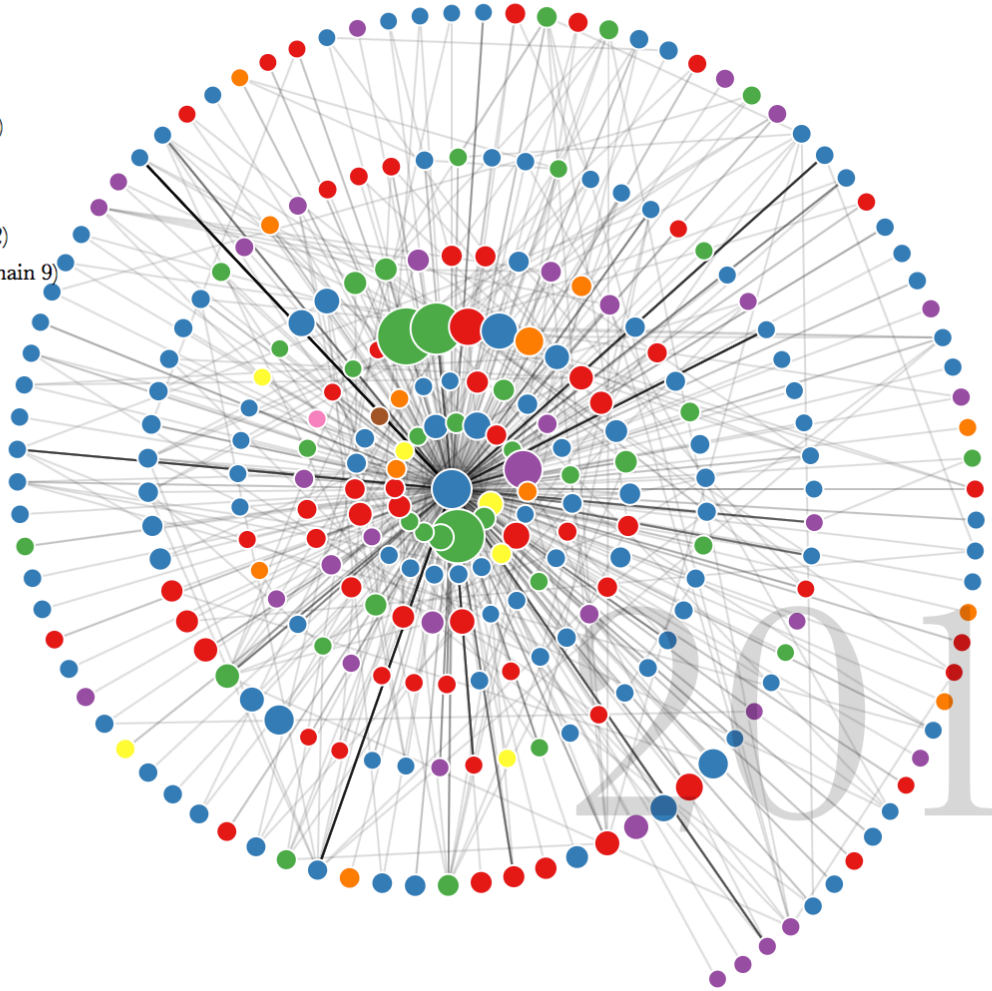
Mark W. Grinstaff

Learn More

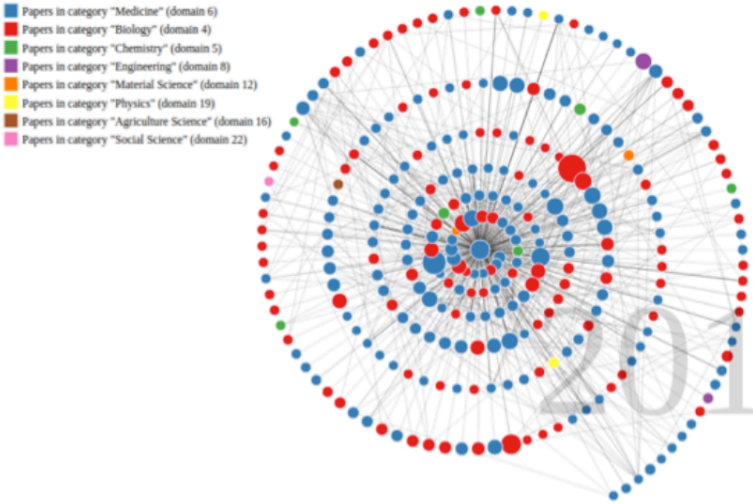
- Papers in category "Chemistry" (domain 5)
- Papers in category "Medicine" (domain 6)
- Papers in category "Biology" (domain 4)
- Papers in category "Material Science" (domain 12)
- Papers in category "Engineering" (domain 8)
- Papers in category "Physics" (domain 19)
- Papers in category "Computer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

Mark W.
Grinstaff

Pew Scholar
1999

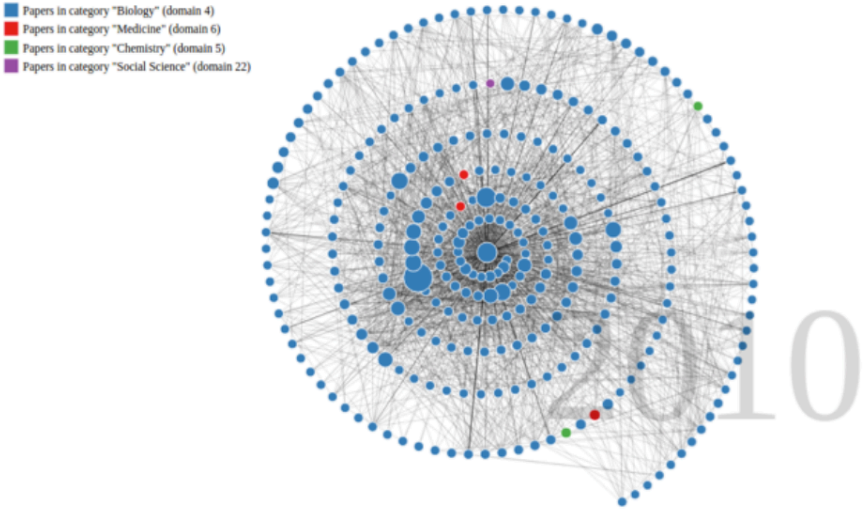


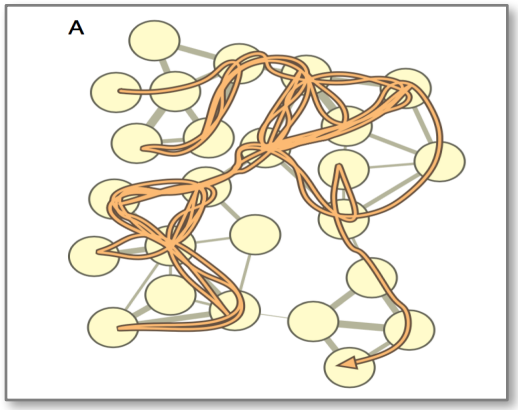
Comparing Authors



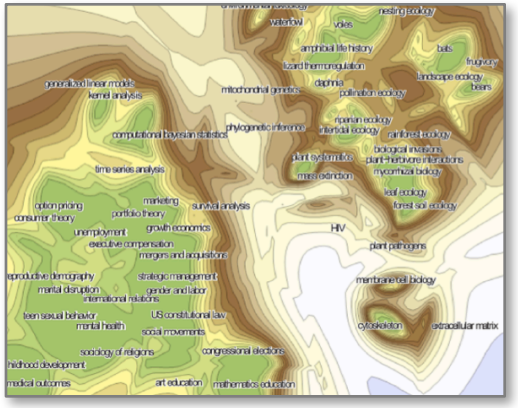
A more sparse network indicates fewer citations between papers shown in the network. This could be a result of the central scholar having impact across a wider set of academic communities.

