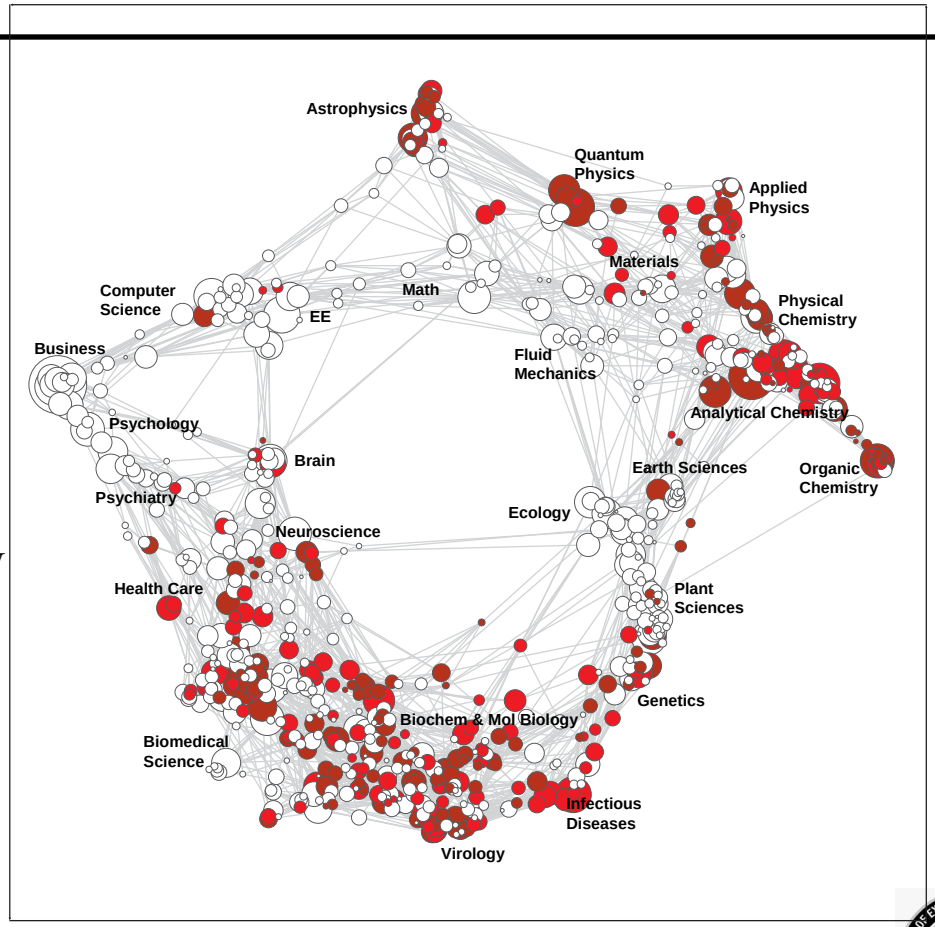
Generation of Large-Scale Maps of Science and Associated Indicators

Kevin W. Boyack
Sandia National Laboratories
May 8, 2006

Slides of national strengths, courtesy of my collaborator, Richard Klavans, SciTech Strategies, Inc.



Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

Collaborators

- Richard Klavans, SciTech Strategies, Berwyn, PA
  - Science maps, indicators, past 3 years
  - National science strategy slides (later) are his
- Katy Börner, Indiana University, SLIS
  - Early science maps, past 5 years
- George Davidson, Shawn Martin, SNL
  - Related studies (genomics, clustering, text mining) using VxInsight, VxOrd

Historical Overview

- Mid-to-late 1990s – Desire to map all of science to better guide the S&T enterprise at Sandia
  - Meetings with ISI
  - Internal funding (LDRD) of VxInsight visualization tool
  - Never got there; limited to maps of ~50,000 documents
- 2003-2005 – Key advances
  - VxOrd graph layout tool structures re-written to enable graph layout of millions of nodes
  - Boyack and Klavans meet in 2003, exchange ideas, and start working together to map "all of science and technology"
  - We can now create yearly maps of the ISI data, link yearly maps together, and calculate indicators at the "cluster of papers" level

Outline

- Goals and uses of science maps
- Background and terminology
- Recent evolution in science mapping (Boyack & Klavans)
  - First validation study at the journal level
  - Three maps at the paper level
  - Overlay of indicators, distributions on map substrates
- Places & Spaces exhibit

Goals of Mapping Science

- Provide a better basis for evaluation and planning of R&D
  - Past efforts break the S&T into few (e.g. 7) categories with little detail
  - Our methods generate indicators of potential impact at "research community" levels, which enable better
    - national comparisons
    - institutional comparisons
    - evaluation of potential partners
    - …
- In general, mapping science is aimed to enable better R&D through:
  - Identification and evaluation of current work in a global context
  - Identification and evaluation of proposed work in a global context
  - Identification of highly-ranked communities in areas related to current work
  - Identification of research entry points (or potential collaborators) and emerging applications in specific areas of focus
  - Identification of opportunity and vulnerability using institutional comparisons
  - Better understanding of the innovation process and better anticipation of future trends

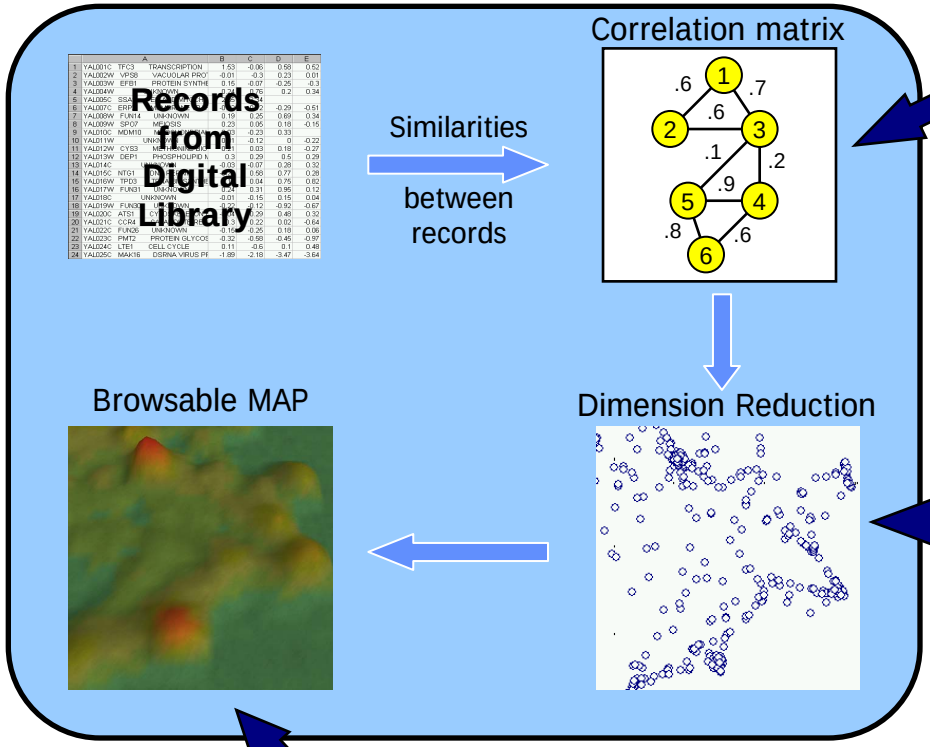
Uses of Science Maps

- Future possibilities
  - Create data mining system and interface to make "research community" data available to all Sandians
    - Find the communities they are currently a part of
    - Explore communities they might want to join
    - Metrics and indicators
  - Explore innovation pathways
    - Add funding and grants, patents, VC holdings, etc. to link the entire idea-to-product pathways
    - Map known innovation case studies onto the data; identify promising graph patterns tied to known innovation
    - Prediction by applying known patterns to current partial patterns
  - Science education through map-based discovery
    - Innovation case studies are discovery-based stories; ways to educate using these maps could be a natural by-product of innovation studies

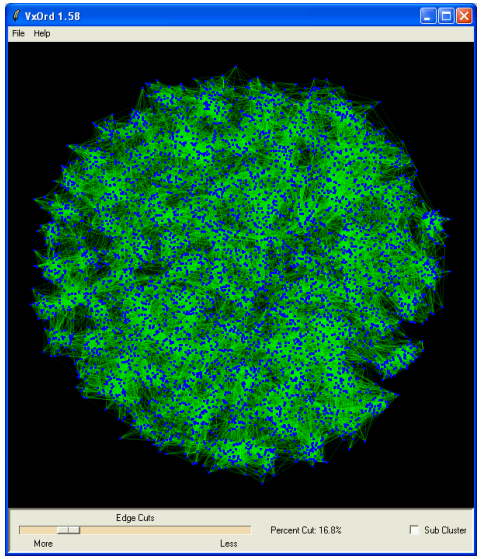
Sandia Mapping Process

Common index values such as Cosine

N_{ij} / sqrt($N_i N_j$)



VxOrd

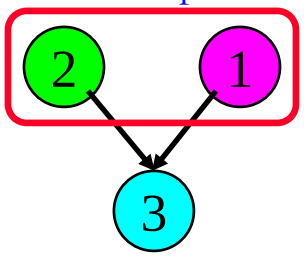


VxInsight

Bibliographic Coupling & Co-citation

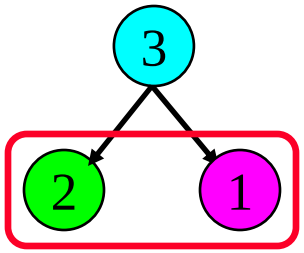
- **Bibliographic coupling** occurs when two papers (or journals) have common elements in their reference lists or bibliographies
  - Used to create maps of current science

PAPERS
1 and 2 are
bib coupled



- **Co-citation** occurs when two papers (or journals) are cited by a subsequent paper
  - Used to create maps of earlier or past science

PAPERS
1 and 2 are
co-cited



Terminology

- Current map – one based on bibliographic coupling where all objects are "current" fileyear papers
- Reference (or base) map – one based on co-citation where all objects are reference papers that were co-cited above a given threshold by the "current" fileyear papers
- Disciplinary map – a journal map
- (Sub)discipline – a cluster of journals in a journal map
- Research community – a cluster of papers on the same topic
  - can be either a current community or a base (reference) community
- Paradigm – a supercluster of reference papers
  - paradigms are created by 3 recursive clusterings of the reference papers

Context is Critical

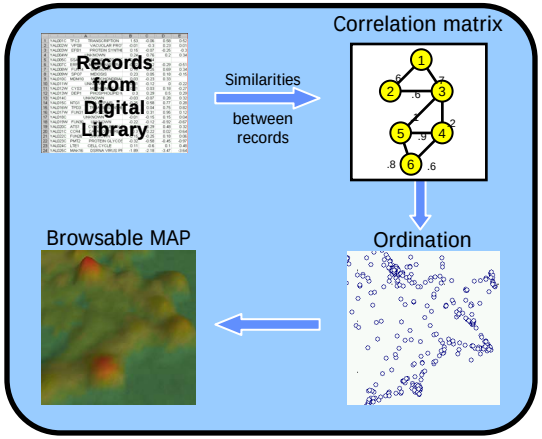
- Science maps provide a quantitative way to break up the literature into communities that are amenable to examination and comparison
- Comparison should be done at the community level or higher (not at the paper level)
- Communities provide context for comparison
  - when comparing institutions, nations, etc., communities and paradigms show relative investment strategies and strengths rather than lumping together uncommon things and reducing to one number
  - this provides for greater richness in the analysis and better characterization as to what is really going on, and how to take advantage of it

Strategy

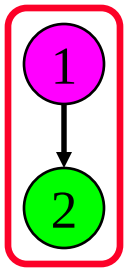
- Develop and **validate** process, methods, and algorithms at small scale (~7k objects)
  - Macromodel
  - Using ISI citation data, create disciplinary maps of science using journals (~7000 titles)
  - Validate using the known journal categorization structure
- Employ validated process, methods, and algorithms at larger scale (~1M objects)
  - Micromodel
  - Create paper-level (~1M annually) maps of science from ISI citation data
  - Validate detailed maps at local structural levels where possible
  - Calculate indicators and metrics at the cluster or community level

Validation Study: Macromodel Process

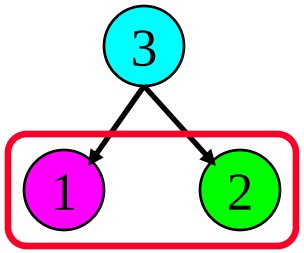
- Identify individual journals
- Calculate similarity between journals from inter-citation data and co-citation data
- Use VxOrd to determine coordinates for each journal
- Generate cluster assignments (k-means)
- Validate against ISI journal category assignments
  - Simple yes/no
  - Entropy measure



Inter-citation
1 cites 2



Co-citation
1 and 2 are co-cited



Journal Citation Data

Inter-citation

Ci,j Citing Journal

	1 Nature	2 Astrophys J.	3 Cell	4 Embo Rep	5 J. Biol Chem	6 Paleoceanography	7 PNAS	8 Science
1	4462	313	1224	2	861	138	1361	2380
2	2015	34264					22	359
3	1258		2033		887		960	1112
4	123			4				
5	11080		10695		43652		14881	8680
6	323					418		147
7	4887	22	3512		5099		7802	4737
8	2330	422	975		586	99	1100	3255

RAWi,j (IC Raw)

	1	2	3	4	5	6	7	8
1		2328	2482	125	11941	461	6248	4710
2							44	781
3					11582		4472	2087
4								
5							19980	9226
6								246
7								5837
8								

IC-Cosine
IC-K50
IC-Jaccard
IC-Pearson

IC RFavg

Co-citation

Fi,j (CC Raw)

	1	2	3	4	5	6	7	8
1		146984	584695	48	775522	19745	746379	714875
2			77		352	2	391	430
3				47	680721	3	589480	461689
4					42		55	28
5						7	1E+06	641486
6							196	8819
7								672839
8								

CC-Cosine
CC-K50
CC-Pearson

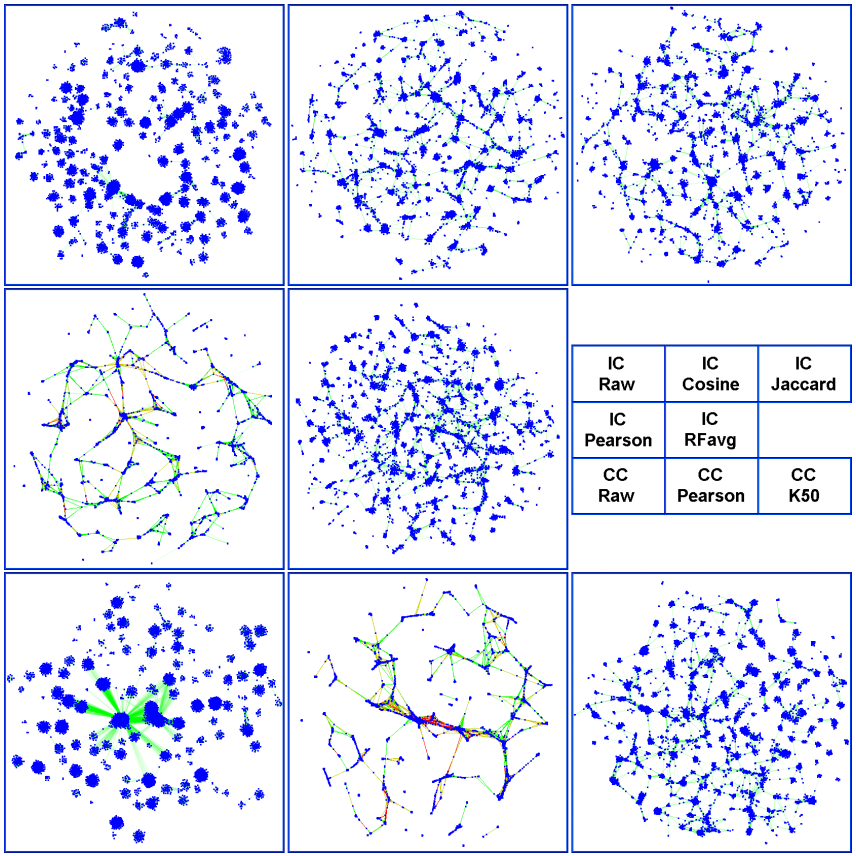
Comparison

ISI category cooccurrence

	1	2	3	4	5	6	7	8
1		0	0	0	0	0	1	1
2			0	0	0	0	0	0
3				1	1	0	0	0
4					1	0	0	0
5						0	0	0
6							0	0
7								1
8								

Macromodel: Different Similarity Metrics

- ISI file year 2000, SCIE and SSCI
- Ten different similarity metrics
  - 6 Inter-citation (raw counts, cosine, modified cosine, Jaccard, RF, Pearson)
  - 4 Co-citation (raw counts, cosine, modified cosine, Pearson)
- Inter-citation gives structure based on current citing patterns
- Co-citation gives structure based on how science is currently used



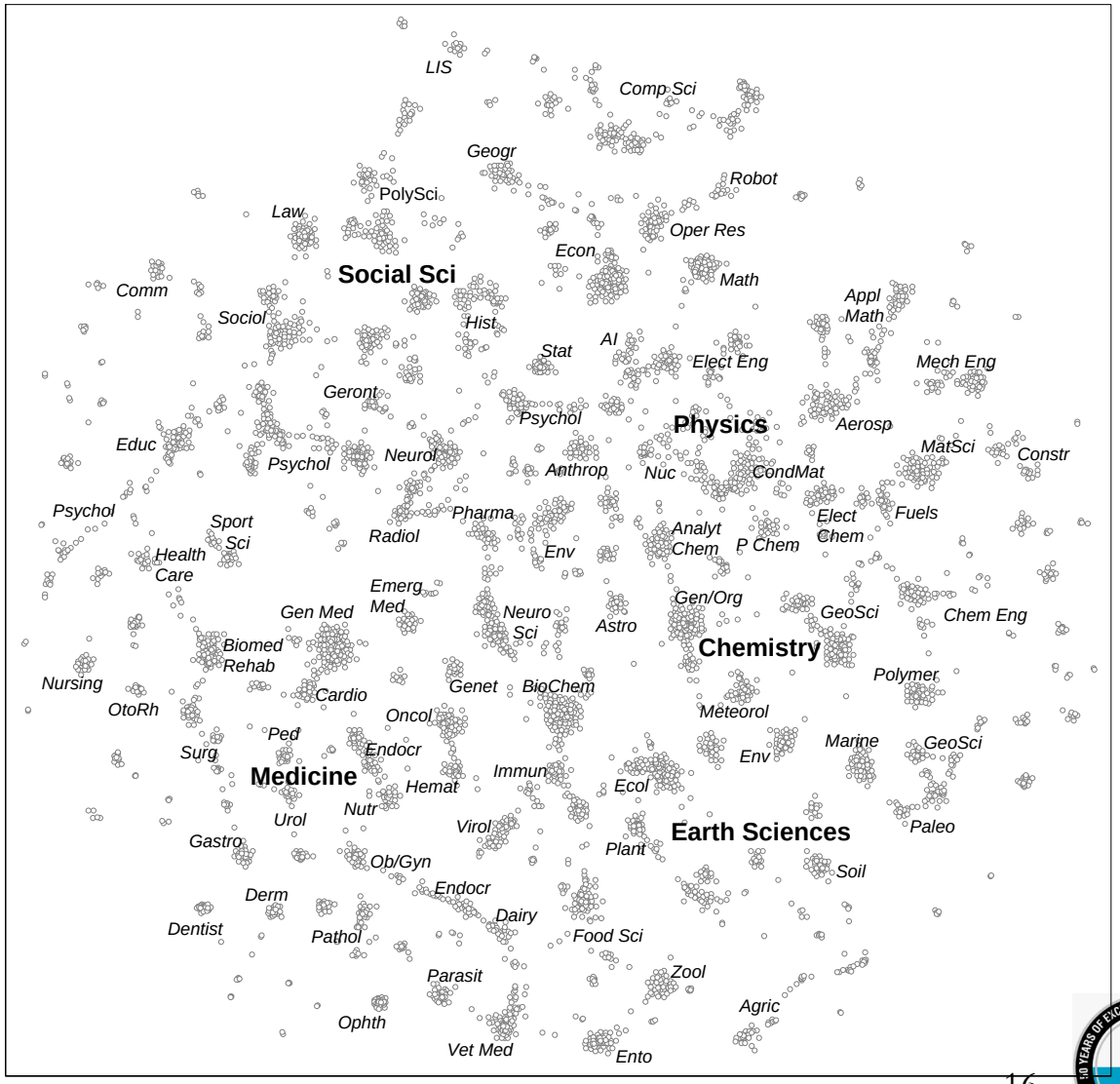
Boyack, K.W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics* 64(3), 351-374.

Accuracy & Validation

- Lots of effort and work to validate and calculate the accuracy of these methods at the journal and paper levels
  - 3 published articles, 1 conference paper

Macromodel: "Best" Map

- Each dot is one journal
- Journals group by discipline
- Labeled by hand



Boyack, K.W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics* 64(3), 351-374.

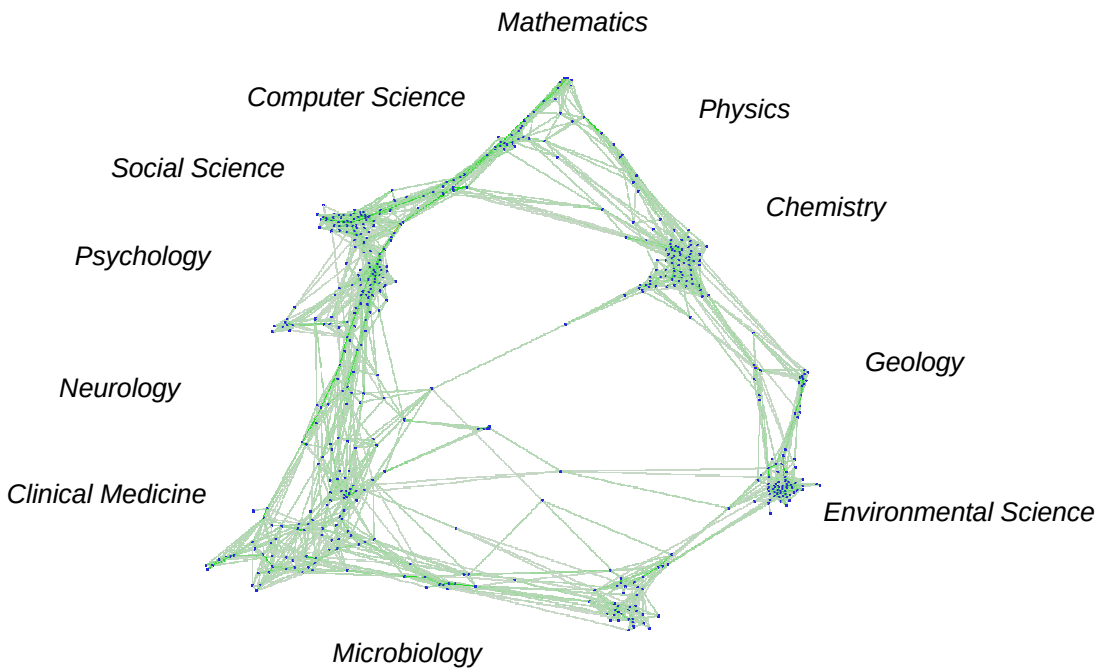
Macromodel: Structural Map

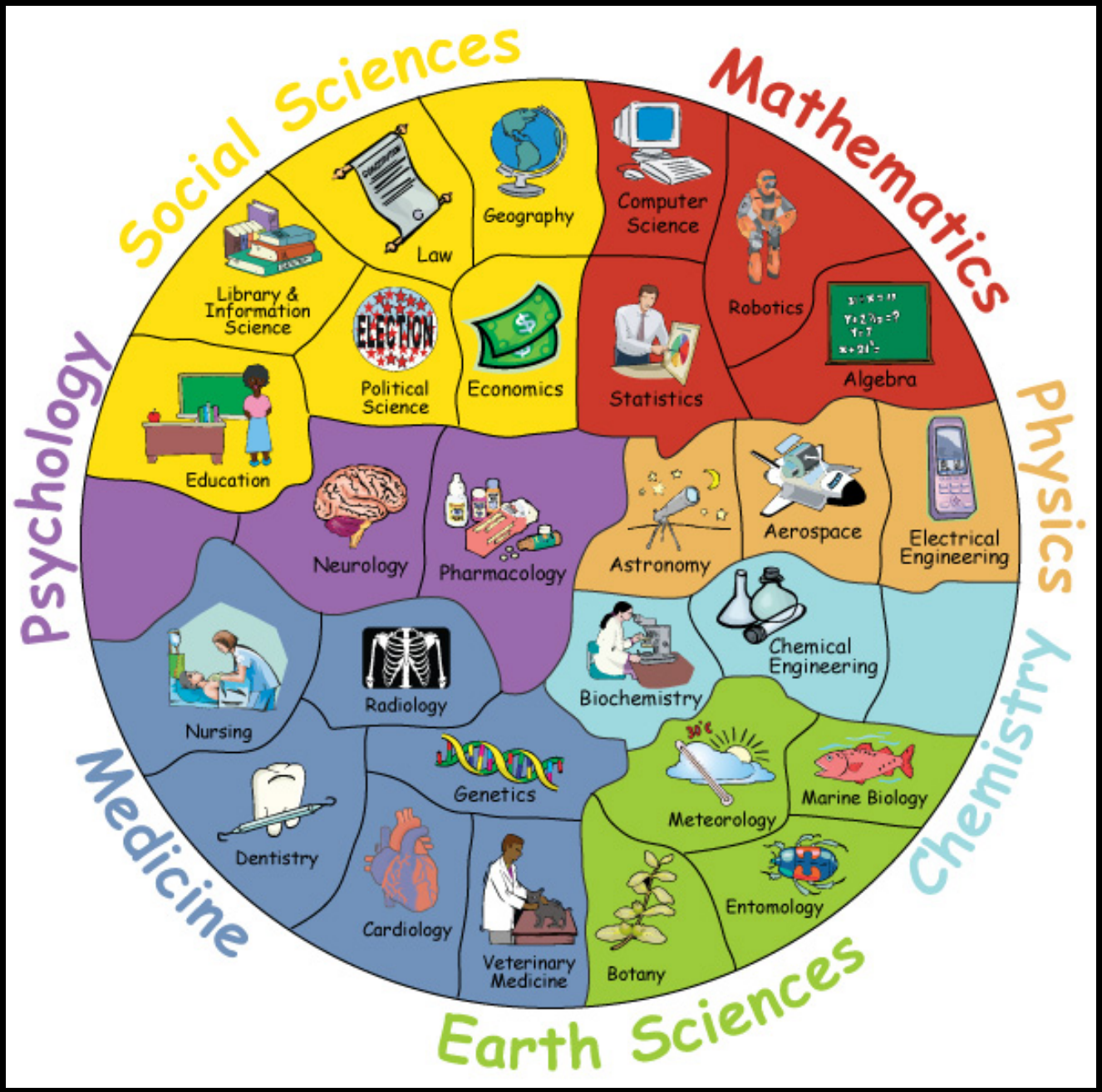
- Clusters of journals denote disciplines
- Lines denote strongest relationships between disciplines
- Enables disciplinary diffusion studies
- This map implies that there is something in the CENTER of science
- Closer inspection suggests this is wrong

Boyack, K.W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics* 64(3), 351-374.

