

MESUR

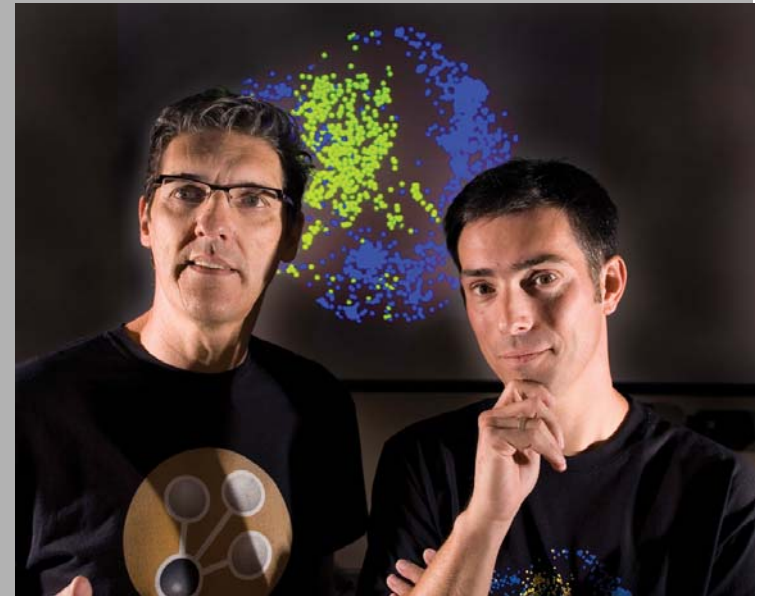
Making Use and Sense of Scholarly Usage Data

<<http://www.mesur.org>>

Johan Bollen - jbollen@lanl.gov

Herbert Van de Sompel - herbertv@lanl.gov

Digital Library Research & Prototyping Team
Research Library
Los Alamos National Laboratory, USA



The MESUR research was funded by the Andrew W. Mellon Foundation

Acknowledgements: Marko A. Rodriguez (LANL), Ryan Chute (LANL), Lyudmila L. Balakireva (LANL), Aric Hagberg (LANL), Luis Bettencourt (LANL)

MESUR in a nutshell

MESUR surveys new approaches to assess scholarly impact

- **Then:** print-era, leads to reliance on citations
 - Citation data: captures only particular type of resource and their authors, lack of detailed contextual information, publication delays, etc.
 - Rate metrics: “counts” and “rates”, total lack of context (who cites what, why, how?)
- **Now:** *online* era = wide variety of resources pre-dominantly accessed online, multiple overlaying online networks connecting resources, large community including practitioners, laypersons etc
- **This enables two innovations in scholarly assessment:**
 1. **Usage data:** real-time, detailed recording of user activities
 2. **Network metrics:** based on context of resource in online network = instead of “how much”, status ~ “who”, “what”, “where”, “when”?

Innovation 1: Usage Data

Metrics based on usage data have significant potential

- Interactions recorded for a **wide variety** of digital scholarly content
- Reflects the activities of **all users** of scholarly information, not only of scholarly authors
- Interactions can be recorded **immediately after *publication***: no double publication delay
- Scale: Total of all existing WoS citations past 100 years = +-650M

hence **COUNTER**, **IKS**, **MESUR**, etc.

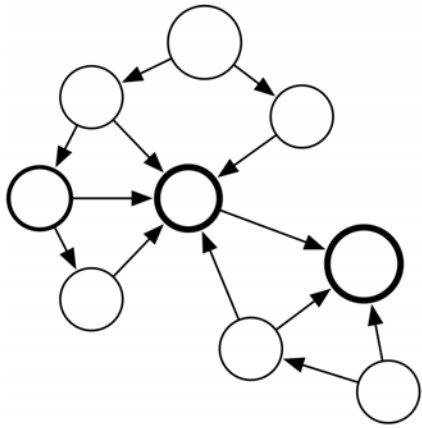
Significant challenges:

- Sampling: whose usage data? Aggregation leads to more representative sample, cf. Bollen (2005,2006)
- Definition: what exactly is usage? Standardized recording, aggregation, processing, tracking across systems, etc. (Bollen, 2005)
- Privacy concerns: protect interests of both institutions and users.
- Reliability: bots, crawlers, noise, etc.

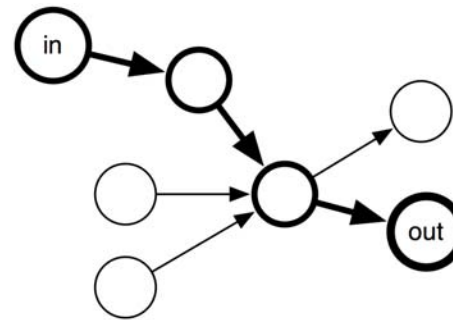
Innovation 2: Network-Based Metrics

... backed by decades of network science ...

Some examples that go beyond “counting” citations/reads/hits/etc:

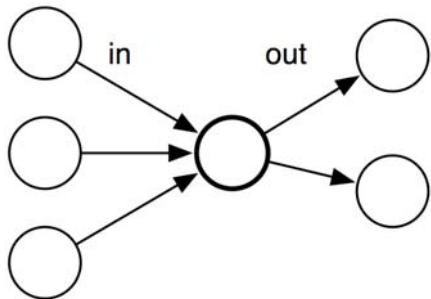


PageRank/Eigenfactor:
Simulate “random walk”
traffic in network



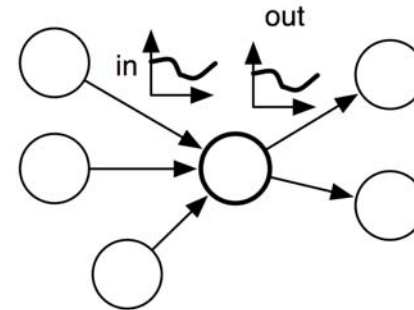
Shortest path centrality:

- Closeness
- Betweenness



Degree centrality

- In-degree
- Out-degree



Distribution parameters:

- In-degree entropy
- Out-degree entropy
- H-index

Readings: Barabasi (2003) “Linked”, Wasserman (1994). Social network analysis.


PageRank computed on Citation Network

ISI IF			PR _w x 10 ³		Y-factor x 10 ²	
rank	value	Journal	value	Journal	value	Journal
1	52.28	ANNU REV IMMUNOL	17.46	J BIOL CHEM	51.15	NATURE
2	37.65	ANNU REV BIOCHEM	16.51	NATURE	47.72	SCIENCE
3	36.83	PHYSIOL REV	16.02	SCIENCE	19.92	NEW ENGL J MED
4	35.04	NAT REV MOL CELL BIO	13.77	PNAS	14.36	CELL
5	34.83	NEW ENGL J MED	8.90	PHYS REV LETT	14.14	PNAS
6	33.95	NAT REV CANCER	5.93	PHYS REV B	11.32	J BIOL CHEM
7	33.06	CANCER J CLIN	5.72	NEW ENGL J MED	8.73	JAMA
8	30.98	NATURE	5.40	ASTROPHYS J	7.83	LANCET
9	30.55	NAT MED	5.39	CELL	7.22	NAT GENET
10	30.17	ANNU REV NEUROSCI	4.90	J AM CHEM SOC	6.26	PHYS REV LETT

2003 JCR, Science Edition
5709 journals

Usage ranking

In-degree			PageRank x 100		Betweenness	
rank	value	journal	value	journal	value	journal
1	4195	SCIENCE	0.160	SCIENCE	0.035	SCIENCE
2	4019	NATURE	0.150	NATURE	0.032	NATURE
3	3562	PNAS	0.130	PNAS	0.020	PNAS
4	2438	J BIOL CHEM	0.100	LNCS	0.017	LNCS
5	2432	LNCS	0.080	J BIOL CHEM	0.006	LANCET

Johan Bollen, Marko A. Rodriguez, and Herbert Van de Sompel. [Journal status](#). Scientometrics, 69(3), December 2006 (DOI:10.1007/s11192-006-0176-z)
 Philip Ball. Prestige is *factored* into journal ratings. Nature **439**, 770-771, February 2006 (DOI:10.1038/439770a)
 Cf: <http://www.eigenfactor.org/>  eigenFACTOR.org
RANKING AND MAPPING SCIENTIFIC KNOWLEDGE

MESUR: scientific, large-scale program to address following issues:

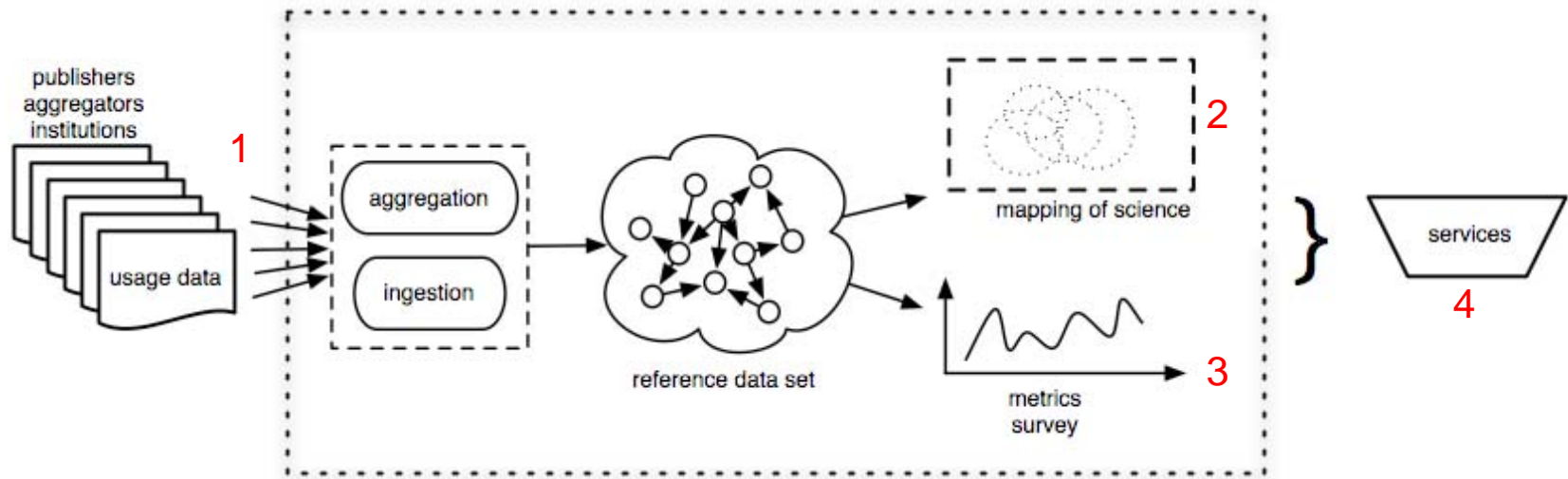
1. Maps and models: study and model structure of usage
2. Metrics: surveying novel means of assessing scholarly impact
3. Services: create public tools for (1) and (2)

How?