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





Science of Mapping



Mapping of Science

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

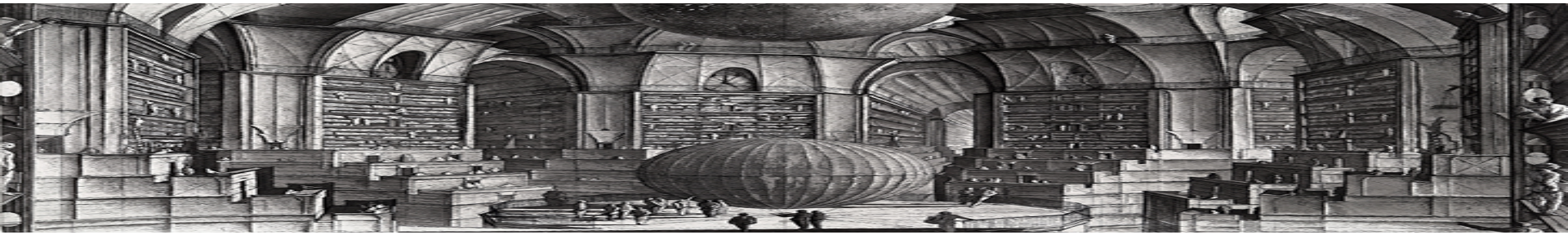
Habitat conservation

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

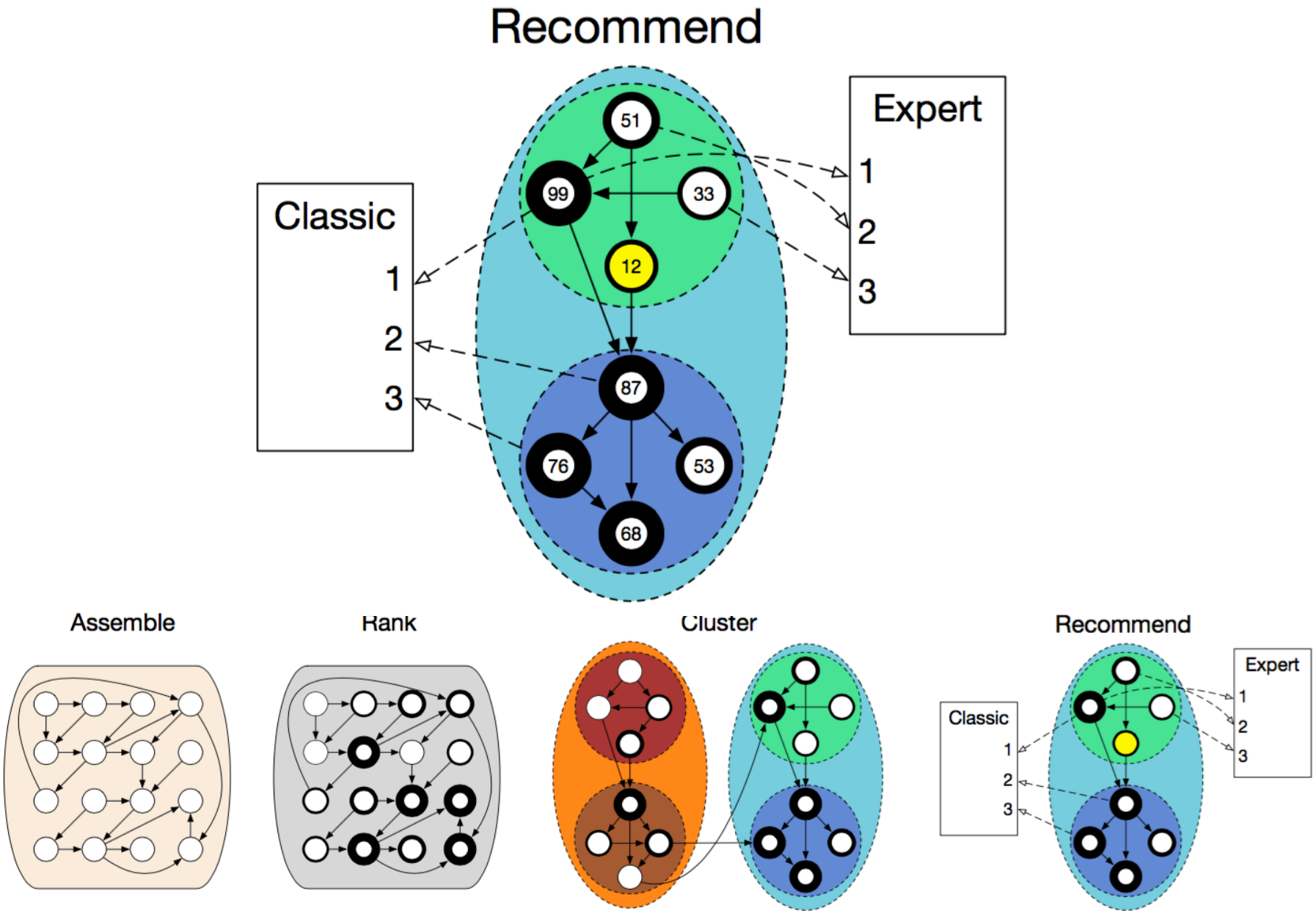
JW

Explore the recommendations

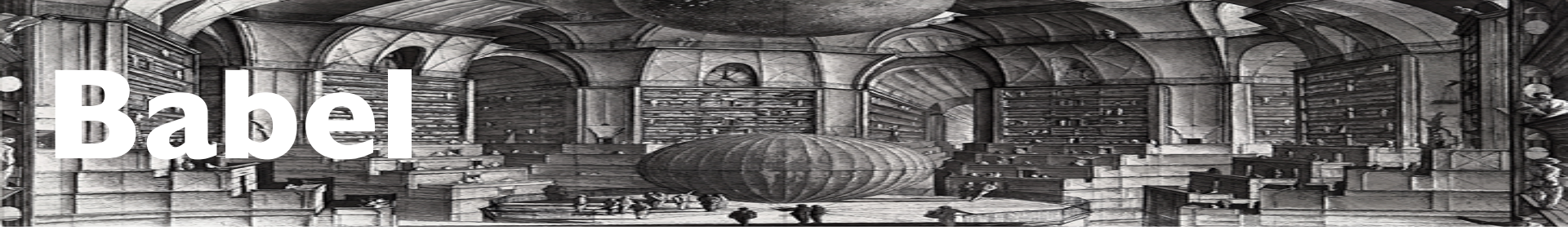
babel.eigenfactor.org



jevinw@uw.edu



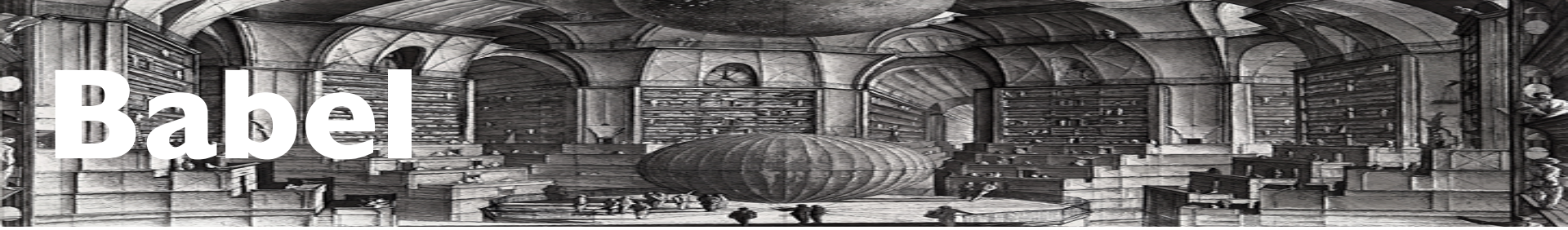
West, Wesley-Smith, Bergstrom (2016) A recommendation system based on hierarchical clustering of an article-level citation network. *IEEE Transactions on Big Data*



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

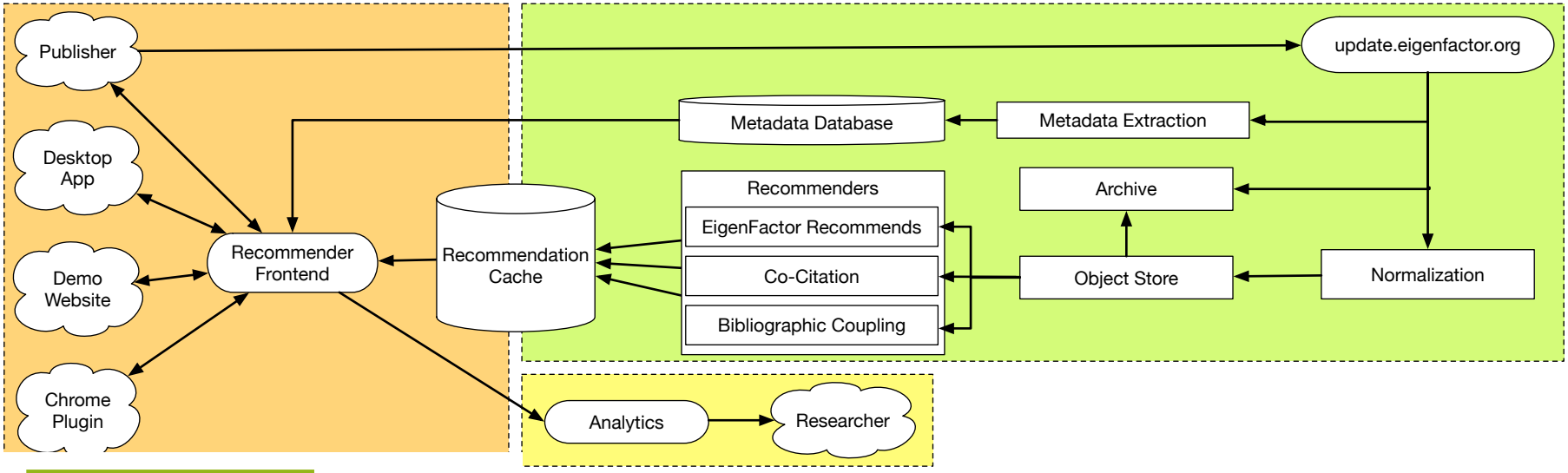
babel.eigenfactor.org



Babel

- Promote creation of better tools for scholarly authoring
- Allows 3rd parties to easily integrate with publishers, researchers
  - Examples: Zotero Plugin, Google Scholar Browser Plugin
- Easily test out new recommendations from other groups
  - Just change your end point!
- Publishers can host their own endpoints, meet their reliability requirements