Macromodel: Clustering Clusters

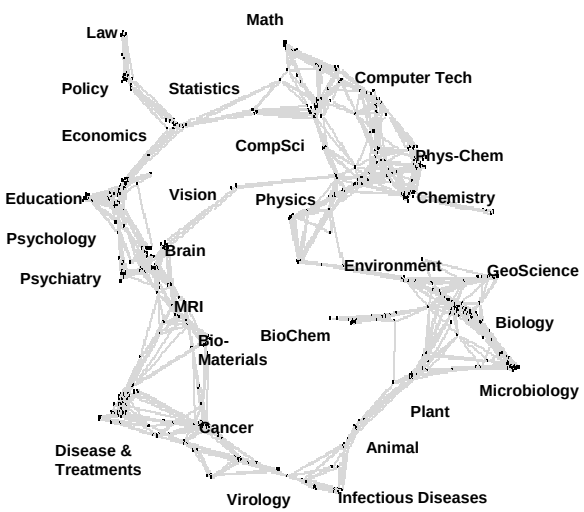
- During ordination, we notice a ring structure
- VxOrd is a space filling algorithm; things in the "middle" have backfilled empty space
- Cluster the clusters to mitigate this effect




Social Sciences
Mathematics
Physics
Chemistry
Earth Sciences
Medicine
Psychology
Geography
Computer Science
Law
Robotics
Library & Information Science
Political Science
Economics
Statistics
Algebra
Education
Neurology
Pharmacology
Astronomy
Aerospace
Electrical Engineering
Chemical Engineering
Biochemistry
Radiology
Nursing
Genetics
Meteorology
Marine Biology
Dentistry
Cardiology
Veterinary Medicine
Botany
Entomology

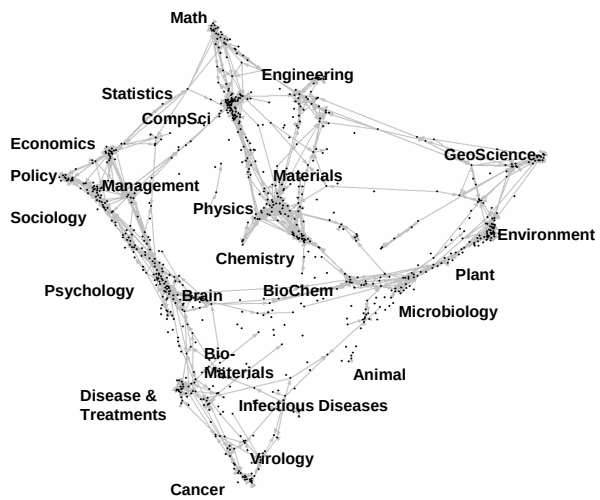
Three Maps

(1)
2002 Science Map
671 Disciplines
(major relationships shown)



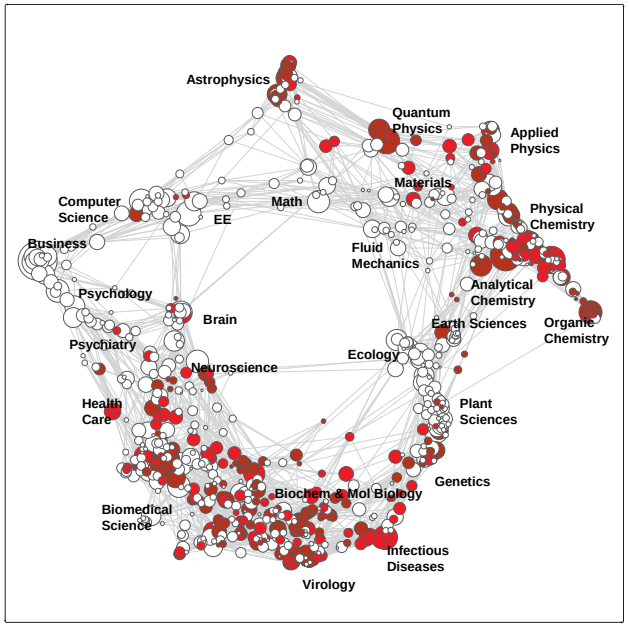
SCIE/SSCI
bibliographic coupling
journals
community overlays

(2)
2003 S&T Map
860 Disciplines
(major relationships shown)



SCIE/SSCI/Proceedings
bibliographic coupling
journals
community overlays

(3)
2003 Paradigm Map
776 Paradigms
(major relationships shown)



SCIE/SSCI
co-citation
papers
3 levels of clustering

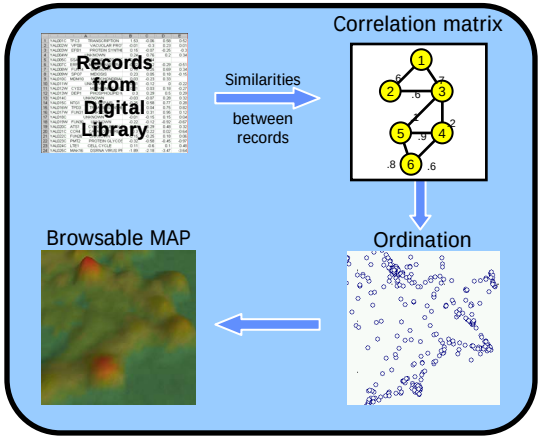
(1-2) Literature Maps: Multi-step Process

- Calculate a "disciplinary model" using journal-level maps of science and technology
  - bibliographic coupling

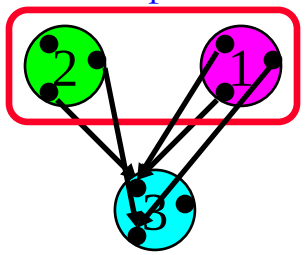
- Calculate a "research community model" (where research communities are clusters of papers), and locate them on the disciplinary map
  - bibliographic coupling

Disciplinary Model

- Identify individual journals (~10,000)
- Calculate similarity between journals from bibliographic coupling of reference data
  - Coupling at the paper level
  - Aggregate coupling counts at the journal level
- Use VxOrd to determine coordinates for each journal
- Generate cluster assignments using a modified single-link algorithm
- Reaggregate coupling counts at the journal cluster level and "cluster the clusters"

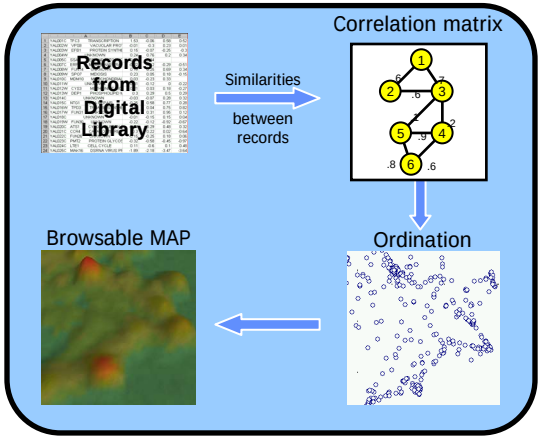


JOURNALS
1 and 2 are coupled

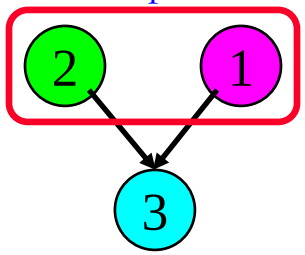


Research Community Model

- Identify individual papers (~1,000,000)
- Calculate similarity between papers from bibliographic coupling of reference data
- Use VxOrd to determine coordinates for each paper
- Generate **research community** (cluster) assignments using a modified single-link algorithm
- Calculated research community positions based on the journal distribution in each community using the coordinates from the disciplinary map

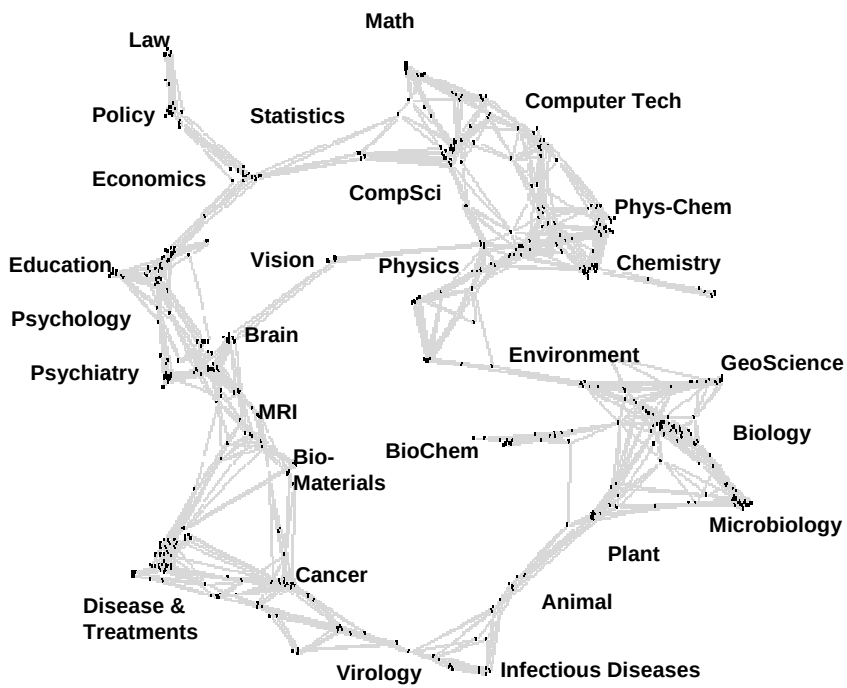


PAPERS
1 and 2 are coupled



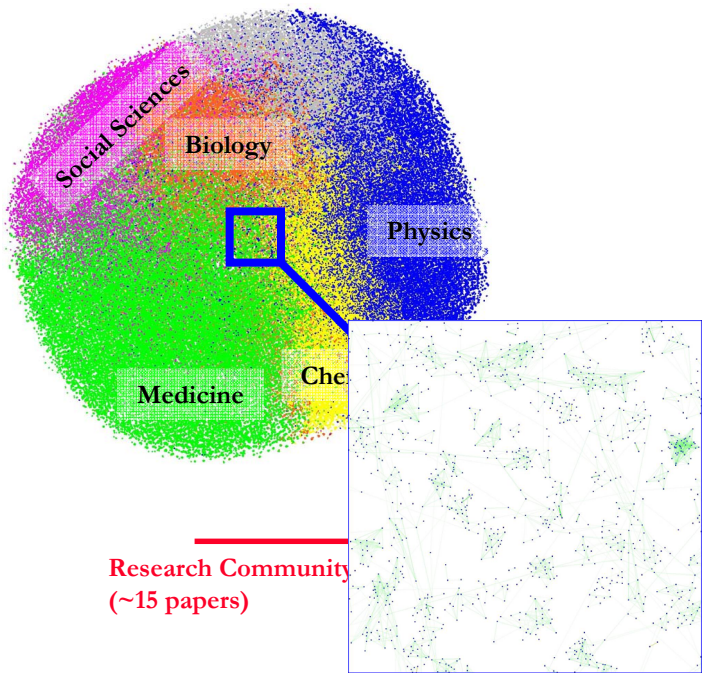
(1) Disciplinary Science Model - Details

- Uses combined SCIE/SSCI from 2002
  - 1.07M papers, 24.5M references, 7300 journals
  - Bibliographic coupling of papers, aggregated to journals
- Initial ordination and clustering of journals gave 671 clusters
- Coupling counts reaggregated at the journal cluster level; ordination of journal clusters
  - (x,y) positions for each journal cluster
  - by association, (x,y) positions for each journal



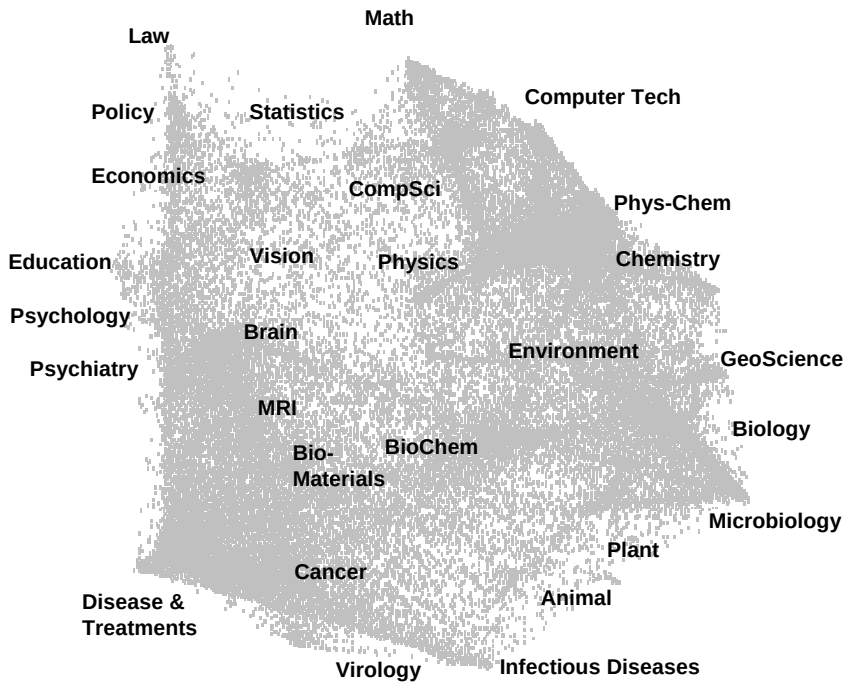
(1) Research Community Model - Details

- Uses combined SCIE/SSCI from 2002
  - 1.07M papers, 24.5M references, 7300 journals
  - Bibliographic coupling of papers
- Initial ordination and clustering of journals gave 96,500 clusters for 731,000 papers
  - The other 340,000 papers are editorials, book reports, etc. that have no citation links
- Cluster positions calculated from journal positions on the disciplinary map



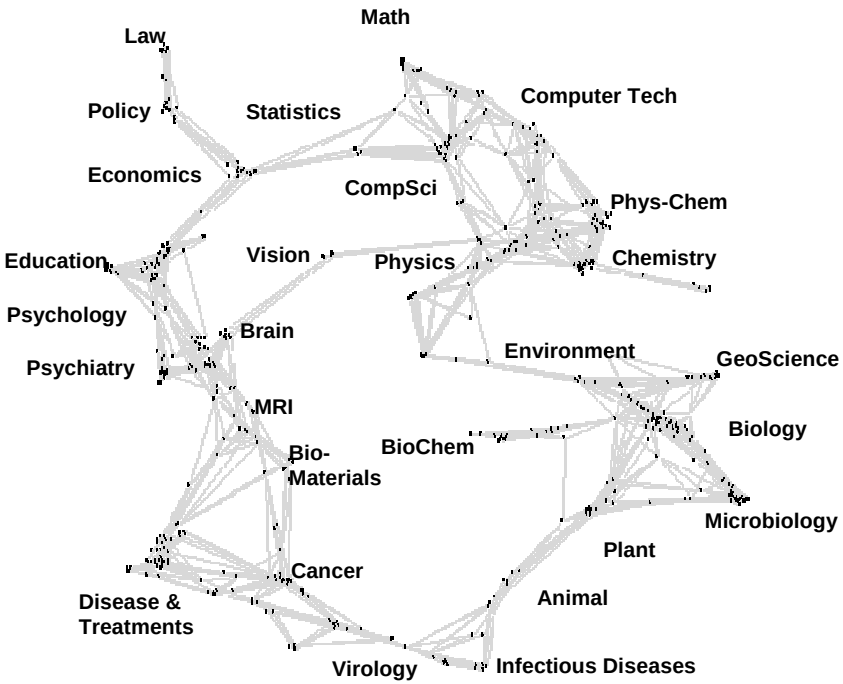
(1) Research Community Model - Details

- Uses combined SCIE/SSCI from 2002
  - 1.07M papers, 24.5M references, 7300 journals
  - Bibliographic coupling of papers
- Initial ordination and clustering of journals gave 96,500 clusters for 731,000 papers
  - The other 340,000 papers are editorials, book reports, etc. that have no citation links
- Cluster positions calculated from journal positions on the disciplinary map

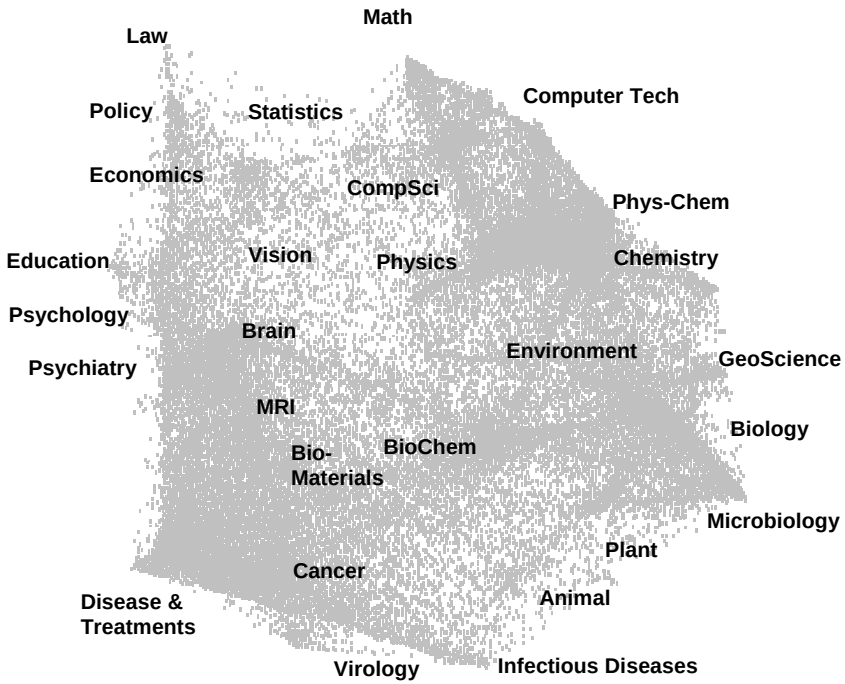


(1) Side-by-Side

671 Disciplines
(major relationships shown)



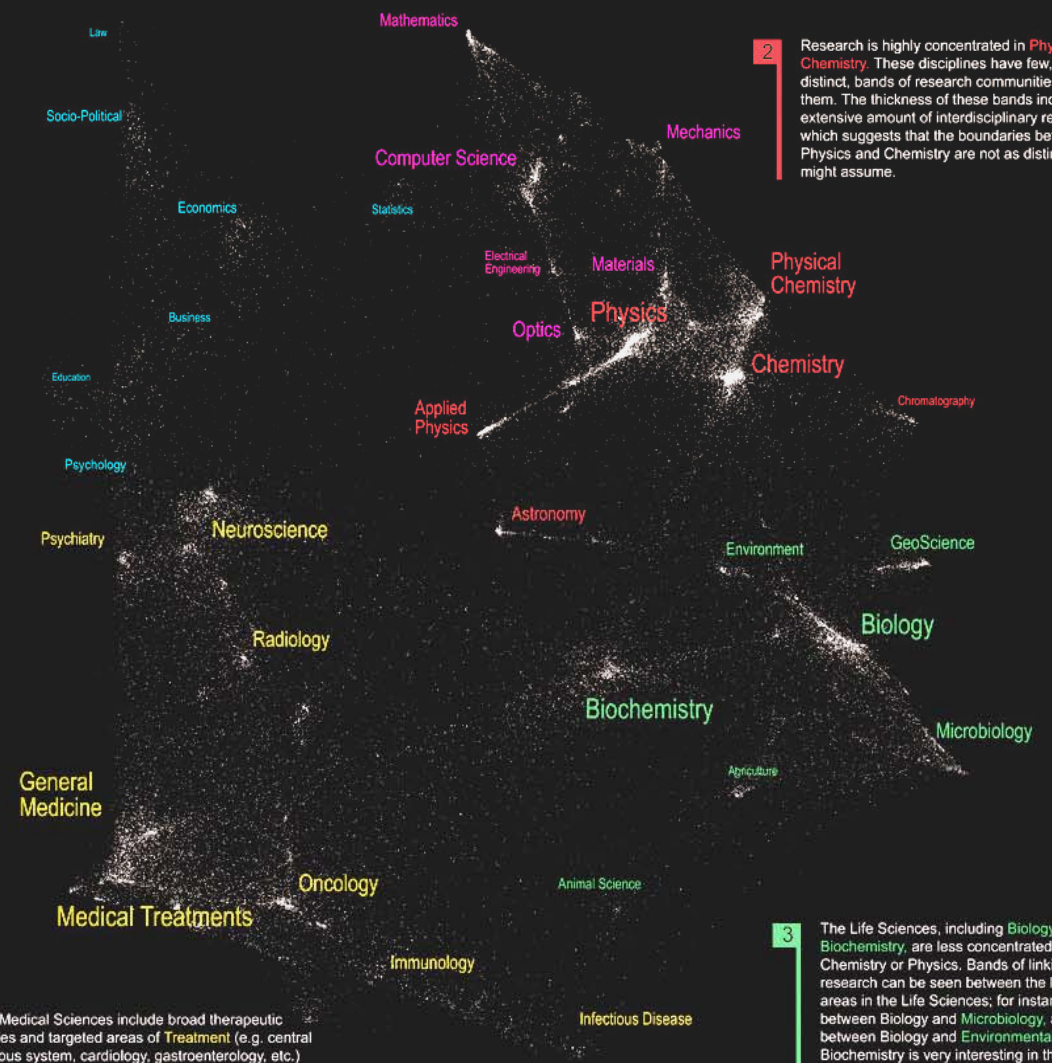
96,500 Research Communities
(no relationships shown)



The Structure of Science

5 The Social Sciences are the smallest and most diffuse of all the sciences. Psychology serves as the link between Medical Sciences (Psychiatry) and the Social Sciences. Statistics serves as the link with Computer Science and Mathematics.

1 Mathematics is our starting point, the purest of all sciences. It lies at the outer edge of the map. Computer Science, Electrical Engineering, and Optics are applied sciences that draw upon knowledge in Mathematics and Physics. These three disciplines provide a good example of a linear progression from one pure science (Mathematics) to another (Physics) through multiple disciplines. Although applied, these disciplines are highly concentrated with distinct bands of research communities that link them. Bands indicate interdisciplinary research.



2 Research is highly concentrated in Physics and Chemistry. These disciplines have few, but very distinct, bands of research communities that link them. The thickness of these bands indicates an extensive amount of interdisciplinary research, which suggests that the boundaries between Physics and Chemistry are not as distinct as one might assume.

4 The Medical Sciences include broad therapeutic studies and targeted areas of Treatment (e.g. central nervous system, cardiology, gastroenterology, etc.) Unlike Physics and Chemistry, the medical disciplines are more spread out, suggesting a more multi-disciplinary approach to research. The transition into Life Sciences (via Animal Science and Biochemistry) is gradual.

3 The Life Sciences, including Biology and Biochemistry, are less concentrated than Chemistry or Physics. Bands of linking research can be seen between the larger areas in the Life Sciences; for instance between Biology and Microbiology, and between Biology and Environmental Science. Biochemistry is very interesting in that it is a large discipline that has visible links to disciplines in many areas of the map, including Biology, Chemistry, Neuroscience, and General Medicine. It is perhaps the most interdisciplinary of the sciences.

We are all familiar with traditional maps that show the relationships between countries, provinces, states, and cities. Similar relationships exist between the various disciplines and research topics in science. This allows us to map the structure of science.