1. Create very large-scale reference data set
 - a) Usage, citation and bibliographic data combined
 - b) Various communities, various collections
2. Investigate usage data:
 1. Significant structure in usage data?
 2. Map and model Science in real-time
3. Compute a variety of metrics:
 1. citation, usage, rate-based, network-based
 2. cross-validate with existing journal metrics
4. Deploy tools to explore usage-based journal metrics

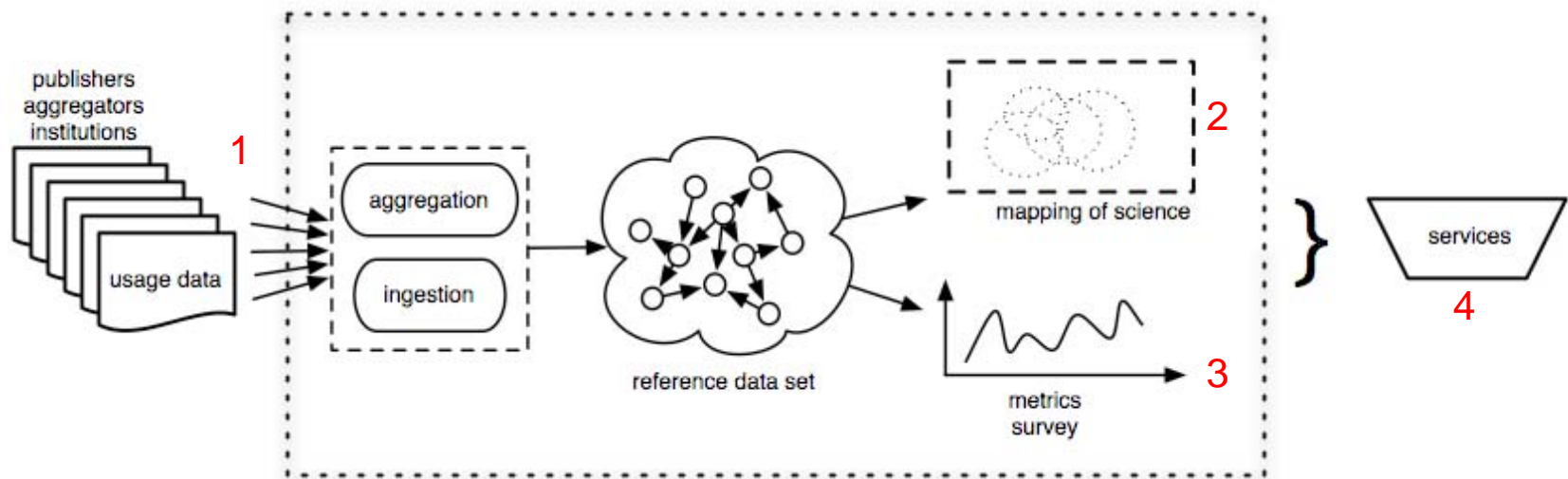
MESUR: Project Phases

- 1) Usage data acquisition
- 2) Structure in usage data - Map of Science
- 3) Metrics based on usage and citation - Compare
- 4) Services



MESUR: Project Phases

- 1) Usage data acquisition
- 2) Structure in usage data - Map of Science
- 3) Metrics based on usage and citation - Compare
- 4) Services



How to Obtain 1,000,000,000 (1B) Usage Events?

Politely ask publishers, aggregators, institutions....

- Attempt to create “good” sample from Publishers, Aggregators, Linking Servers, Proxy Servers:
 - BMC, Blackwell, UC, CSU (23), EBSCO, ELSEVIER, EMERALD, INGENTA, JSTOR, LANL, MIMAS/ZETOC, THOMSON, UPENN, UTEXAS (9)
- Strict agreements regarding confidentiality of data
- Resulting scale: > 1,000,000,000 usage events
- Period: 2002-2007, but mostly 2006
- Span:
 - > 50M articles ; > 100,000 journals (inc. newspapers, magazines,...)

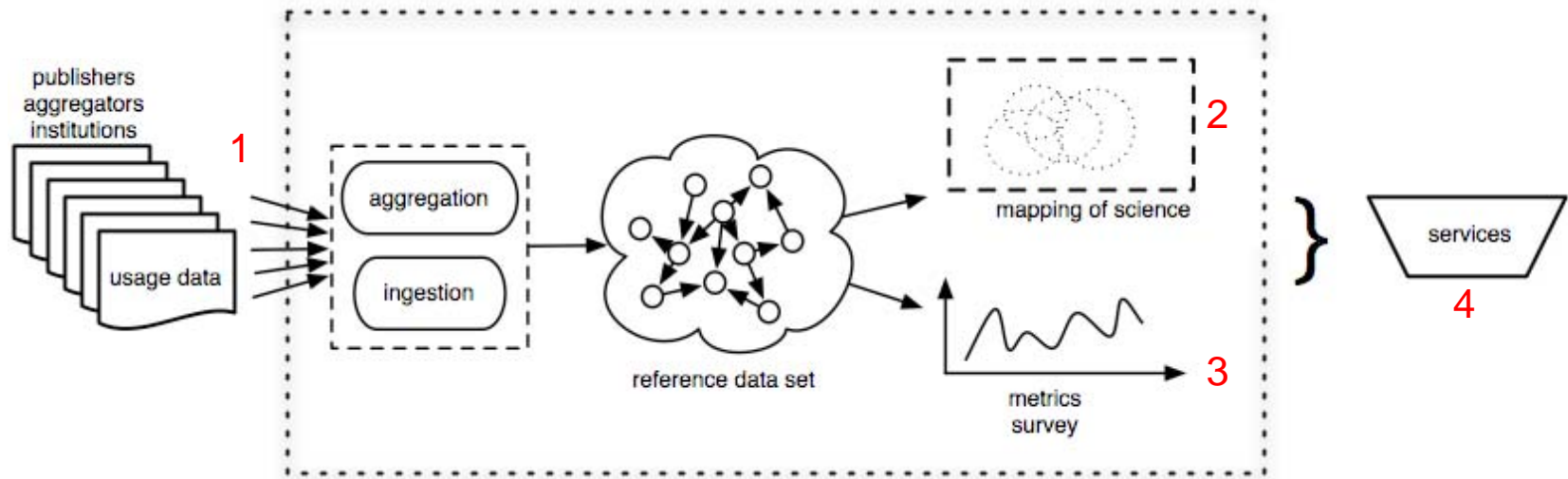
Some Minimal Requirements for Usage Data

Not just usage statistics, we need more detailed information to construct usage-based networks and models:

- Article level usage events
- Fields:
 - unique session ID,
 - date/time,
 - unique document ID and/or metadata,
 - request type
- MESUR took care of conversion, normalization, etc in-house

MESUR: Project Phases

- 1) Usage data acquisition
- 2) Structure in usage data – Maps and models of Science
- 3) Metrics based on usage and citation - Compare
- 4) Services

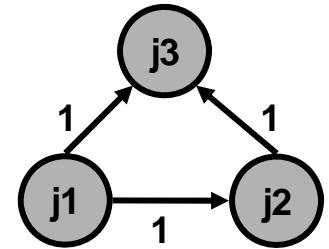
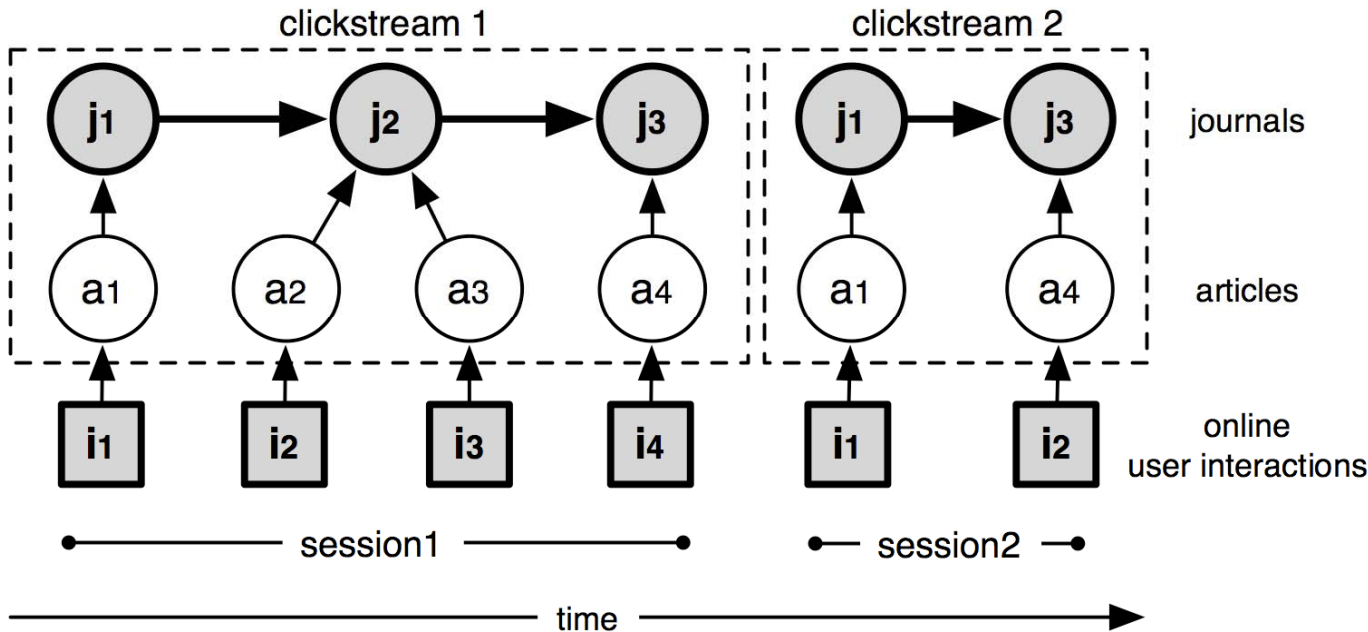


Data set: subset of MESUR

- Common time period:
 - March 1st 2006 - February 1st 2007
 - Thomson Scientific (Web of Science), Elsevier (Scopus), JSTOR, Ingenta, University of Texas (9 campuses, 6 health institutions), and California State University (23 campuses)
- 346,312,045 usage events
- 97,532 serials (many of which not journals)

Domain	Usage	UC Degrees	JCR
Natural Science	37%	39%	92.8%
Social Sciences	45%	46%	7.2%
Humanities	14%	15%	

Generating a usage network



Same session ~ documents relatedness

- Same session, same user: common interest
- Frequency of co-occurrence = degree of relationship
- Normalized: conditional probability