Babel Architecture



swagger

✔ Processed with no error

/recommendation/{publisher}/{paper_id}

GET /recommendation/{publisher}/{paper_id}

Description

Generates recommendations for `paper_id`

Parameters

Name	Located in	Description	Required	Schema
paper_id	path	Publisher assigned identifier of a paper	Yes	string
publisher	path	Publisher to perform this operation on	Yes	string
client_id	query	Identifier provided by the platform to clients to track client usage.	No	string
limit	query	Maximum number of recommendations to return	No	integer
algorithm	query	Algorithm to generate recommendations with. If not provided a random algorithm will be used.	No	string

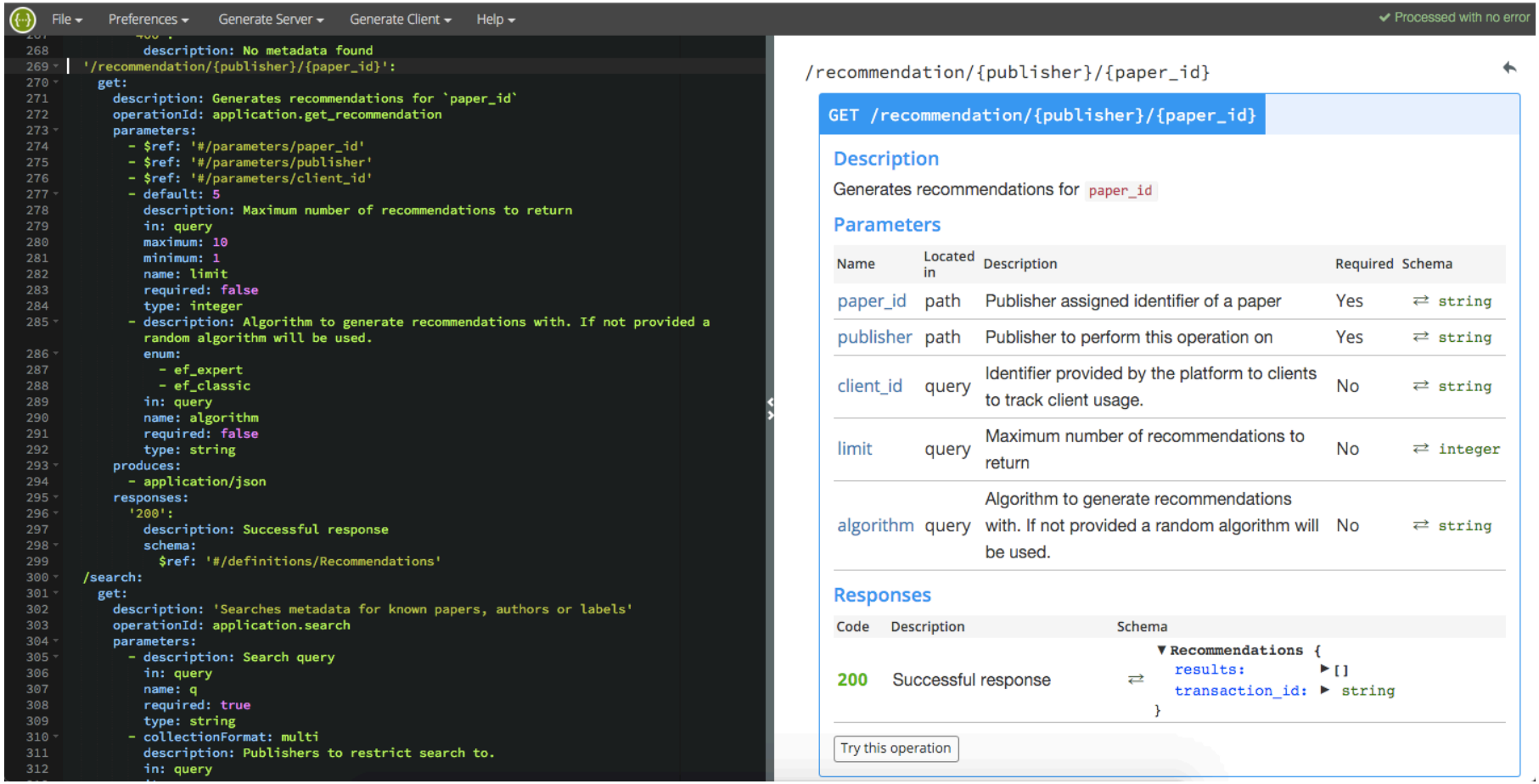
Responses

Code	Description	Schema
200	Successful response	▼ Recommendations { results: ▶ [] transaction_id: ▶ string }

Try this operation

Application	Flask — web development, one drop at a time
Package	docker
Deploy	AWS Elastic Bean Stalk

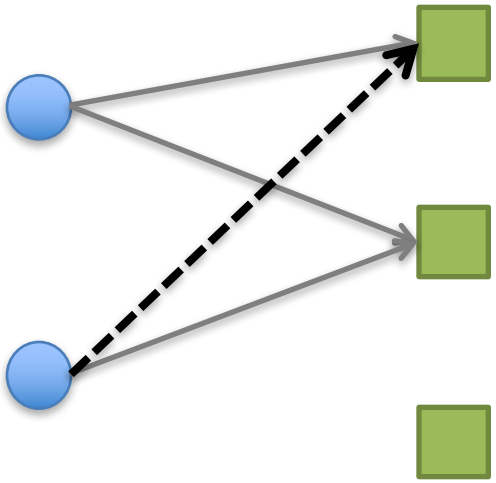
Open API


File
Preferences
Generate Server
Generate Client
Help
Processed with no error
description: No metadata found
'/recommendation/{publisher}/{paper_id}':
get:
description: Generates recommendations for `paper_id`
operationId: application.get_recommendation
parameters:
- $ref: '#/parameters/paper_id'
- $ref: '#/parameters/publisher'
- $ref: '#/parameters/client_id'
- default: 5
description: Maximum number of recommendations to return
in: query
maximum: 10
minimum: 1
name: limit
required: false
type: integer
- description: Algorithm to generate recommendations with. If not provided a random algorithm will be used.
enum:
- ef_expert
- ef_classic
in: query
name: algorithm
required: false
type: string
produces:
- application/json
responses:
'200':
description: Successful response
schema:
$ref: '#/definitions/Recommendations'
/search:
get:
description: 'Searches metadata for known papers, authors or labels'
operationId: application.search
parameters:
- description: Search query
in: query
name: q
required: true
type: string
- collectionFormat: multi
description: Publishers to restrict search to.
in: query
/recommendation/{publisher}/{paper_id}
GET /recommendation/{publisher}/{paper_id}
Description
Generates recommendations for paper_id
Parameters
Name
Located in
Description
Required
Schema
paper_id
path
Publisher assigned identifier of a paper
Yes
string
publisher
path
Publisher to perform this operation on
Yes
string
client_id
query
Identifier provided by the platform to clients to track client usage.
No
string
limit
query
Maximum number of recommendations to return
No
integer
algorithm
query
Algorithm to generate recommendations with. If not provided a random algorithm will be used.
No
string
Responses
Code
Description
Schema
200
Successful response
Recommendations {
results: []
transaction_id: string
}
Try this operation