One of the first maps of science was developed at the Institute for Scientific Information over 30 years ago. It identified 41 areas of science from the citation patterns in 17,000 scientific papers. That early map was intriguing, but it didn't cover enough of science to accurately define its structure.

Things are different today. We have enormous computing power and advanced visualization software that make mapping of the structure of science possible. This galaxy-like map of science (left) was generated at Sandia National Laboratories using an advanced graph layout routine (VxOrd from the citation patterns in 800,000 scientific papers published in 2002. Each dot in the galaxy represents one of the 96,000 research communities active in science in 2002. A research community is a group of papers (9 on average) that are written on the same research topic in a given year. Over time, communities can be born, continue, split, merge, or die.

The map of science can be used as a tool for science strategy. This is the terrain in which organizations and institutions locate their scientific capabilities. Additional information about the scientific and economic impact of each research community allows policy makers to decide which areas to explore, exploit, abandon, or ignore.

We also envision the map as an educational tool. For children, the theoretical relationship between areas of science can be replaced with a concrete map showing how math, physics, chemistry, biology and social studies interact. For advanced students, areas of interest can be located and neighboring areas can be explored.

Nanotechnology

Most research communities in nanotechnology are concentrated in Physics, Chemistry, and Materials Science. However, many disciplines in the Life and Medical Sciences also have nanotechnology applications.

Proteomics

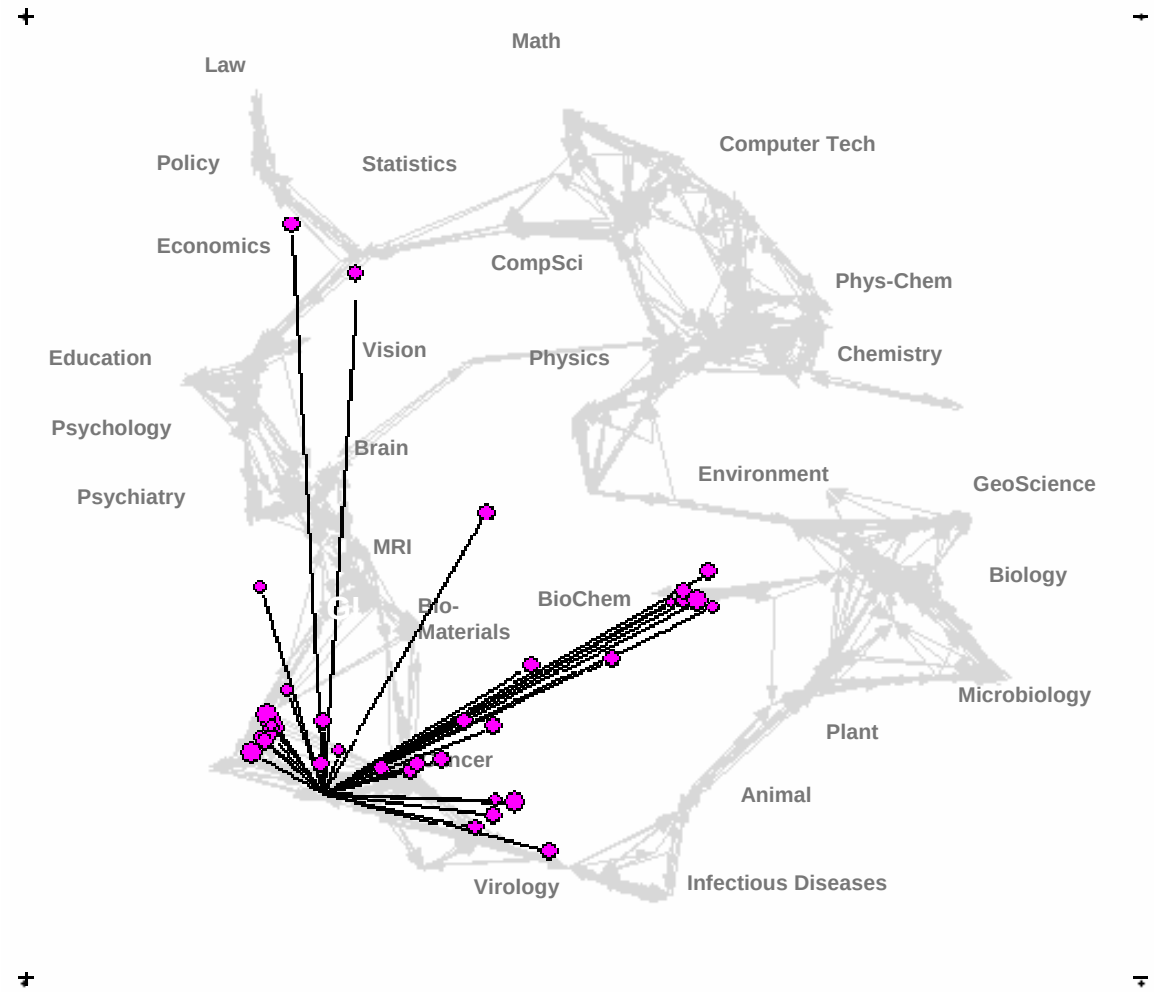
Research communities in proteomics are centered in Biochemistry. In addition, there is a heavy focus in the tools section of chemistry, such as Chromatography. The balance of the proteomics communities are widely dispersed among the Life and Medical Sciences.

Pharmacogenomics

Pharmacogenomics is a relatively new field with most of its activity in Medicine. It also has many communities in Biochemistry and two communities in the Social Sciences.

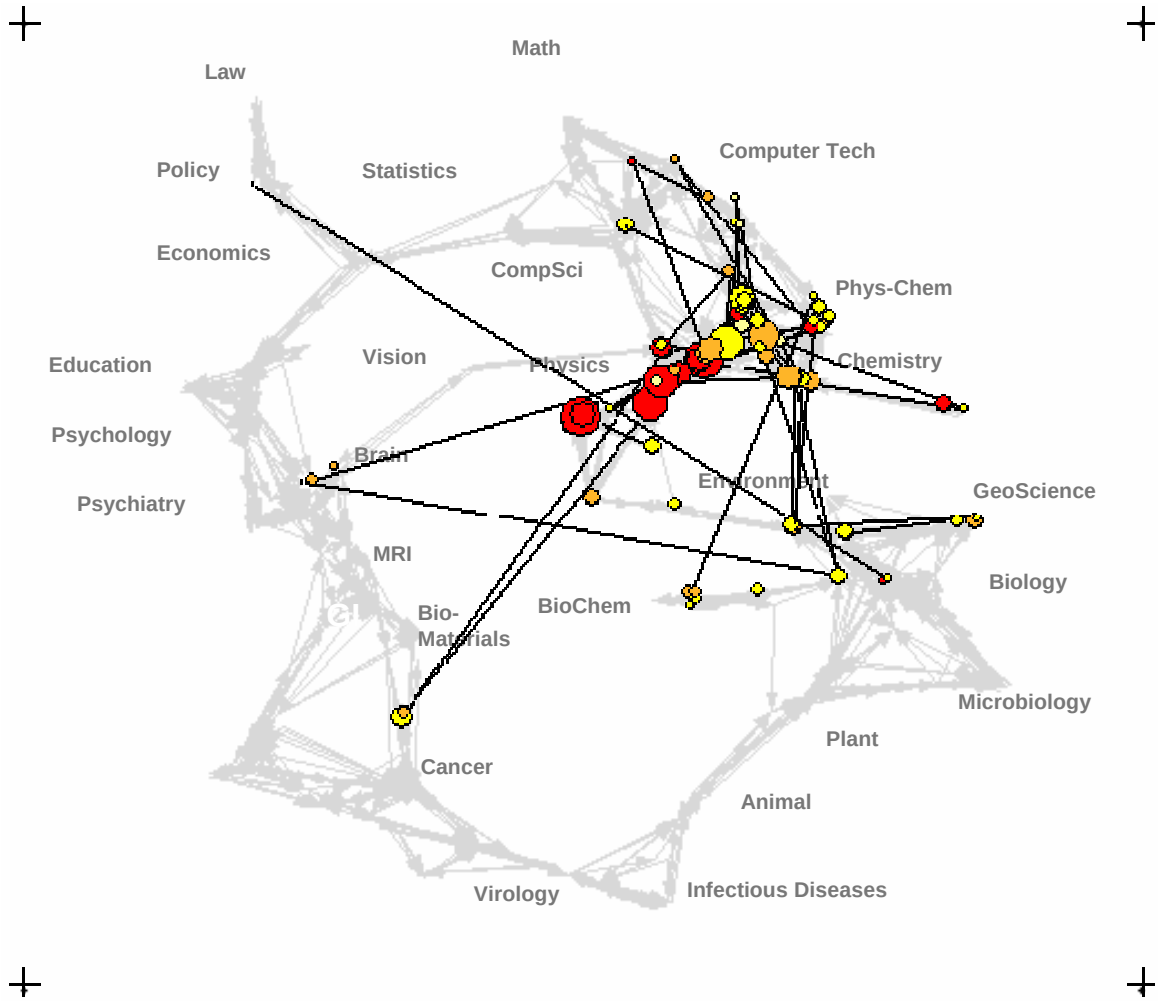
(1) Identifying Opportunities/Threats

(36 Research Communities that will impact GI Research…
that GI Researchers are least likely to be aware of)



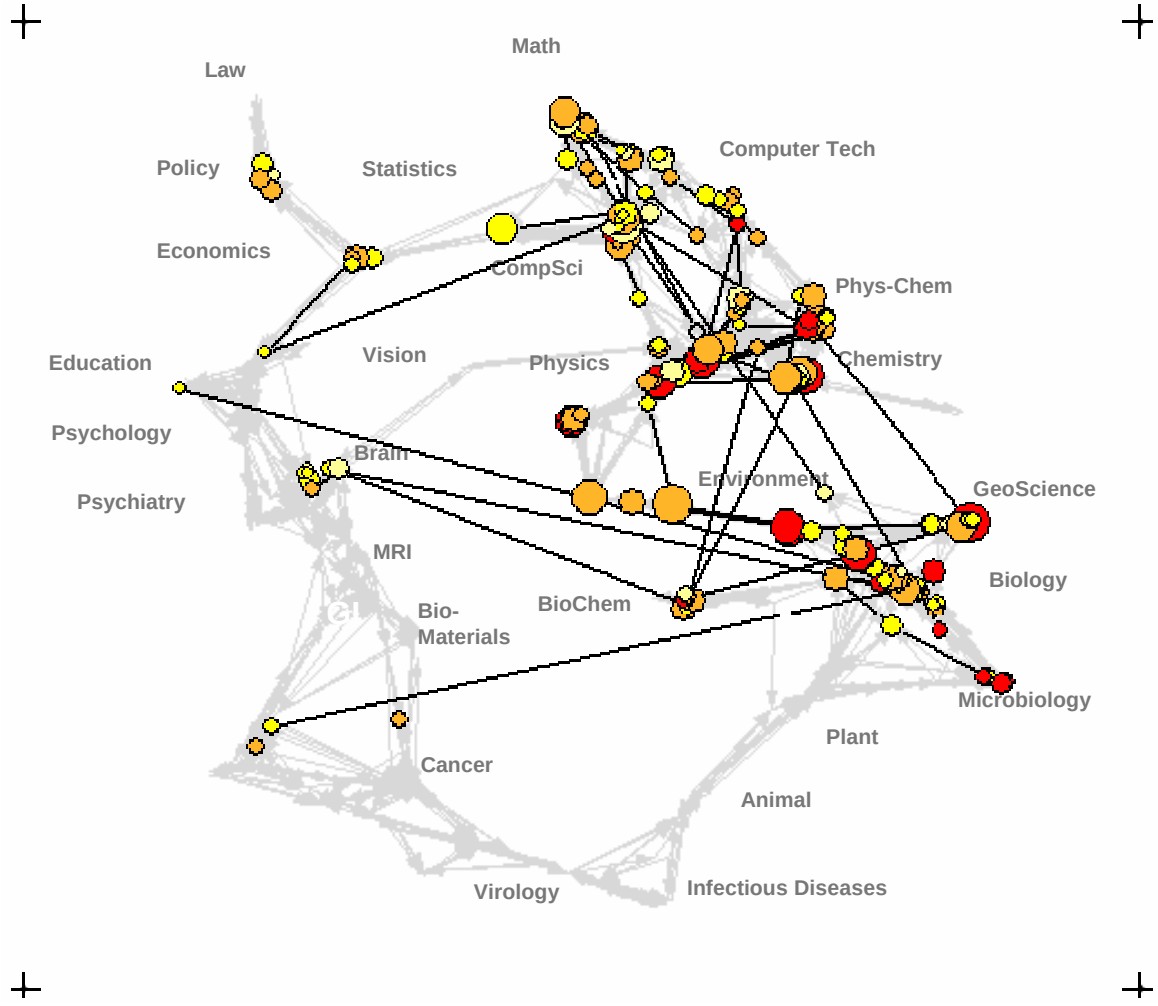
(1) Identifying Core Competency

Funding patterns of the US Department of Energy (DOE)



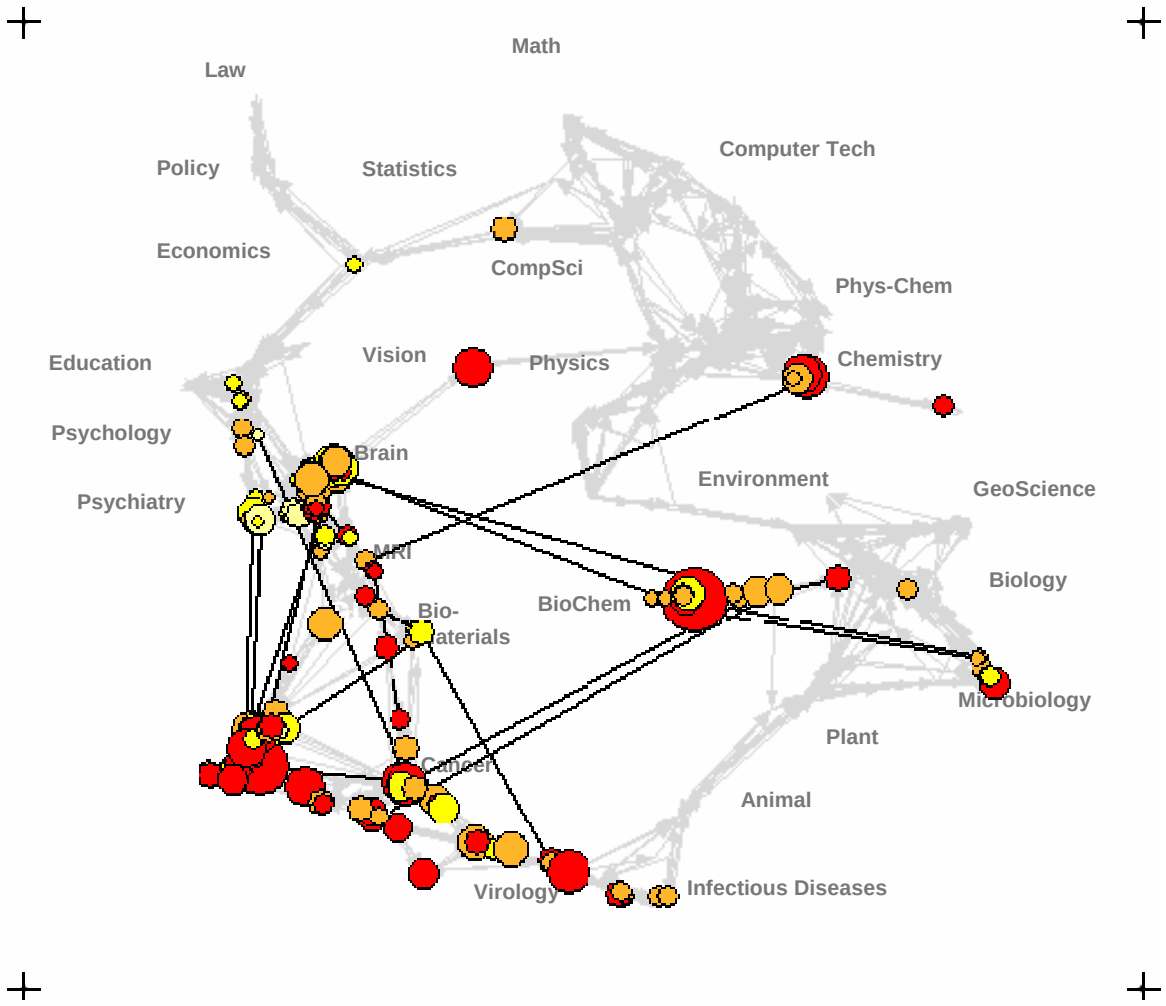
(1) Identifying Core Competency

Funding patterns of the National Science Foundation (NSF)



(1) Identifying Core Competency

Funding patterns of the National Institutes of Health (NIH)



(1) Funding Overlays

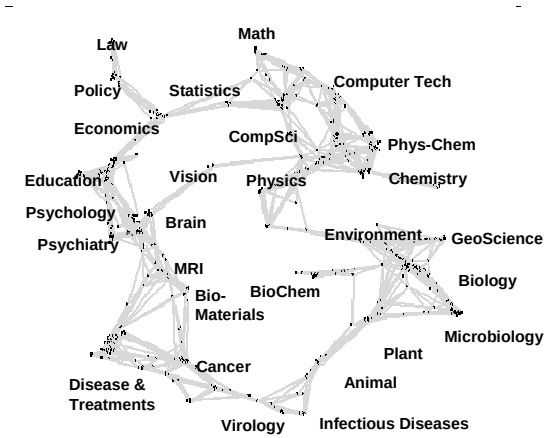
- DOE – simple lookup of author institutions (LANL, LLNL, SNL, ANL, BNL, LBL, …)
- NIH and NSF: Used RaDiUS data from 1999
  - matched funded PI and institution from 1999 funding data
  - to first author and institution from the 2002 ISI publication data
  - not comprehensive, but representative
  - undercounts actual funding profiles (e.g. if PI wasn't first author, if publications appeared earlier or later than 2002, …)
  - some false hits – could use text analysis to narrow further

Two Indicators

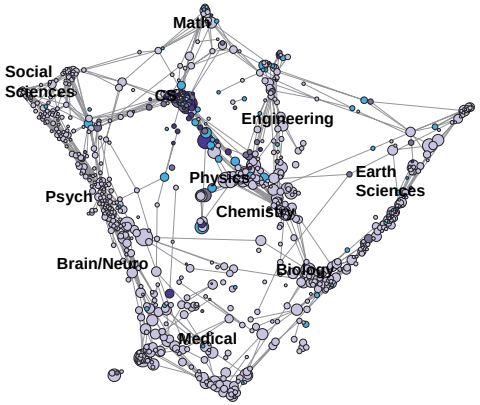
- Scientific vitality – a measure of the speed at which a group of researchers reaches consensus about major improvements
  - calculated from the reference ages of a community of papers
  - identifies communities that build upon more recent work
  - high vitality communities are quicker to incorporate recent research findings
- Patent yield
  - identifies communities that yield patents, and thus contribute directly to the economy
  - can be expressed in patents/paper

(2) Disciplinary S&T Model - Details

- Uses combined SCIE/SSCI/ISI Proceedings from 2003
  - 7445 journals, 1198 proceedings
  - Journals, proceedings treated equivalently
  - Bibliographic coupling of papers, aggregated to journals
- Initial ordination and clustering of journals gave 852 clusters
- Same two step process as used on the science map
- Each paper cluster assigned to a journal cluster; comparisons made on the structural map



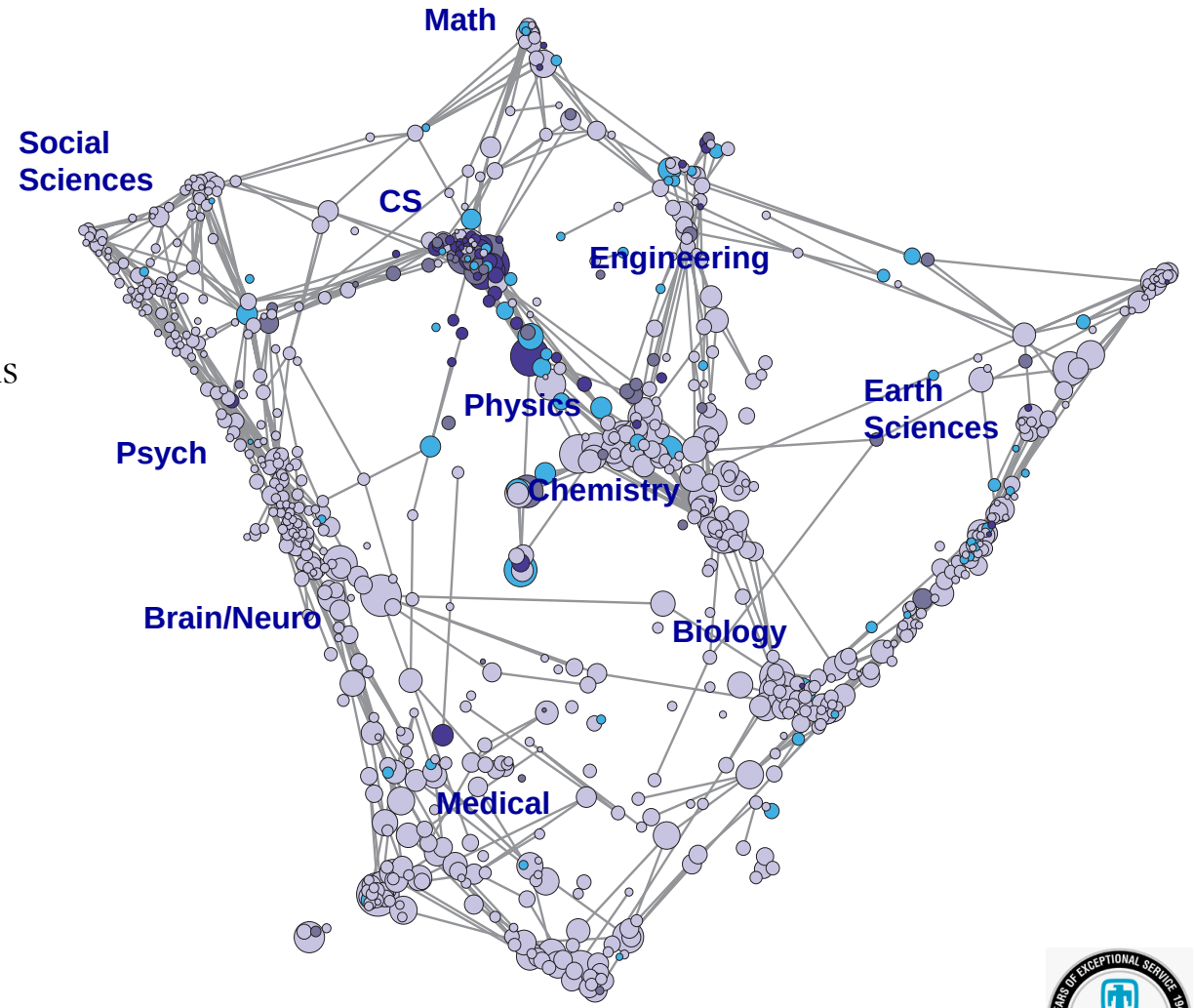
Science only (jnls)



S&T (jnls + conf)