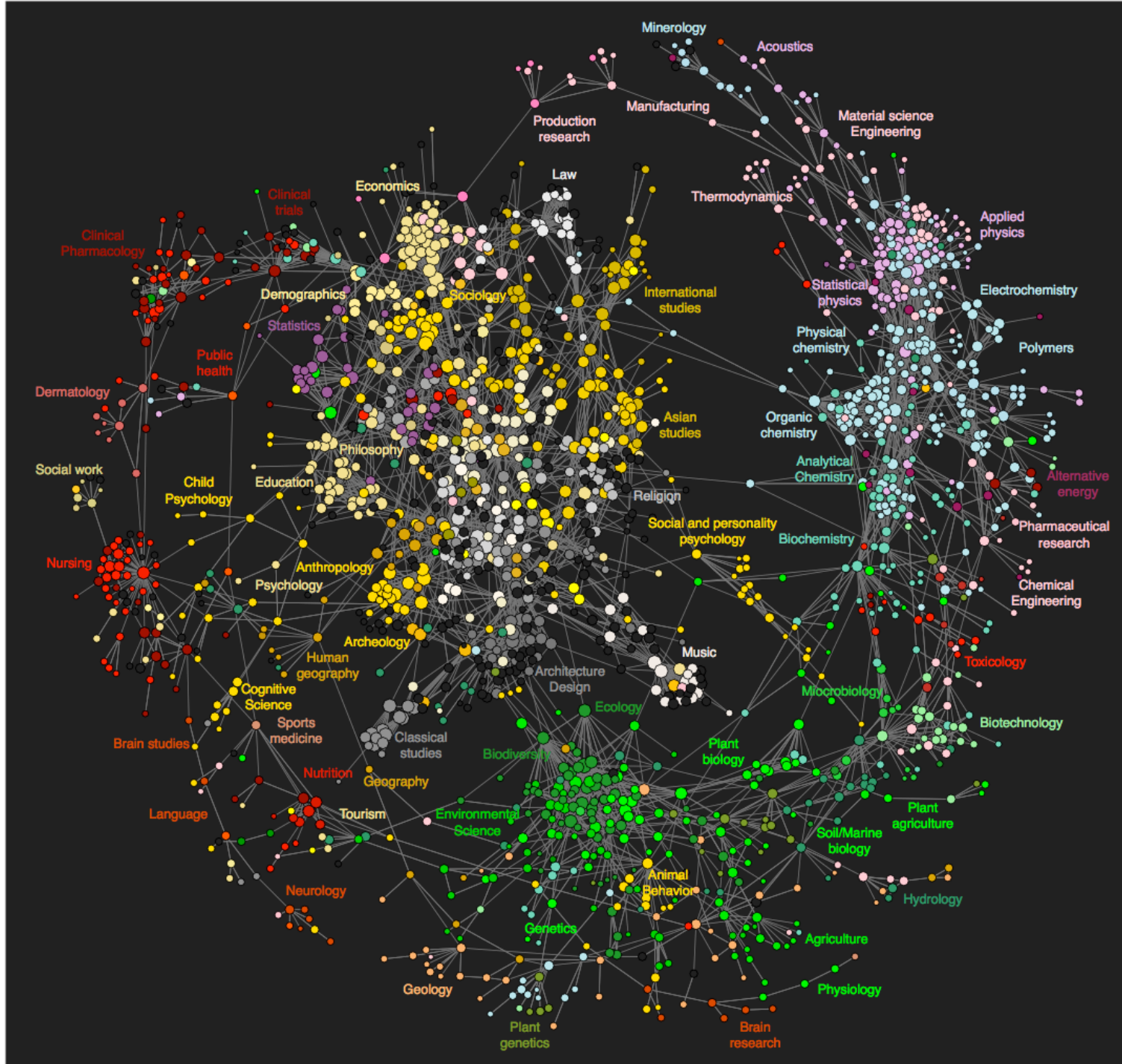
Note: not something we invented

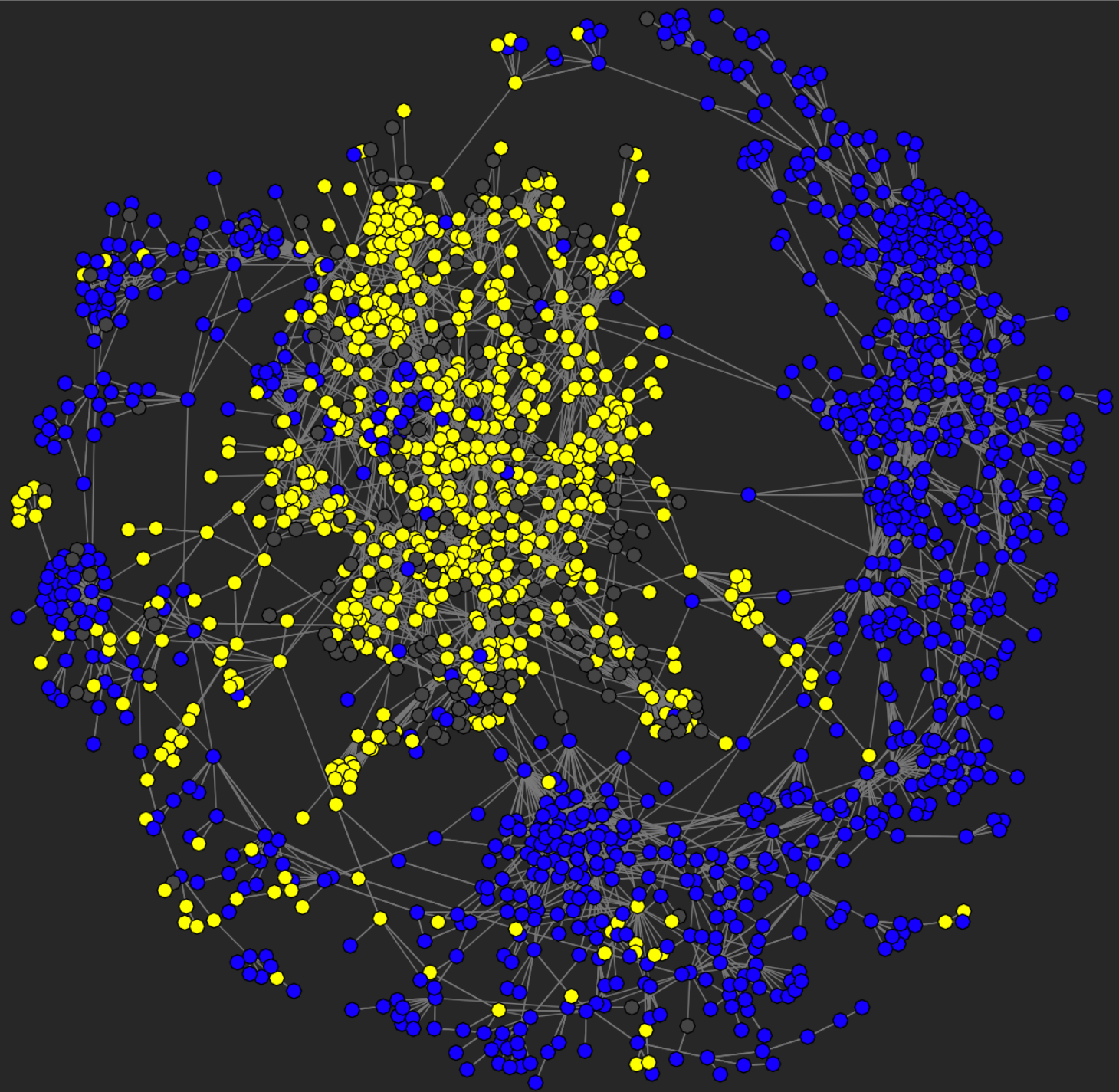
- Association rule learning in data mining
- Cf. Netflix, Amazon recommendations

Resulting network:

Table 5. Sample of journals pairs with high $P(v_i, v_j)$

v_i	v_j	$p(v_i, v_j)$	$N(v_i, v_j)$	$N(v_i)$
American Journal of International Law	International Organization	0.0207	9,292	448,034
	International Affairs	0.0184	8,254	
	International and Comparative Law Quarterly	0.0171	7,654	
	Foreign Policy	0.0167	7,500	
	American Political Science Association	0.0140	6,291	
Journal of Educational Sociology	American Journal of Sociology	0.0334	2,790	83,419
	Journal of Higher Education	0.0303	2,529	
	Journal of Negro Education	0.0286	2,389	
	American Sociological Review	0.0276	2,303	
	Social Forces	0.0249	2,076	
Surface Science	Physical Review B	0.0704	2,555	36,282
	Applied Surface Science	0.0341	1,239	
	Physical Review Letters	0.0339	1,230	
	Journal of Chemical Physics	0.0333	1,207	
	Applied Physics Letters	0.0327	1,188	
Journal of Organic Chemistry	Journal of the American Chemical Society	0.0873	4,141	47,439
	Tetrahedron Letters	0.0865	4,105	
	Tetrahedron	0.0602	2,857	
	Organic Letters	0.0532	2,526	
	Angewandte Chemie	0.0305	1,448	
Ecological Applications	Ecology	0.0965	13,659	141,481
	Conservation Biology	0.0524	7,408	
	Bioscience	0.0215	3,043	
	Annual Review of Ecology and Systematics	0.0215	3,043	
	Clinical and Experimental Allergy	0.0191	2,699	
Annals of Mathematics	American Journal of Mathematics	0.0705	5,392	76,526
	American Mathematical Monthly	0.0579	4,432	
	PNAS	0.0156	1,195	
	Econometrica	0.0082	624	
	Mathematics Magazine	0.0077	587	





From static to dynamic maps: visualizing trends in science

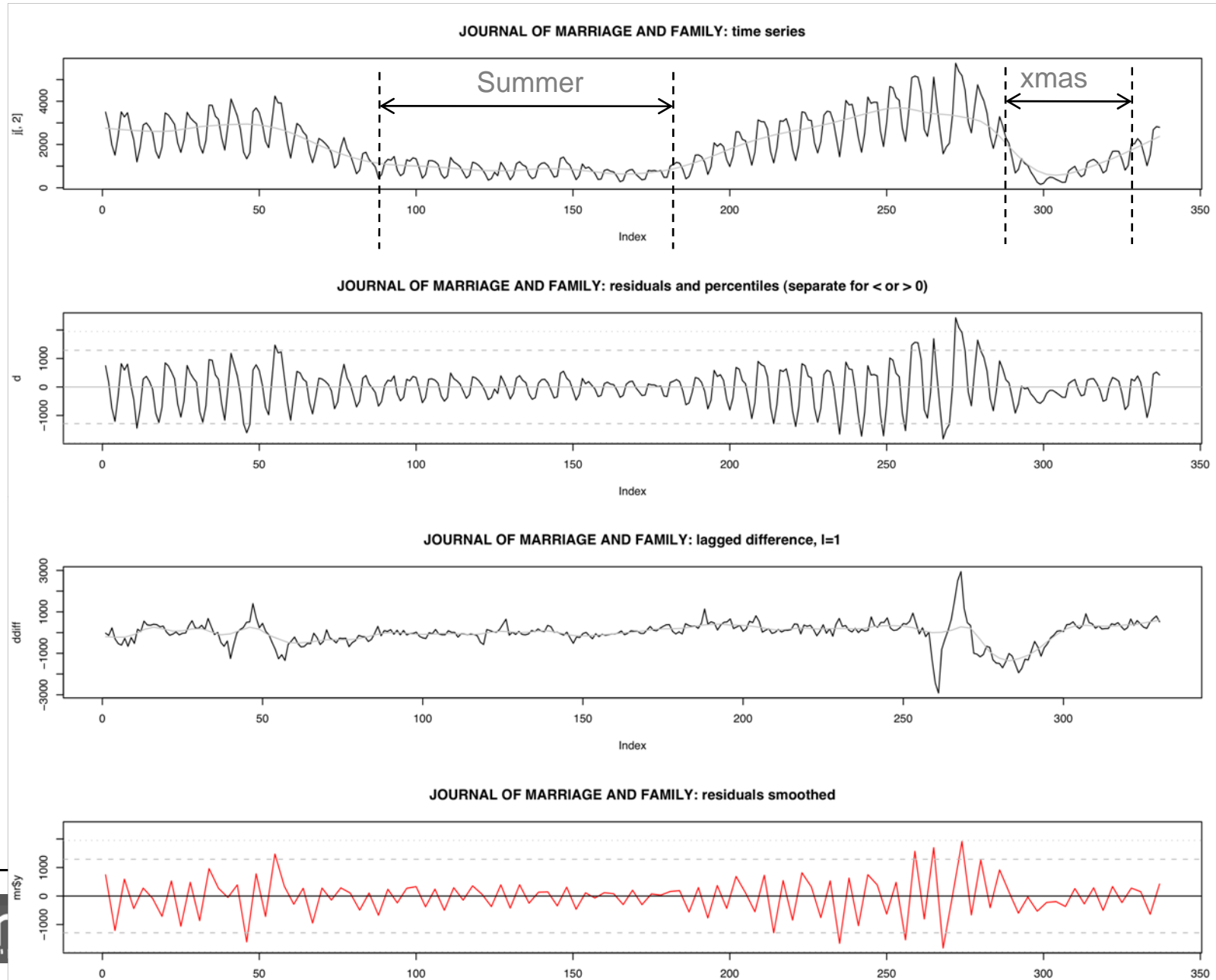
Existing maps:

- Static, focused on particular period of time
- Clusters: “centers of attention”
- Attention may shift over time reflecting changes in scientific interests
- Build-up of attention over time: future areas of innovation?