JW

A-B Testing

Usage-based

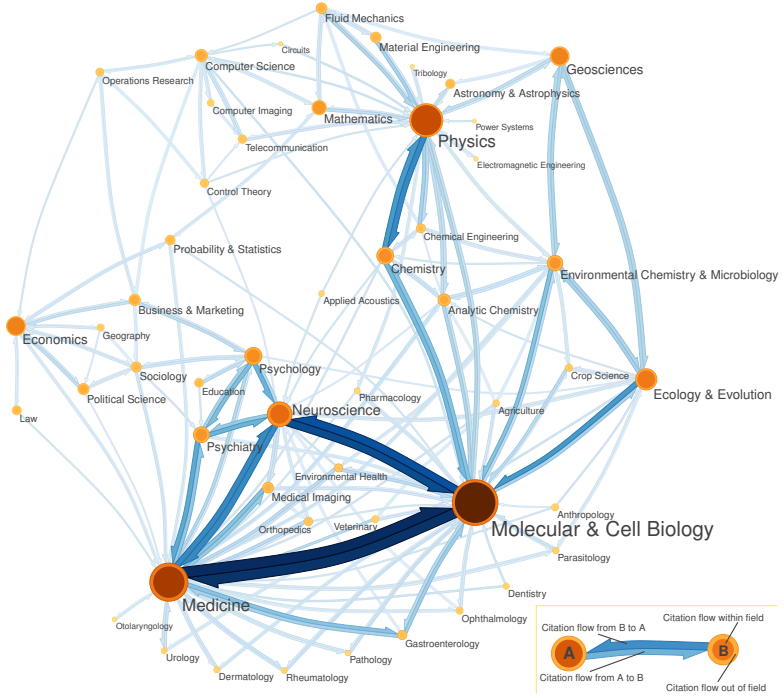


People

Papers

Utilizes download paper of similar readers

Citation-based



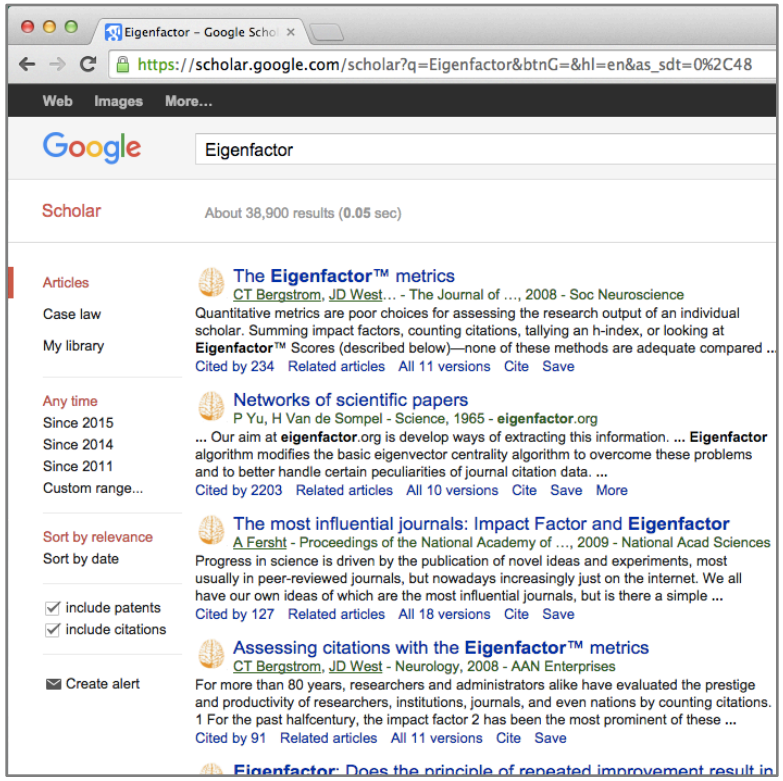
Utilizes hierarchical structure of citation graph and the relative position of papers

I. Wesley-Smith, R. Dandrea, J.D. West (2015) An Experimental Platform for Scholarly Article recommendation. *Information Retrieval (BIR2015)*

Browser Plugins

Firefox

chrome



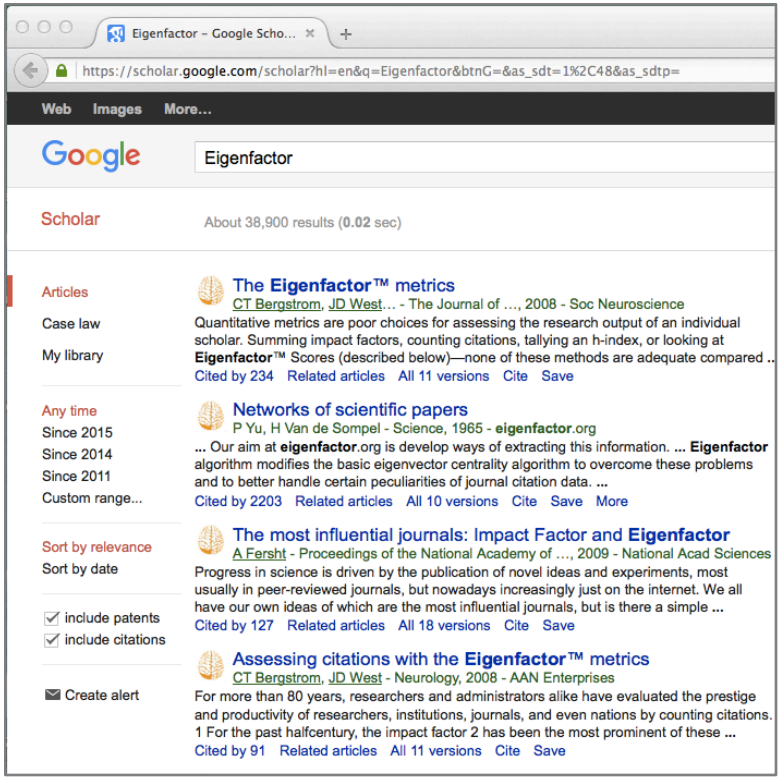
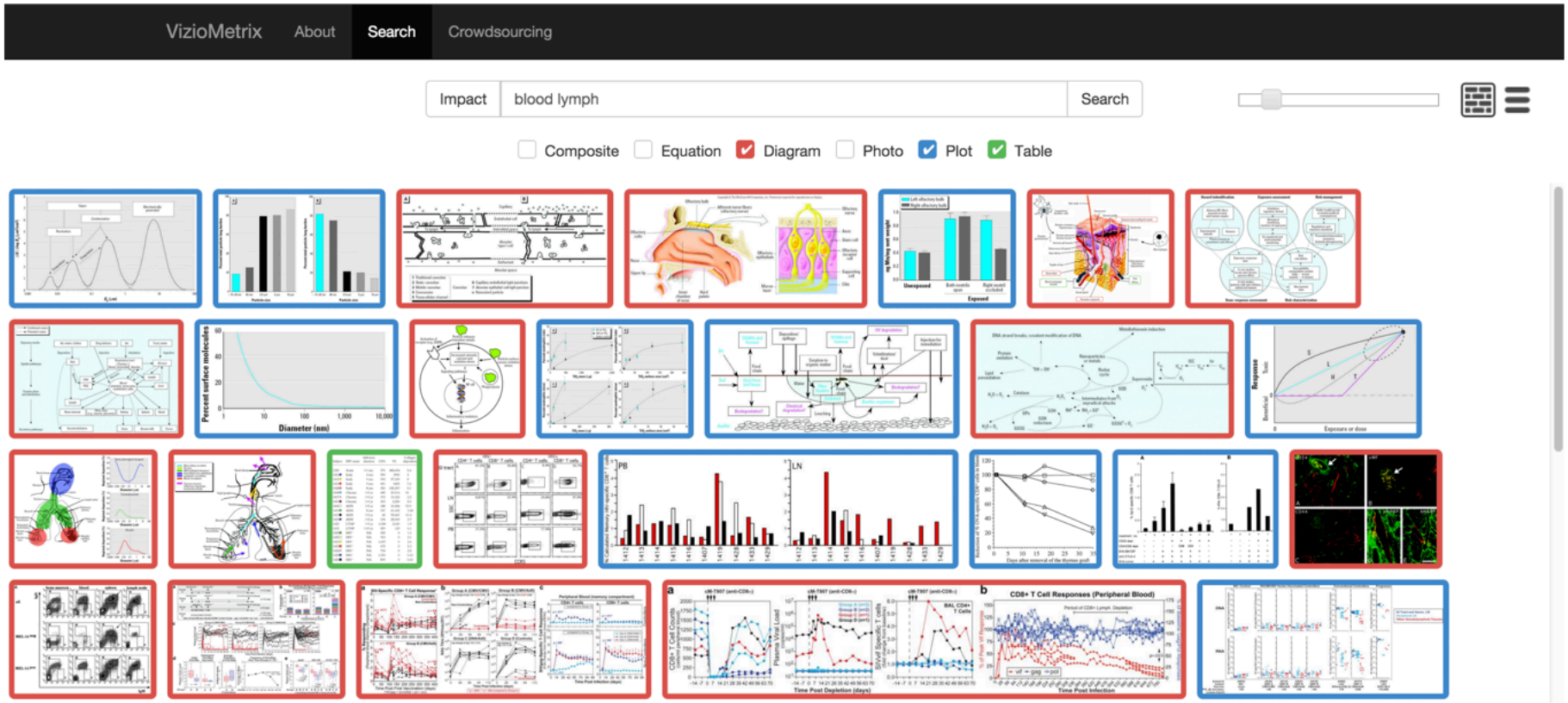