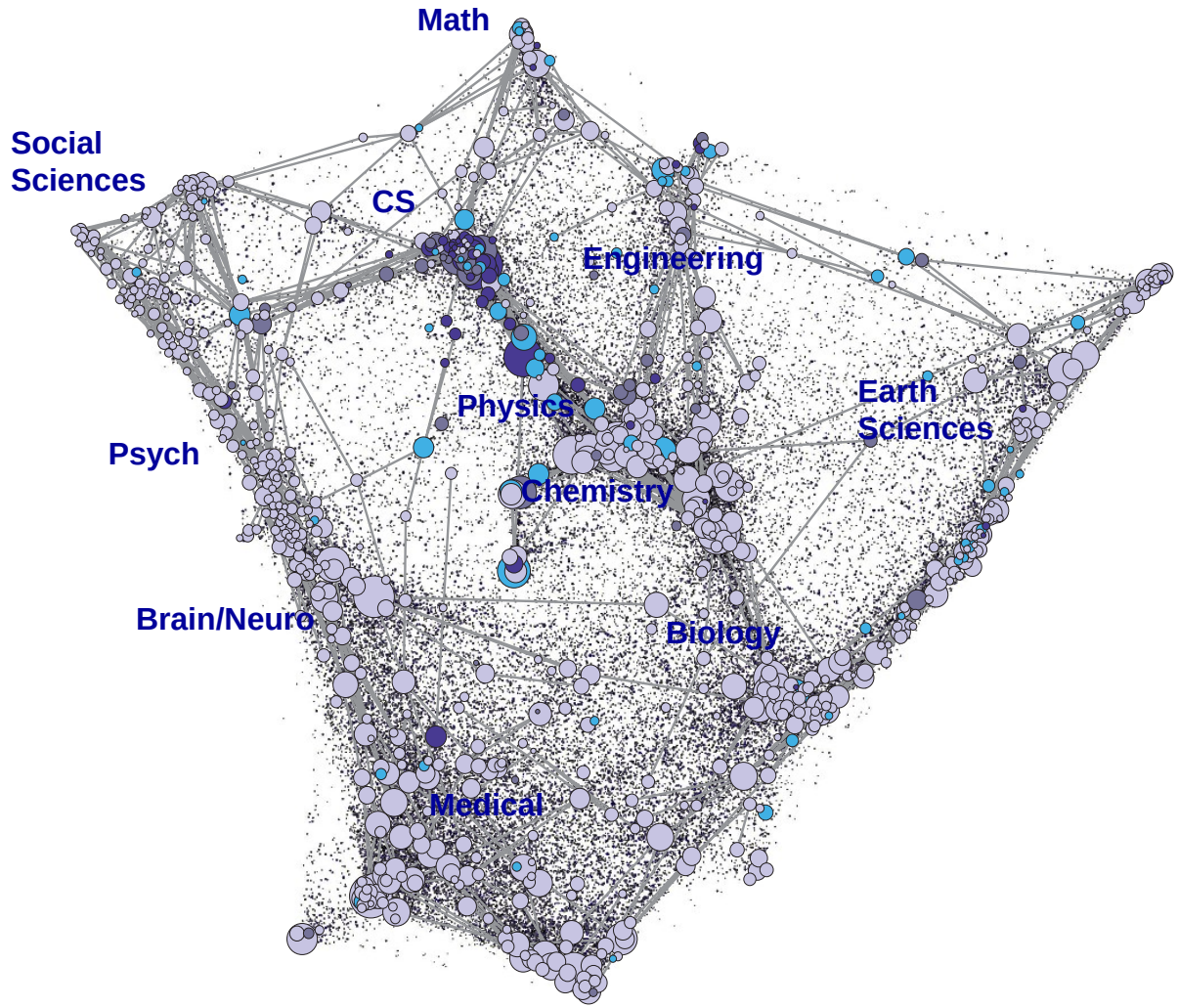
(2) Disciplinary S&T Model - Details

- Uses combined SCIE/ SSCI/ISI Proceedings from 2003
  - 7445 journals, 1198 proceedings
  - journals, proceedings treated equivalently
  - Bib coupling of journals
- Initial ordination and clustering of journals gave 852 clusters
- Coupling counts reaggregated at the journal cluster level; ordination of journal clusters
  - (x,y) positions for each journal cluster, journal



(2) Research Community Model - Details

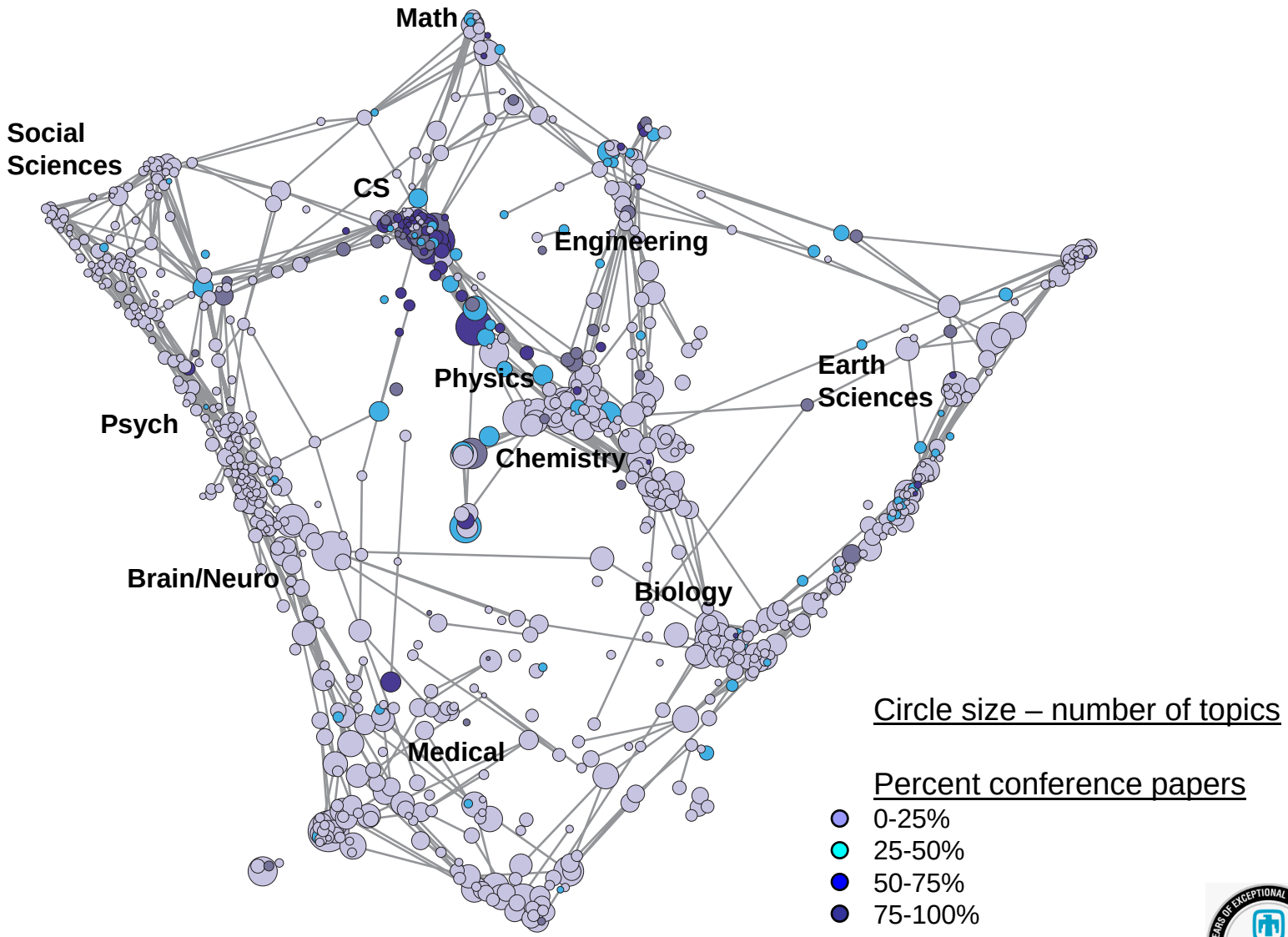
- Uses combined SCIE/ SSCI/ISI Proceedings from 2003
  - 997,775 papers from 8643 sources
  - Bib coupling of papers
- Initial ordination and clustering of journals gave 117,433 clusters / communities
- Cluster positions calculated using journal positions from the disciplinary map



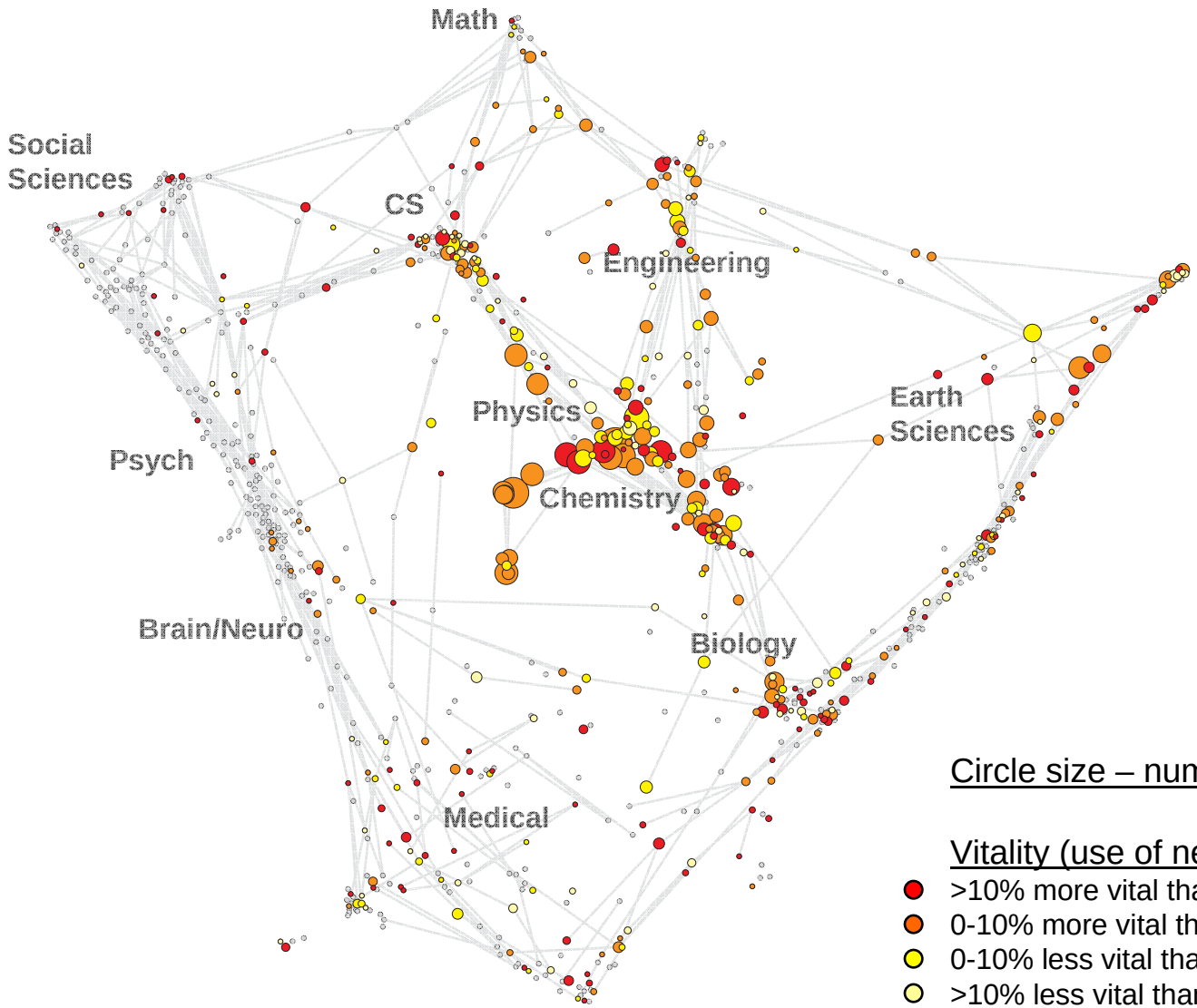
Communities Have a Topic Focus

- Summary for cluster community 85407
- ----------
- Papers Found in Community:
- ----------
- Journal Paper Title Weight
- VACCINE **Vaccine** prevention of **meningococcal disease**, coming soon? 6)
- ARCH PATHOL LAB MED Evaluation, of a diagnostic polymerase chain reaction assay for **Neisseria meningitidis** in North America and field experience during an **outbreak** 5)
- BRIT MED J Acquisition of W135 **meningococcal carriage** in Hajj pilgrims and **transmission** to household contacts: prospective study 5)
- EMERG INFECT DIS Clonal groupings in serogroup X **Neisseria meningitis** 4)
- J CLIN PATHOL Ultrasound enhanced detection of individual **meningococcal serogroups** by latex immunoassay 4)
- J INFECT DIS Genotype-specific **carriage** of **Neisseria meningitidis** in Georgia counties with hyper- and hyposporadic rates of meningococcal disease 4)
- LANCET Development of **vaccines** against **meningococcal disease** 4)
- VACCINE Avidity maturation following **vaccination** with a **meningococcal** recombinant hexavalent PorA OMV vaccine in UK infants 4)
- J CLIN MICROBIOL 5 ' exonuclease assay for detection of serogroup Y **Neisseria meningitidis** 4)
- EPIDEMIOL INFECT Increase in **meningococcal disease** associated with the emergence of a novel ST-11 variant of serogroup C Neisseria meningitidis in Victoria, Australia, 1999-2000 4)
- HOSP MED **Epidemiology** of **meningococcal disease** 4)
- INFECT IMMUN Effect of **vaccination** with carrier protein on response to **meningococcal** C conjugate vaccines and value of different immunoassays as predictors of protection 4)

(2) Worldwide S&T



(2) DOE Profile

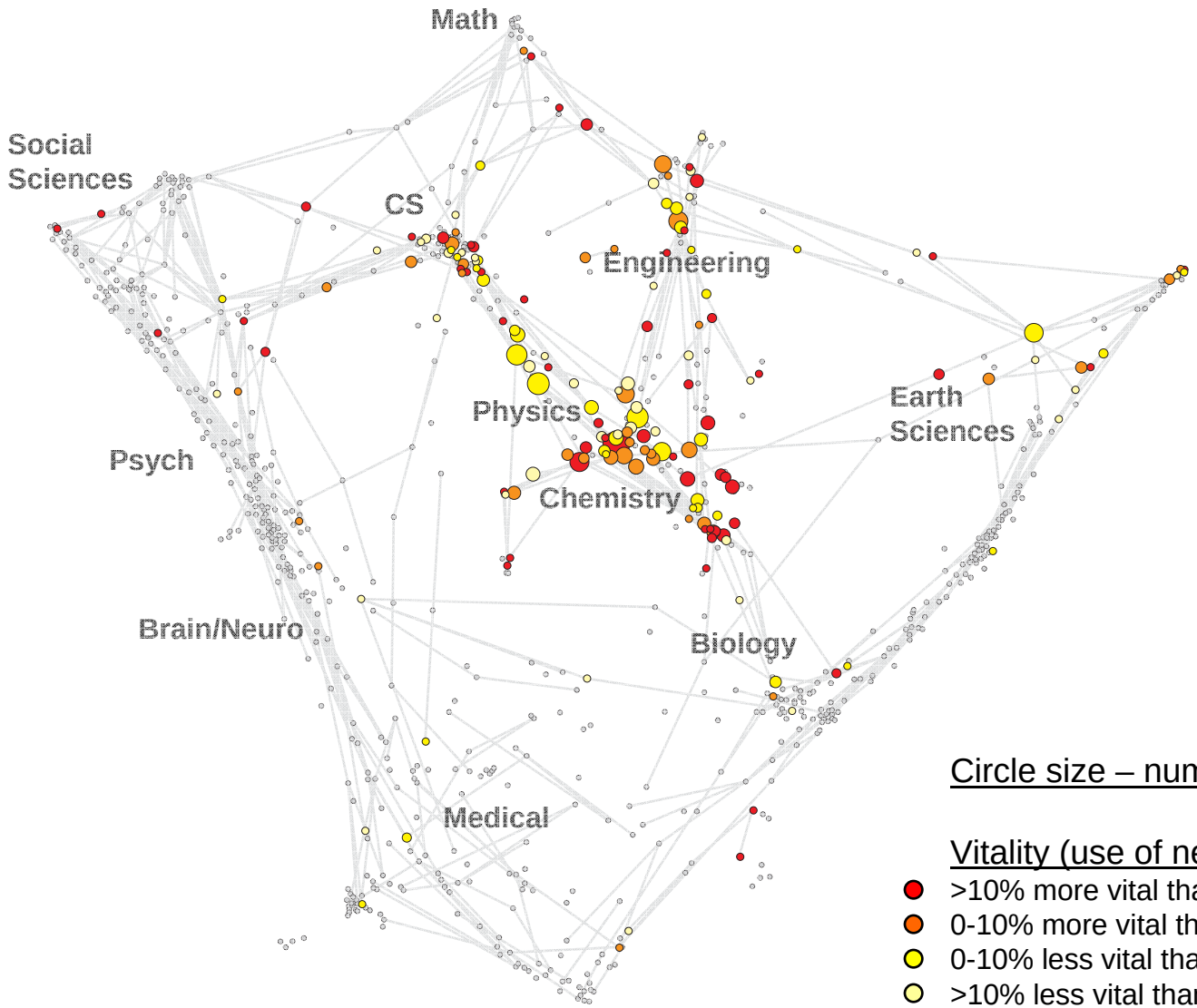


Circle size – number of topics

Vitality (use of newer ideas)

- >10% more vital than world
- 0-10% more vital than world
- 0-10% less vital than world
- >10% less vital than world

(2) Sandia Profile

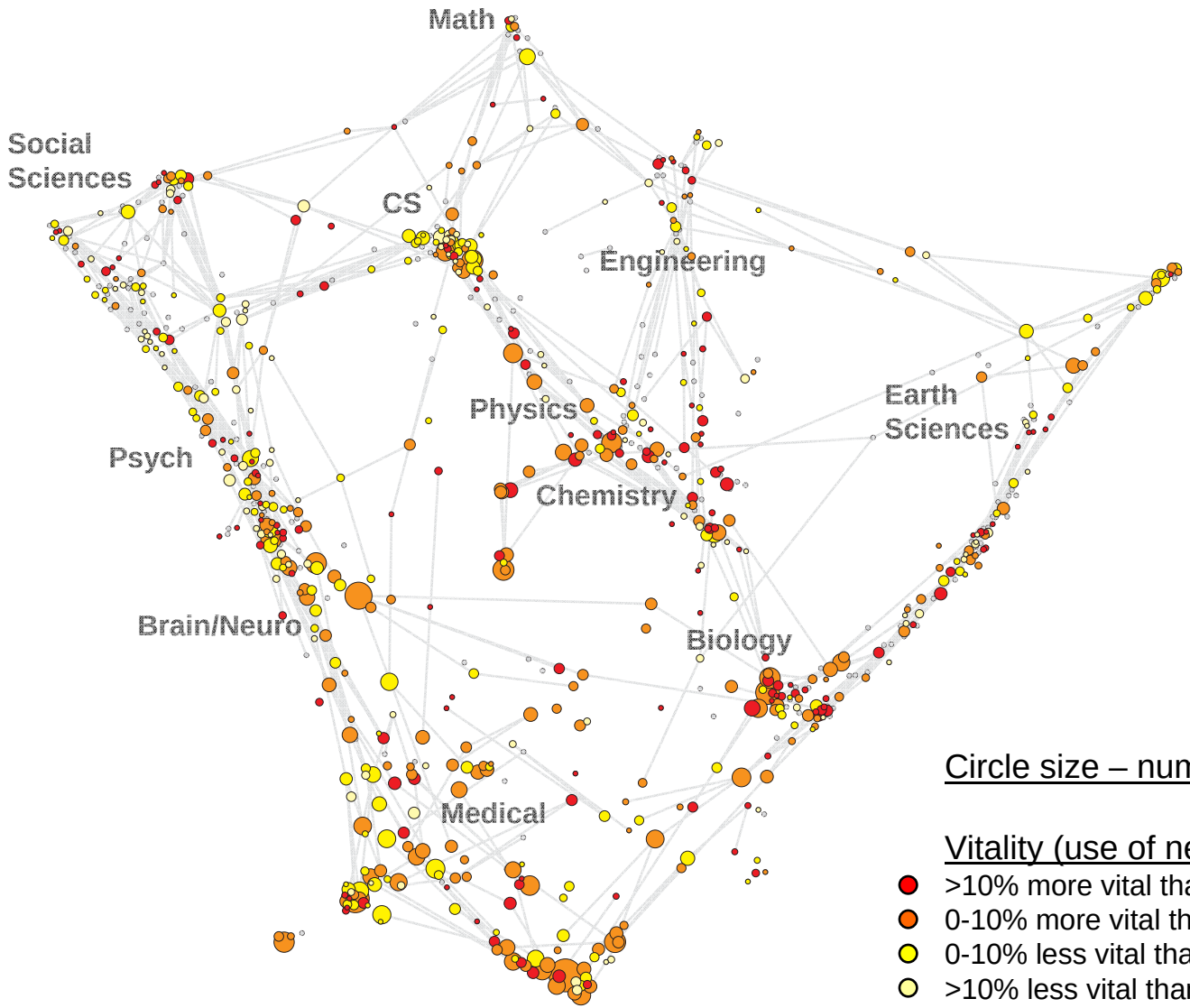


Circle size – number of topics

Vitality (use of newer ideas)

- >10% more vital than world
- 0-10% more vital than world
- 0-10% less vital than world
- >10% less vital than world

(2) Univ Texas (System) Profile

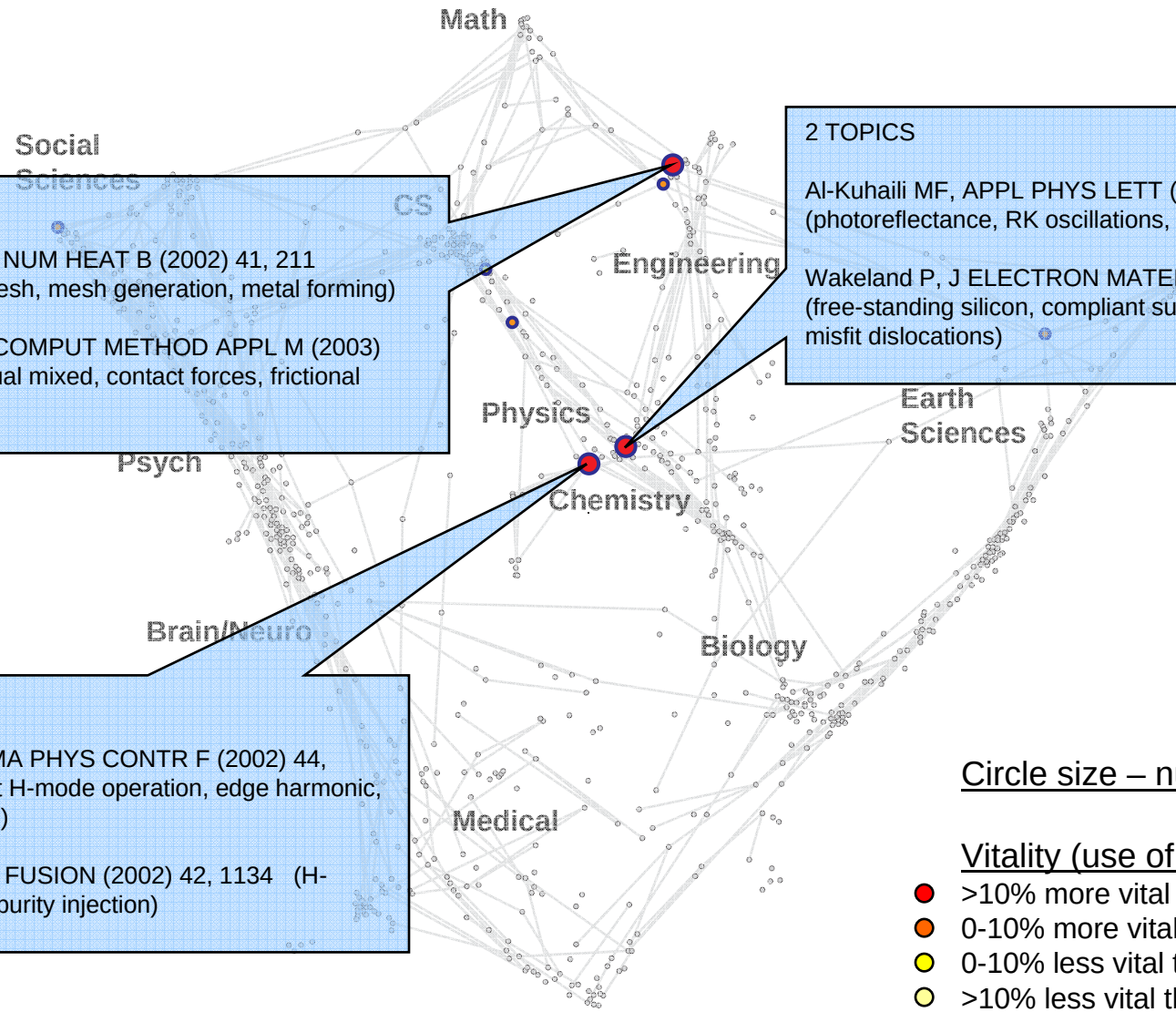


Circle size – number of topics

Vitality (use of newer ideas)

- >10% more vital than world
- 0-10% more vital than world
- 0-10% less vital than world
- >10% less vital than world

(2) Existing SNL/UT Collaborations



2 TOPICS

Douglass RW, NUM HEAT B (2002) 41, 211 (hexahedral mesh, mesh generation, metal forming)

Bessette GC, COMPUT METHOD APPL M (2003) 192, 1649 (dual mixed, contact forces, frictional contact …)

2 TOPICS

Al-Kuhaili MF, APPL PHYS LETT (2003) 82, 1203 (photoreflectance, RK oscillations, electron gas …)

Wakeland P, J ELECTRON MATER (2003) 32, 836 (free-standing silicon, compliant substrates, SiGe film, misfit dislocations)

2 TOPICS

Burrell KH, PLASMA PHYS CONTR F (2002) 44, A0253 (quiescent H-mode operation, edge harmonic, MHD oscillation …)