Examine changes in activity over time

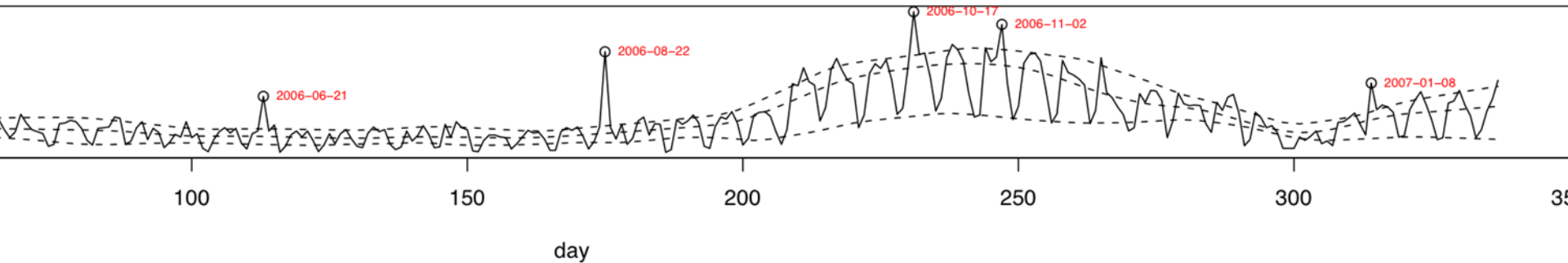
- Model usage over time
- Identify “unexpected” build-ups of activity
- Extrapolate to future for prediction?

