Conducting research on emerging innovation systems through bibliometric analysis

Jan Youtie[1,3,4], Philip Shapira[2, 3,4], & Alan Porter[3,4]

RTTA 1 - Research and Innovation Systems Analysis Group @ Georgia Tech

Manchester International Summer School, June 2014

1. Georgia Tech Enterprise Innovation Institute, Atlanta, GA, USA
2. Manchester Institute of Innovation Research, University of Manchester, UK
3. School of Public Policy, Georgia Institute of Technology, Atlanta, GA, USA
4. Center for Nanotechnology in Society (CNS-ASU)

Learning Objectives

- Examine how to define emerging technology in bibliometric terms
- Link bibliometric methods to research questions
- Provide introduction into practical bibliometric analytical techniques

Overview of Session

1. What is research and innovation analysis? What types of anticipatory questions does it address?
2. Research questions and analyses
3. Defining an emerging technology for bibliometric analysis

CNS-ASU: Research-Train-Engage-Partner to understand governance of new technologies

***Anticipatory governance** is a data-driven movement: measurements are increasingly used, not just to evaluate outcomes and hold actors accountable, but to manage, legislate, and do course-corrections in real time.*

Real-Time Technology Assessment

1. Research and Innovation Systems Analysis (RISA)
2. Public Opinion and Values (POV)
3. Deliberation and Participation (D&P)
4. Reflexivity Assessment and Evaluation (RAE)

Thematic Research Clusters

1. Equity and Responsibility
2. (2005-2010) Human Identity, Enhancement, & Biology

   (2011-2015) Nanotech and the City

1. Research and Innovation Systems Assessment

- Objective
  - Provide evidence bases for anticipatory governance with unobtrusive structured data
- Types of Anticipatory Questions:
  - Overall: Who is doing what in nano research and innovation?
    - What, when: How can nano be characterized?
    - With whom: How will nano be organized? Who will be the major players?
    - Where: Where will nanotechnology develop?
    - With what implications?

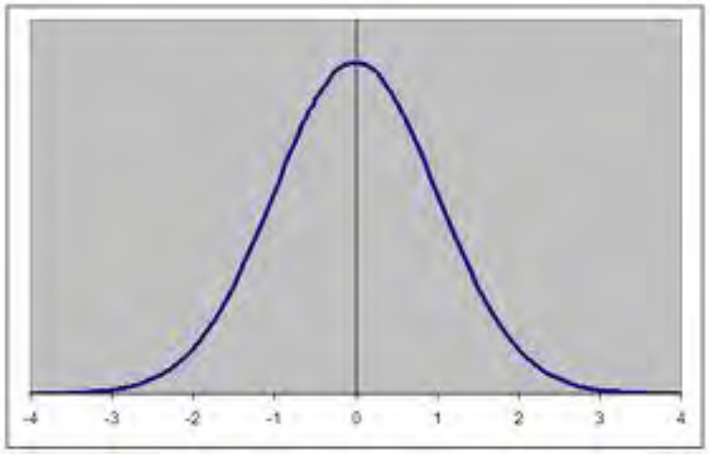
2. Research Questions

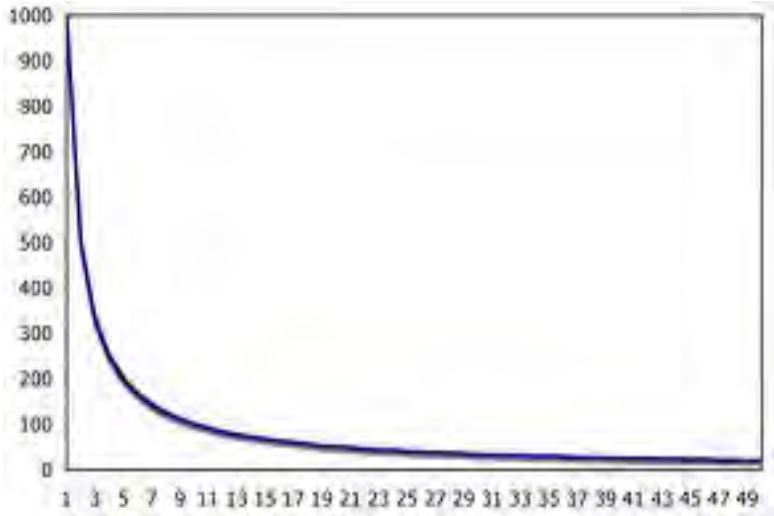
- How fast is the field growing (trend analysis)? What research is the most influential (citation analysis)?
- Where are the greatest concentrations of research activity (GIS analysis)?
- Which researchers are collaborating (network analysis)?
- How multidisciplinary is a set of publications? How many applications are involved in a set of patents? (science, patent mapping)

Types of Analyses

- Counts
- Time series
- GIS maps
- Network maps
- Disciplinary analyses

Publications, Citation Counts





- Zip's Law: Frequency of a term is inversely proportional to its rank
  - A small number occur often
  - A large number occur rarely
  - There are some medium frequency terms
- Means have no meaning→percentiles more useful (Wagner, Leydesdorff 2012)

What do Citations Mean?

- Citations are commonly viewed as a measure of quality. Why would an author include a citation in a paper other than it is the “best work”?

Citations

- Differences by discipline – fractional counting
- Position in the article (central v. peripheral, negative, discarded)
- # of co-authors
- Self-citation (more important in shorter horizons)
- Hot/Immediate v. Sleeping Beauties (Glanzel et al. 2006)

Common Citation Measures

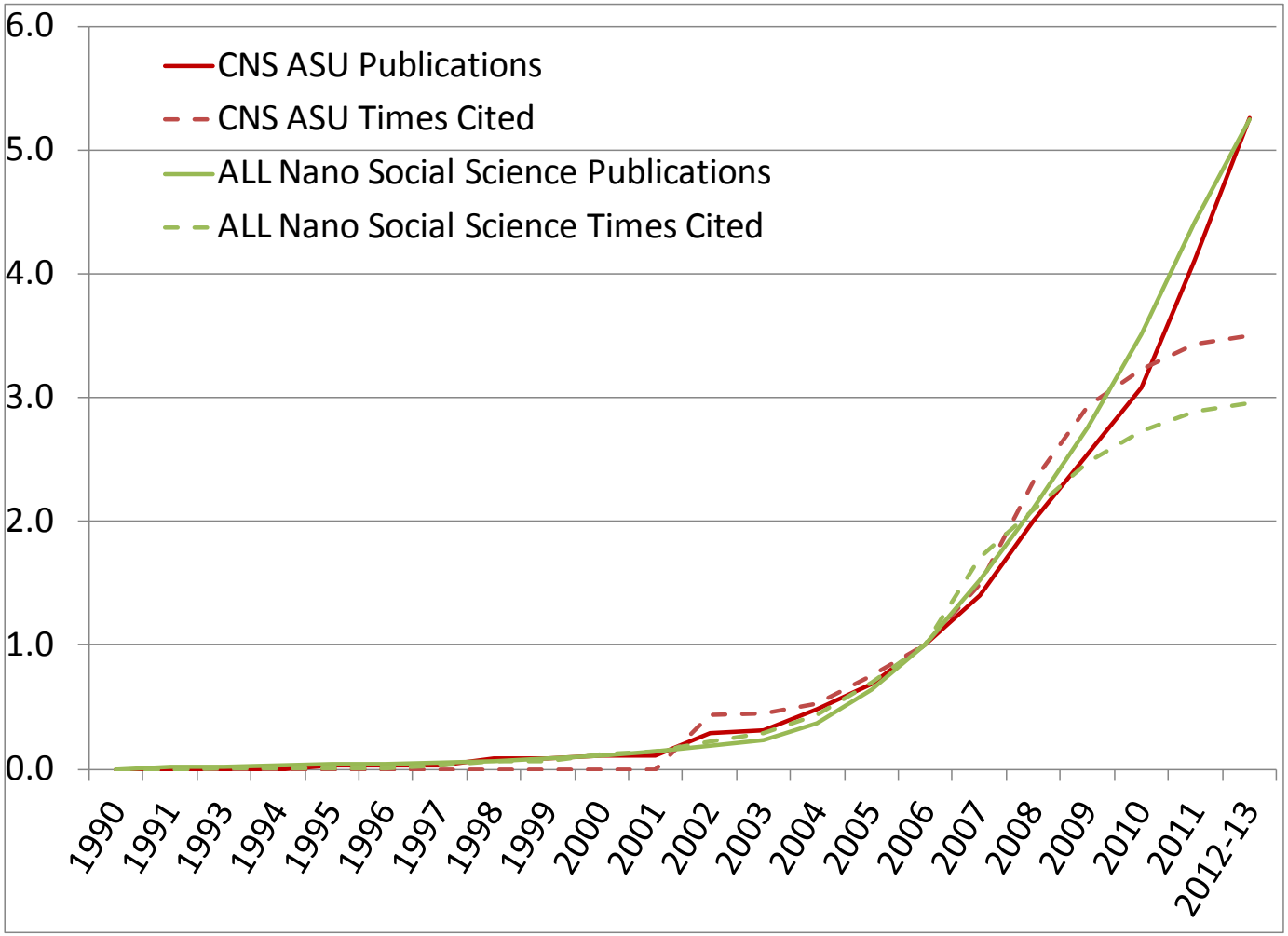
- H-Index

Rank	Article	Citations
1	Article	240
2	Article	145
3	Article	72
4	Article	68
5	Article	67
6	Article	59
7	Article	43
8	Article	36
9	Article	28
10	Article	27
11	Article	25
12	Article	23
13	Article	21
14	Article	21
15	Article	19
16	Article	16
17	Article	12

- Journal Impact Factor (Garfield 2006)
  - # citations/# citable works in 2 year period
  - Relative importance of journal
  - Used by Thomson Reuters Journal Citation Reports

Time Series:

Publication Output, Citations: CNS ASU and All Nano Social Science (normalized, year 2006=1)