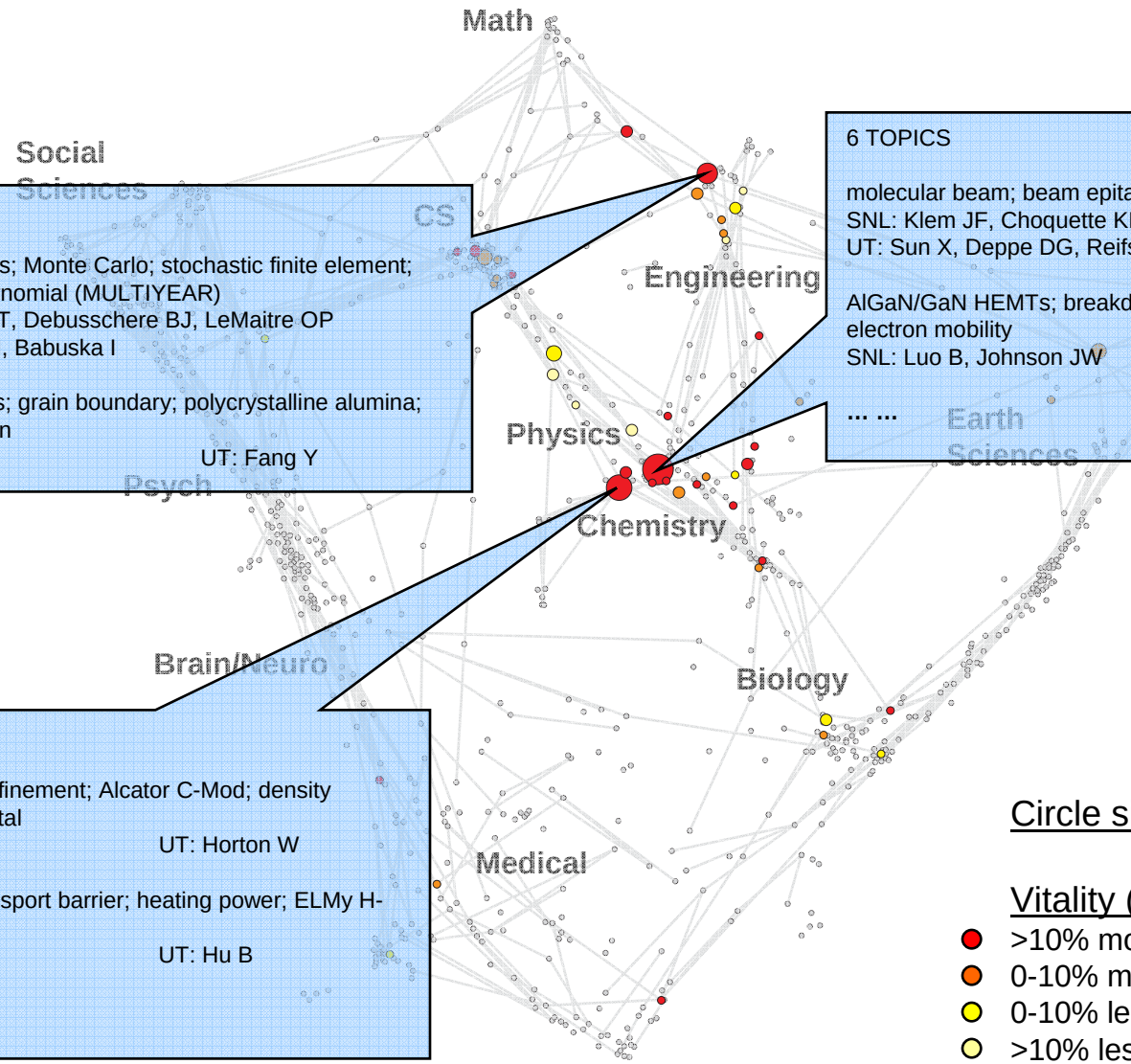
Colchin RJ, NUCL FUSION (2002) 42, 1134 (H-mode, L-mode, impurity injection)

Circle size – number of topics

Vitality (use of newer ideas)

- >10% more vital than world
- 0-10% more vital than world
- 0-10% less vital than world
- >10% less vital than world

(2) Potential SNL/UT Collaborations



2 TOPICS

polynomial chaos; Monte Carlo; stochastic finite element; generalized polynomial (MULTIYEAR)
SNL: Reagan MT, Debusschere BJ, LeMaitre OP
UT: Jefferys WH, Babuska I

residual stresses; grain boundary; polycrystalline alumina; stress distribution
SNL: Vedula VR UT: Fang Y

6 TOPICS

molecular beam; beam epitaxy; GaAsN layers; PL intensity
SNL: Klem JF, Choquette KD
UT: Sun X, Deppe DG, Reifsnider JM

AlGaN/GaN HEMTs; breakdown voltage; power density; electron mobility
SNL: Luo B, Johnson JW UT: Wong MM

… …

4 TOPICS

H mode; energy confinement; Alcator C-Mod; density profiles; edge pedestal
SNL: Mahdavi MA UT: Horton W

magnetic shear; transport barrier; heating power; ELMy H-mode
SNL: Gohil P UT: Hu B

… …

Circle size – number of topics

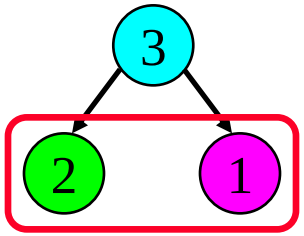
Vitality (use of newer ideas)

- >10% more vital than world
- 0-10% more vital than world
- 0-10% less vital than world
- >10% less vital than world

(3) Article Maps: Multi-step Process

- Successive clusterings of reference papers based on co-citations
  - first clustering
    - 820,000 papers > 53,000 communities
  - second clustering
    - 53,000 communities > 6,140 level1 clusters
  - third clustering
    - 6,140 level1 clusters > 776 level2 clusters

PAPERS
1 and 2 are
coupled

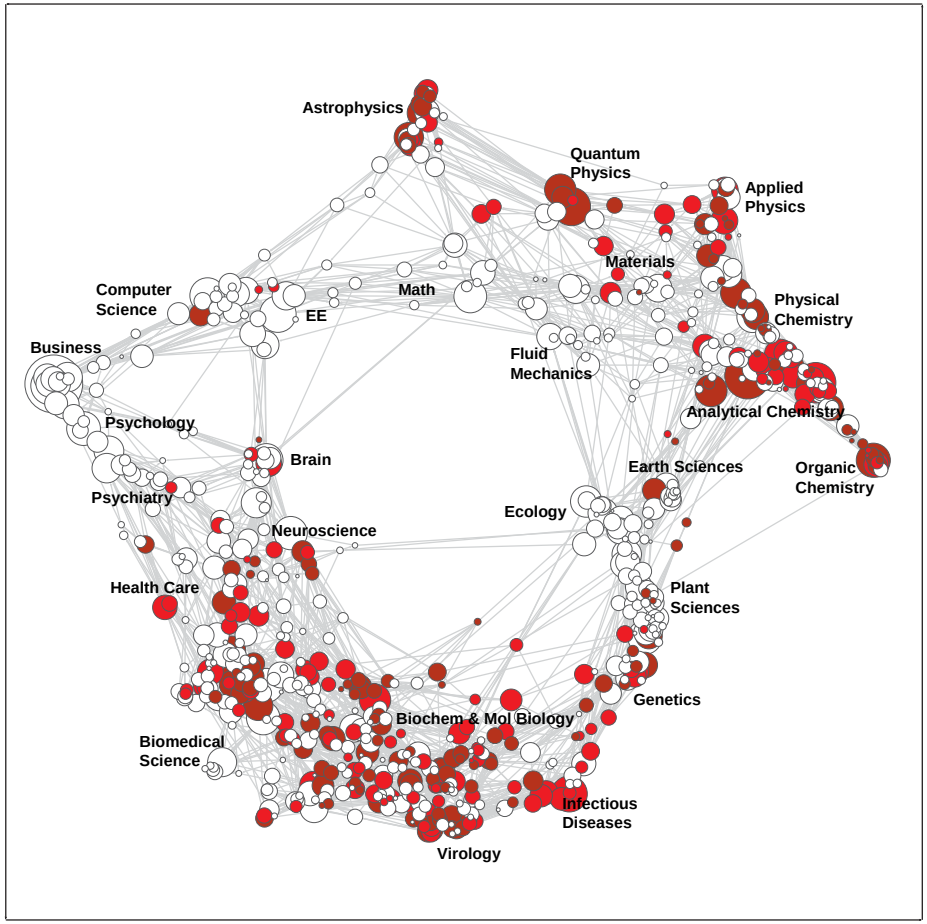


  - VxOrd used to generate layout for 776 "paradigms"

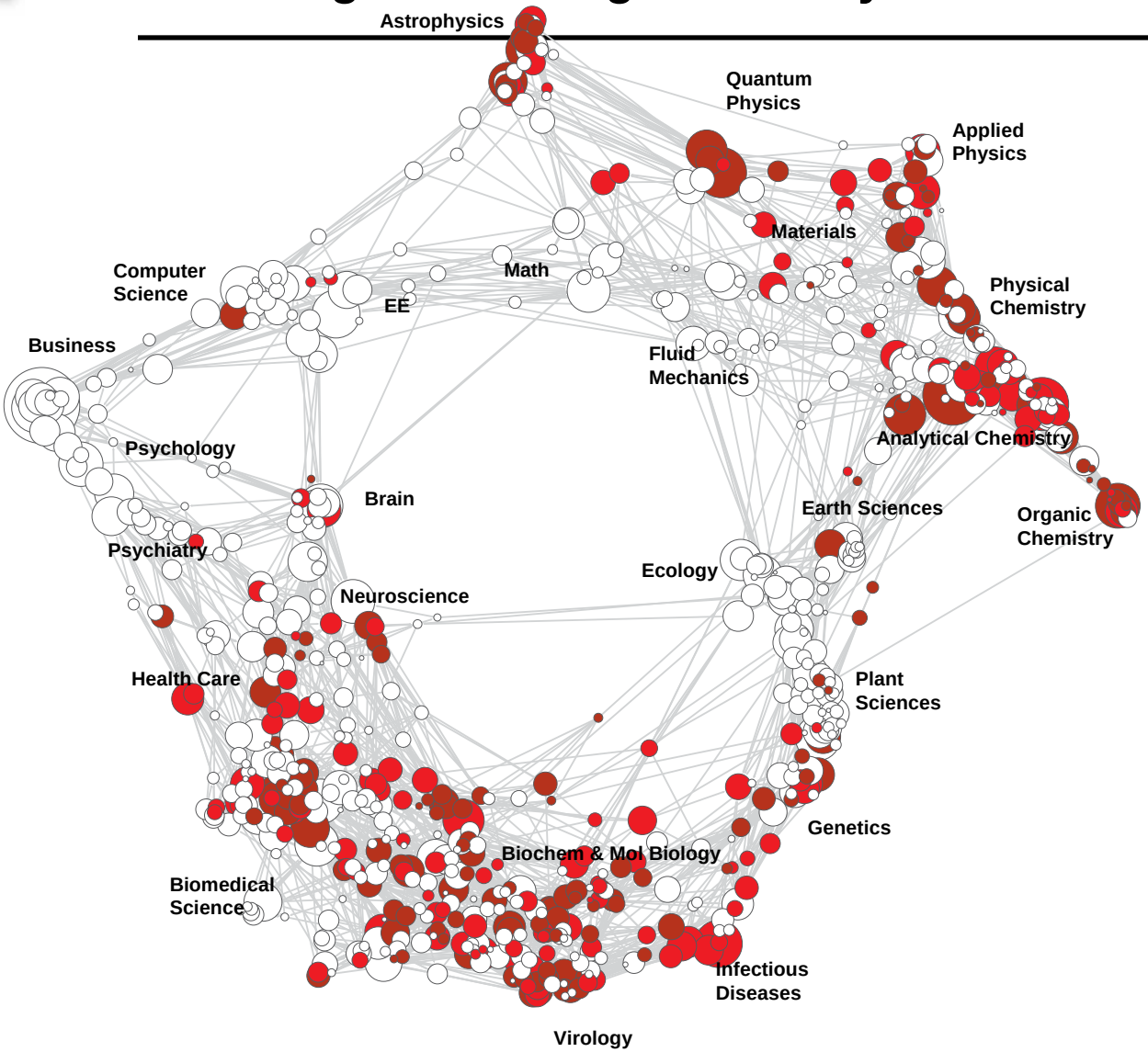
  - Current papers assigned to paradigms
  - Patents linked to paradigms

(3) Latest Model (Klavans) - Details

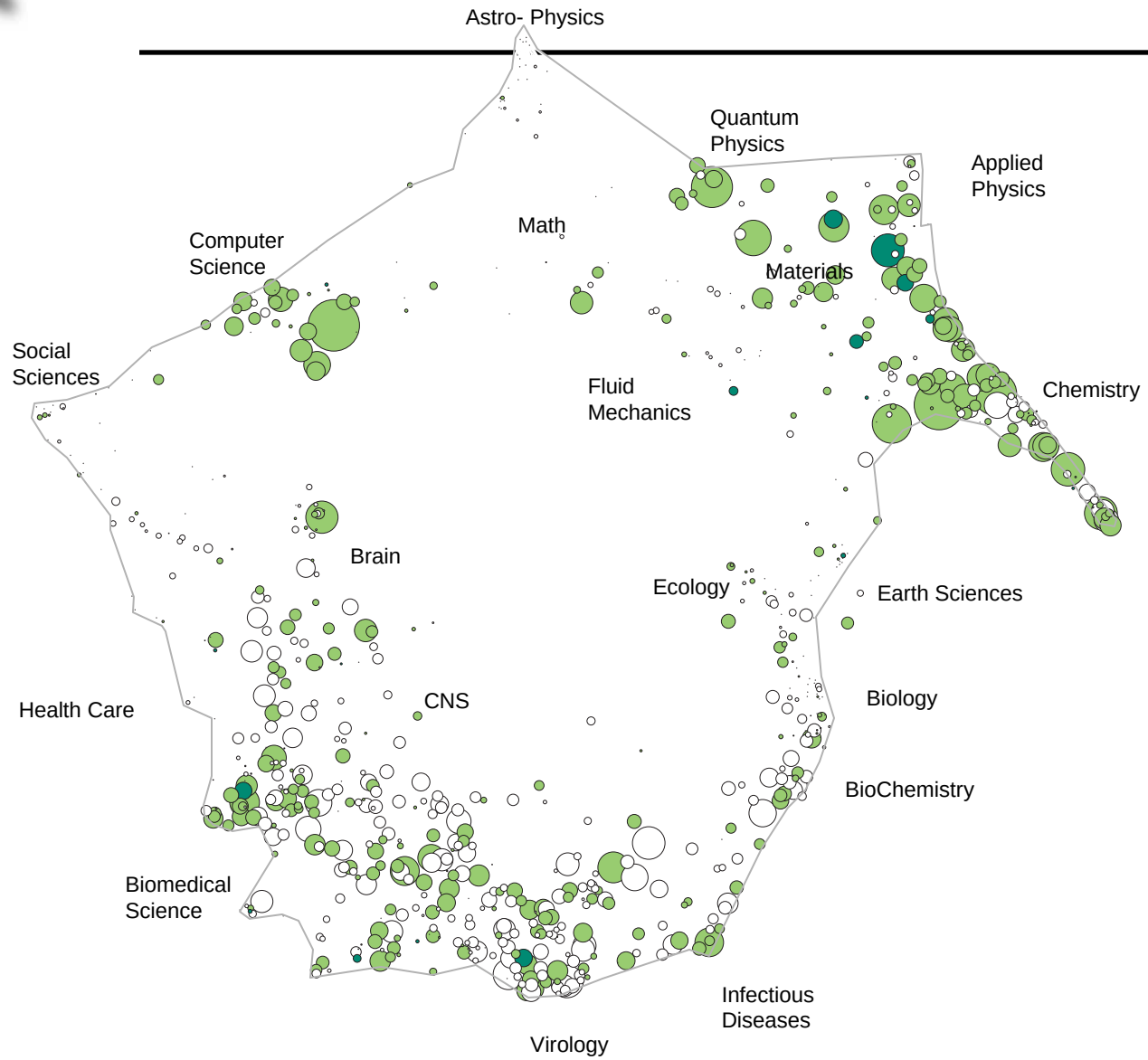
- Uses combined SCIE/ SSCI from 2003
- 3 levels of clustering used
  - 820k reference papers
  - 53k communities
  - 6100 level1 clusters
  - 776 paradigms
  - 760k current papers
- Gives a structure of scientific paradigms rather than disciplines
- Indicators calculated at the "community" level
- Aggregated results displayed at the "paradigm" level
- Red scale shows average vitality of different paradigms



Paradigms with Higher Vitality

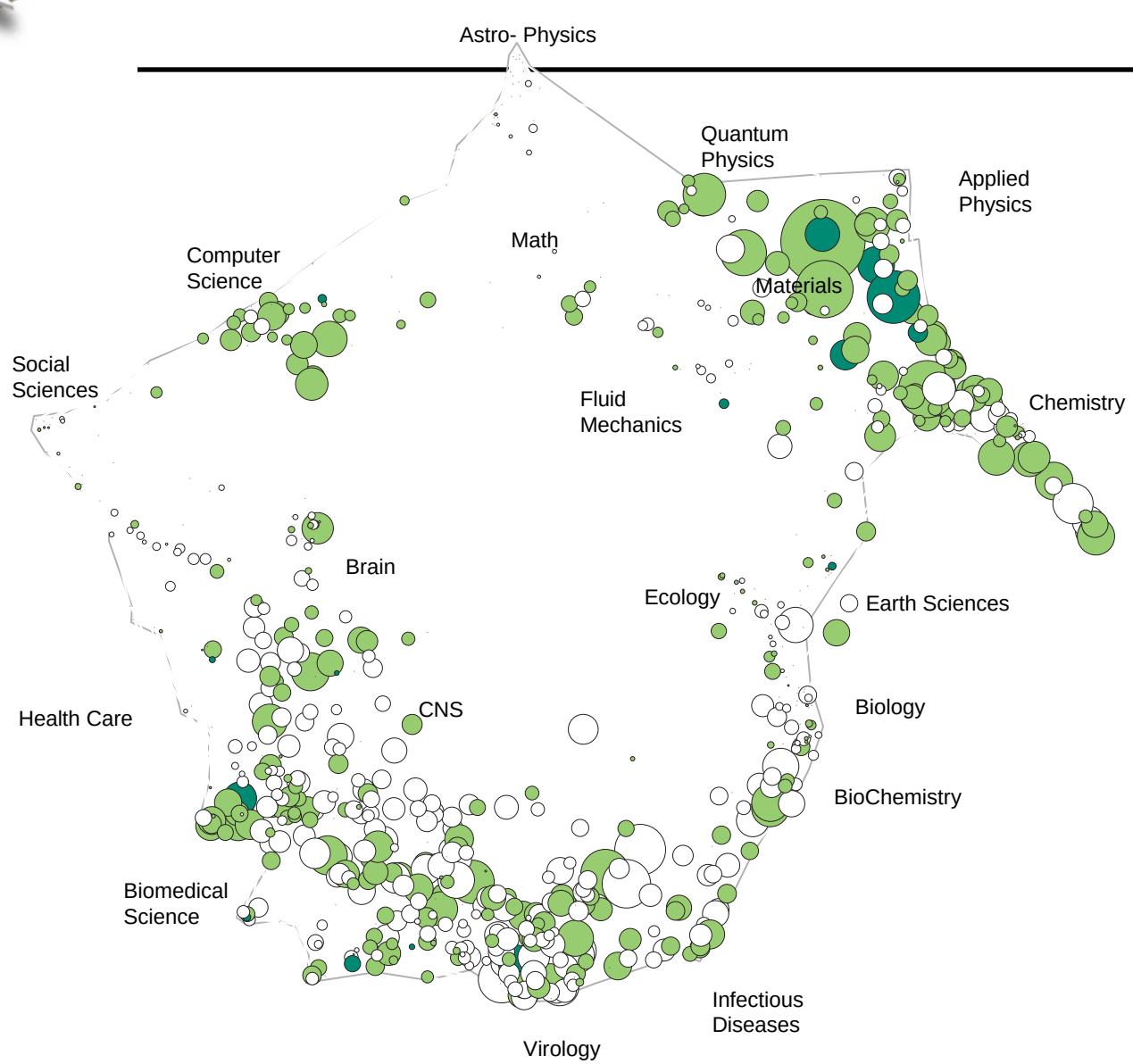


Patents that Build on Scientific Paradigms

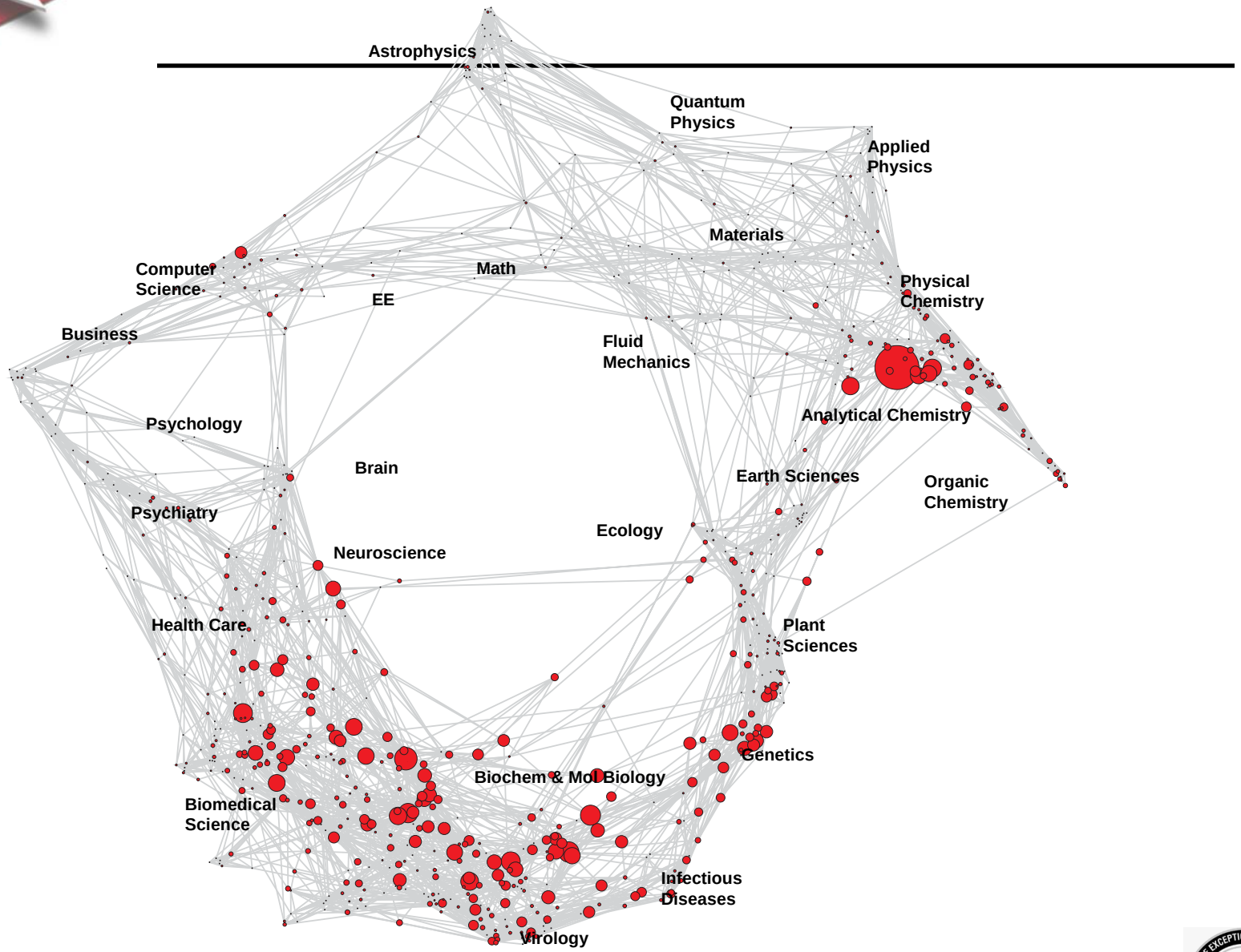


Size= Number of Patents
Color=Patent Quality (# Citations – Expected # Citations)/Number of Patents