Ebb and flow of scientific attention

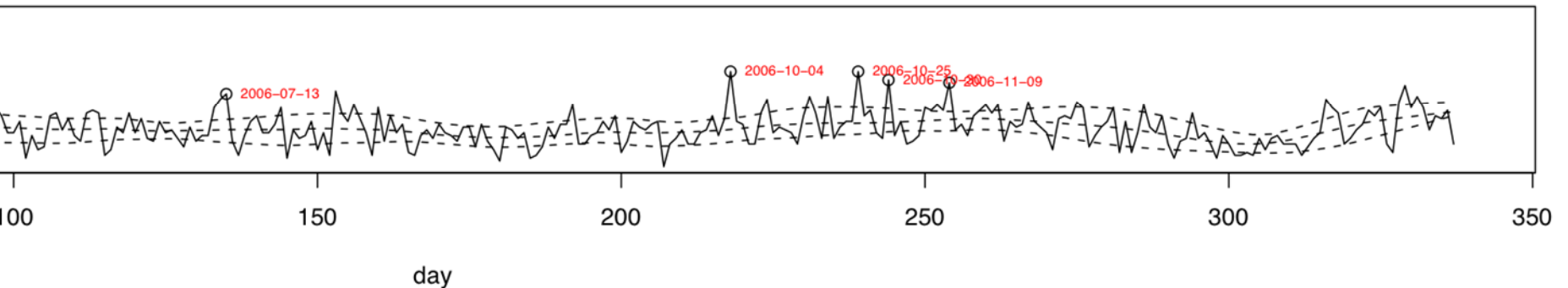


Bursts: unexpected peaks in collective attention

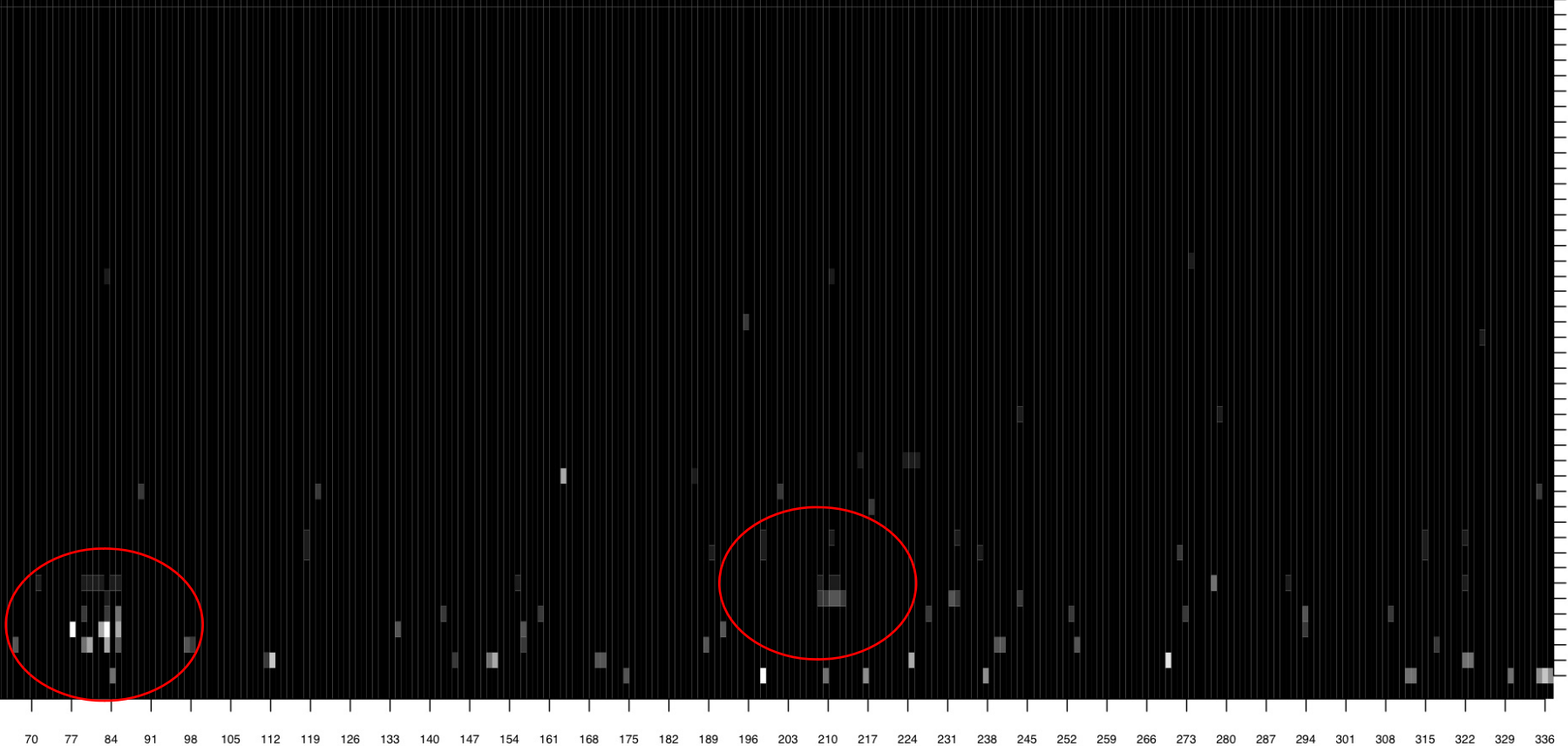
LANDSCAPE ECOLOGY

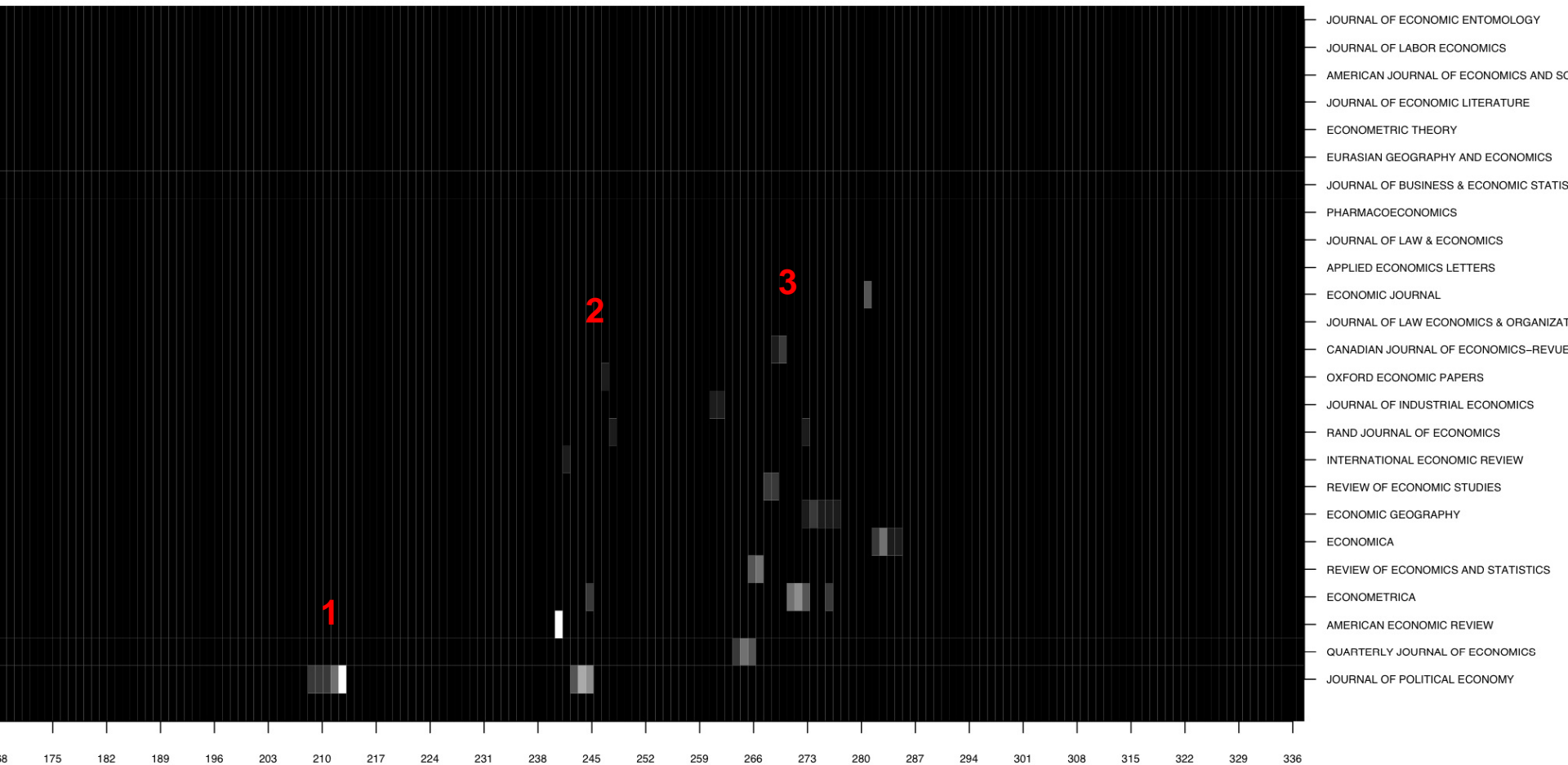


MARINE ECOLOGY



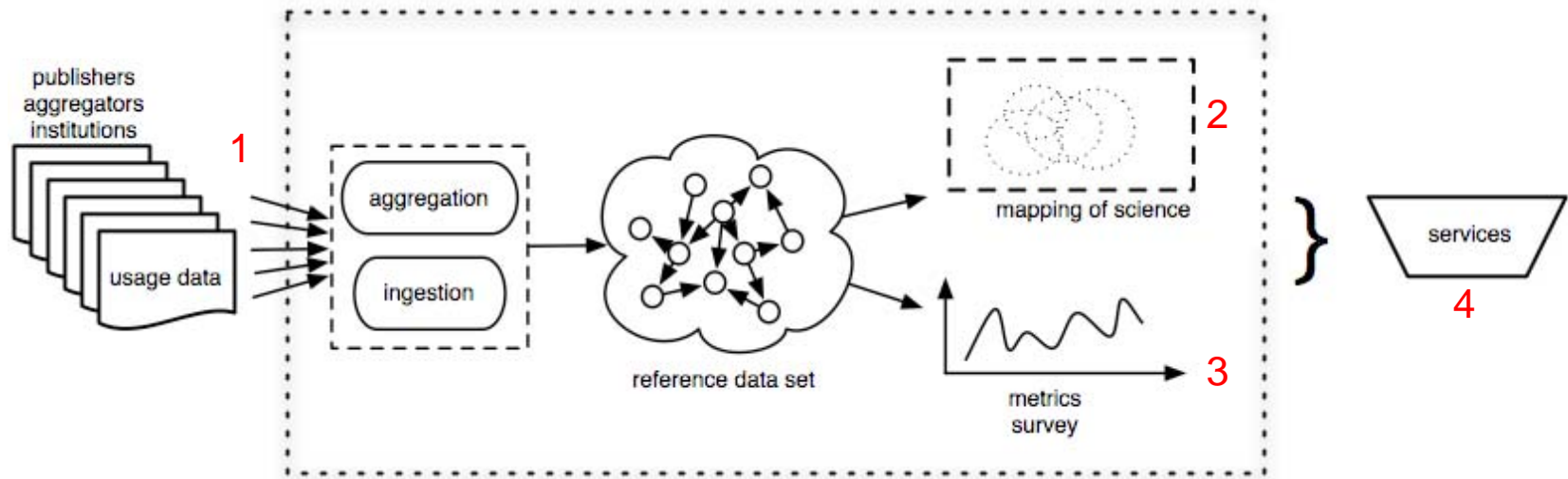
COMMUNITY GENETICS
INTERNATIONAL
KOREAN JOURNAL
GENETICS IN ME
ADVANCES IN GE
JOURNAL OF GE
CURRENT GENET
GENETICS SELEC
AMERICAN JOUR
MUTATION RESE
ANNALES DE GE
GENETIC EPIDEM
GENETIC ENGINE
AMERICAN JOUR
EUROPEAN JOUR
MOLECULAR AN
EUROPEAN JOUR
CURRENT OPINIO
BMC GENETICS
GENETICS AND M
PHARMACOGENET
GENETIC TESTIN
PSYCHIATRIC GE
NEUROGENETICS
GENETIC SOCIAL
CYTOGENETIC A
JOURNAL OF NE
GENETICAL RES
JOURNAL OF GE
GENES & GENET
JOURNAL OF HU
ANNUAL REVIEW
HUMAN GENETIC
RUSSIAN JOURN
AMERICAN JOUR
IMMUNOGENETIC
TRENDS IN GEN
ANNUAL REVIEW
PLOS GENETICS
CANCER GENETI
JOURNAL OF ME
BIOCHEMICAL J
AMERICAN JOUR
EUROPEAN JOUR
NEW GENETICS
SILVAE GENETIC
GENETICA
NATURE GENETI
CONSERVATION
BEHAVIOR GENE
MOLECULAR GE
FUNGAL GENETI
MOLECULAR PHY
JOURNAL OF AN
ANNALS OF ASS
CLINICAL GENET
ANIMAL GENETIC
TWIN RESEARCH
HUMAN MOLEC





MESUR: Project Phases

- 1) Usage data acquisition
- 2) Structure in usage data - Map of Science
- 3) Metrics based on usage and citation - Compare
- 4) Services



Metrics Computed for Usage and Citation Data

ID	Type	Measure	Source
1	Citation	Scimago Journal Rank	Scimago/Scopus
2	Citation	Immediacy Index	JCR 2007
3	Citation	Closeness	JCR 2007
4	Citation	Cites per doc	Scimago/Scopus
5	Citation	Journal Impact Factor	JCR 2007
6	Citation	Closeness centrality	JCR 2007
7	Citation	Out-degree centrality	JCR 2007
8	Citation	Out-degree centrality	JCR 2007
9	Citation	Degree Centrality	JCR 2007
10	Citation	Degree Centrality	JCR 2007
11	Citation	H-Index	Scimago/Scopus
12	Citation	Scimago Total cites	Scimago/Scopus
13	Citation	Journal Cite Probability	JCR 2007
14	Citation	In-degree centrality	JCR 2007
15	Citation	In-degree centrality	JCR 2007
16	Citation	PageRank	JCR 2007
17	Citation	PageRank	JCR 2007
18	Citation	PageRank	JCR 2007
19	Citation	PageRank	JCR 2007
20	Citation	Y-factor	JCR 2007
21	Citation	Betweenness centrality	JCR 2007
22	Citation	Betweenness centrality	JCR 2007
23	Citation	<i>Citation Half-Life</i>	<i>JCR 2007</i>
24	Usage	Closeness centrality	MESUR 2007
25	Usage	Closeness centrality	MESUR 2007
26	Usage	Degree centrality	MESUR 2007
27	Usage	PageRank	MESUR 2007
28	Usage	PageRank	MESUR 2007
29	Usage	In-degree centrality	MESUR 2007
30	Usage	Out-degree centrality	MESUR 2007
31	Usage	PageRank	MESUR 2007
32	Usage	PageRank	MESUR 2007
33	Usage	Betweenness centrality	MESUR 2007
34	Usage	Betweenness centrality	MESUR 2007
35	Usage	Degree centrality	MESUR 2007
36	Usage	Out-degree centrality	MESUR 2007
37	Usage	In-degree centrality	MESUR 2007
38	Usage	Journal Use Probability	MESUR 2007
39	Usage	<i>Usage Impact Factor</i>	<i>MESUR 2007</i>

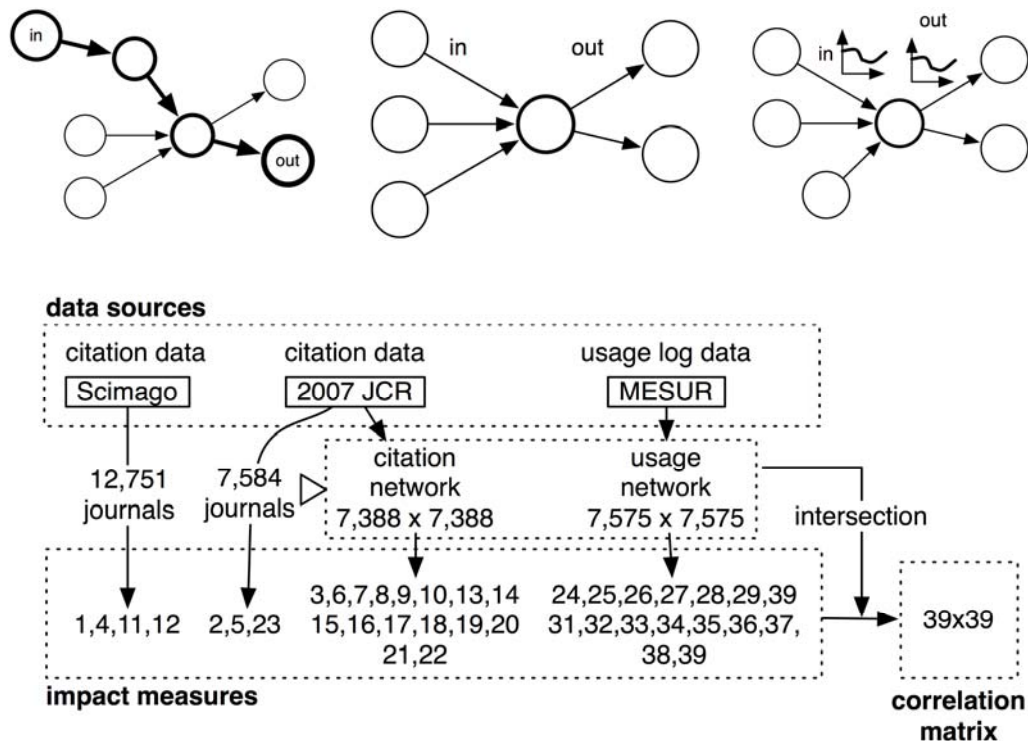


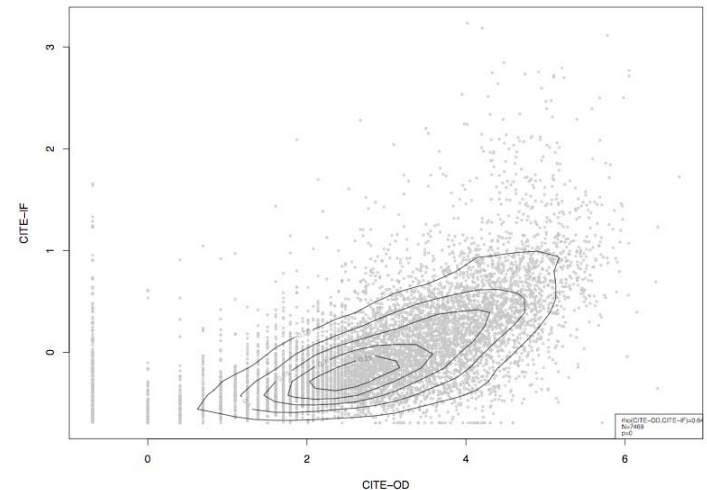
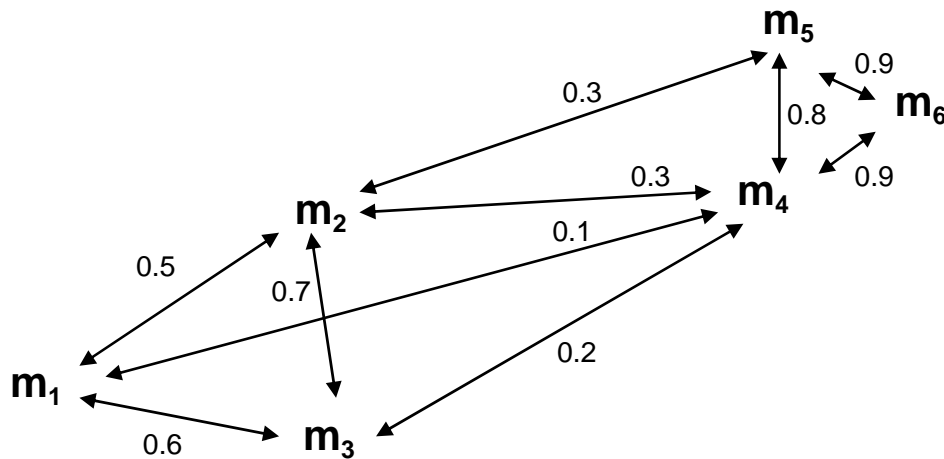
Fig. 4. Schematic representation of data sources and processing. Impact measure identifiers refer to Table 1.