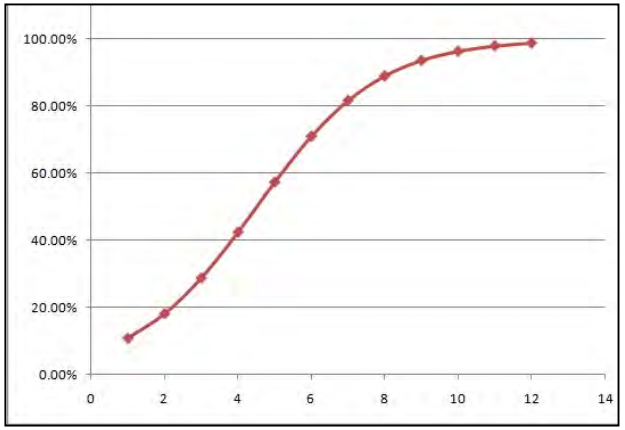
Source: Youtie, Carley, et al., 2013, 1760 publications 1990-2013 (March), from Web of Science and Scopus; citations of 177 books/book chapters from Google Scholar

Time Series Issues

- Messy data
  Which date?
  - Pubs: accepted, online, published
  - Patents: application, filing, priority, publication, family,

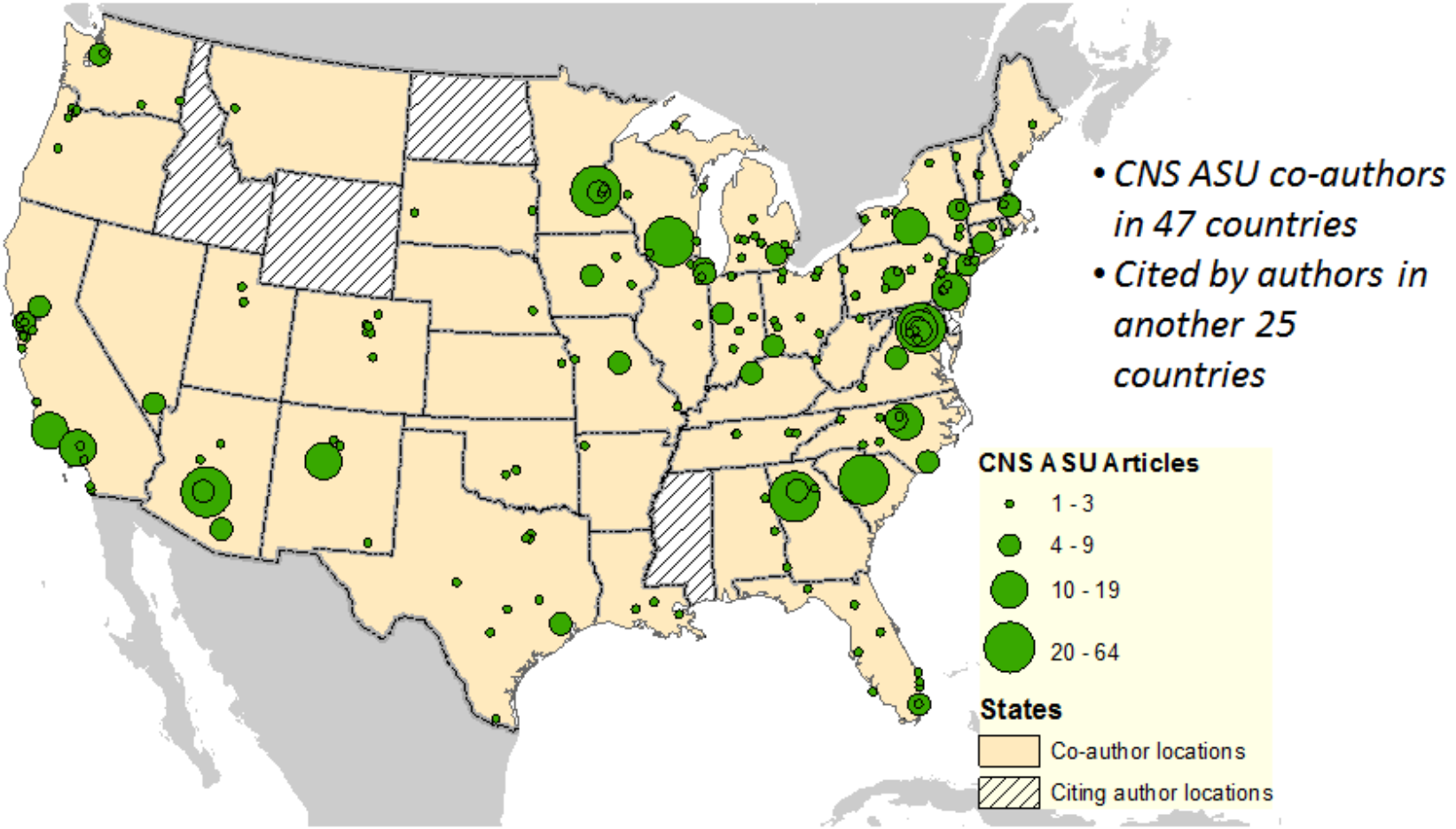
Publication Date
2009-03-12; 2009-06-11
4/23/2009
9/18/2008
2008-05-15; 2008-10-23

- Extrapolations from known data
  - Linear
  - Exponential
  - Fisher-Pry/Gompertz (S curve)
- Overall upward trends



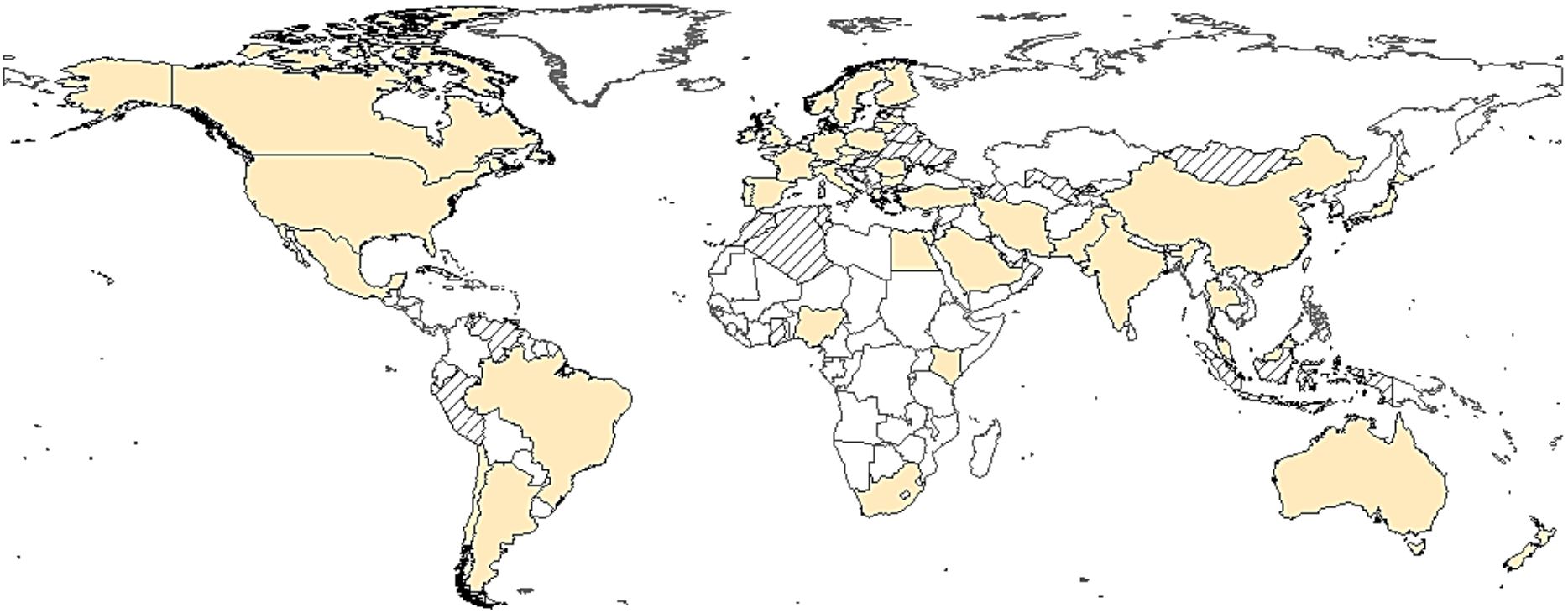
CNS ASU Articles: co-authors in 45 states; cited by authors in all 50 states

(CNS ASU co-author, citing author locations for co-terminus US)