Figure-Centric Search Engine

viziometrics.org



Questions

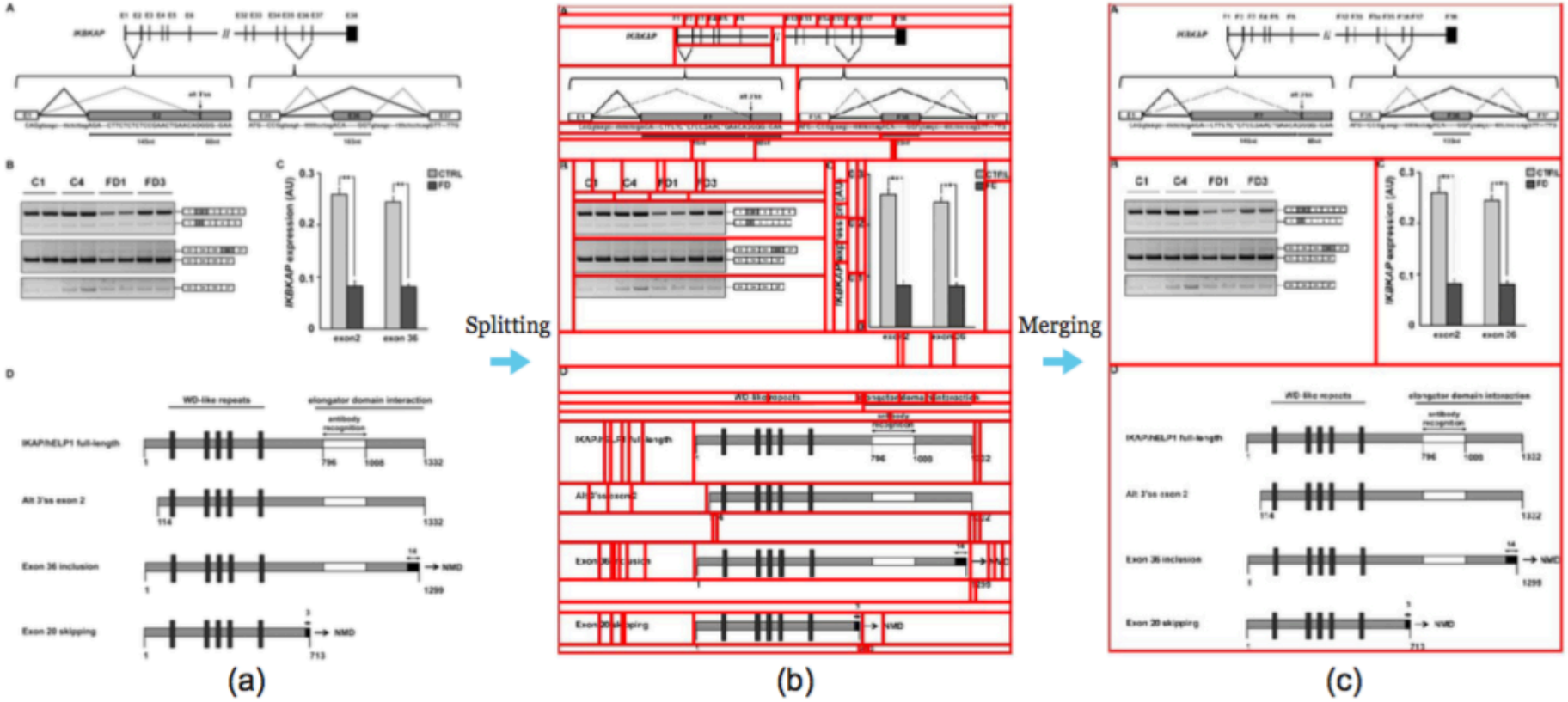
- How do patterns of encoding visual information in the literature vary across disciplines?
- How have patterns of encoding visual information in the literature evolved over time?
- Is there any link between patterns of encoding visual information and scientific impact?

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

650,000 papers

5 million figures

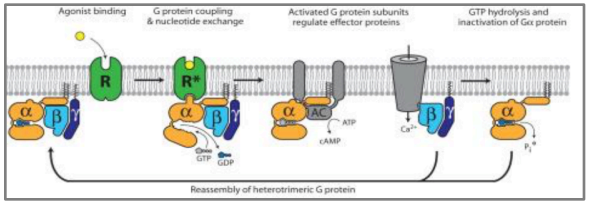
Composite Figure Dismantling



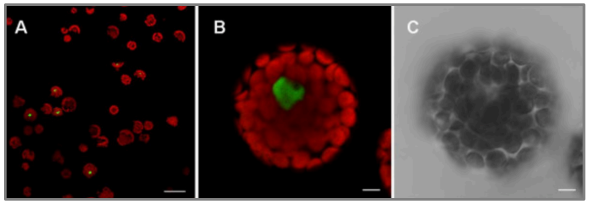
Lee et al (2015)

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

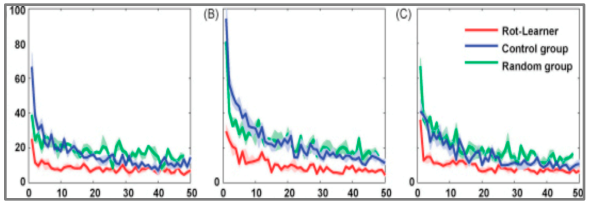
Equations (394)



Schematics (769)



Photos (782)

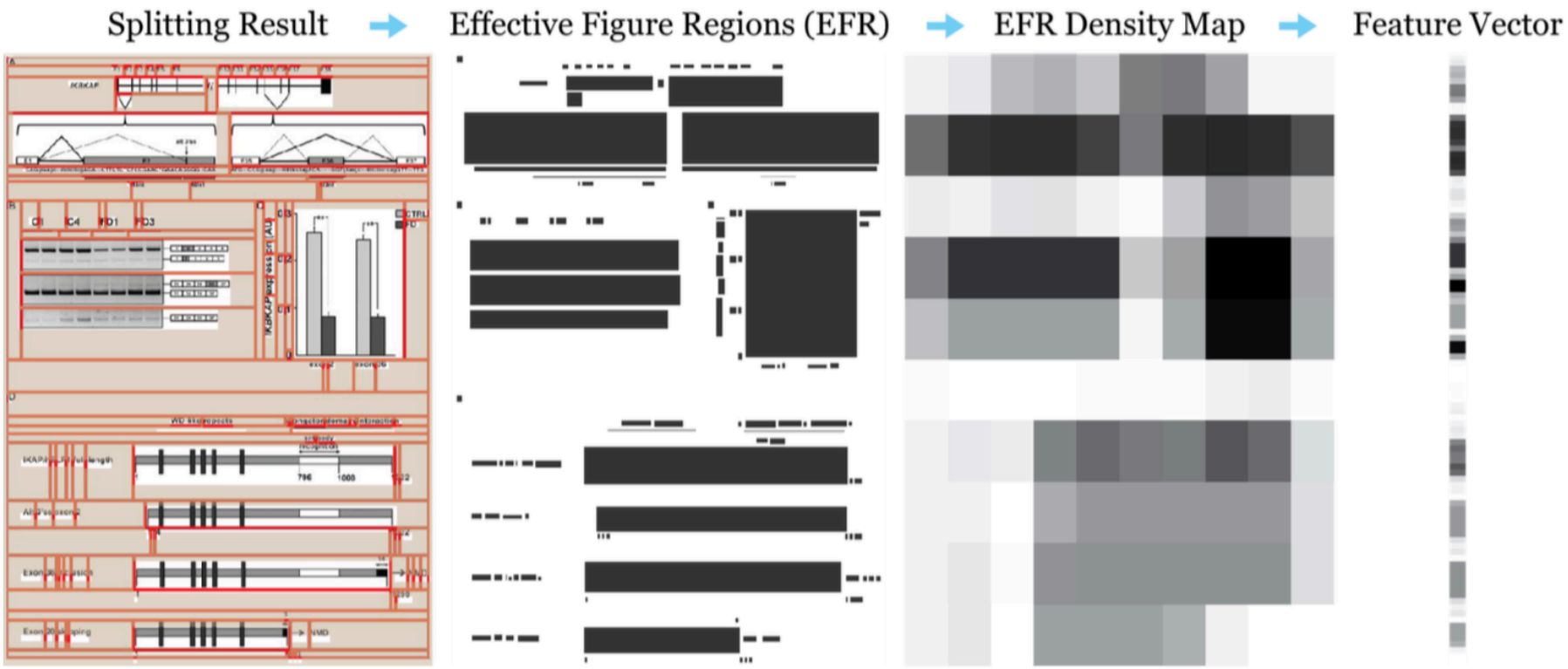


Plots (890)