Patent/Publication Ratios

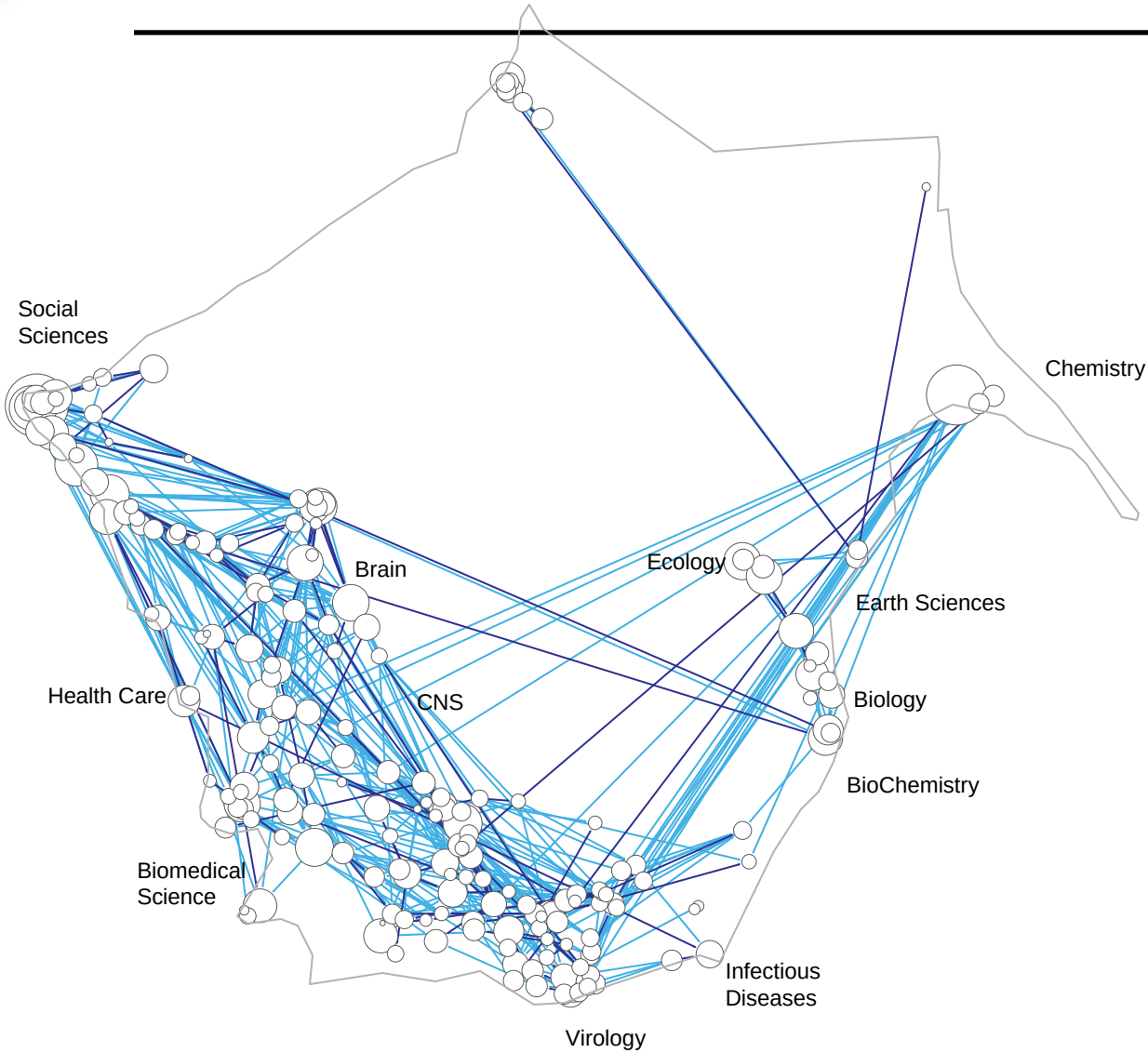


Size= #Patents/#Papers (all Paradigms with at least 100 papers)
Color=Patent Quality (see prior slide)

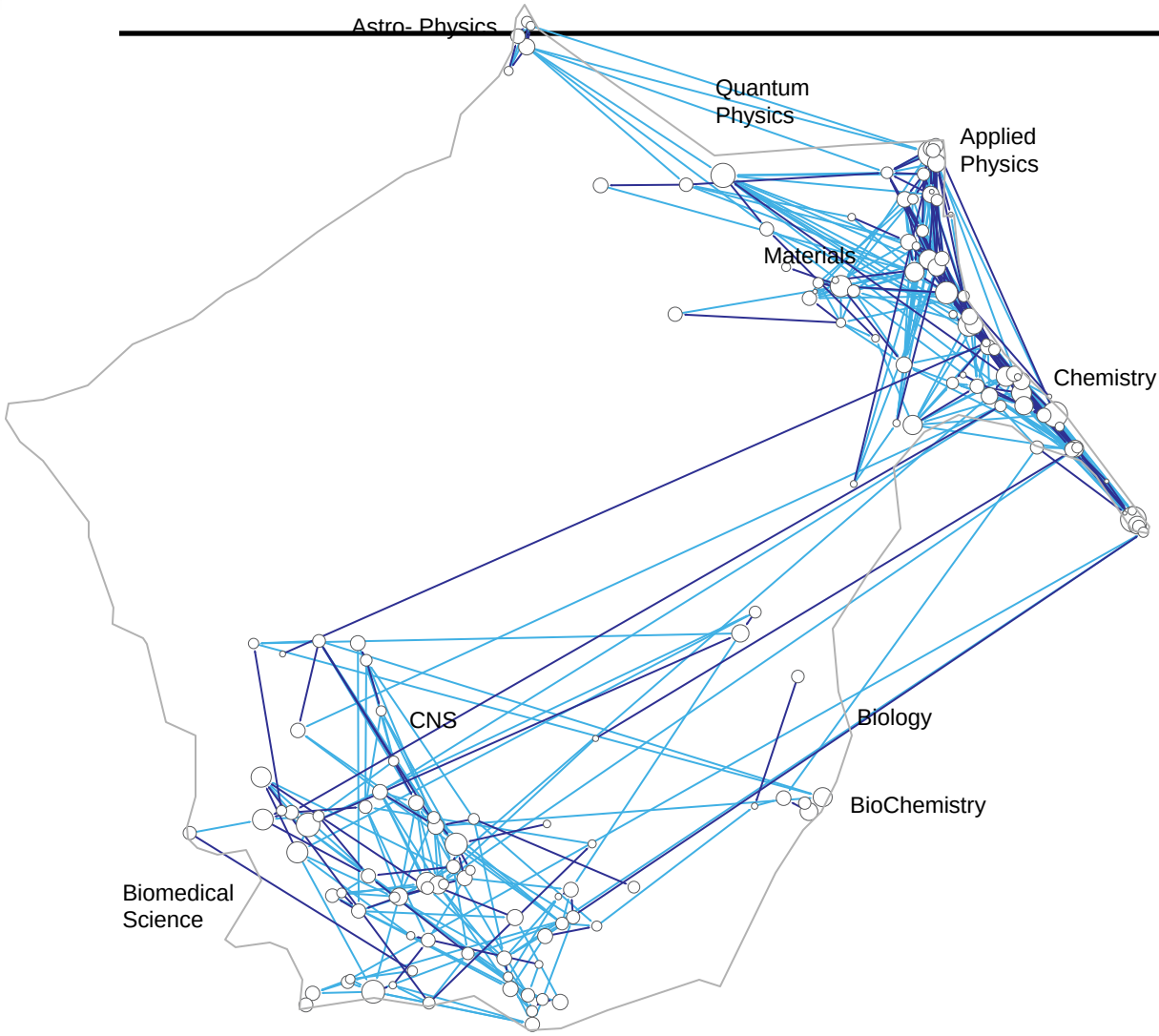


Chem-bio interface

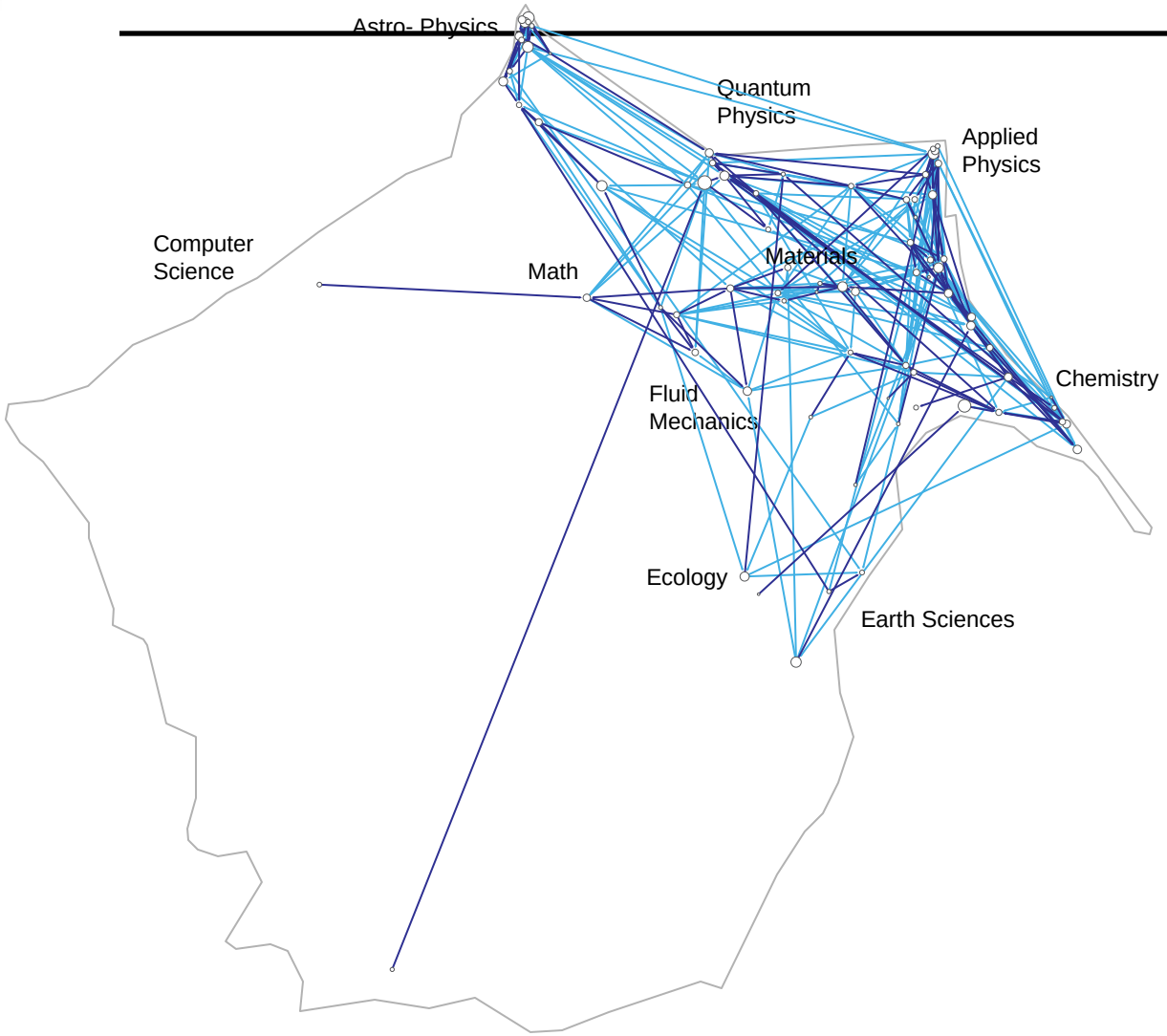
US


Social
Sciences
Chemistry
Brain
Ecology
Earth Sciences
Health Care
CNS
Biology
BioChemistry
Biomedical
Science
Infectious
Diseases
Virology

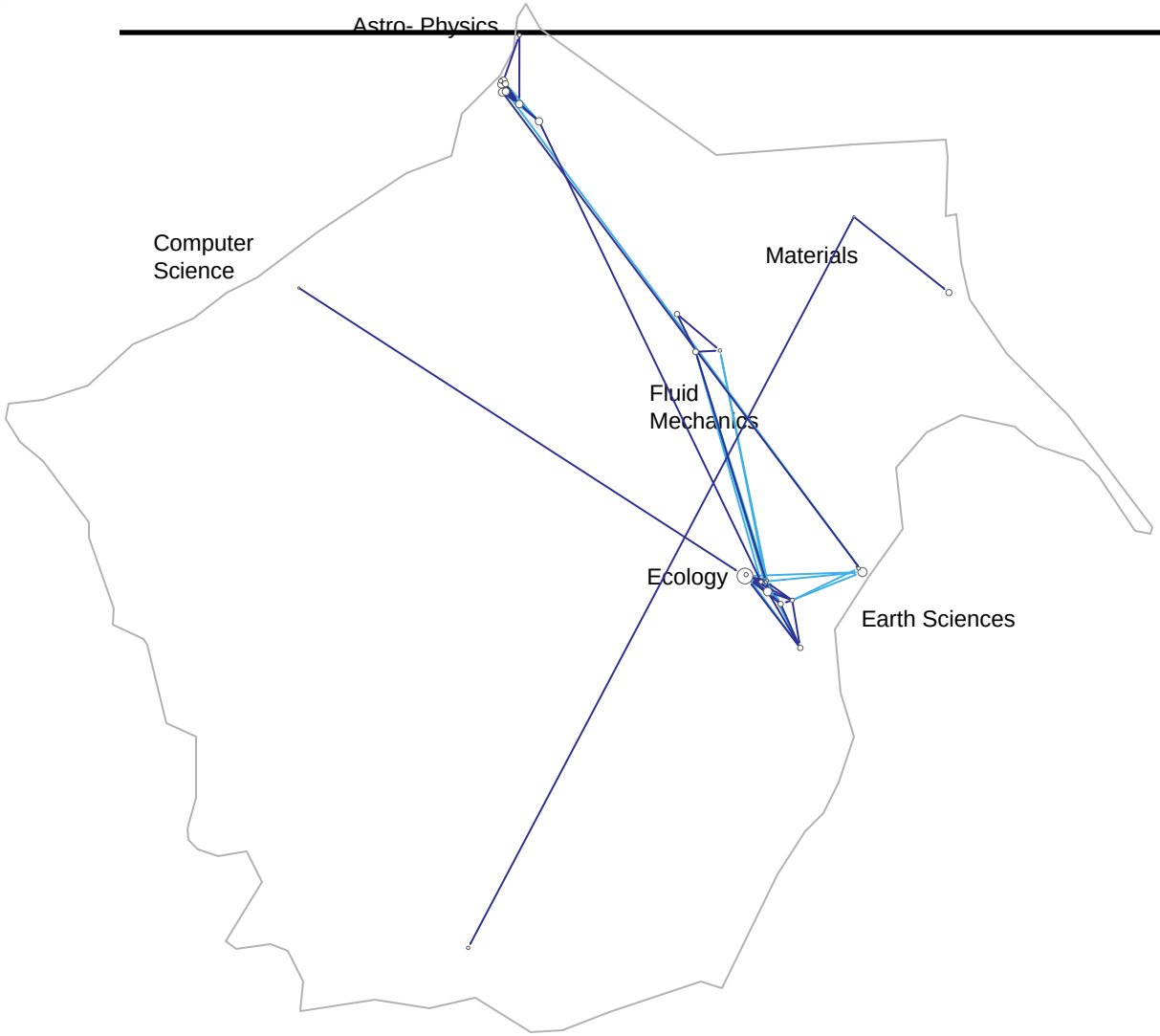
Japan

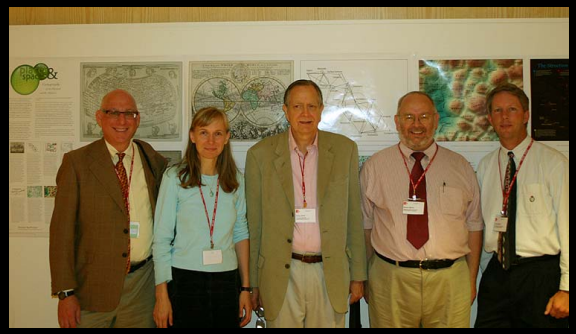

Astro- Physics
Quantum Physics
Applied Physics
Materials
Chemistry
CNS
Biology
BioChemistry
Biomedical Science

DOE Labs (79 Paradigms)



NASA (27 Paradigms)





Places & Spaces: Mapping Science

a science exhibit that introduces people to maps of sciences, their makers and users.

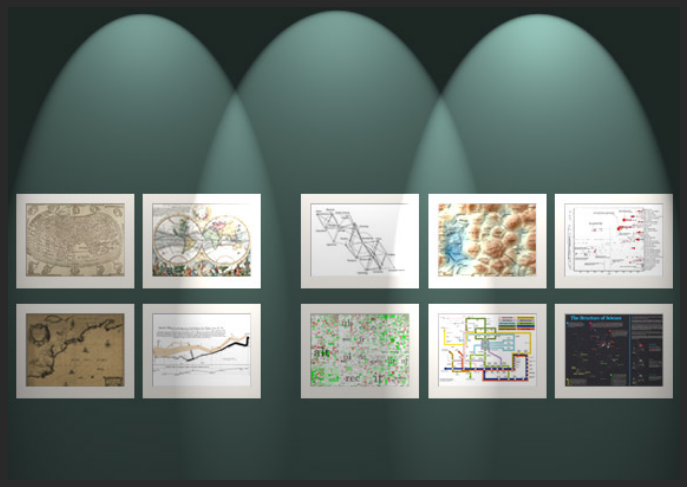
Exhibit Curators:

Dr. Katy Börner &

Deborah MacPherson

The Power of Maps

Four Early Maps of Our World
VERSUS
Six Early Maps of Science



(1st Iteration of Places & Spaces Exhibit - 2005)

Insculptum est per Johannem Schnitzer de Arnsheim
CAVRVS CHORVS VEL IAPIX SIVE ARGESTES
CIRCVS VEL TRESIIAS
SEPTENTRIO VEL APARCTIAS
AQVILO VEL BOREAS
CECIAS APELIOTES
FAVONIVS ZEPHIRVS
SVBSOLANVS
AFRICVS VEL LIBS
LIBONOTVS EVROAVSTER
AVSTER VEL NOTVS
EVRNOTVS
VLTVRNVS EVRVS
mare glaciale
EVROPA
ASIA
LIBIA INTERIOR
AFFRICA
ETHIOPIA INTERIOR
MARE INDICVM
MARE
PRASODVM
Terra incognita secundum ptholomeum
Tropicus Capricorni
Gradus longitudinis a occidente in orientem

A New Map of the WHOLE WORLD with the Trade winds

According to ye latest and most Exact Observations By H. Moll Geographer

In this Maps is inserted A View of ye General & Coasting Trade-Winds, Monsoons or ye Shifting Trade winds Note that ye Arrows among ye Lines shew ye Course of those General & Coasting Winds. and ye Arrows in ye void Spaces shew ye Course of ye Shifting Trade-winds, and ye Abbreviation Sept. &c. Shew ye Times of ye Year when such Winds Blow.

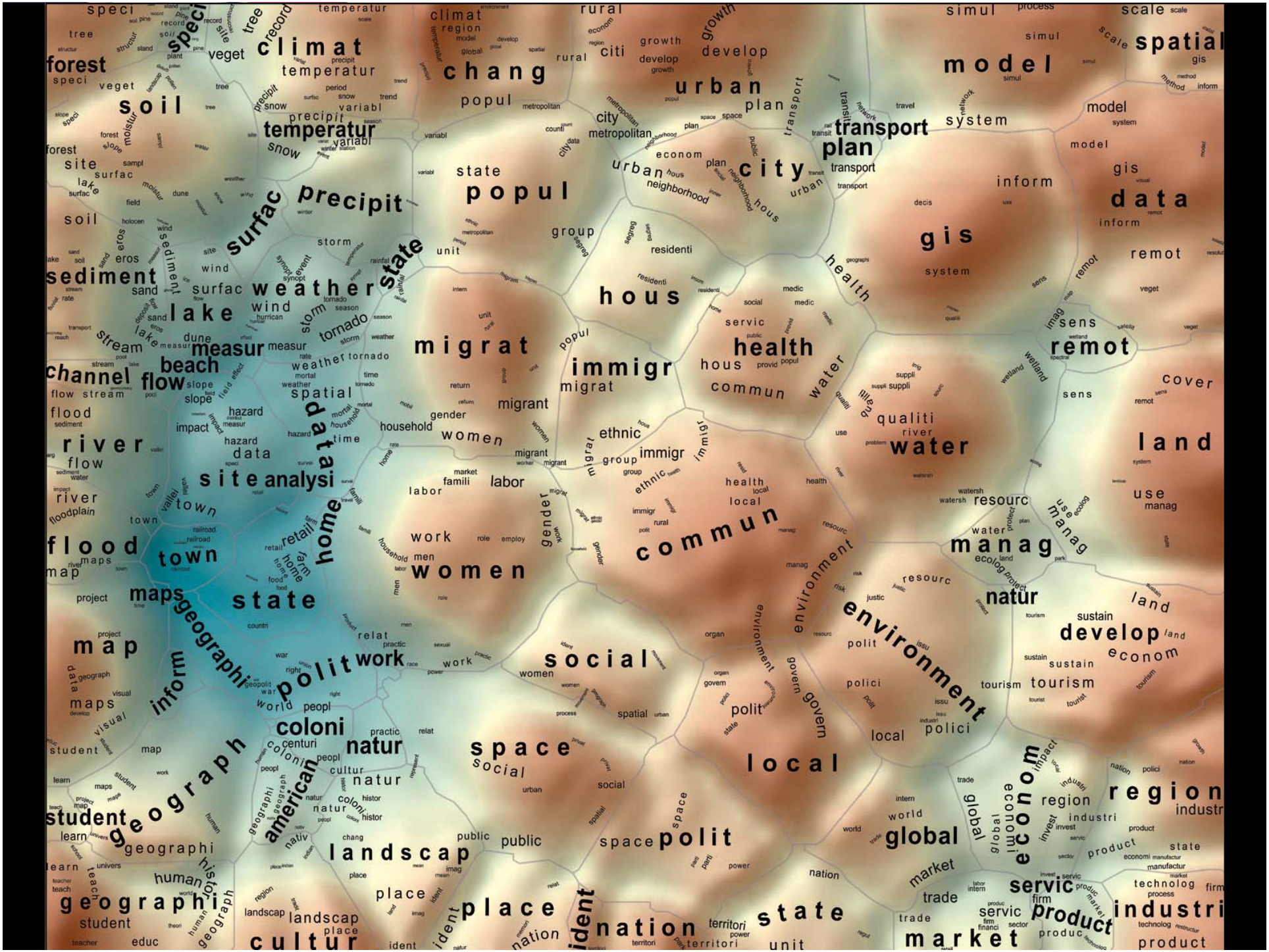
The Signs of the Zodiack, The First 6 are Northern, the other Southern Signs

♈ Aries. March	♌ Leo. July	♐ Sagittarius. November
♉ Taurus. April	♍ Virgo. August	♑ Capricornus. Decemb.
♊ Gemini. May	♎ Libra. September	♒ Aquarius. January
♋ Cancer. June	♏ Scorpio. October	♓ Pisces. February

Printed for Thos. Bowles Print and Map Seller next ye Chapter House in St. Pauls Church yard; and John Bowles Print and Map Seller at the Black Horse in Cornhill London.

How would a map of science look?

What metaphors would work best?

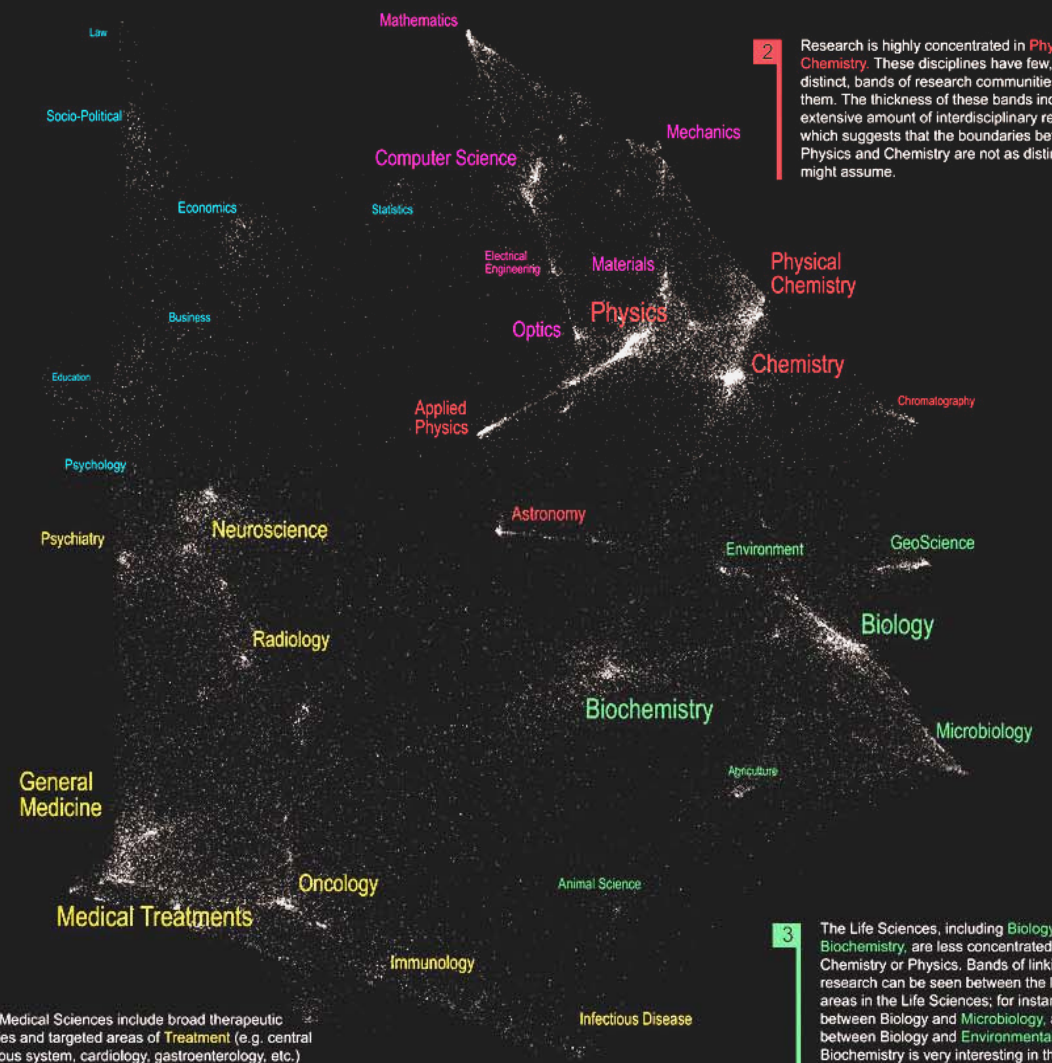


speci
climat
temperatur
chang
rural
citi
growth
develop
urban
plan
transport
model
simul
process
scale
spatial
forest
soil
precipit
popul
city
metropolitan
system
data
inform
gis
remot
sediment
surfac
weather
state
hous
health
lake
measur
beach
channel
flow
migrat
immigr
commun
water
land
river
site
analysi
home
flood
town
maps
state
women
manag
natur
develop
map
geographi
polit
work
social
environment
coloni
natur
space
local
region
student
geograph
american
landscap
polit
global
econom
servic
product
industri
geographi
place
nation
state
market
cultur

The Structure of Science

5 The Social Sciences are the smallest and most diffuse of all the sciences. Psychology serves as the link between Medical Sciences (Psychiatry) and the Social Sciences. Statistics serves as the link with Computer Science and Mathematics.