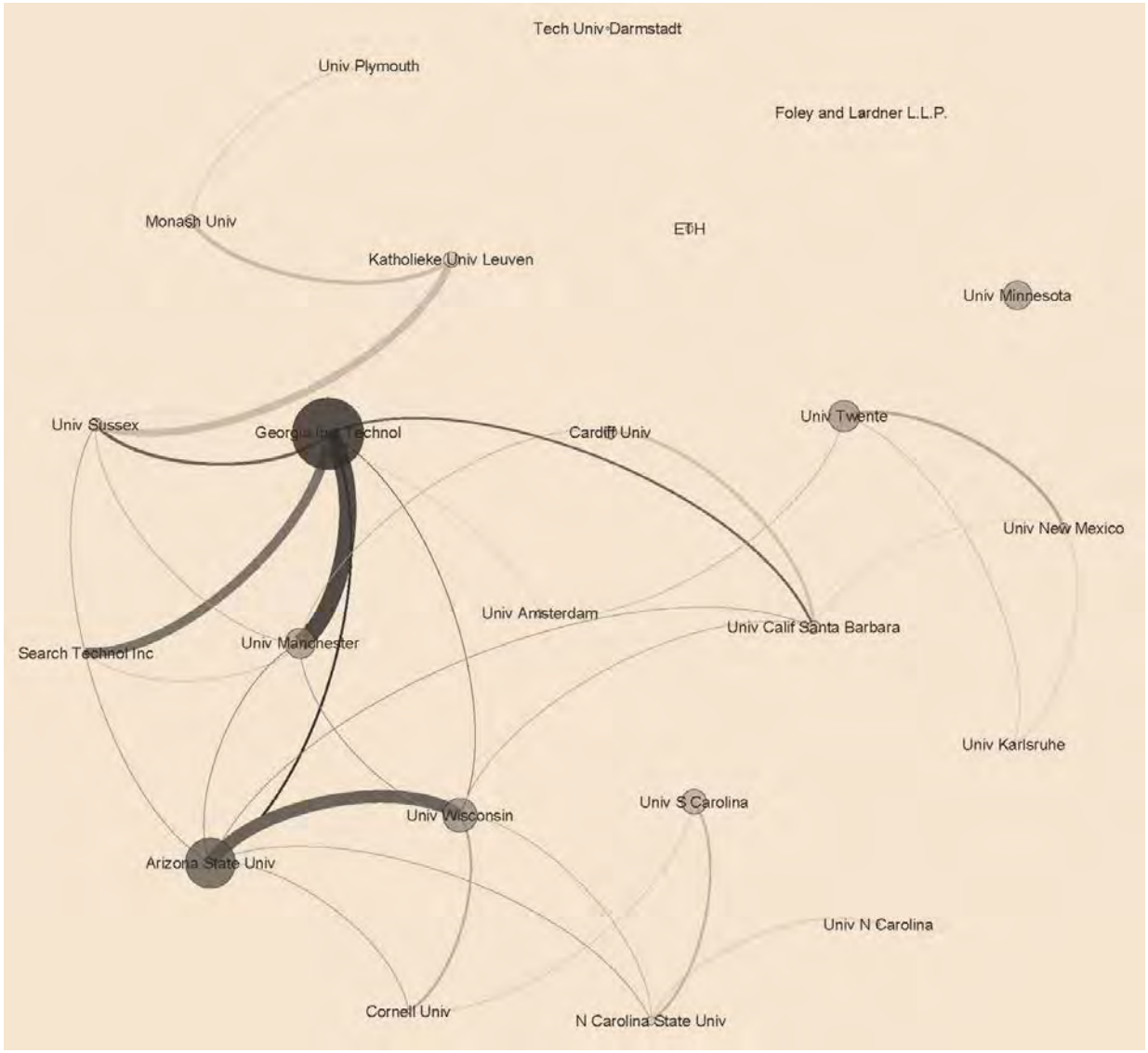
Source: Youtie, Carley, et al., 2013, 1760 publications 1990-2013 (March); 11,189 citing articles from Web of Science and Scopus

But international impact more difficult to distinguish



GIS Map Issues

- Linking information to geographies (batch geocoding websites)
  - Geocoding lat/long: 53.4874992,-2.2458899
- Grouping nearby places that are socio-economically tied
- Dots v. shading
- Labeling

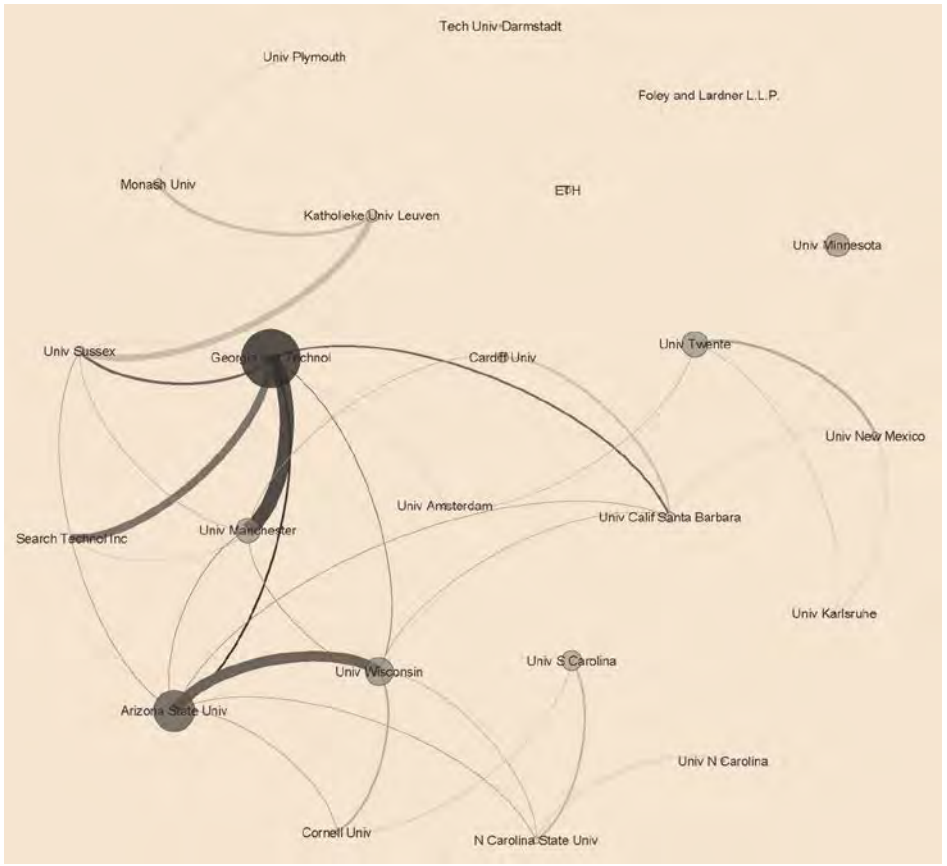


Network Maps

(Co-authorships by Institution in Nano Social Science Publications, 10+ publications in WoS, Scopus)

Source: Youtie, Carley, et al., 2013, 401 publications 1990-2013 (March), from Web of Science and Scopus

Some Network Measures

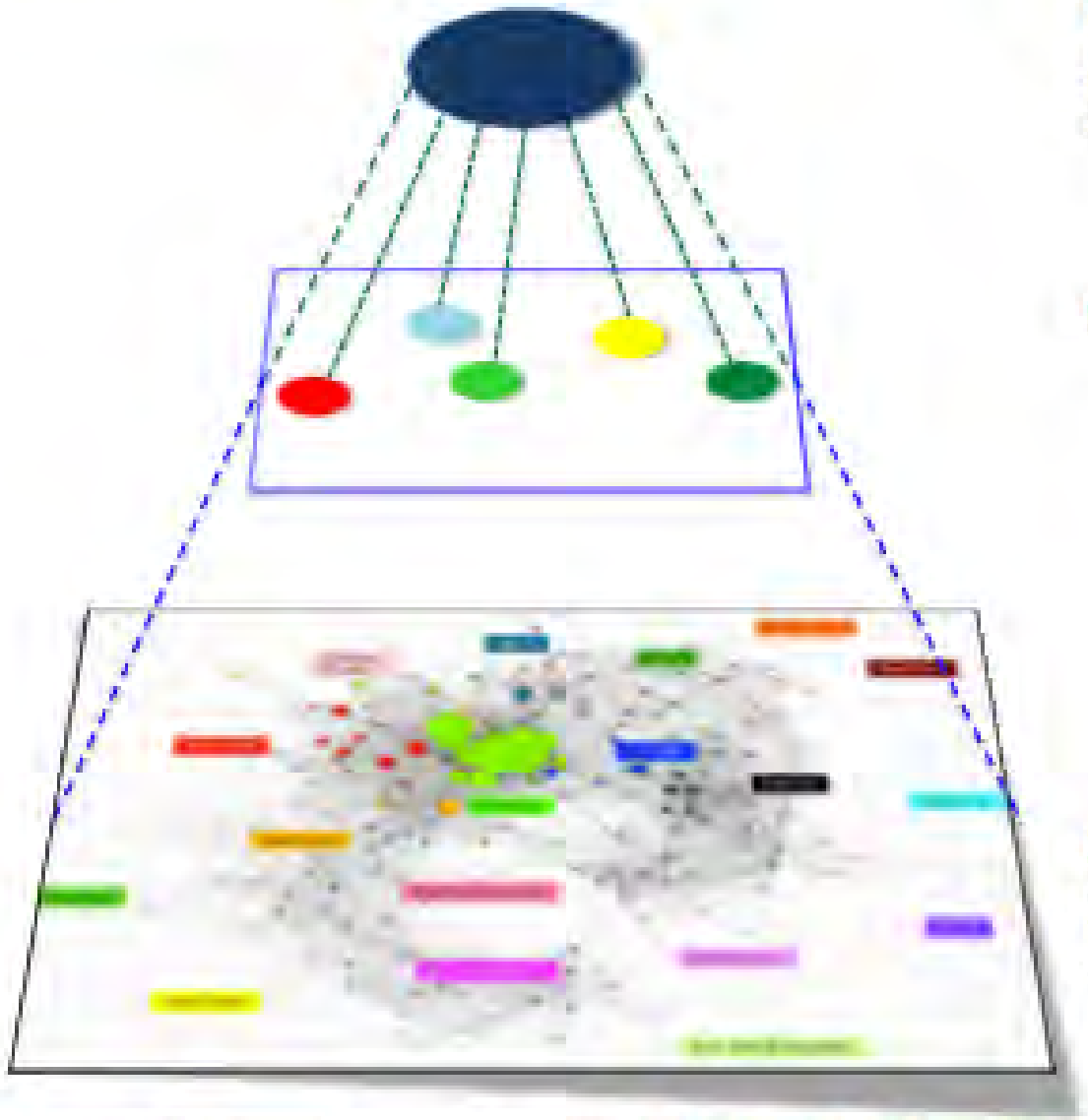


- Degree (# direct connections)
  - Arizona State U: 7
  - Georgia Inst. Tech: 6
  - Univ Manchester: 6
  - Univ Wisconsin: 6
- Betweenness (broker)
  - Univ. of Sussex: 58
  - Arizona State U: 39
  - UC Santa Barbara: 24
  - North Carol. State: 23