	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

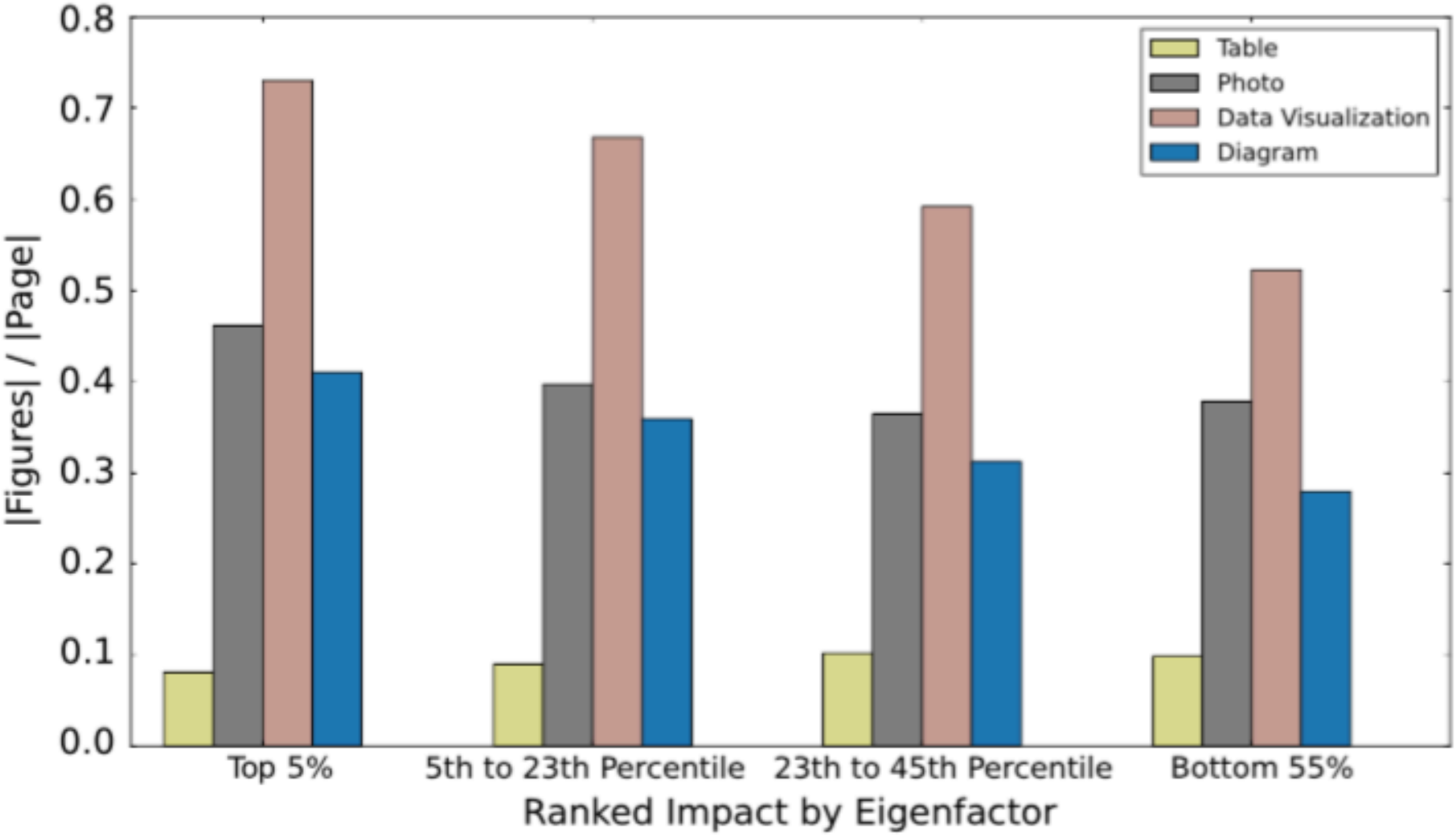
Tables (436)

Feature Extraction

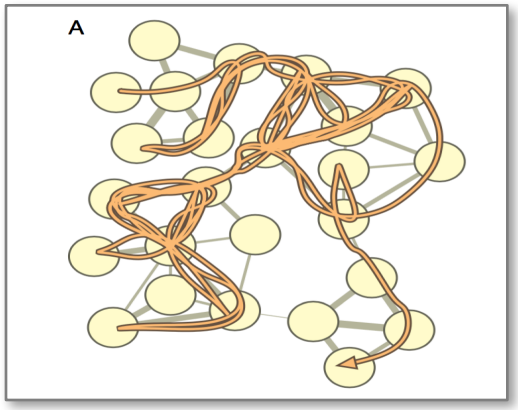


Lee et al . (2016) Viziometrics: Analyzing Visual Information in the Scientific Literature. https://www.arxiv.org/pdf/1605.04951.pdf

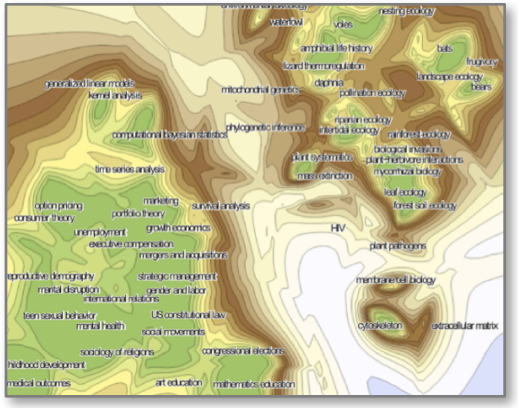
Impact versus Figure Density



Lee et al . (2016) Viziometrics: Analyzing Visual Information in the Scientific Literature. https://www.arxiv.org/pdf/1605.04951.pdf