Bollen J, Van de Sompel H, Hagberg A, Chute R. 2009 A principal component analysis of 39 scientific impact measures. <http://arxiv.org/abs/0902.2183> Submitted to PLoS ONE

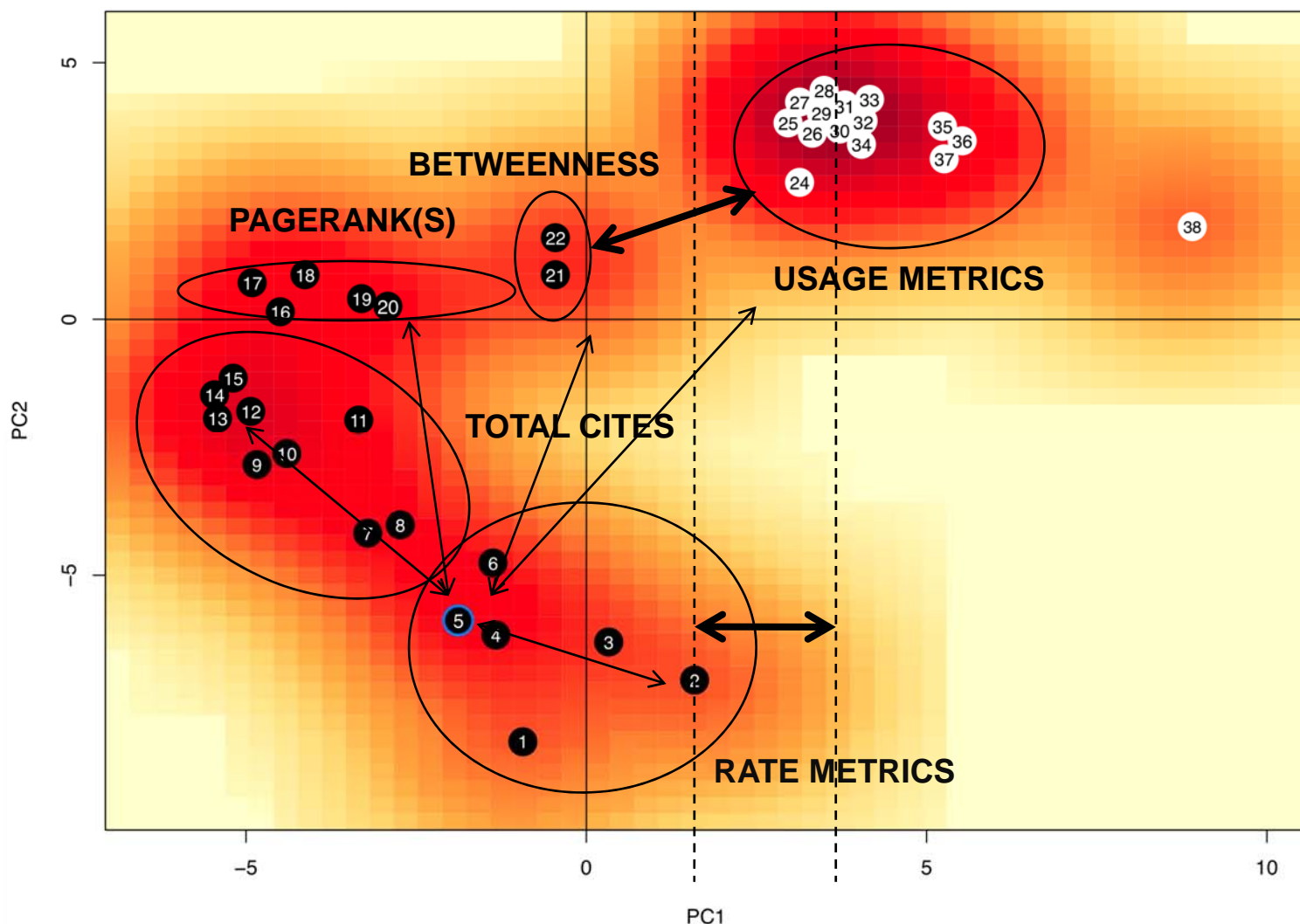
Metric Correlations: Metric Maps

$$R_{10 \times 10} = \begin{pmatrix} 1.00 & 0.71 & 0.77 & 0.52 & 0.79 & 0.55 & 0.69 & 0.63 & 0.60 & 0.18 \\ 0.71 & 0.99 & 0.52 & 0.69 & 0.79 & 0.85 & 0.49 & 0.44 & 0.49 & 0.22 \\ 0.77 & 0.52 & 1.00 & 0.62 & 0.63 & 0.39 & 0.70 & 0.73 & 0.68 & 0.20 \\ 0.52 & 0.69 & 0.62 & 1.00 & 0.68 & 0.78 & 0.49 & 0.56 & 0.65 & 0.06 \\ 0.79 & 0.79 & 0.63 & 0.68 & 1.00 & 0.82 & 0.66 & 0.62 & 0.66 & 0.15 \\ 0.55 & 0.85 & 0.39 & 0.78 & 0.82 & 1.00 & 0.40 & 0.40 & 0.50 & 0.13 \\ 0.69 & 0.49 & 0.70 & 0.49 & 0.66 & 0.40 & 1.00 & 0.89 & 0.85 & 0.53 \\ 0.63 & 0.44 & 0.73 & 0.56 & 0.62 & 0.40 & 0.89 & 1.00 & 0.97 & 0.45 \\ 0.60 & 0.49 & 0.68 & 0.65 & 0.66 & 0.50 & 0.85 & 0.97 & 1.00 & 0.42 \\ 0.18 & 0.22 & 0.20 & 0.06 & 0.15 & 0.13 & 0.53 & 0.45 & 0.42 & 1.00 \end{pmatrix}$$

19: Citation PageRank
5: Journal Impact Factor
22: Citation Betweenness
6: Citation Closeness
11: Citation H-index
1: Citation Scimago Journal Rank
31: Usage PageRank
34: Usage Betweenness
24: Usage Closeness
39: Usage Impact Factor



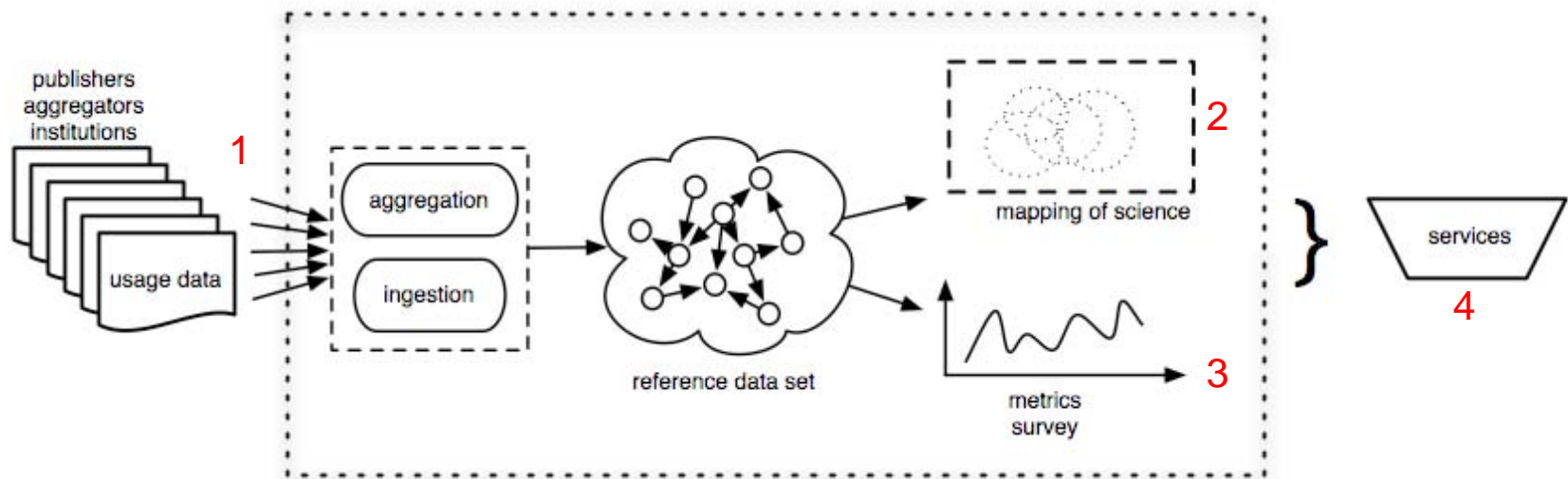
The MESUR Metrics Map




ID	Type	Measure
1	Citation	Scimago Journal Rank
2	Citation	Immediacy Index
3	Citation	Closeness
4	Citation	Cites per doc
5	Citation	Journal Impact Factor
6	Citation	Closeness centrality
7	Citation	Out-degree centrality
8	Citation	Out-degree centrality
9	Citation	Degree Centrality
10	Citation	Degree Centrality
11	Citation	H-Index
12	Citation	Scimago Total cites
13	Citation	Journal Cite Probability
14	Citation	In-degree centrality
15	Citation	In-degree centrality
16	Citation	PageRank
17	Citation	PageRank
18	Citation	PageRank
19	Citation	PageRank
20	Citation	Y-factor
21	Citation	Betweenness centrality
22	Citation	Betweenness centrality
23	Citation	Citation Half-Life
24	Usage	Closeness centrality
25	Usage	Closeness centrality
26	Usage	Degree centrality
27	Usage	PageRank
28	Usage	PageRank
29	Usage	In-degree centrality
30	Usage	Out-degree centrality
31	Usage	PageRank
32	Usage	PageRank
33	Usage	Betweenness centrality
34	Usage	Betweenness centrality
35	Usage	Degree centrality
36	Usage	Out-degree centrality
37	Usage	In-degree centrality
38	Usage	Journal Use Probability
39	Usage	Usage Impact Factor

MESUR: Project Phases

- 1) Usage data acquisition
- 2) Structure in usage data - Map of Science
- 3) Metrics based on usage and citation - Compare
- 4) Services



MESUR Services – <http://www.mesur.org/services/>



MESUR: science maps and rankings from large-scale usage data

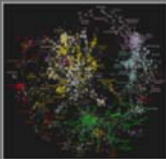
Services: [Maps](#) [Rankings](#) [Documentation](#) [Demos](#)

Search a domain, e.g. [biology](#)


The [MESUR project](#) studies science from large-scale usage data collected from some of the world's most significant publishers, aggregators and university consortia.

MESUR services

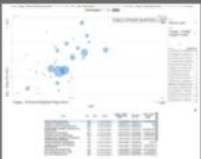
Maps of science




Explore interactive maps of science generated from large-scale usage data, including impact rankings provided for journals in the map (requires Java).
Featured in Nature News, Wired, the New York Times and many other venues.