Network Map Issues

- Spacing, mapping algorithm (edges crossing)
- Trying to extract meaning out of the visualization (not just a pretty picture)
- Make unit of analysis clear (nodes and edges)
- Adding labels, colors, weights, sizes, legend
- Very descriptive – no way to think of causality without longitudinal analysis

The global map of science and overlay

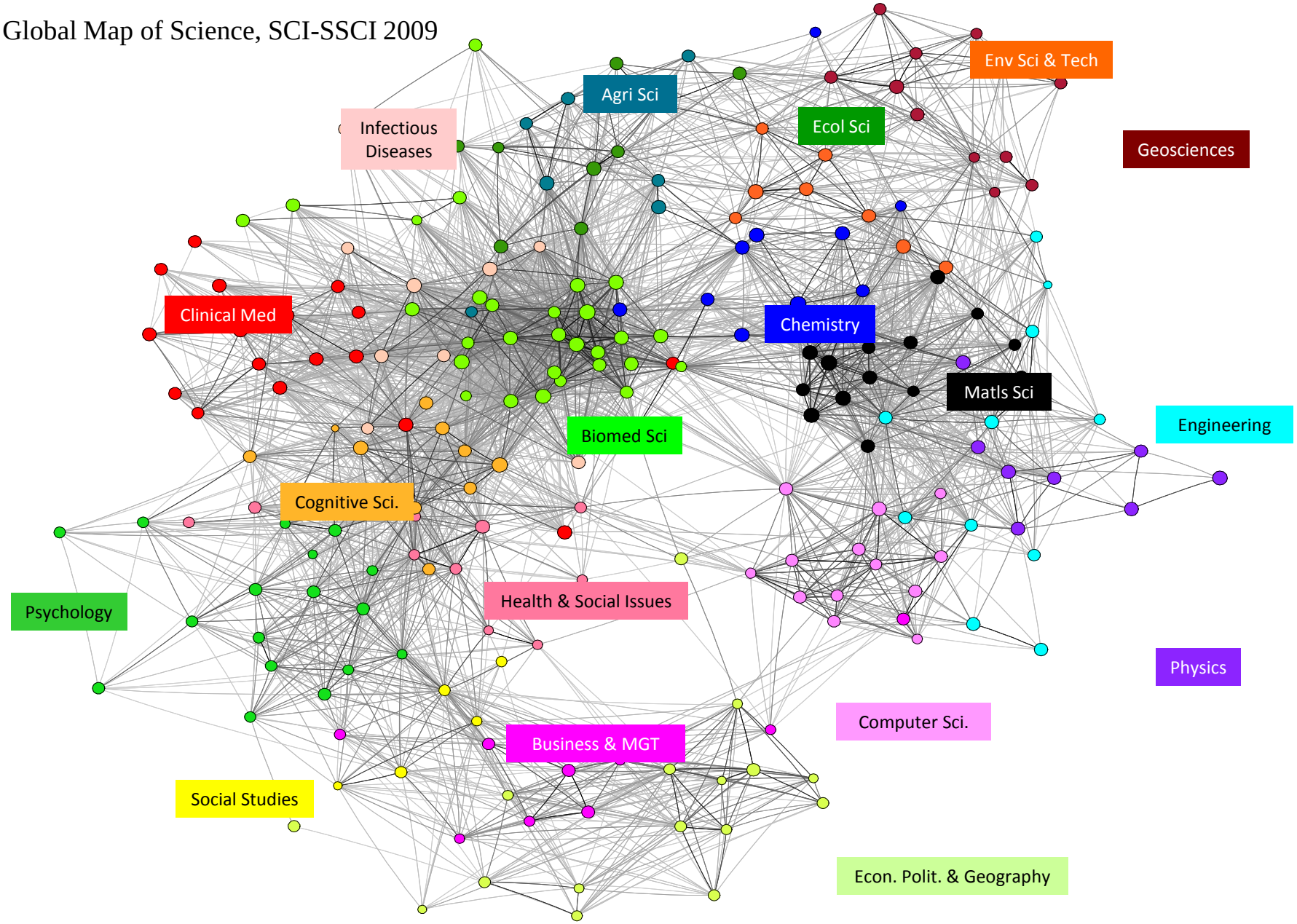


Entity
(Organization, Technology, Scientific Field, etc.)

Overlay

Basemap
(Organizations, Technologies, Scientific Fields, etc.)

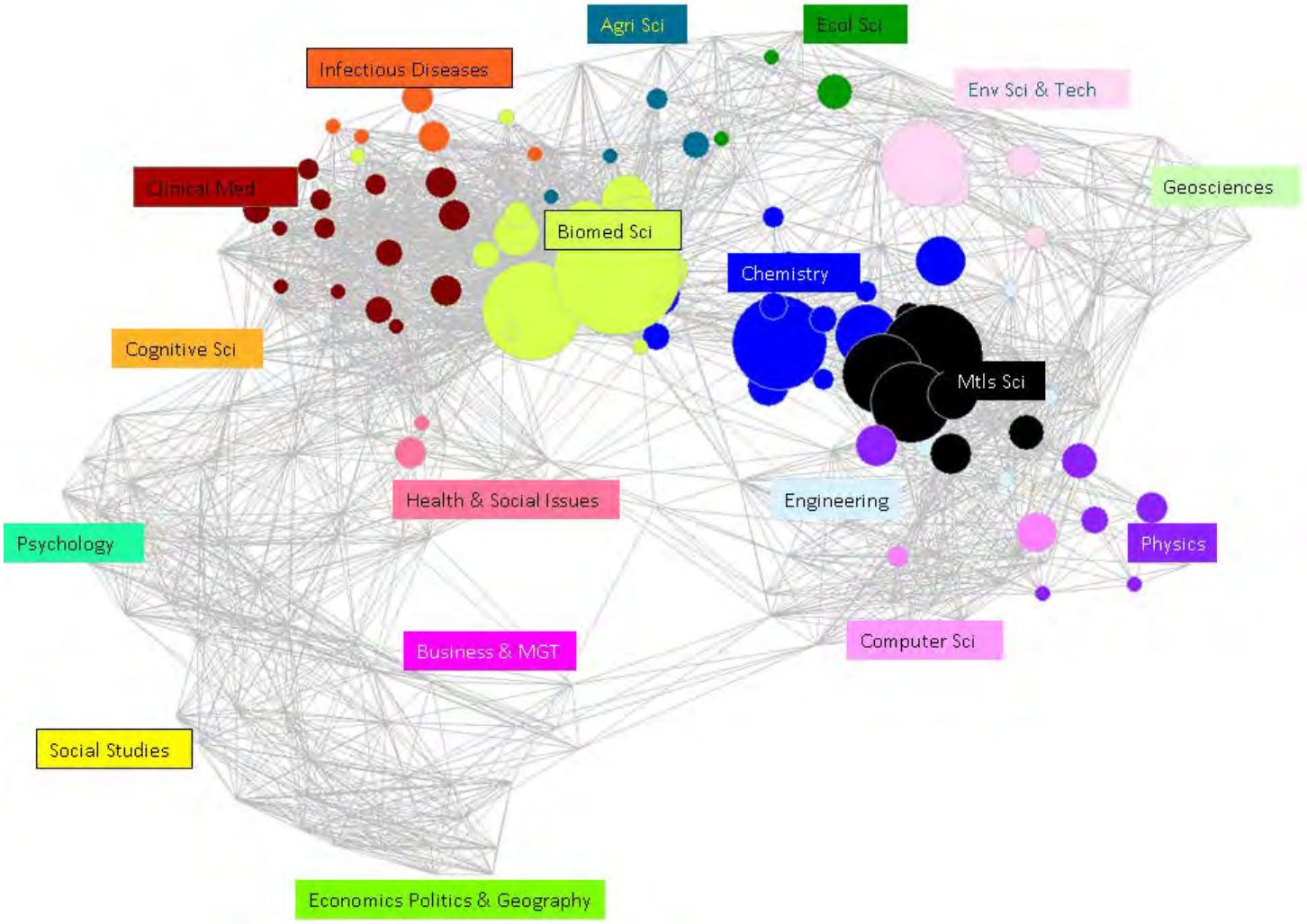
Global Map of Science, SCI-SSCI 2009



Rafols, Porter and Leydesdorff (2009)