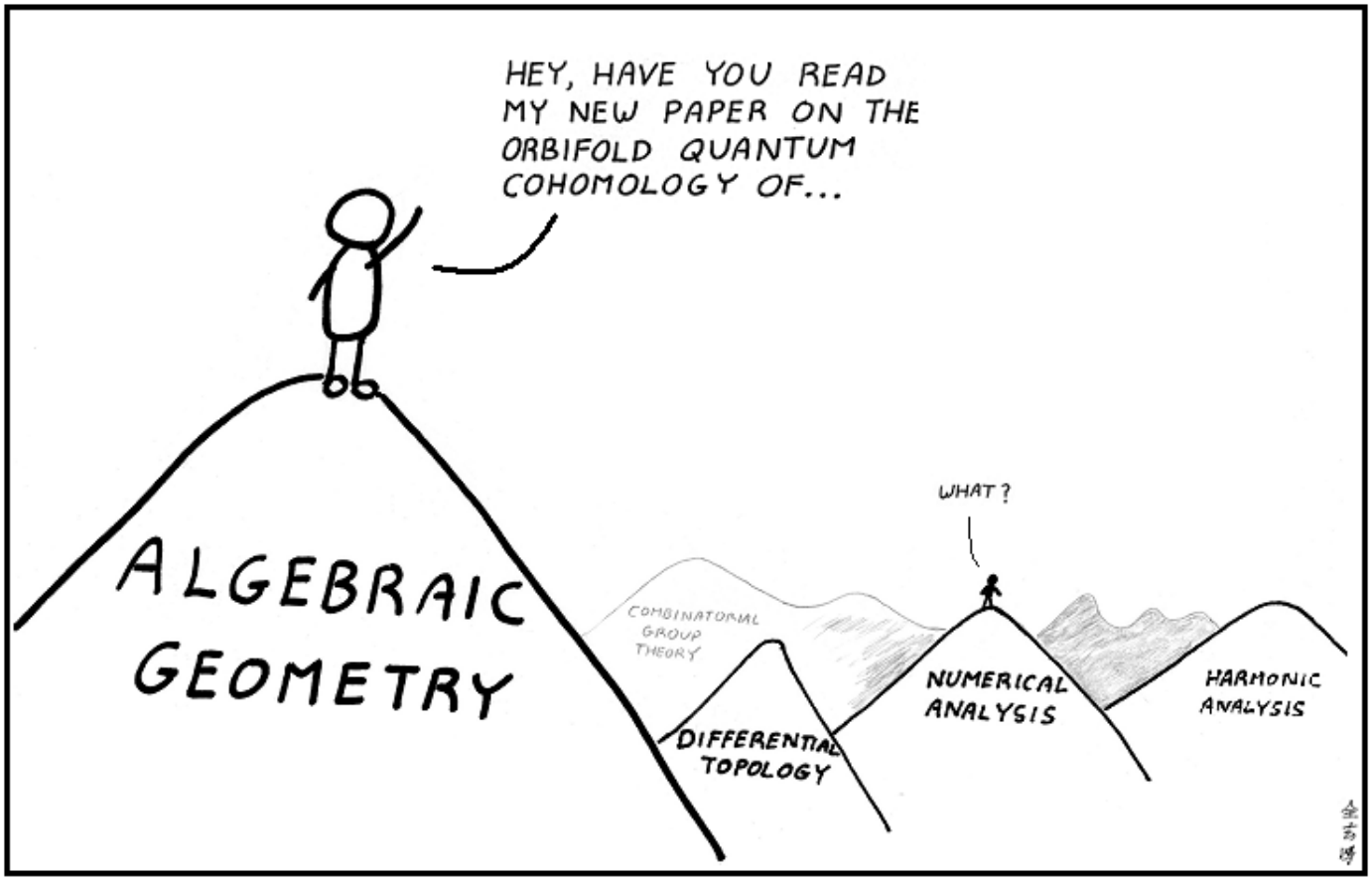


Science of Mapping

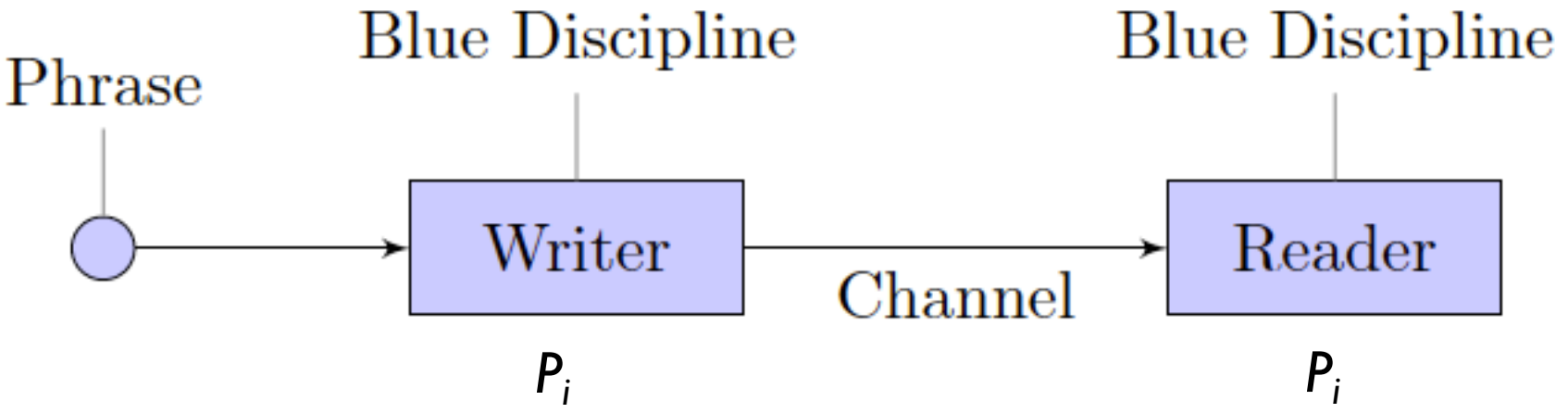


Mapping of Science

The jargon barriers of science



The Landscape of Modern Mathematics



$\boldsymbol{X}$ ~ space of all phrases

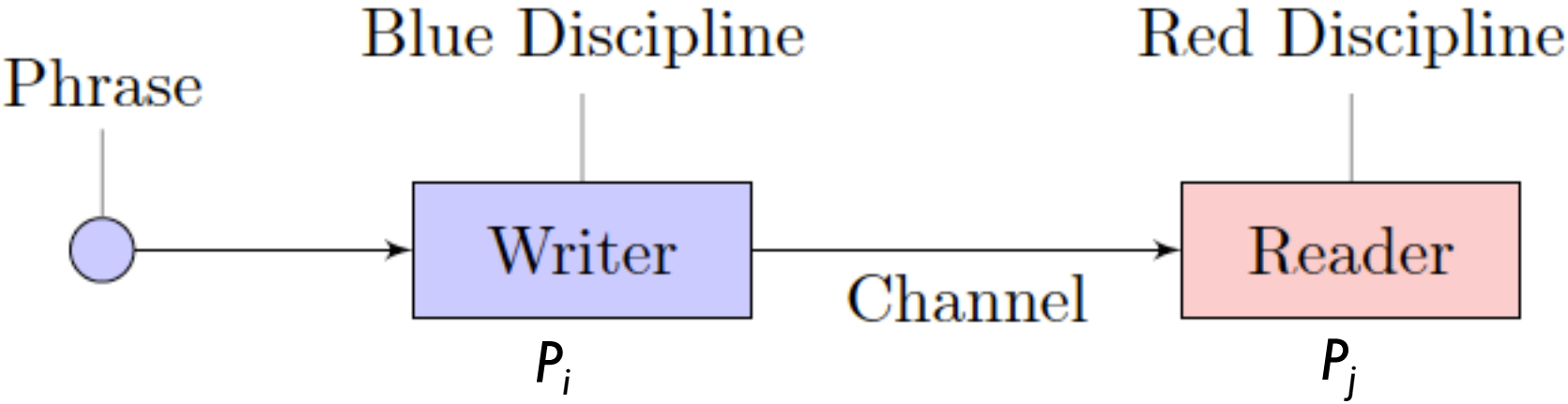
P_i ~ probability distribution over χ_i with values $x \in \boldsymbol{X}$