1 Mathematics is our starting point, the purest of all sciences. It lies at the outer edge of the map. Computer Science, Electrical Engineering, and Optics are applied sciences that draw upon knowledge in Mathematics and Physics. These three disciplines provide a good example of a linear progression from one pure science (Mathematics) to another (Physics) through multiple disciplines. Although applied, these disciplines are highly concentrated with distinct bands of research communities that link them. Bands indicate interdisciplinary research.



2 Research is highly concentrated in Physics and Chemistry. These disciplines have few, but very distinct, bands of research communities that link them. The thickness of these bands indicates an extensive amount of interdisciplinary research, which suggests that the boundaries between Physics and Chemistry are not as distinct as one might assume.

4 The Medical Sciences include broad therapeutic studies and targeted areas of Treatment (e.g. central nervous system, cardiology, gastroenterology, etc.) Unlike Physics and Chemistry, the medical disciplines are more spread out, suggesting a more multi-disciplinary approach to research. The transition into Life Sciences (via Animal Science and Biochemistry) is gradual.

3 The Life Sciences, including Biology and Biochemistry, are less concentrated than Chemistry or Physics. Bands of linking research can be seen between the larger areas in the Life Sciences; for instance between Biology and Microbiology, and between Biology and Environmental Science. Biochemistry is very interesting in that it is a large discipline that has visible links to disciplines in many areas of the map, including Biology, Chemistry, Neuroscience, and General Medicine. It is perhaps the most interdisciplinary of the sciences.

We are all familiar with traditional maps that show the relationships between countries, provinces, states, and cities. Similar relationships exist between the various disciplines and research topics in science. This allows us to map the structure of science.

One of the first maps of science was developed at the Institute for Scientific Information over 30 years ago. It identified 41 areas of science from the citation patterns in 17,000 scientific papers. That early map was intriguing, but it didn't cover enough of science to accurately define its structure.

Things are different today. We have enormous computing power and advanced visualization software that make mapping of the structure of science possible. This galaxy-like map of science (left) was generated at Sandia National Laboratories using an advanced graph layout routine (VxOrd from the citation patterns in 800,000 scientific papers published in 2002. Each dot in the galaxy represents one of the 96,000 research communities active in science in 2002. A research community is a group of papers (9 on average) that are written on the same research topic in a given year. Over time, communities can be born, continue, split, merge, or die.

The map of science can be used as a tool for science strategy. This is the terrain in which organizations and institutions locate their scientific capabilities. Additional information about the scientific and economic impact of each research community allows policy makers to decide which areas to explore, exploit, abandon, or ignore.

We also envision the map as an educational tool. For children, the theoretical relationship between areas of science can be replaced with a concrete map showing how math, physics, chemistry, biology and social studies interact. For advanced students, areas of interest can be located and neighboring areas can be explored.

Nanotechnology

Most research communities in nanotechnology are concentrated in Physics, Chemistry, and Materials Science. However, many disciplines in the Life and Medical Sciences also have nanotechnology applications.

Proteomics

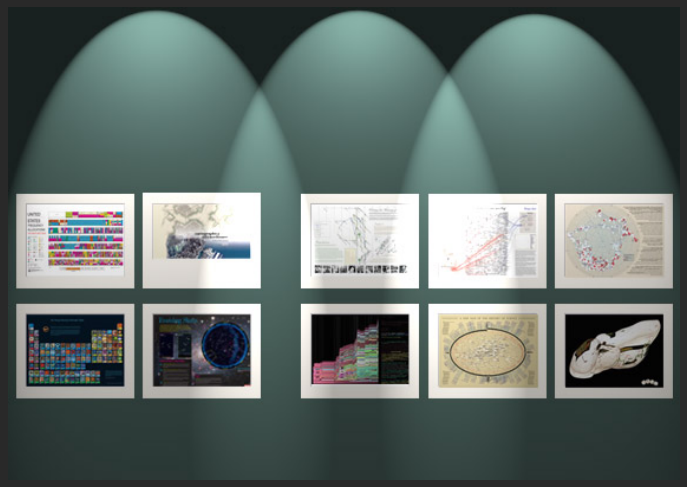
Research communities in proteomics are centered in Biochemistry. In addition, there is a heavy focus in the tools section of chemistry, such as Chromatography. The balance of the proteomics communities are widely dispersed among the Life and Medical Sciences.

Pharmacogenomics

Pharmacogenomics is a relatively new field with most of its activity in Medicine. It also has many communities in Biochemistry and two communities in the Social Sciences.

The Power of Reference Systems

Four Existing Reference Systems
VERSUS
Six Potential Reference Systems of Science



(2nd Iteration of Places & Spaces Exhibit - 2006)

The Visual Elements Periodic Table

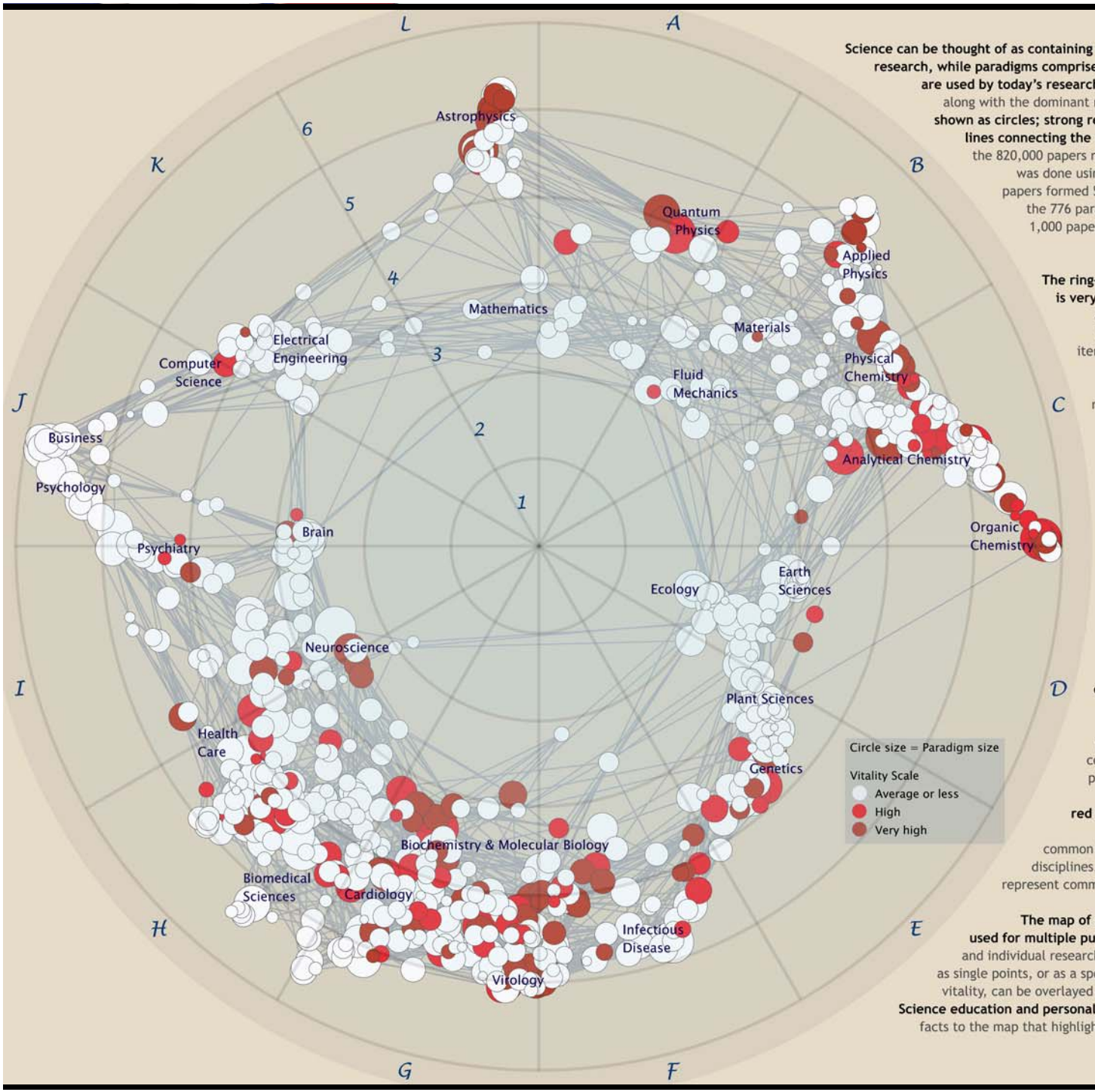
This chart shows the 111 currently known and officially named elements that comprise the Periodic Table (IUPAC 2004). Each element is represented visually by an image produced for the Visual Elements project.

The Periodic Table is an arrangement of all known elements in order of increasing atomic number. The Periodic Table fits all the elements, with their widely diverse physical and chemical properties, into a logical pattern. There are eighteen vertical columns in the table which divide the elements into groups. Elements within a group have closely related physical properties. Horizontal rows list the elements in order of their increasing mass and are called series or periods. Properties of elements change in a systematic way through a period.

Visual Elements is an arts and science collaborative project supported by the Royal Society of Chemistry which aims to explore and reflect upon the diversity of elements that comprise matter in as unique and innovative manner as possible. All the images displayed here, together with screensavers, postcards and chemical data for each element can be viewed on the Visual Elements web site, hosted by the RSC.

Visit the periodic table on the web at:
www.chemsoc.org/viselements

© Murray Robertson/Royal Society of Chemistry 1999-2006



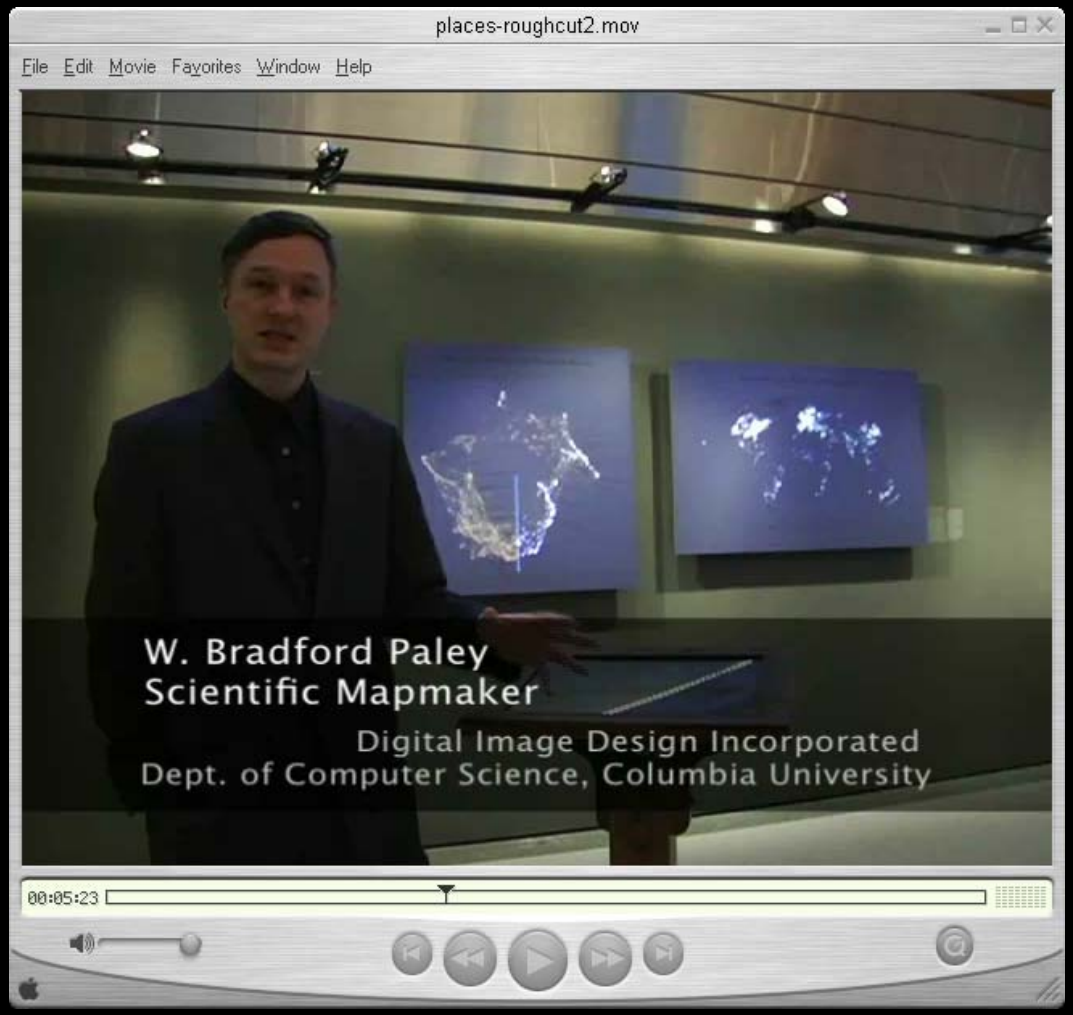
Science can be thought of as containing themes and paradigms. Themes are areas of current research, while paradigms comprise the dominant tool sets and existing knowledge that are used by today's researchers. This map shows **776 major paradigms in science** along with the dominant relationships between these paradigms. **Paradigms are shown as circles; strong relationships between paradigms are indicated by the lines connecting the circles.** The map was created by recursively clustering the 820,000 papers referenced most often in 2003. Clustering at each level was done using VxOrd, a force-directed graph layout routine. These papers formed 53,000 clusters, 6,100 higher-level clusters, and finally the 776 paradigms. Although each paradigm contains, on average, 1,000 papers, some are larger and some are smaller, as shown by different sized circles on the map.

The ring-like structure that is formed by scientific paradigms is very robust. We find similar structures for different years, and for maps generated from scientific journals. "The Structure of Science", a galaxy map shown in the first iteration of Places & Spaces, is a map based on clustering of scientific journals, with superimposition of papers on the journal structure, whereas this map was generated directly from highly-cited papers. "The Structure of Science" shows current science in a disciplinary context, while this map can show the breadth of disciplines that contribute to single paradigms.

Because of the robust nature of the structure of science and its paradigms, we have placed our 776 scientific paradigms within a **reference system containing 12 radial slices and 6 rings.** This allows the position of each paradigm to be codified and available for lookup; for instance ***Fluid Mechanics* paradigms are in grid B3**.

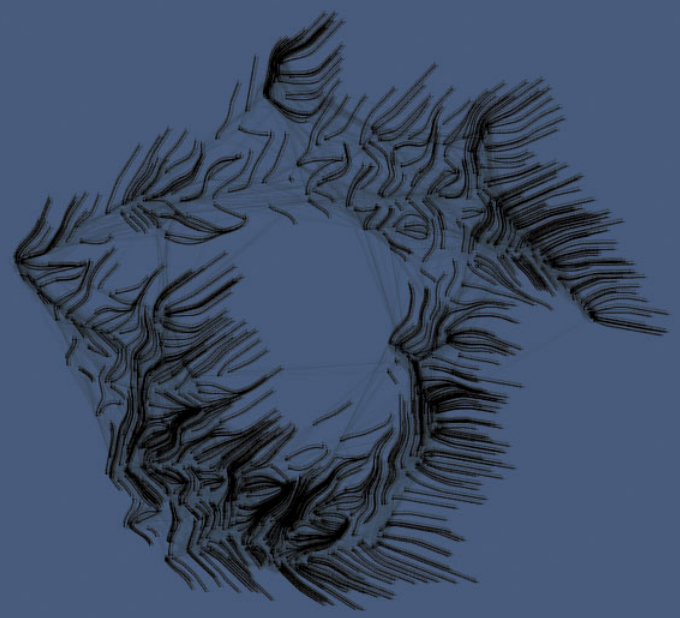
We have also calculated and displayed the vitality of each paradigm. ***Vitality* is a measure of the speed at which a group of researchers reaches consensus about major improvements.** Paradigms are constantly being improved, but it usually takes years to reach consensus about which improvements are major. The white circles represent communities where consensus is reached relatively slowly. This is a common phenomenon in the social sciences, ecological sciences, computer sciences, and mathematics disciplines. **The red circles represent communities of researchers where consensus is reached relatively rapidly.** This is more common in physics, chemistry, biochemistry, and many medical disciplines. Very dark circles (such as those in *Astrophysics*, *L5-6*) represent communities where consensus is reached extremely quickly.

The map of scientific paradigms and its reference system can be used for multiple purposes. Countries, industries, companies, universities, and individual researchers can all **locate themselves within the map**, either as single points, or as a specific collection of paradigms. Various metrics, such as vitality, can be overlayed on this reference system to highlight specific impacts. **Science education and personal discovery** can also be enhanced by linking stories and facts to the map that highlight scientific history, current advances and relationships between scientific paradigms.

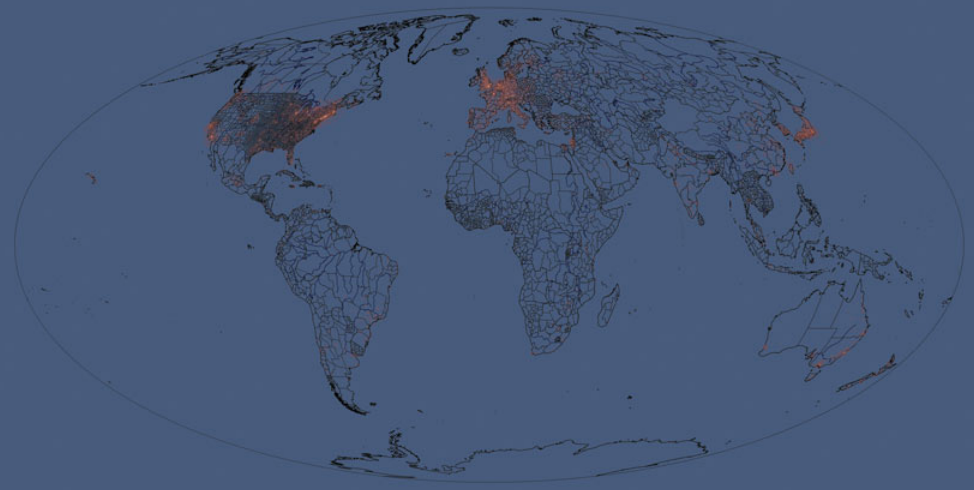


"Places & Spaces: Mapping Science"
on display at the NYPL Science, Industry, and Business Library
Madison/34th, New York City
April 3rd - August 31st, 2006.

Topic Map: How Scientific Paradigms Relate



Geographic Map: Where Science Gets Done



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

Nanotechnology

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

All Topics
Sweep through all 776 scientific paradigms

Nanotechnology
Science on the tiny scale of molecules

Sustainability
The science behind our long-term hopes

Biology & Chemistry
The interface between these two vital fields

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

Francis H. C. CRICK
Co-discovered DNA's double helix

Albert EINSTEIN
Revitalized physics with Relativity theories

Michael E. FISHER
Models critical phase transitions of matter

Susan T. FISKE
Connects perception and stereotypes

Joshua LEDERBERG
Pioneer in bacterial genetic mechanisms

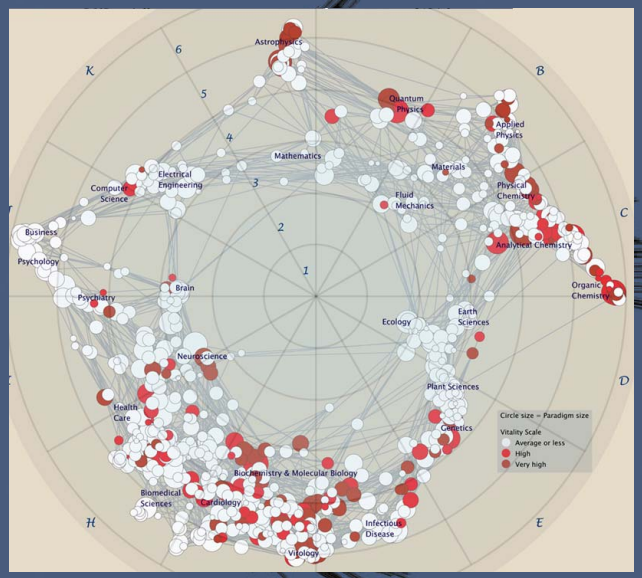
Derek J. de Solla PRICE
Known as the "Father of Scientometrics"

Richard N. ZARE
Uses laser chemistry in molecular dynamics

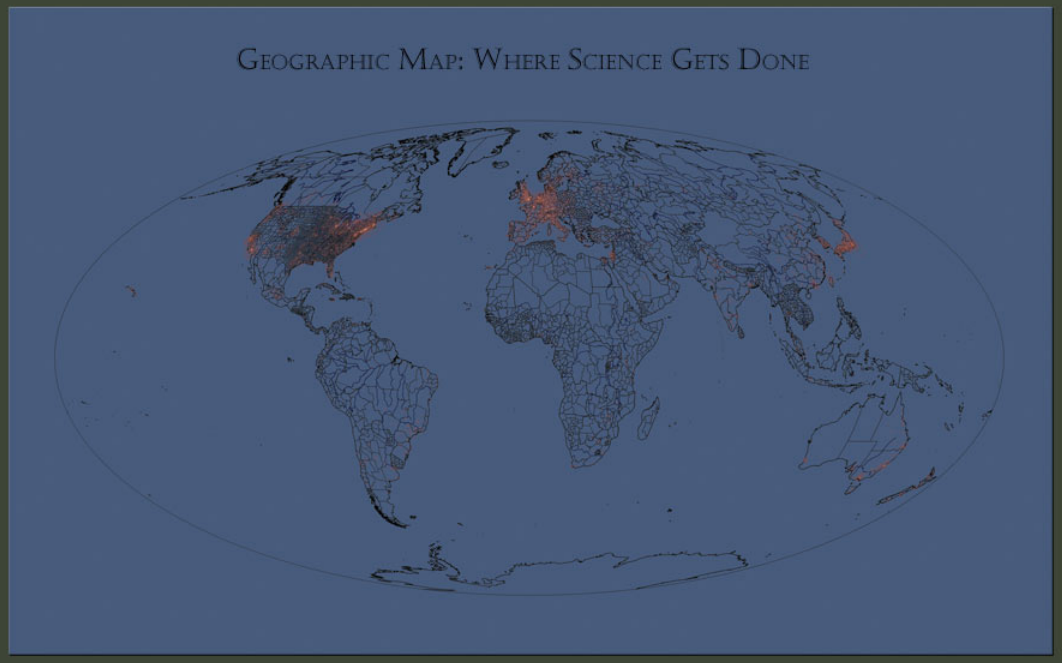
About this display
People & organizations that helped create it

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

Topic Map: How Scientific Paradigms Relate



Geographic Map: Where Science Gets Done



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

Nanotechnology

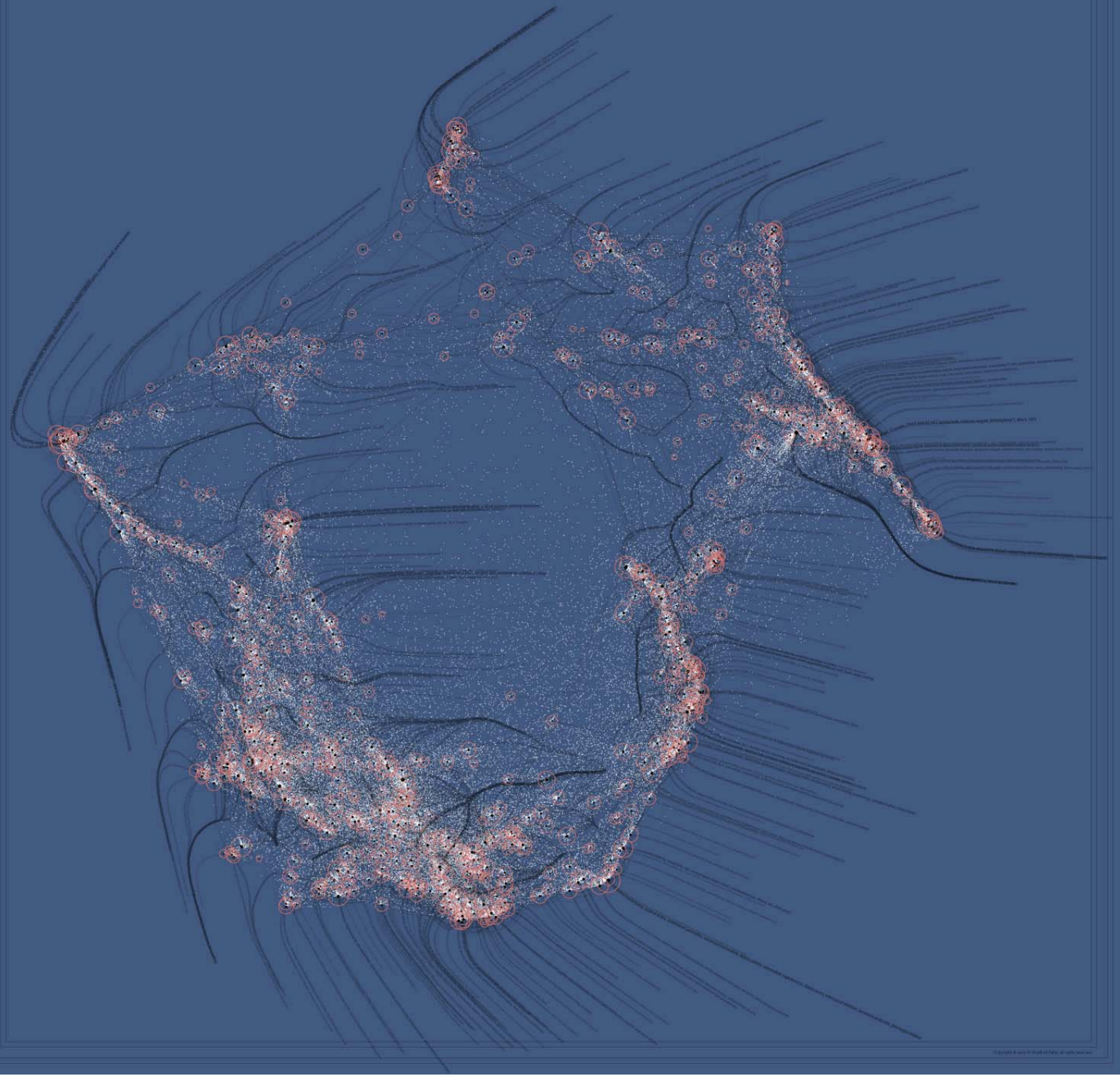
This overlay shows the distribution of nanotechnology within the paradigms of science. The majority of current work in nanotechnology takes places in physics, chemistry, and materials science, at the upper right portion of the map. However, an increasing amount of nanotechnology is being applied in the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