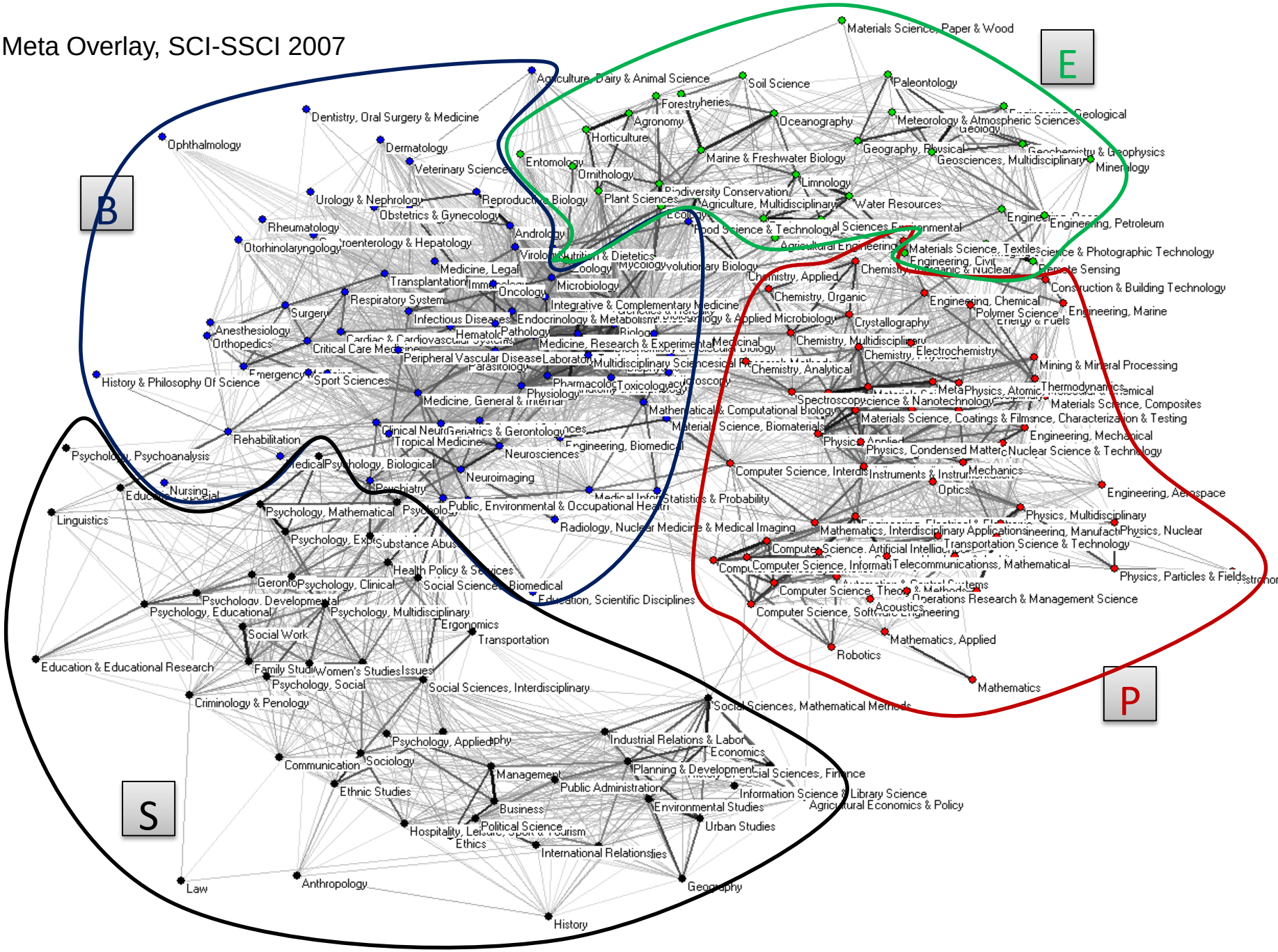
Pajek

Nano EHS Map of Science



Youtie et al., 2011

Meta Overlay, SCI-SSCI 2007



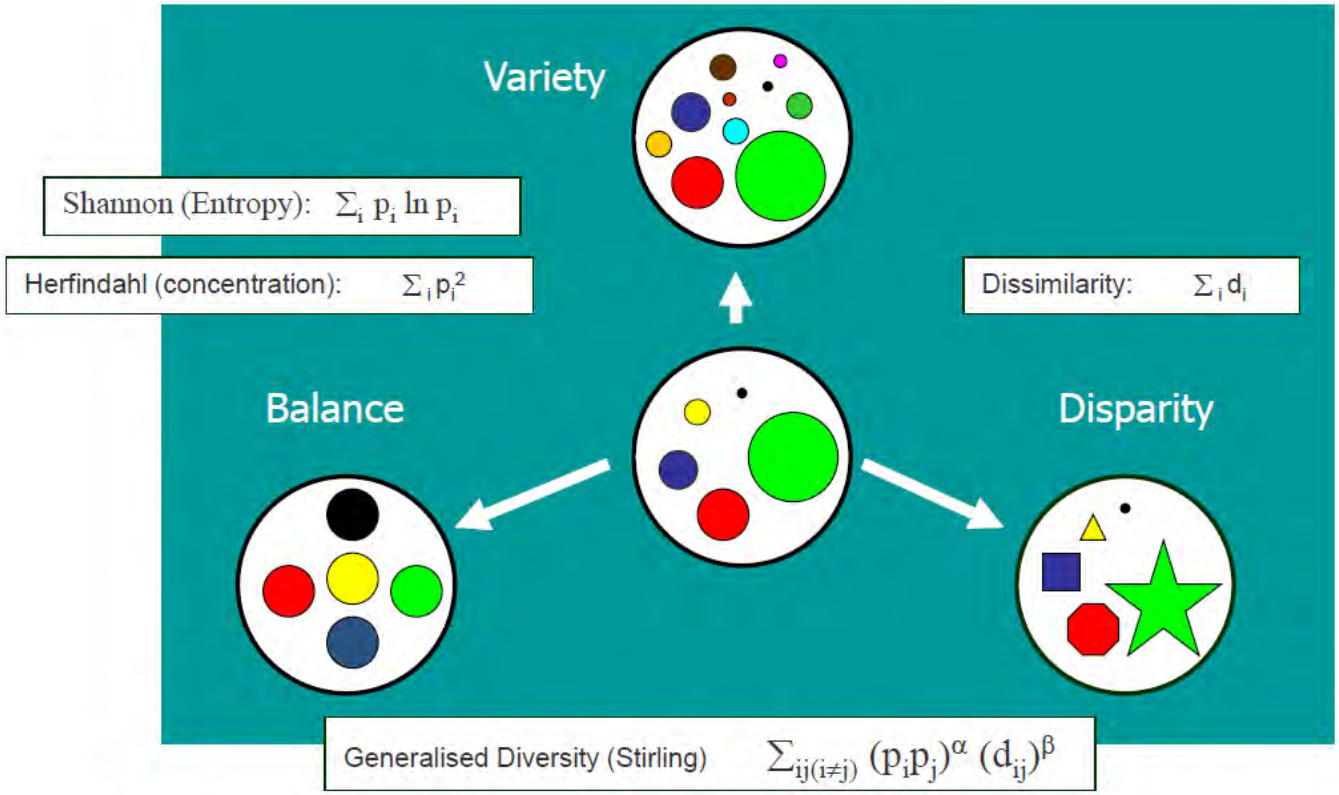
Indicators of Diversity, Coherence

Heuristics of diversity
(Stirling, 1998; 2007)
(Rafols and Meyer, 2009)

Diversity:
'attribute of a system whose elements may be apportioned into categories'

Characteristics:
Variety: Number of distinctive categories
Balance: Evenness of the distribution
Disparity: Degree to which the categories are different.

[** Shannon & Herfindahl do not include Disparity]



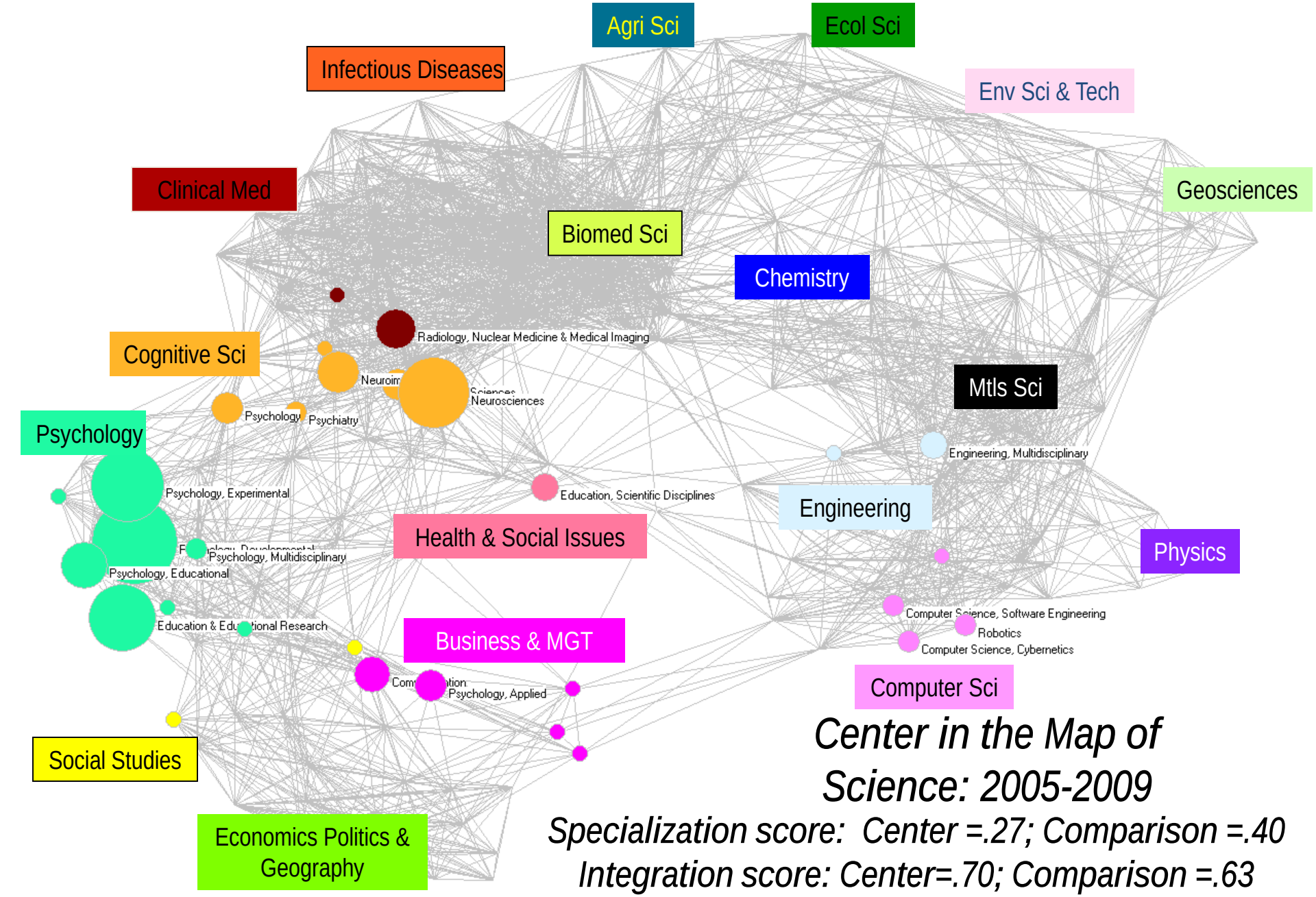
Indicators of Diversity, Coherence

- "Integration" – breadth of WOSCs drawn upon
- "Specialization" – concentration of publication activity
- "Diffusion" – diversity of WOSCs citing the research

Web of Science Category--WOSC
Management
Information Science & Library Science
History & Philosophy Of Science
Business
Multidisciplinary Sciences
Planning & Development
Ethics
Chemistry, Multidisciplinary
Computer Science, Interdisciplinary Appli.
Materials Science, Multidisciplinary
Nanoscience & Nanotechnology
Social Issues
Engineering, Industrial

Indicators of Diversity, Coherence

- Specialization
  - Body of multiple papers (by author, organization)
  - Types of journals/WOSC
  - Range:
    - 0=less specialized (more diverse)
    - 1=more concentrated
- Integration
  - Single paper or body of papers
  - Types of references
  - Range:
    - 0=less integrated (less diverse)
    - 1=more integrated (more diverse)



Center in the Map of Science: 2005-2009

Specialization score: Center =.27; Comparison =.40

Integration score: Center=.70; Comparison =.63

Map source: Rafols, I., Meyer, M. (2009) Diversity and Network Coherence as indicators of interdisciplinarity: case studies in bionanoscience. Scientometrics, 81(2), in print; Leydesdorff, L., Rafols, I. (2009) A Global Map of Science Based on the ISI Subject Categories. Journal of the American Society for Information Science and Technology, 60(2), 348-362.