 [to maps](#)

Interactive journal ranking service

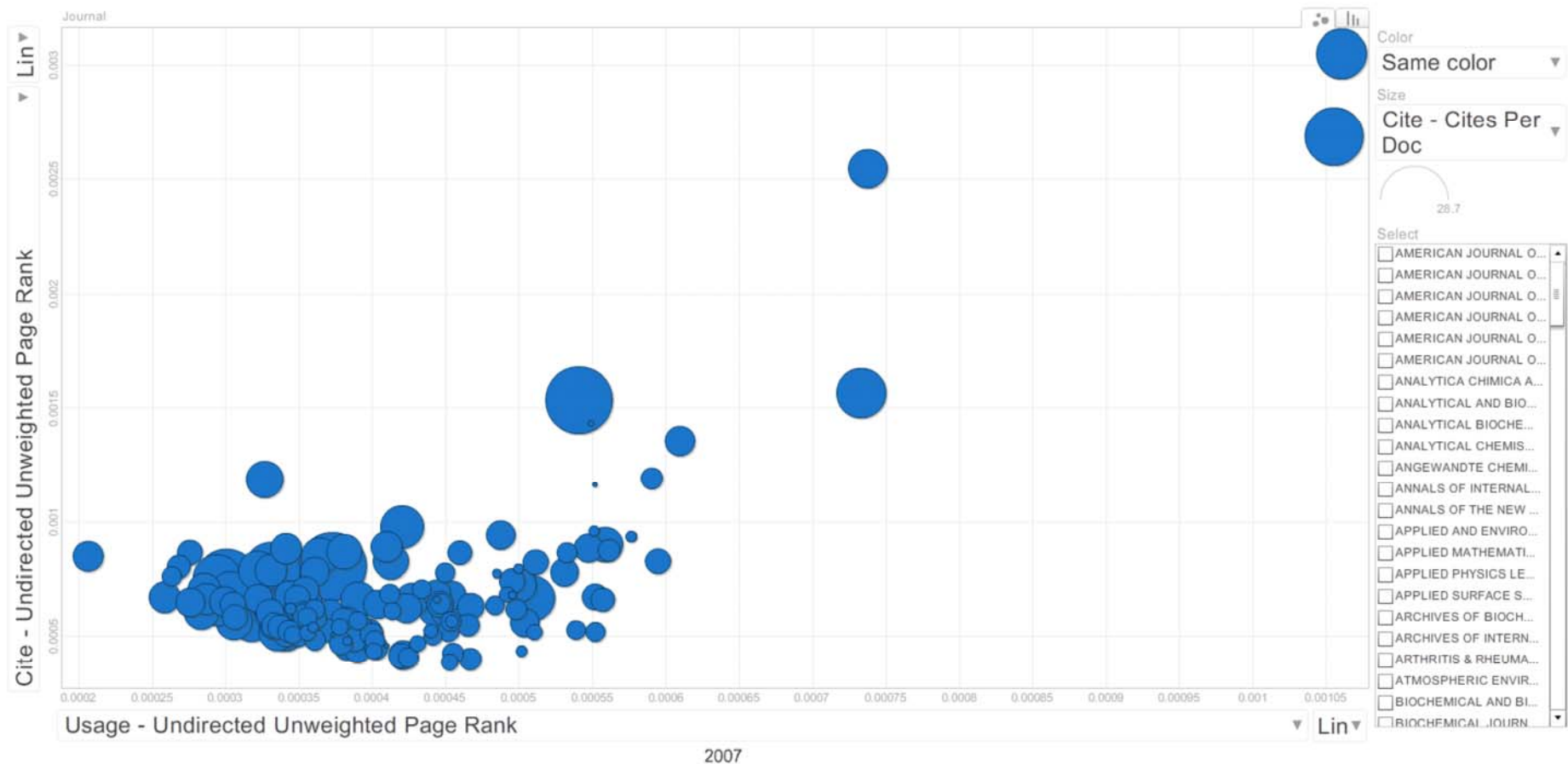


An interactive journal ranking service that allows you explore the top journals in a domain according to a variety of different impact metrics derived from MESUR's usage data collection.

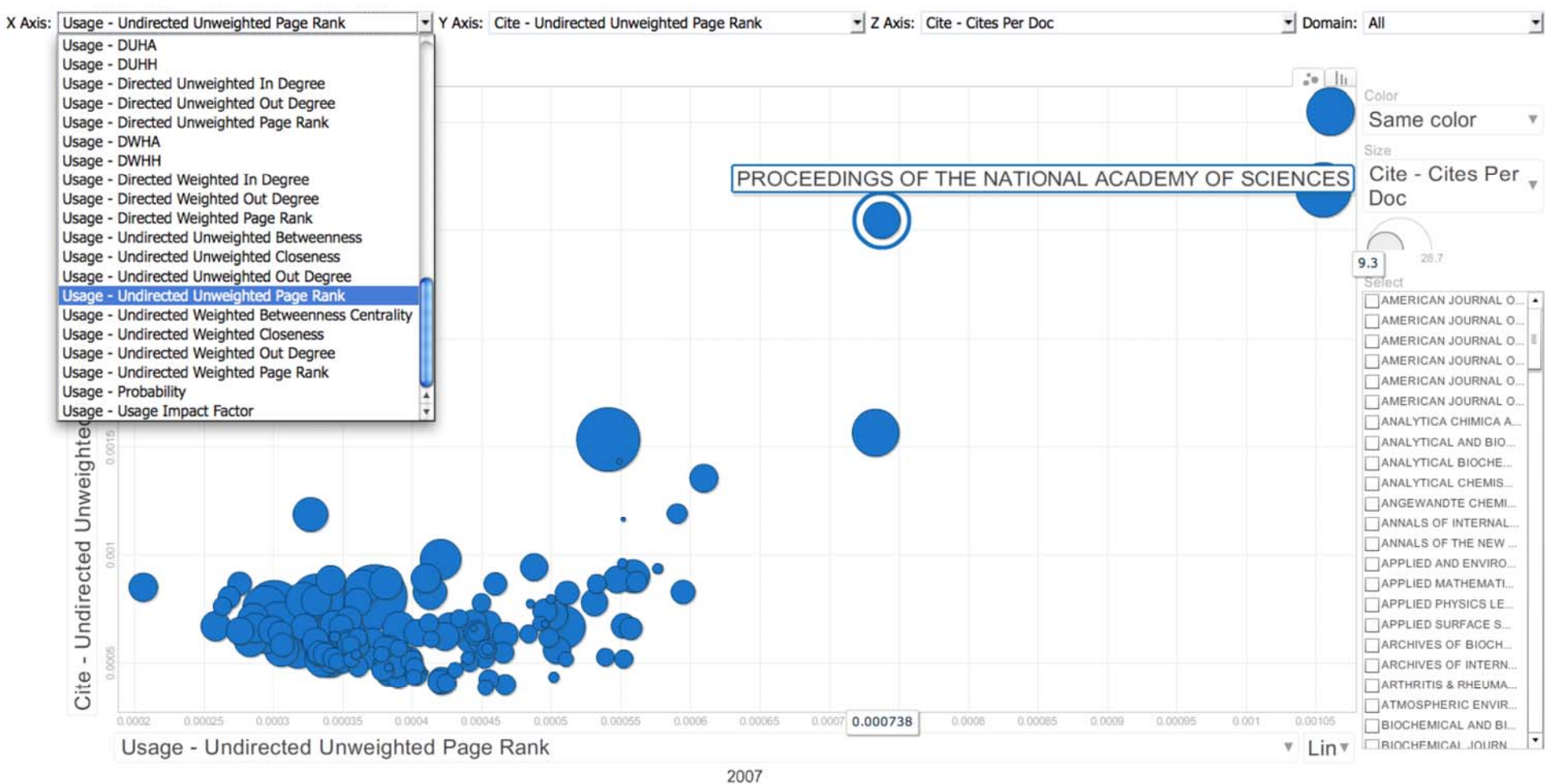
 [go to journal ranks](#)

Announcement:
MESUR has received an NSF grant to pursue...

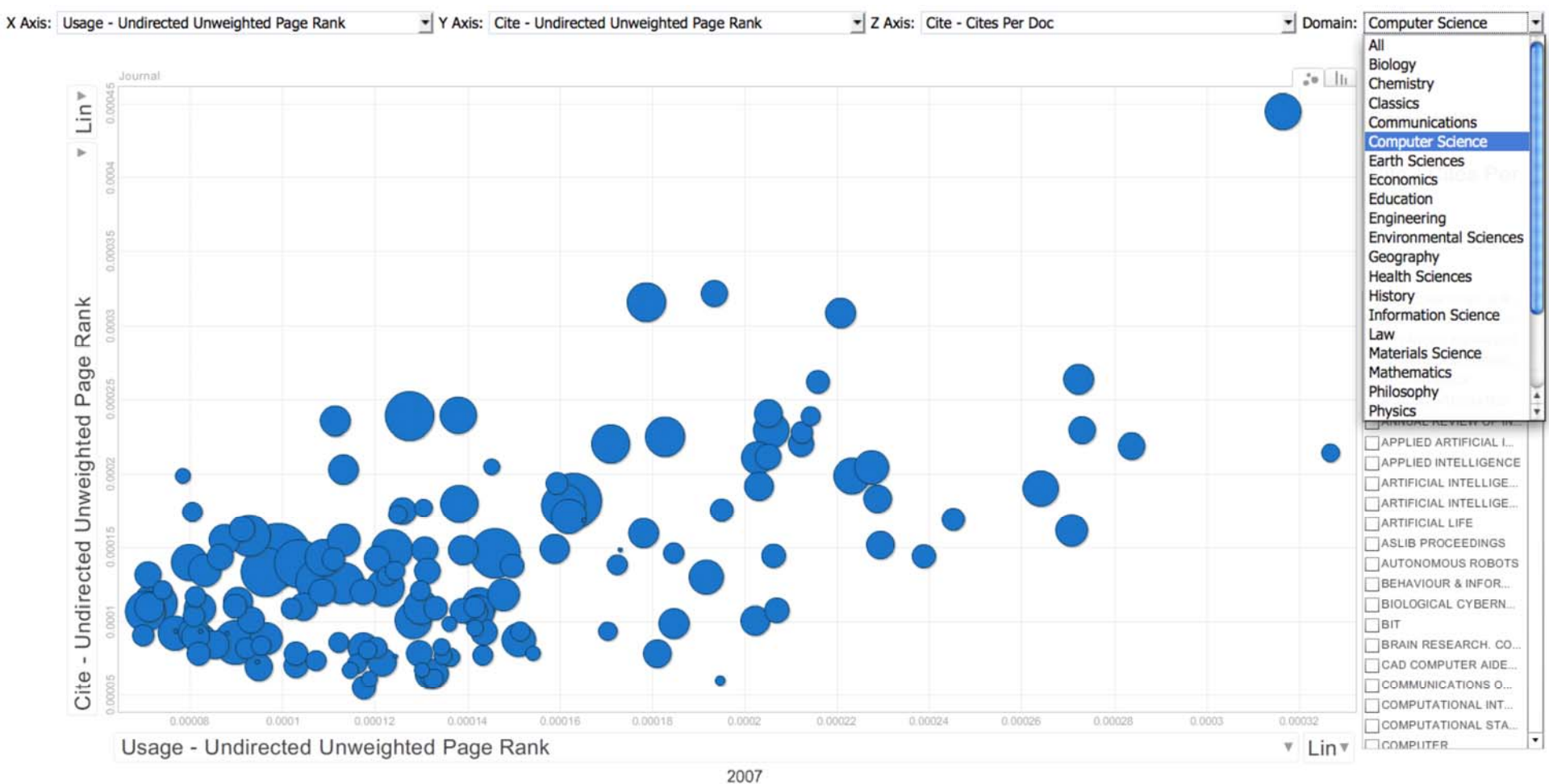
Press:
Discussion of map of science in EOS, a prominent Belgian science magazine.