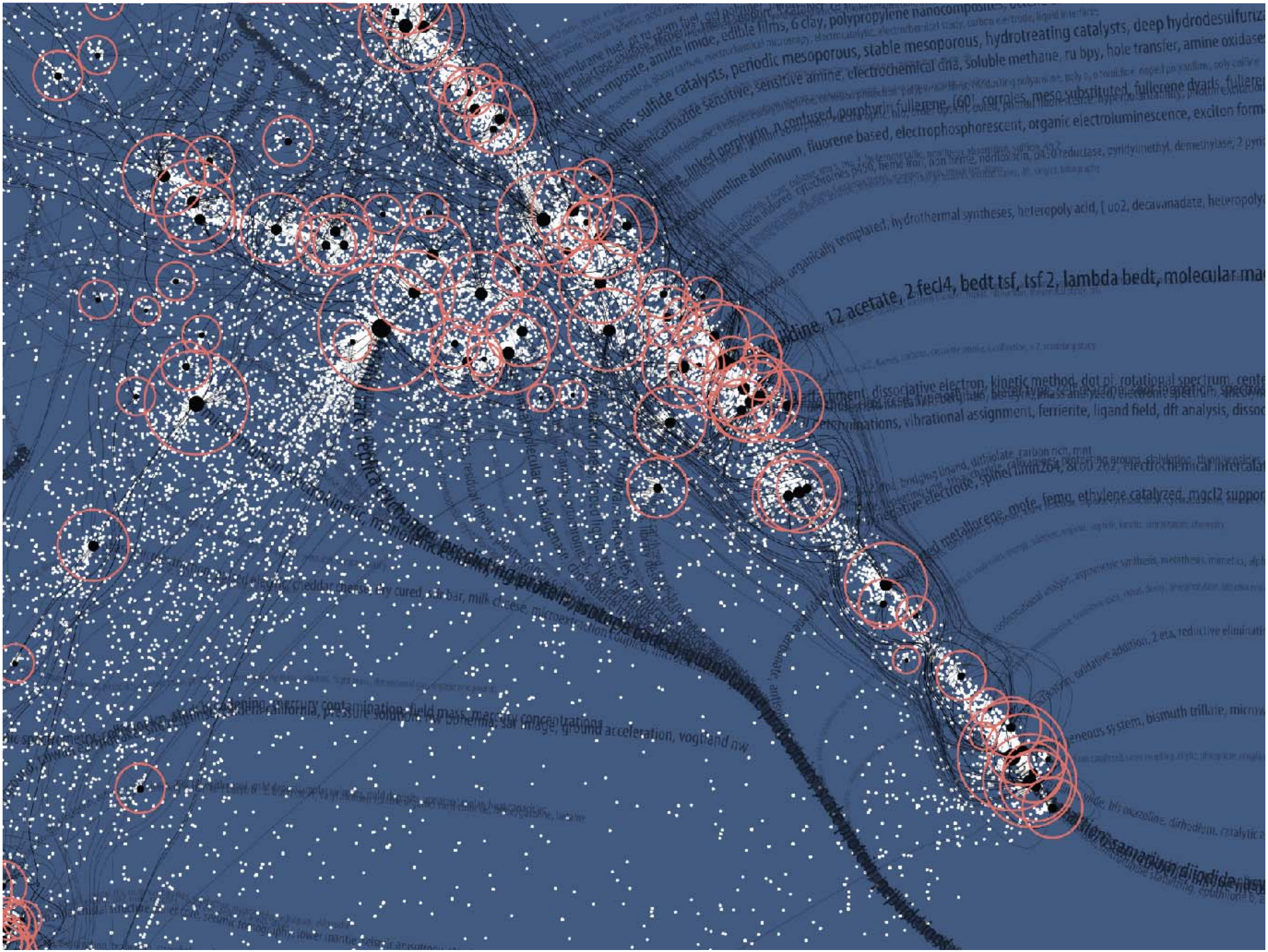
All Topics *Sweep through all 776 scientific paradigms*	**Nanotechnology** *Science on the tiny scale of molecules*	**Francis H. C. CRICK** *Co-discovered DNA's double helix*	**Albert EINSTEIN** *Revitalized physics with Relativity theories*	**Michael E. FISHER** *Models critical phase transitions of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

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

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

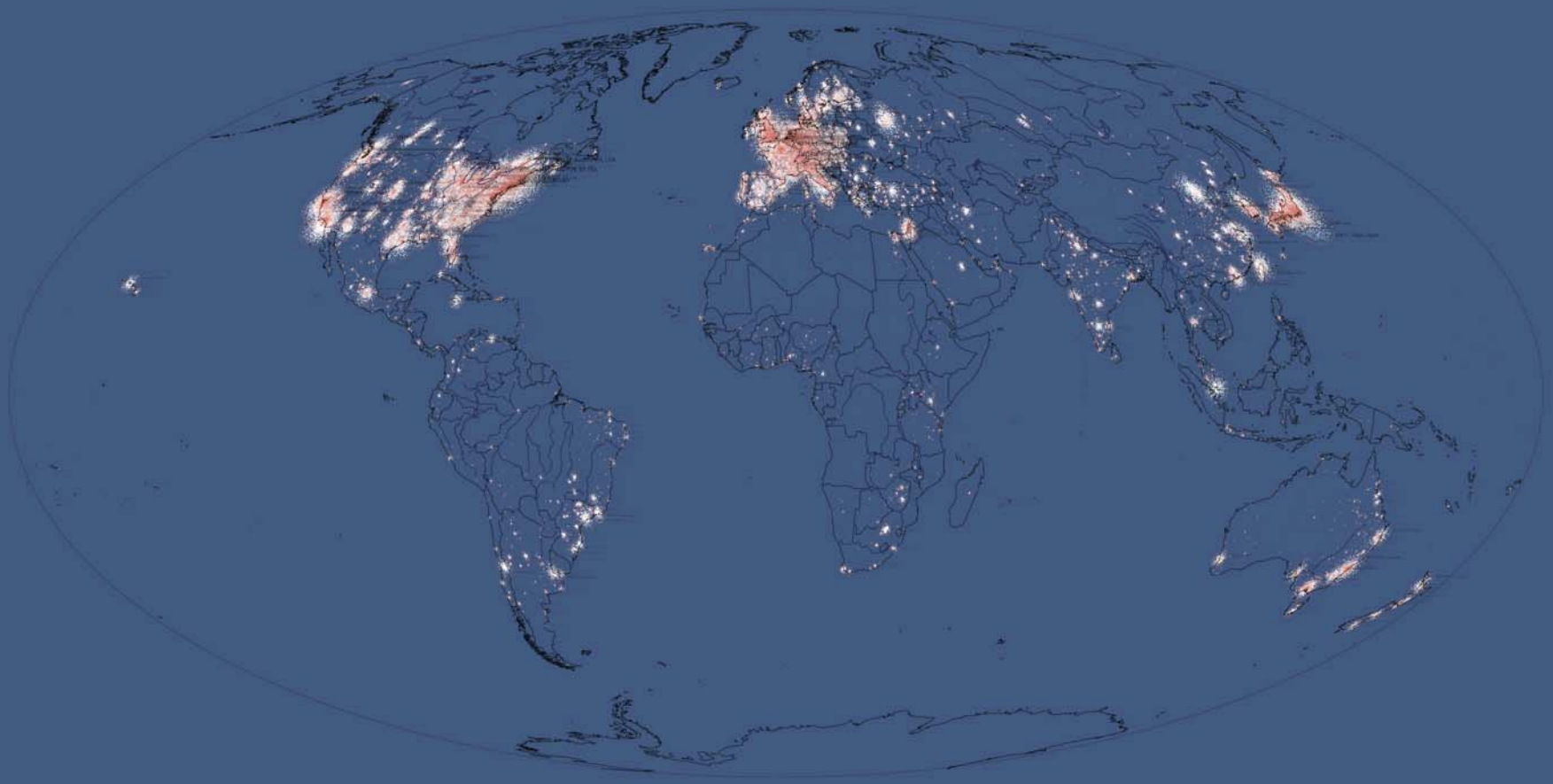
Topic Map: How Scientific Paradigms Relate

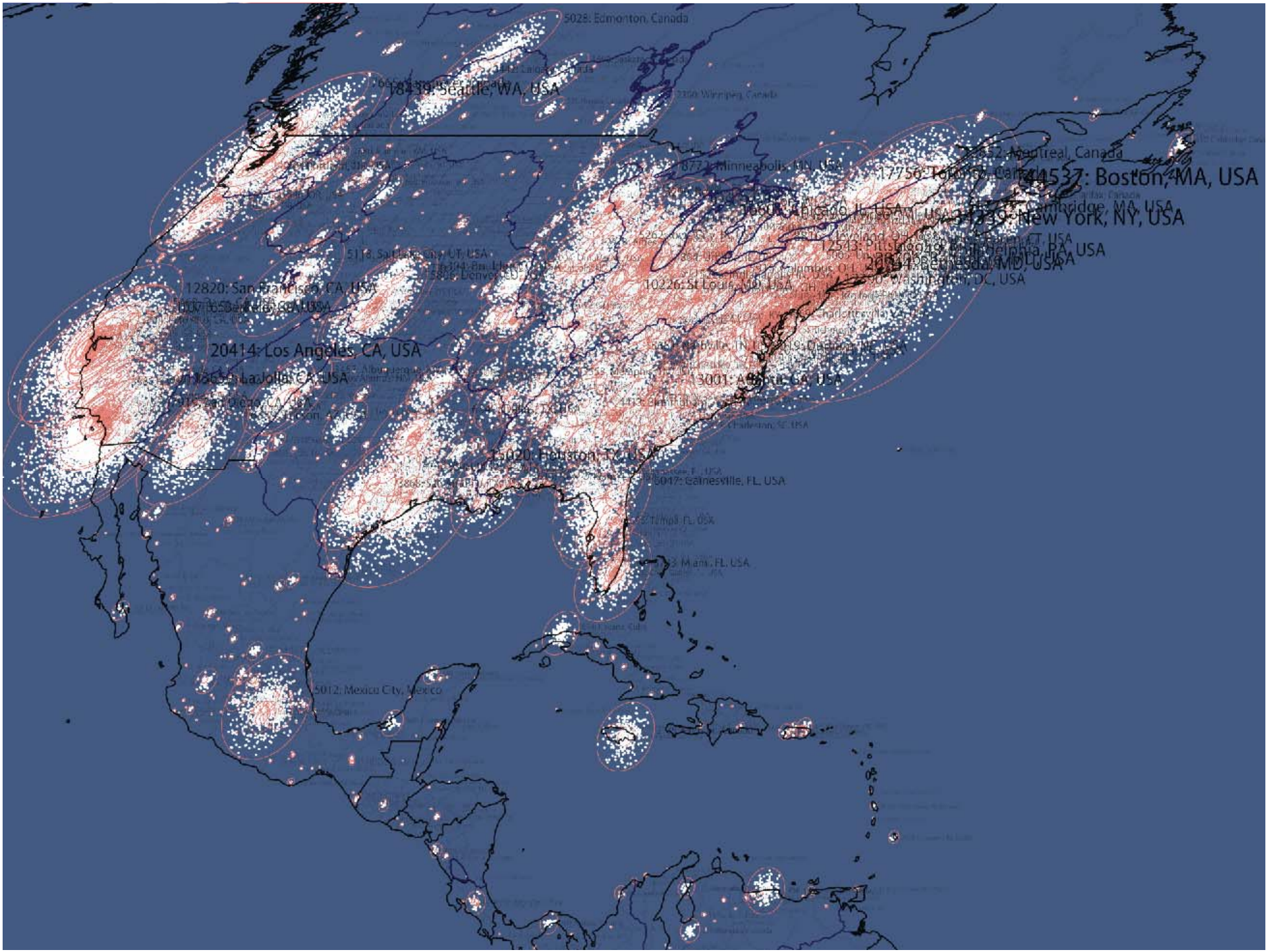


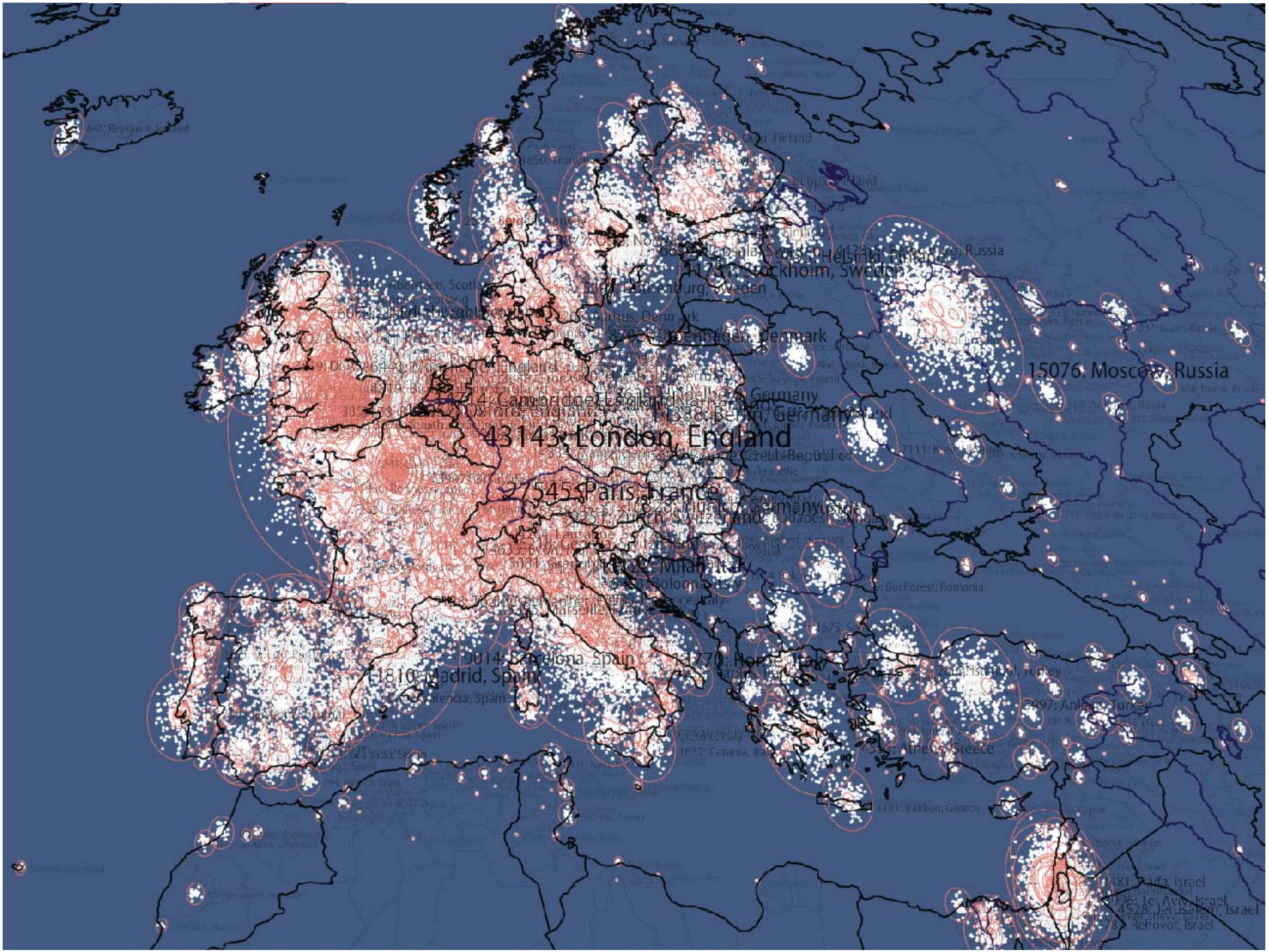


sulfide catalysts, periodic mesoporous, stable mesoporous, hydrotreating catalysts, deep hydrodesulfuriz
electrochemical dna, soluble methane, ru bpy, hole transfer, amine oxidase
fluorene based, electrophosphorescent, organic electroluminescence, exciton form
hydrothermal syntheses, heteropoly acid, [uo2, decavanadate, heteropoly
12 acetate, 2 fecl4, bedt tsf, tsf 2, lambda bedt, molecular ma
dissociative electron, kinetic method, dot pl, rotational spectrum, cente
vibrational assignment, ferrierite, ligand field, dft analysis, dissoc
metallorene, mofe, femo, ethylene catalyzed, mgcl2 suppor
cheddar cheese, dry cured, air bar, milk cheese, microextraction, cuprid, kinetic
mercury contamination, field mass, mercury concentrations
ground acceleration, vogtland nw

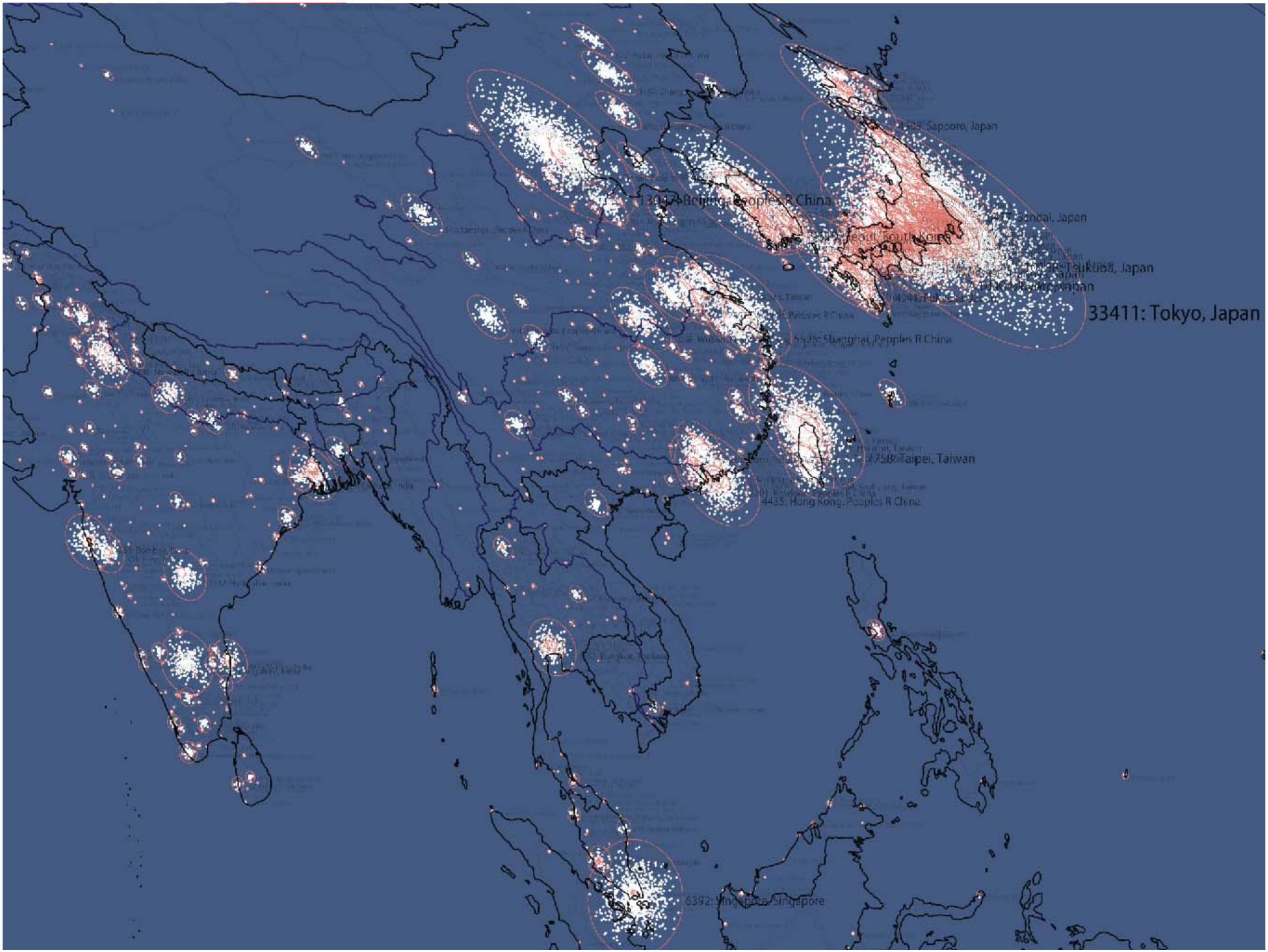
Geographic Map: Where Science Gets Done




5028: Edmonton, Canada
18439: Seattle, WA, USA
8772: Minneapolis, MN, USA
44537: Boston, MA, USA
Cambridge, MA, USA
New York, NY, USA
12543: Pittsburgh, PA, USA
12820: San Francisco, CA, USA
20414: Los Angeles, CA, USA
10226: St Louis, MO, USA
13001: Atlanta, GA, USA
5020: Houston, TX, USA
5047: Gainesville, FL, USA
5012: Mexico City, Mexico



15076: Moscow, Russia
43143: London, England
27545: Paris, France
Copenhagen, Denmark
Berlin, Germany
Milan, Italy
Bologna, Italy
Madrid, Spain
Athens, Greece
Iraklion, Greece
Bucharest, Romania
Rehovot, Israel


33411: Tokyo, Japan
Taipei, Taiwan
Hong Kong, Peoples R China
Singapore

Final Statements

- Maps of science have come a long way recently
- We believe that the ring structure of science is real, and emergent from some fundamental properties of the way nature is organized (here's a research topic for someone)
- Science maps provide substrates to overlay a variety of actionable information
- Indicators at the community level enable decision making
  - For a graduate student – which community should I join?
  - For a funding agency – which communities should I fund?
  - …

Related Publications

- Klavans, R. & Boyack, K. W. (2006, in press). Quantitative evaluation of large maps of science. *Scientometrics.*
- Boyack, K. W., Klavans, R. & Börner, K. (2005). Mapping the backbone of science. *Scientometrics,* 64(3), 351-376.
- Klavans, R. & Boyack, K. W. (2006). Identifying a better measure of relatedness for mapping science. *Journal of the American Society for Information Science and Technology*, 57(2), 251-263.
- Boyack, K. W., & Rahal, N. (2005, in press). Evaluation of LDRD investment areas at Sandia. *Technological Forecasting and Social Change*.
- Boyack, K. W., Mane, K. & Börner, K. (2004). Mapping Medline papers, genes and proteins related to melanoma research. *IEEE Information Visualization 2004*, 965-971.
- Boyack, K. W. (2004). Mapping knowledge domains: Characterizing PNAS. *Proceedings of the National Academy of Sciences* 101(S1), 5192-5199.
- Boyack, K. W., & Börner, K. (2003). Indicator-assisted evaluation and funding of research: Visualizing the influence of grants on the number and quality of research papers. *Journal of the American Society for Information Science and Technology* 54(5), 447.
- Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology* 37, 179-255.
- Boyack, K. W., Wylie, B. N., & Davidson, G. S. (2002). Information visualization, human-computer interaction, and cognitive psychology: Domain visualizations. *Lecture Notes in Computer Science* 2539, 145-160.
- Boyack, K. W., Wylie, B. N., & Davidson, G. S. (2002). Domain visualization using VxInsight for science and technology management. *Journal of the American Society for Information Science and Technology,* 53(9), 764-774.
- Davidson, G. S., Wylie, B. N., & Boyack, K. W. (2001). Cluster stability and the use of noise in interpretation of clustering. *Proc. IEEE Information Visualization 2001*, 23-30.
- Boyack, K.W., Wylie, B.N., Davidson, G.S. & Johnson, D.K., *Analysis of patent databases using VxInsight*. Presented at New Paradigms in Information Visualization and Manipulation 2000, McLean, VA, Nov. 10, 2000.