Patent Maps – Based on Patent Classes

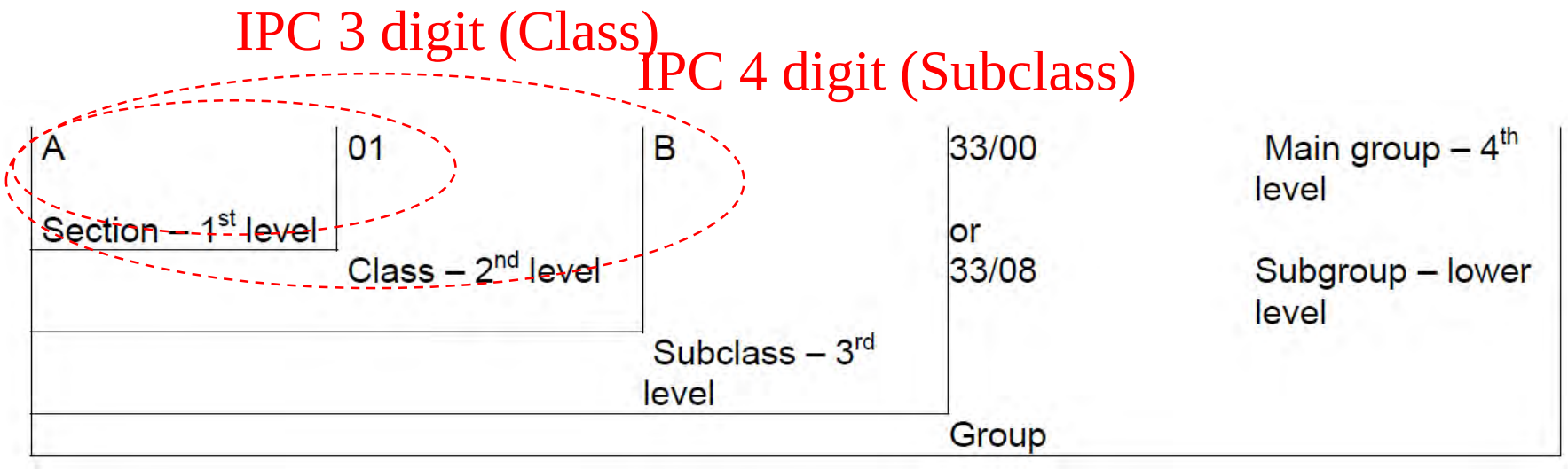
- Sections

A	HUMAN NECESSITIES
B	PERFORMING OPERATIONS; TRANSPORTING
C	CHEMISTRY; METALLURGY
D	TEXTILES; PAPER
E	FIXED CONSTRUCTIONS
F	MECHANICAL ENGINEERING; LIGHTING; HEATING; WEAPONS; BLASTING
G	PHYSICS
H	ELECTRICITY
Y	GENERAL TAGGING OF NEW TECHNOLOGICAL DEVELOPMENTS; GENERAL TAGGING OF CROSS-OVER TECHNOLOGIES SPANNING OVER SEVERAL SECTIONS OF THE IPC

- Classifications in the same broad category are more similar than those in different broad categories
  - Counts
  - Shares
  - Herfindahl

Source: http://www.wipo.int/classifications/ipc/en/

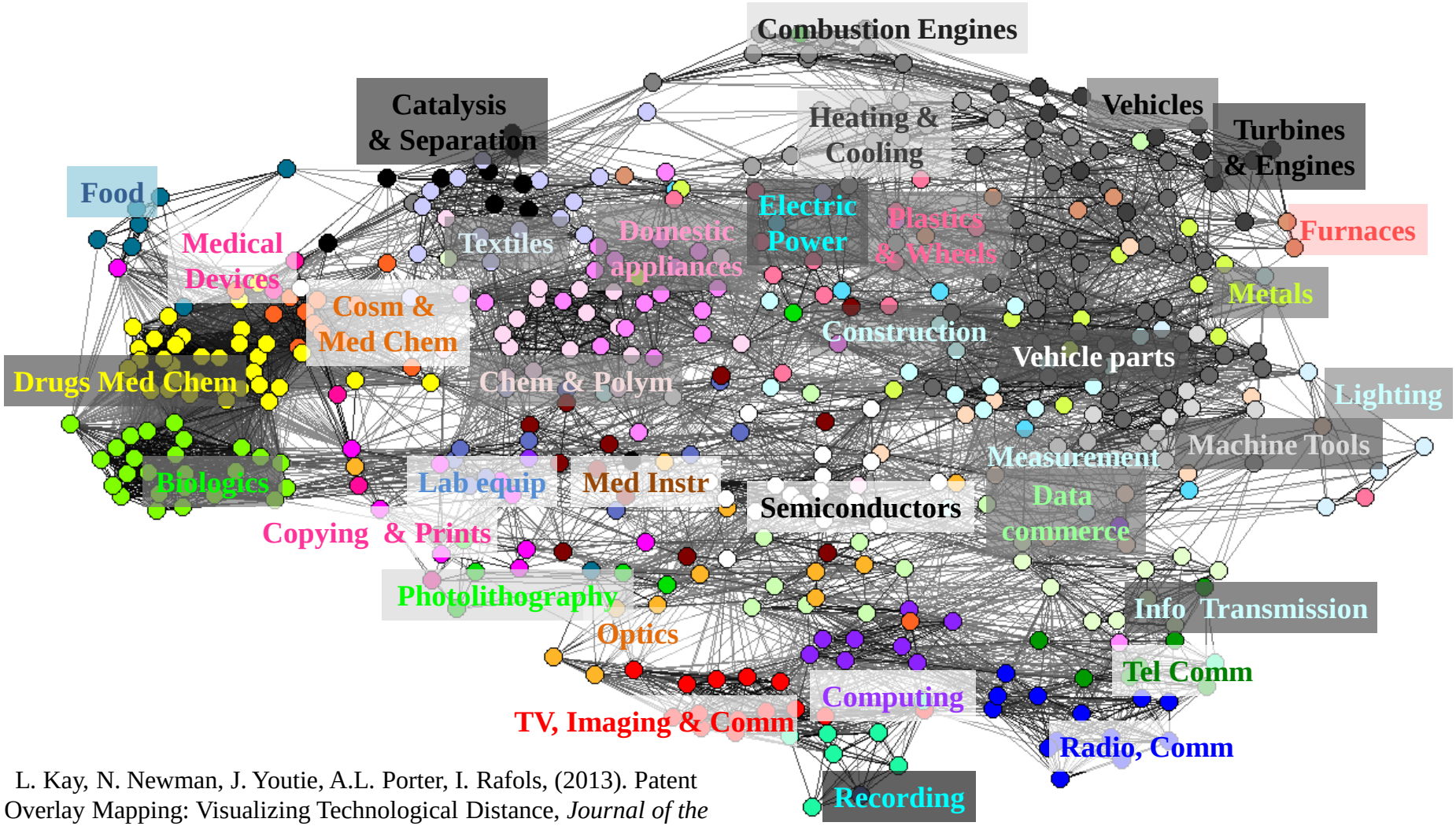
Patent Maps – Based on Patent Classes



Source: Guide to the IPC (2013), p. 6.