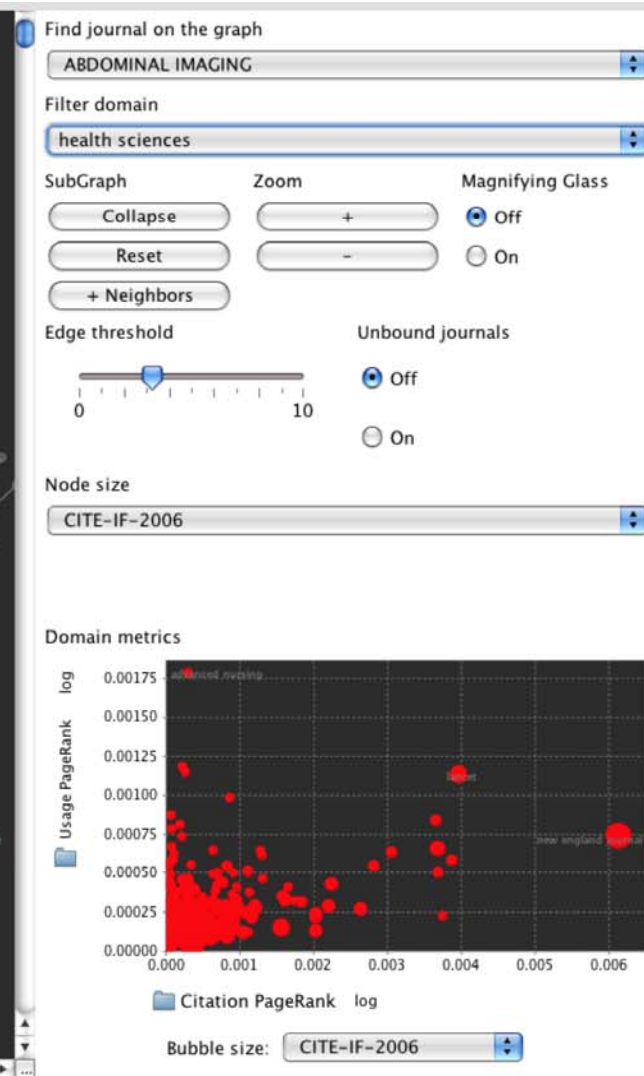
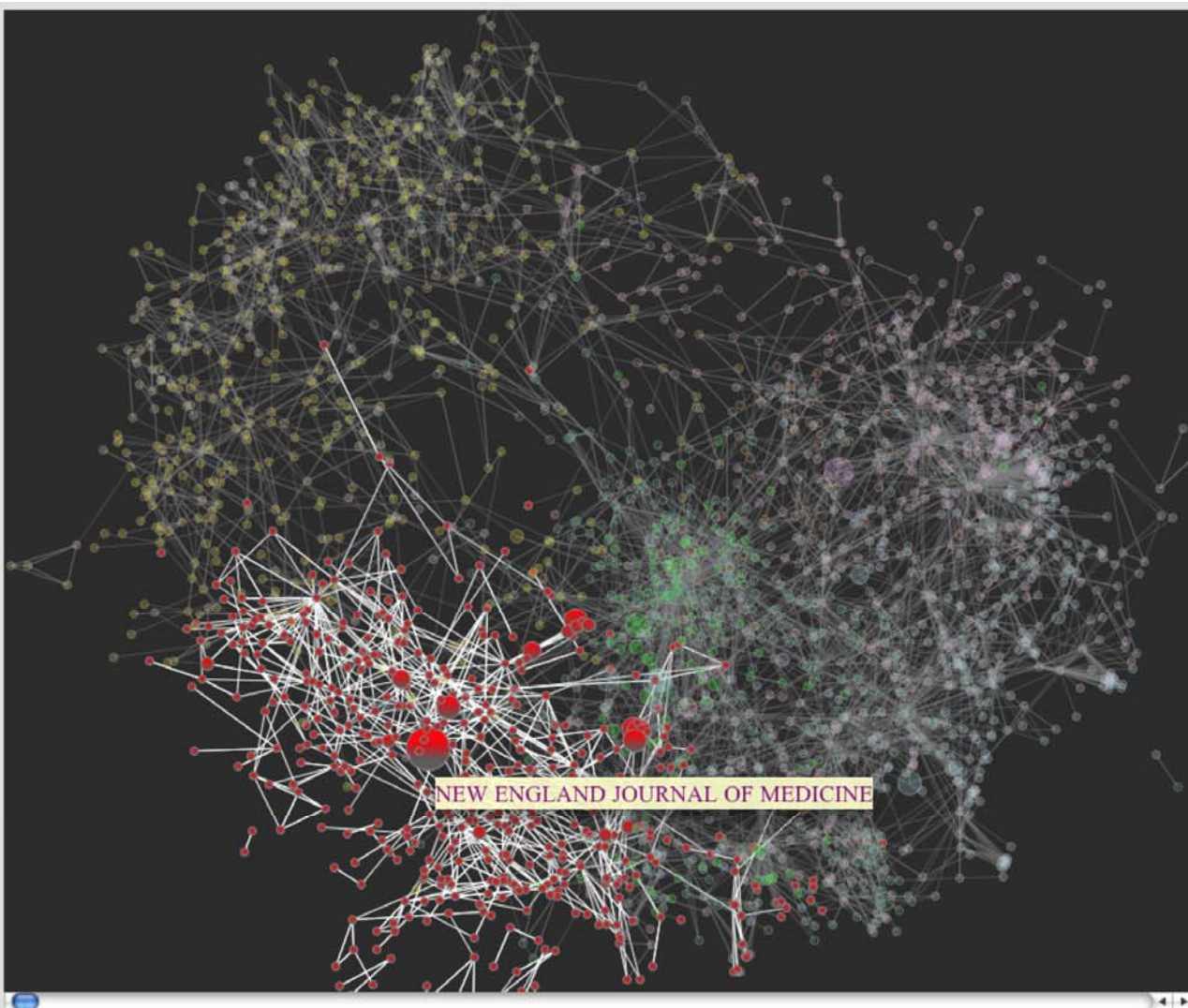
Journal	Year	Rank	Domain	Usage - Undirected Unweighted Page Rank	Cite - Undirected Unweighted Page Rank	Cite - Cites Per Doc
SCIENCE	2007	1	science	0.001060366	0.0030490425	15.4600000381
NATURE	2007	2	science	0.0010552662	0.0026860067	20.7700004578
PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES	2007	3	science	0.0007375684	0.0025442564	9.2700004578
LANCET	2007	4	health sciences	0.0007331272	0.0015647924	15.0600004196
NEW ENGLAND JOURNAL OF MEDICINE	2007	5	health sciences	0.0005408772	0.0015339617	27.7099999845
JOURNAL OF BIOLOGICAL CHEMISTRY	2007	6	chemistry	0.0006096485	0.0013558392	5.4699997902
JAMA THE JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION	2007	7	health sciences	0.0005490104	0.0014342124	0.2000000003



Journal	Year	Rank	Domain	Usage - Undirected Unweighted Page Rank	Cite - Undirected Unweighted Page Rank	Cite - Cites Per Doc
SCIENCE	2007	1	science	0.001060366	0.0030490425	15.4600000381
NATURE	2007	2	science	0.0010552662	0.0026860067	20.7700004578
PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES	2007	3	science	0.0007375684	0.0025442564	9.2700004578
LANCET	2007	4	health sciences	0.0007331272	0.0015647924	15.0600004196
NEW ENGLAND JOURNAL OF MEDICINE	2007	5	health sciences	0.0005408772	0.0015339617	27.7099990845
JOURNAL OF BIOLOGICAL CHEMISTRY	2007	6	chemistry	0.0006096485	0.0013558392	5.4699997902
JAMA THE JOURNAL OF THE AMERICAN MEDICAL ASSOCIATION	2007	7	health sciences	0.0005490104	0.0014342124	0.200000003



Journal	Year	Rank	Domain	Usage - Undirected Unweighted Page Rank	Cite - Undirected Unweighted Page Rank	Cite - Cites Per Doc
JOURNAL OF COMPUTATIONAL PHYSICS	2007	1	computer science	0.0003164522	0.0004445359	2.6700000763
PATTERN RECOGNITION	2007	2	computer science	0.0002722307	0.0002634107	1.8500000238
MATHEMATICAL AND COMPUTER MODELLING	2007	3	computer science	0.0003267292	0.0002140573	0.6399999857
COMMUNICATIONS OF THE ACM	2007	4	computer science	0.000220716	0.0003084775	1.8400000334
COMPUTERS & CHEMICAL ENGINEERING	2007	5	computer science	0.0002729736	0.000229205	1.4400000572
THEORETICAL COMPUTER SCIENCE	2007	6	computer science	0.0001934019	0.0003216609	1.3999999762
EXPERT SYSTEMS WITH APPLICATIONS	2007	7	computer science	0.0002836945	0.0002185048	1.4400000572



MESUR: the good ...

After 2 years of MESUR:

- Scientific exploration of metrics for scholarly evaluation
- Creation of large-scale reference data set
- Mapping science from the viewpoint of users: there **is** structure!
- Variety of Metrics that cover various aspects of scholarly impact and prestige
- MESUR dataset contains many more pearls for future research
- Foundation for future continued research program:
 - Longitudinal studies
 - Models of collective behavior of scientists

MESUR: the bad and the ugly ...

Scalability of the approach:

- Lengthy negotiations to obtain log data
- No infrastructure standards (yet): Recording, aggregating, normalization, ingestion, de-duplication,...
- No generally accepted policies: privacy, property, ...
- No census data: when is a sample large and representative enough?

Quality control:

- Bots, Crawlers (detectable but never perfect)
- Cheating, manipulation (easier with usage statistics than network metrics)

Acceptance:

- Network-based usage metrics require session information. This is overlooked! As a result, will we end up with usage-based statistics only?
- “As simple as possible, but not more simple!”

Publications related to MESUR

Johan Bollen, Herbert Van de Sompel, Aric Hagberg, Luis Bettencourt, Ryan Chute, Marko A. Rodriguez, Lyudmila Balakireva. **Clickstream data yields high-resolution maps of science.** PLoS One, March 2009.

Johan Bollen, Herbert Van de Sompel, Aric Hagberg, Ryan Chute. **A principal component analysis of 39 scientific impact measures.** [arXiv.org/abs/0902.2183](https://arxiv.org/abs/0902.2183) (accepted for publication in PLoS ONE)

Johan Bollen, Herbert Van de Sompel, and Marko A. Rodriguez. **Towards usage-based impact metrics: first results from the MESUR project.** In Proceedings of the Joint Conference on Digital Libraries, Pittsburgh, June 2008

Marko A. Rodriguez, Johan Bollen and Herbert Van de Sompel. **A Practical Ontology for the Large-Scale Modeling of Scholarly Artifacts and their Usage,** In Proceedings of the Joint Conference on Digital Libraries, Vancouver, June 2007

Johan Bollen and Herbert Van de Sompel. **Usage Impact Factor: the effects of sample characteristics on usage-based impact metrics.** (cs.DL/0610154)

Johan Bollen and Herbert Van de Sompel. **An architecture for the aggregation and analysis of scholarly usage data.** In Joint Conference on Digital Libraries (JCDL2006), pages 298-307, June 2006.

Johan Bollen and Herbert Van de Sompel. **Mapping the structure of science through usage.** Scientometrics, 69(2), 2006.

Johan Bollen, Marko A. Rodriguez, and Herbert Van de Sompel. **Journal status.** Scientometrics, 69(3), December 2006 (arxiv.org:cs.DL/0601030)

Johan Bollen, Herbert Van de Sompel, Joan Smith, and Rick Luce. **Toward alternative metrics of journal impact: a comparison of download and citation data.** Information Processing and Management, 41(6):1419-1440, 2005.