- 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



cross entropy

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

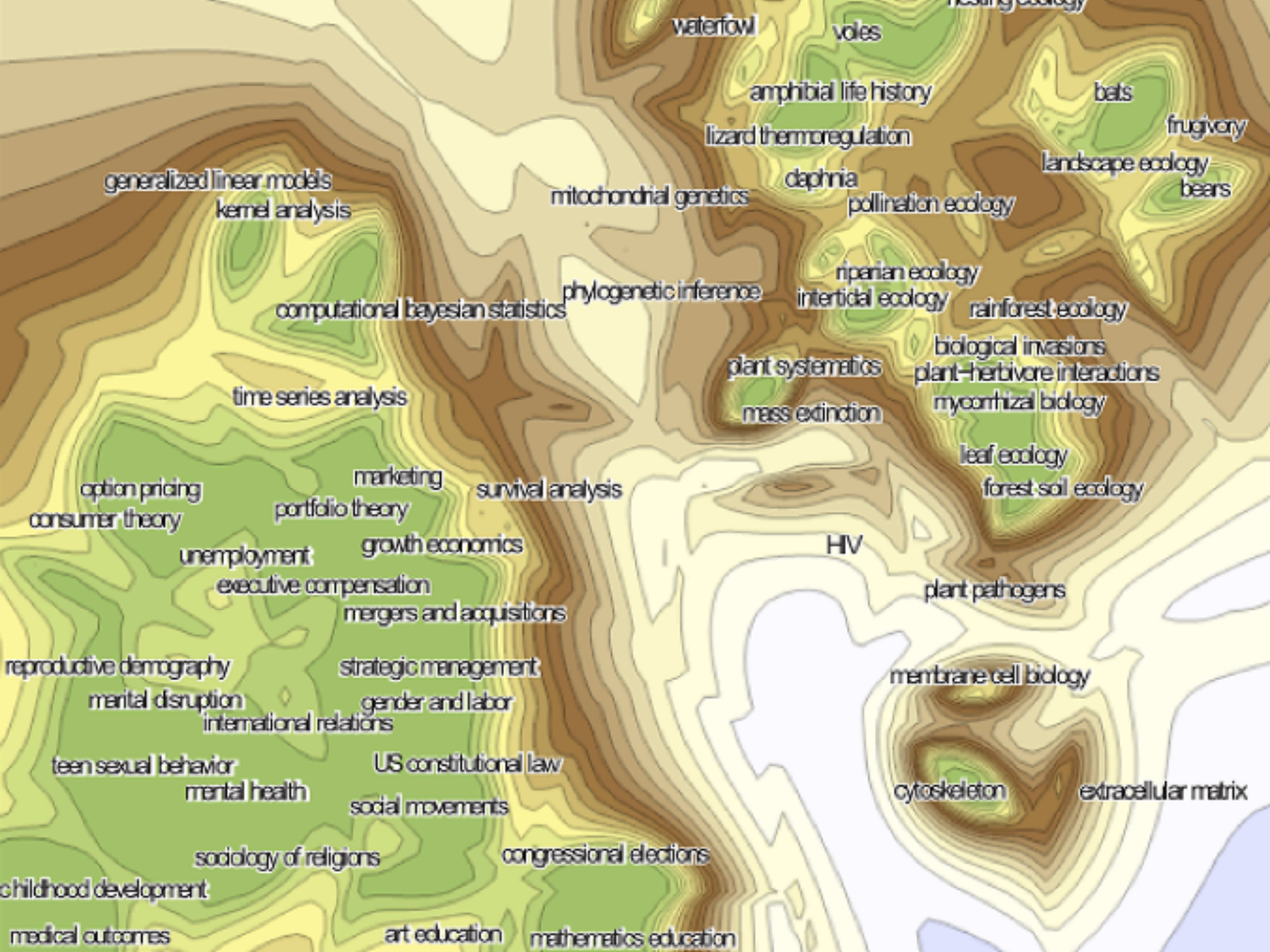
efficiency of communication

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

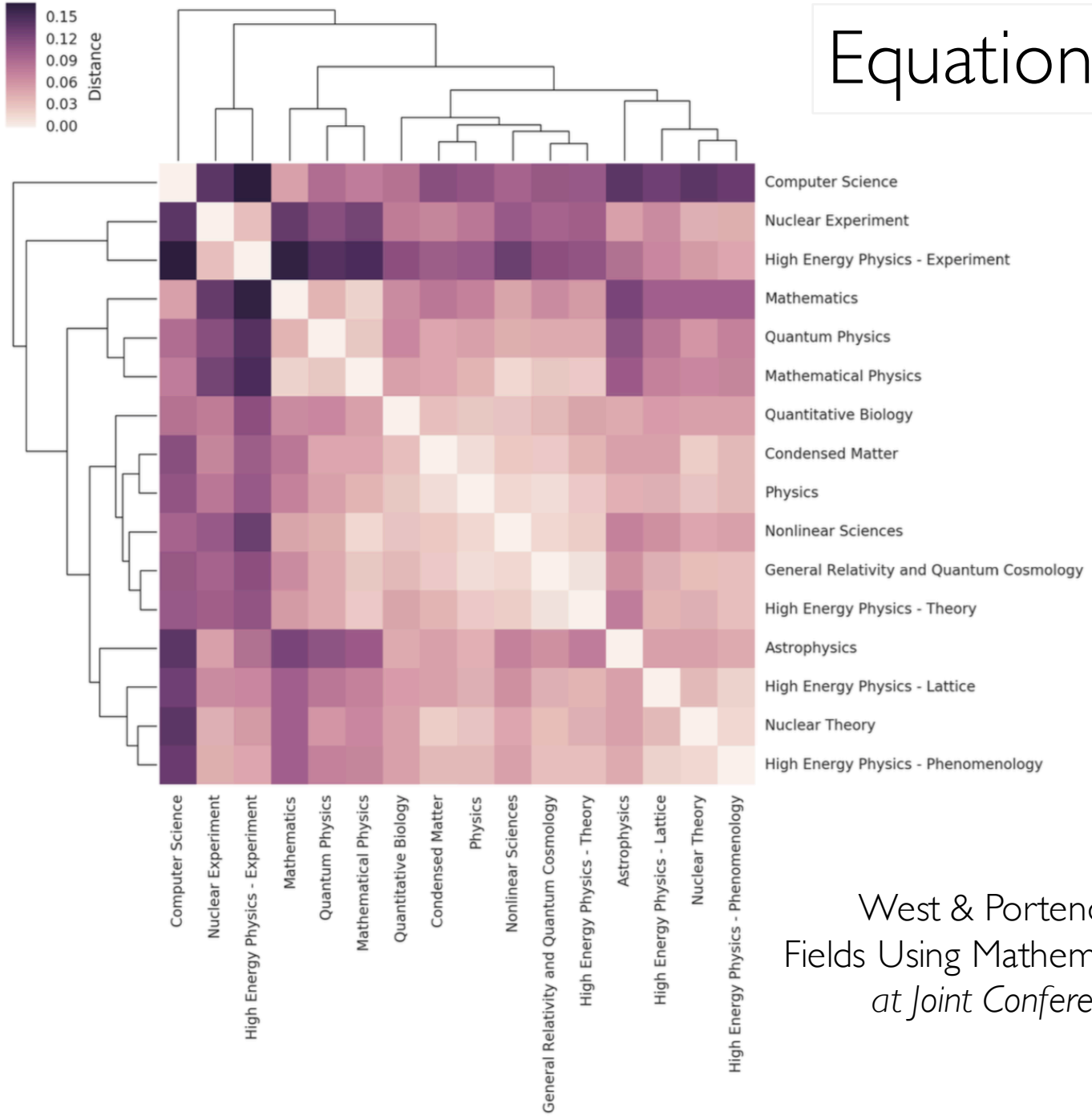
cultural hole

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

Vilhena (2014) **Finding Cultural Holes: How Structure and Culture Diverge in Networks of Scholarly Communication.** Sociological Science


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

Equation Distance



West & Portenoy (2016) Delineating Fields Using Mathematical Jargon. *BIRNDL at Joint Conference on Digital Libraries*

Summary

- Study the *Science of Science*
- Assemble knowledge graph into machine readable formats
- Ask questions about the origin and evolution of ideas and fields, interdisciplinarity, impact assessment and sociology of science
- Building statistical and visualization tools that improve navigation, make relevant connections and facilitate knowledge discovery
- 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

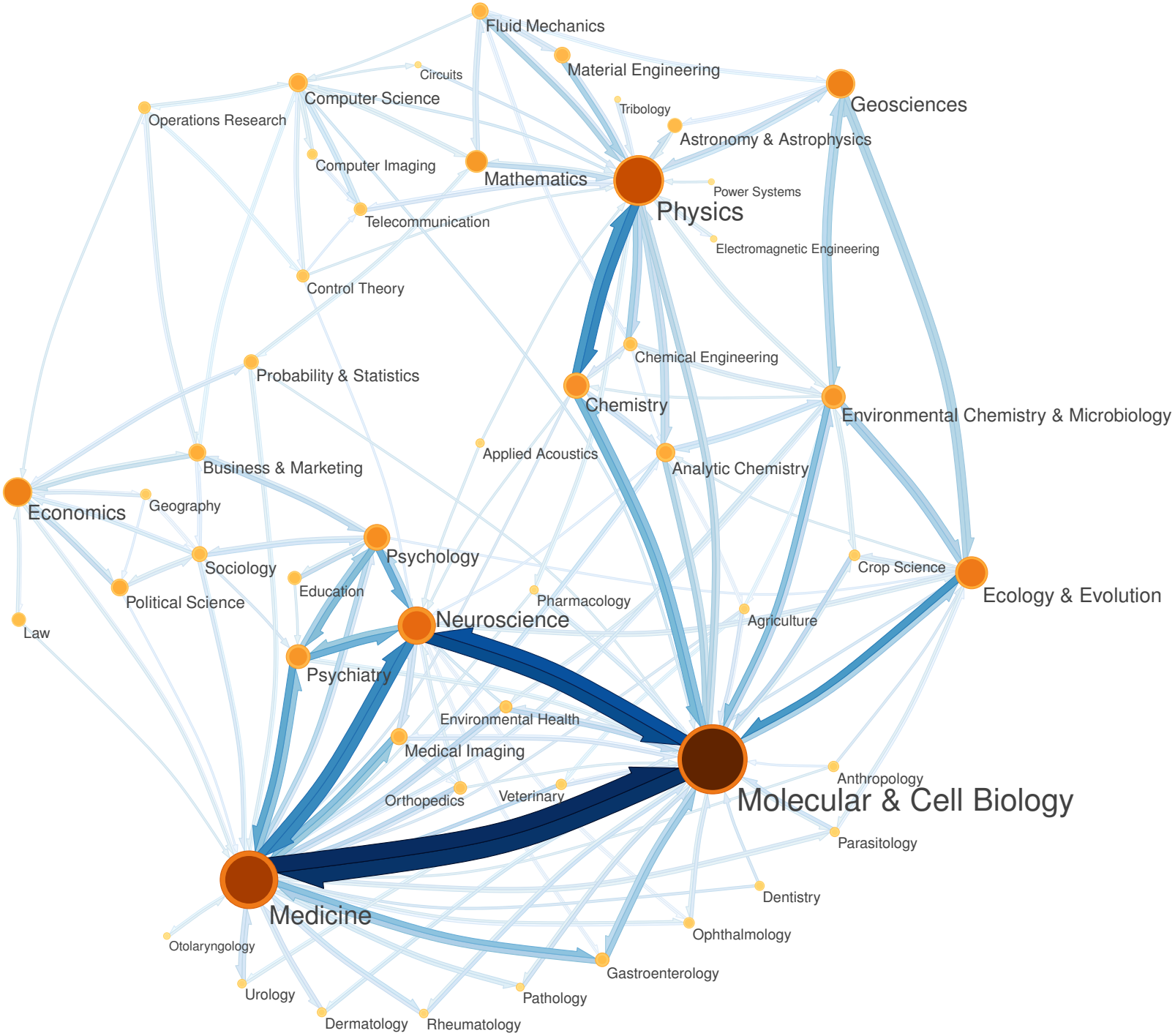
Poshen Lee, CSE, University of Washington

Jevin West

jevinw@uw.edu

jevinwest.org

Eigenfactor.org


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
Neuroscience
Law
Agriculture
Psychiatry
Environmental Health
Medical Imaging
Anthropology
Orthopedics
Veterinary
Molecular & Cell Biology
Parasitology
Dentistry
Medicine
Ophthalmology
Otolaryngology
Gastroenterology
Urology
Pathology
Dermatology
Rheumatology