Global Base Map of Patents

(35 factors/dimensions color coded)

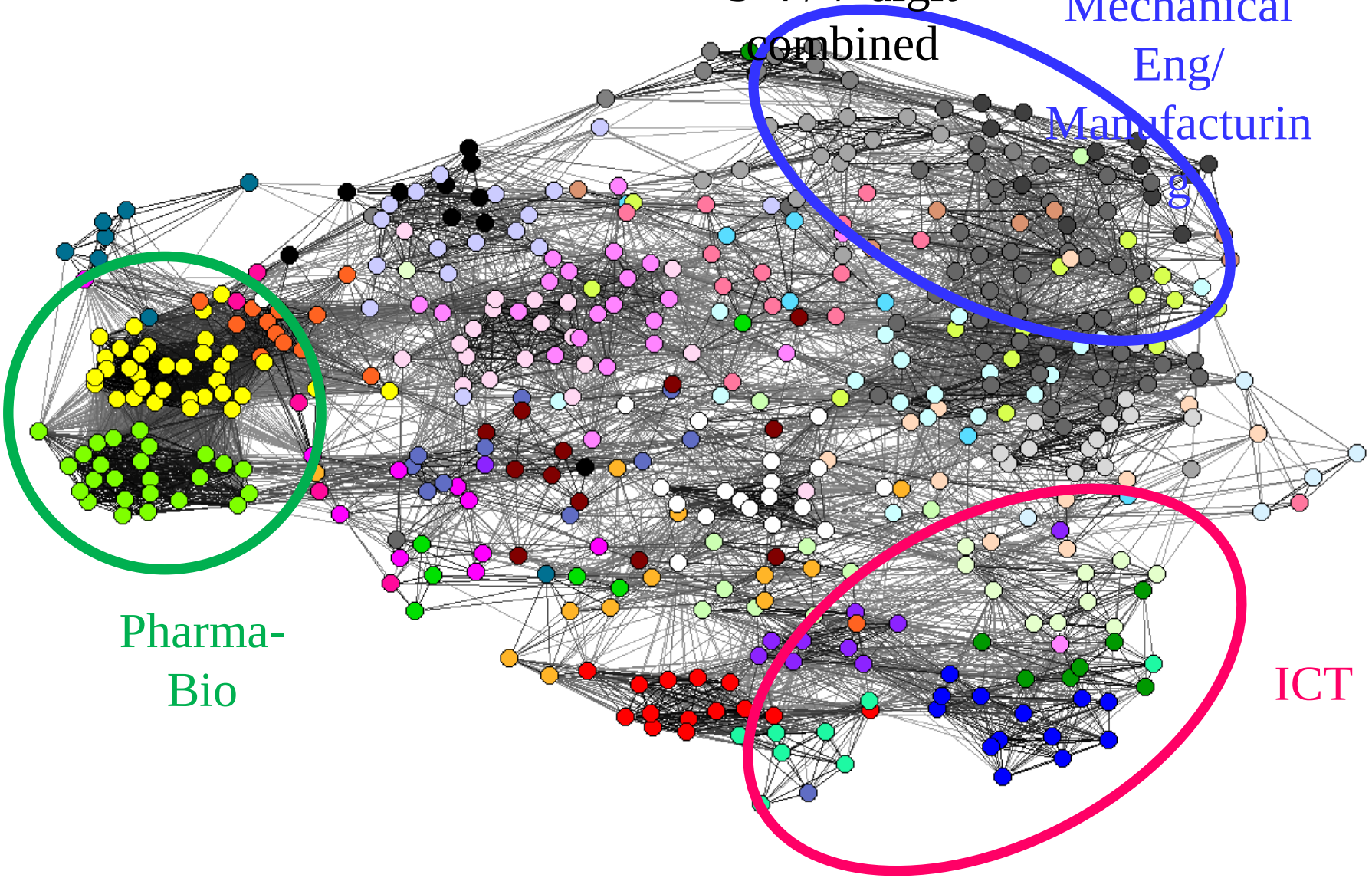


L. Kay, N. Newman, J. Youtie, A.L. Porter, I. Rafols, (2013). Patent Overlay Mapping: Visualizing Technological Distance, *Journal of the American Society for Information Science and Technology*

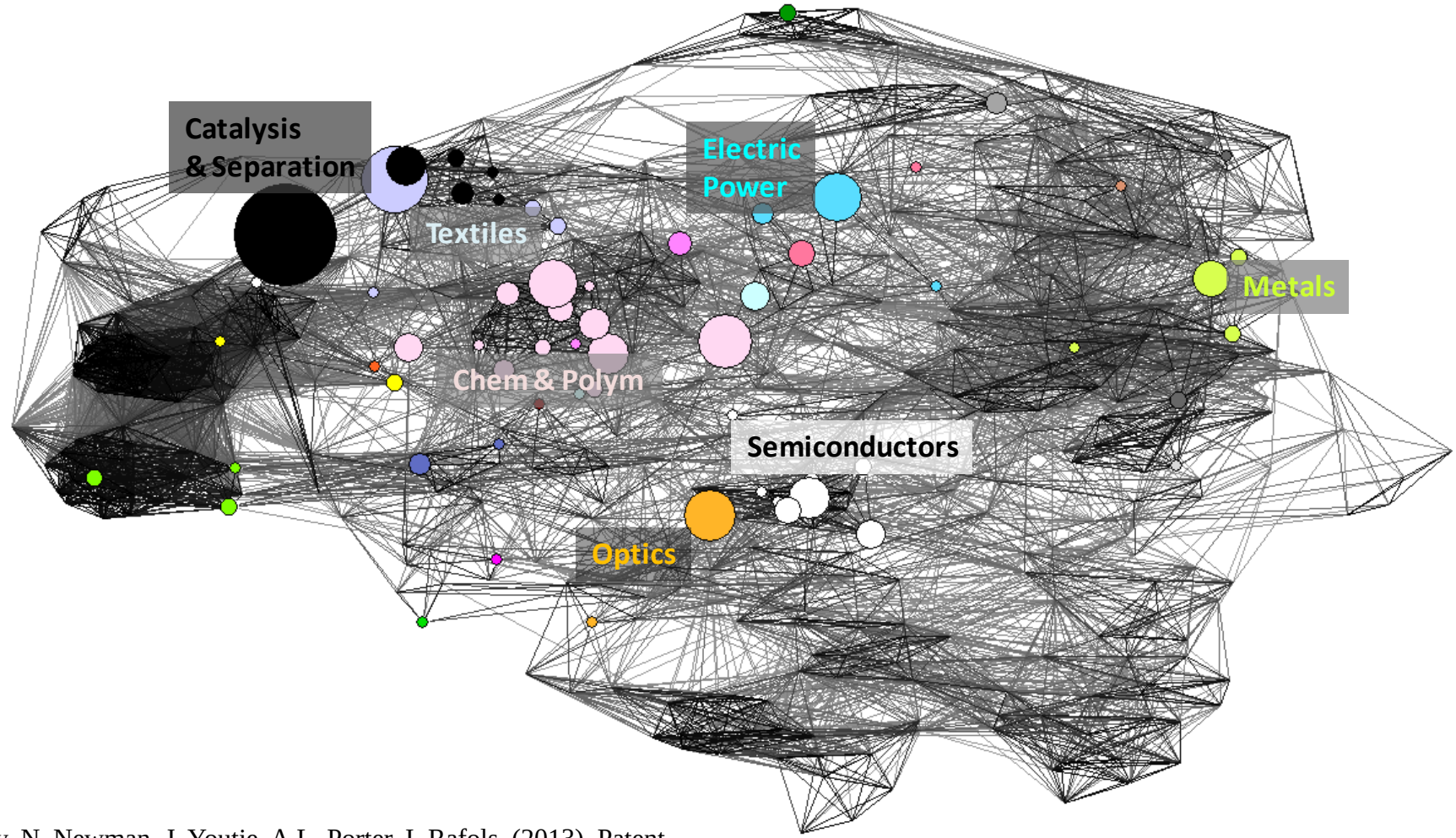
Patent Map has 3 dimensions

PatStat 2000-2006
760,000 pats
IPC classes
3-4 / 7 digit
combined

Cosine Citation Similarity
Pajek (Kamada-Kawai)

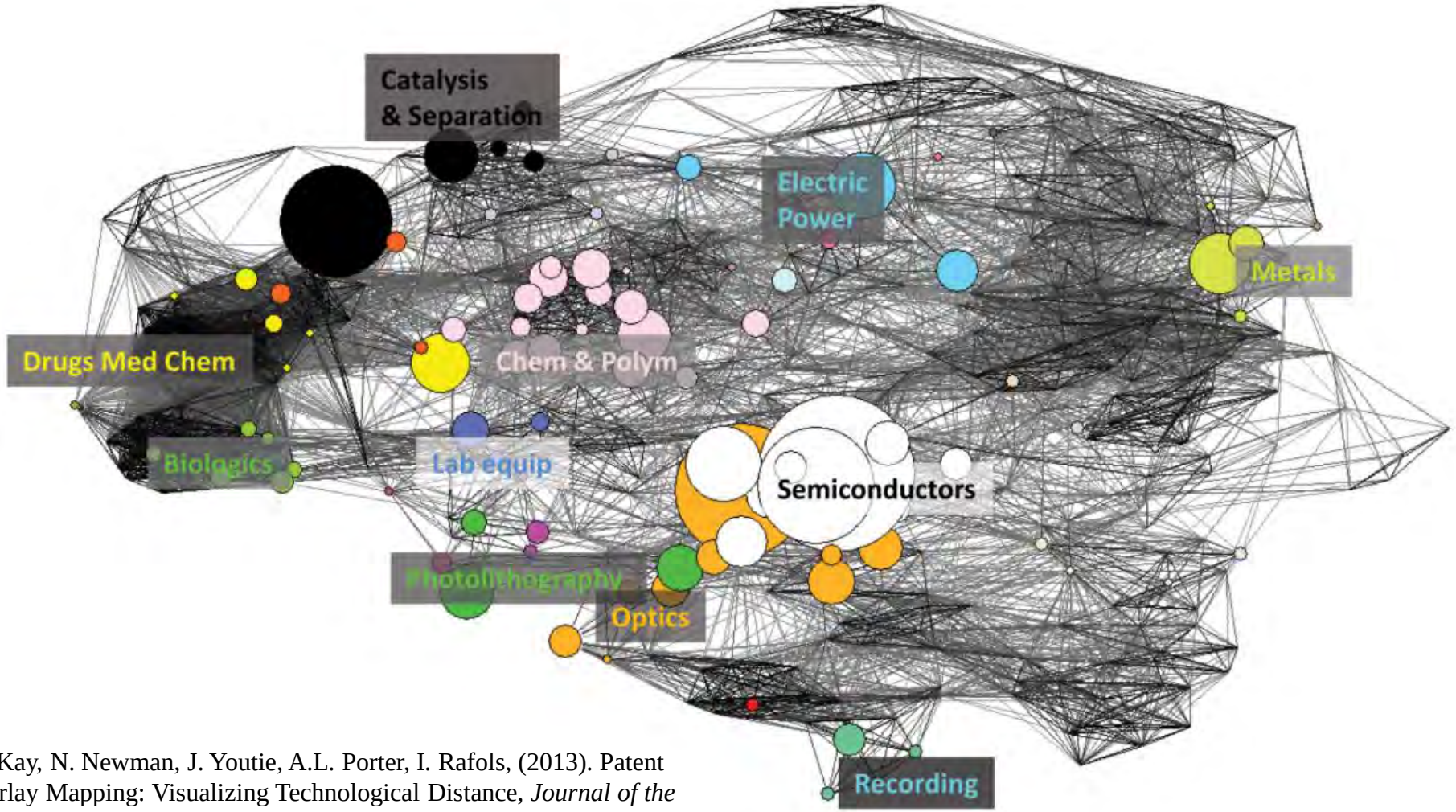


Graphene Overlay Patent Map



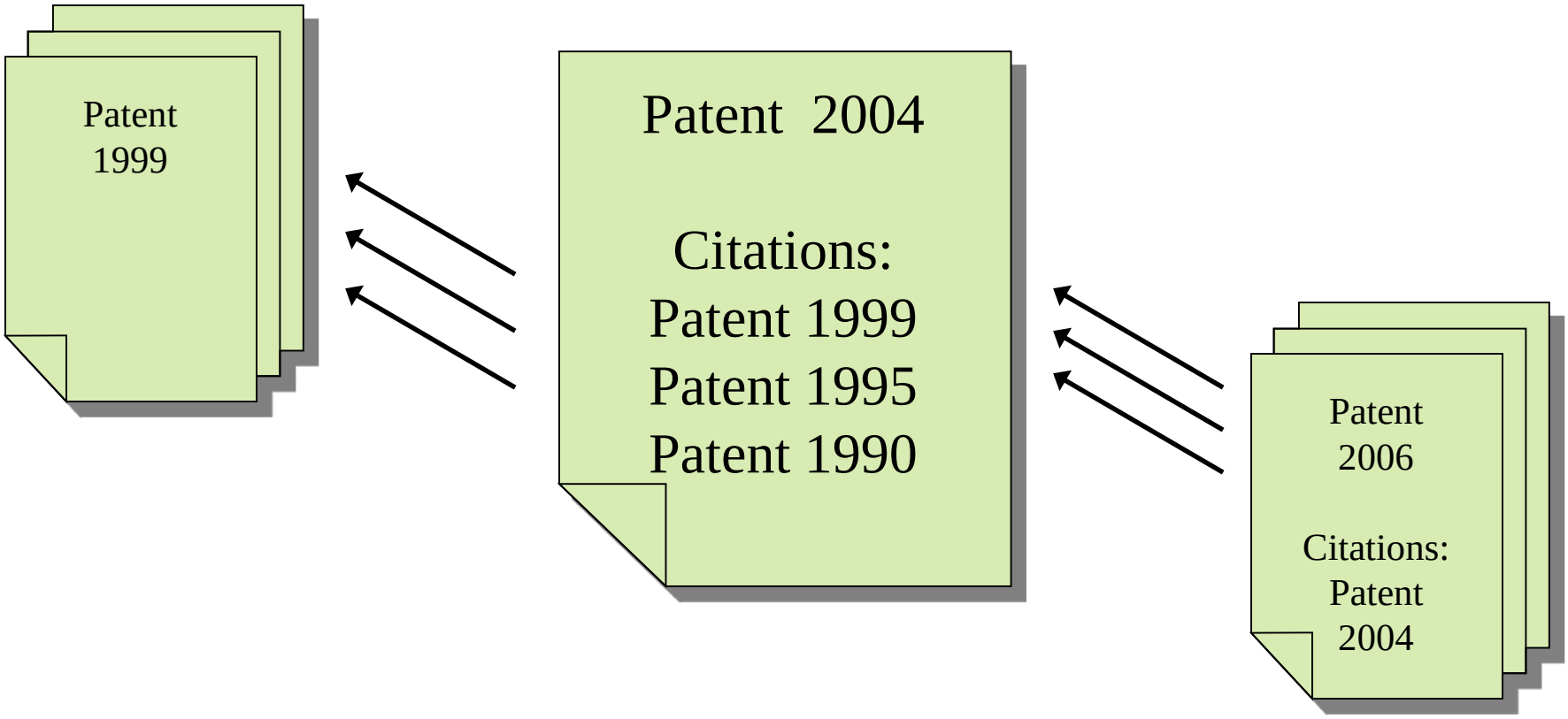
L. Kay, N. Newman, J. Youtie, A.L. Porter, I. Rafols, (2013). Patent Overlay Mapping: Visualizing Technological Distance, *Journal of the American Society for Information Science and Technology*

Samsung Overlay Patent Map



L. Kay, N. Newman, J. Youtie, A.L. Porter, I. Rafols, (2013). Patent Overlay Mapping: Visualizing Technological Distance, *Journal of the American Society for Information Science and Technology*

Citations: Backward and Forward



Patents as Evidence: Pervasiveness; "Generic Function"

- $$Generality \equiv G_i = 1 - \sum_j^{n_i} s_{ij}^2$$
  - S_{ij} = share of patent *i's* forward citations in class *j*
- Forward looking Herfindahl concentration index of the general use of the patented technology
  - As G→1 suggests more general adoption
- Problematical, due to limited lags for new, nanotechnology patents

Herfindahl Example: Nanotechnology Patent Assignees in Georgia, USA

Assignees 2000-9	Patents	Share	Square
Georgia Tech	34	0.48	0.229
Emory University	6	0.08	0.007
Fitel USA Corp	6	0.08	0.007
Nanoventions, Inc	6	0.08	0.007
nGimat	5	0.07	0.005
Selecto, Inc	3	0.04	0.002
ADM Corp.	2	0.03	0.001
Micron Optics, Inc	2	0.03	0.001
Nanolumens	2	0.03	0.001
Furukawa Elect. N. Amer.	1	0.01	0.000
Johnson R&D	1	0.01	0.000
K2 Concepts, Inc	1	0.01	0.000
Nanoscale Materials, Inc	1	0.01	0.000
Technology Resources	1	0.01	0.000
	HHI	**=**	**0.261**

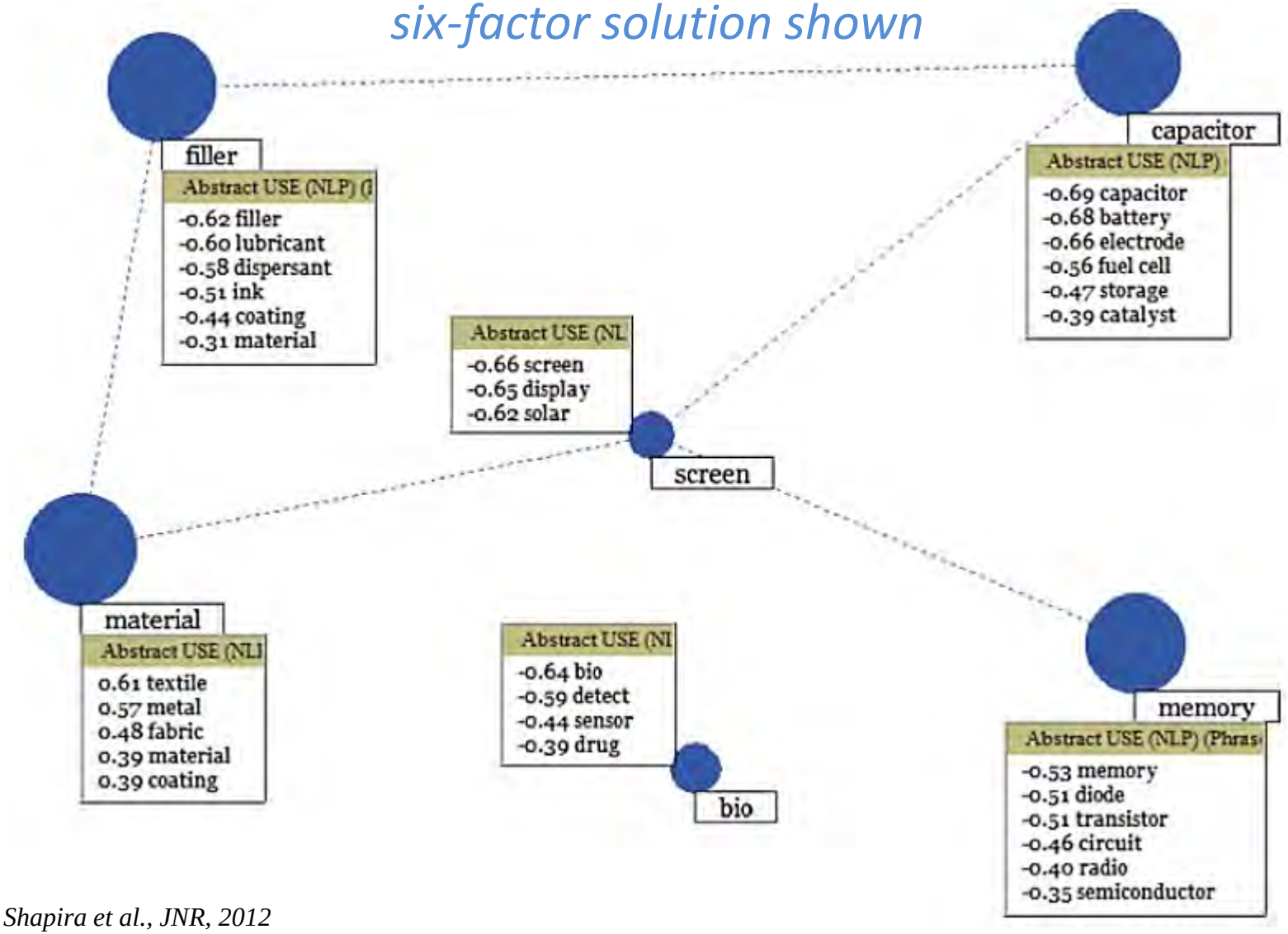
How do we find a pattern in text?

- Co-word Bibliometrics/Co-Occurrence
  - If a two words appear together in the same document, they might be connected.
  - If the same two words appear together in many documents, they are related.

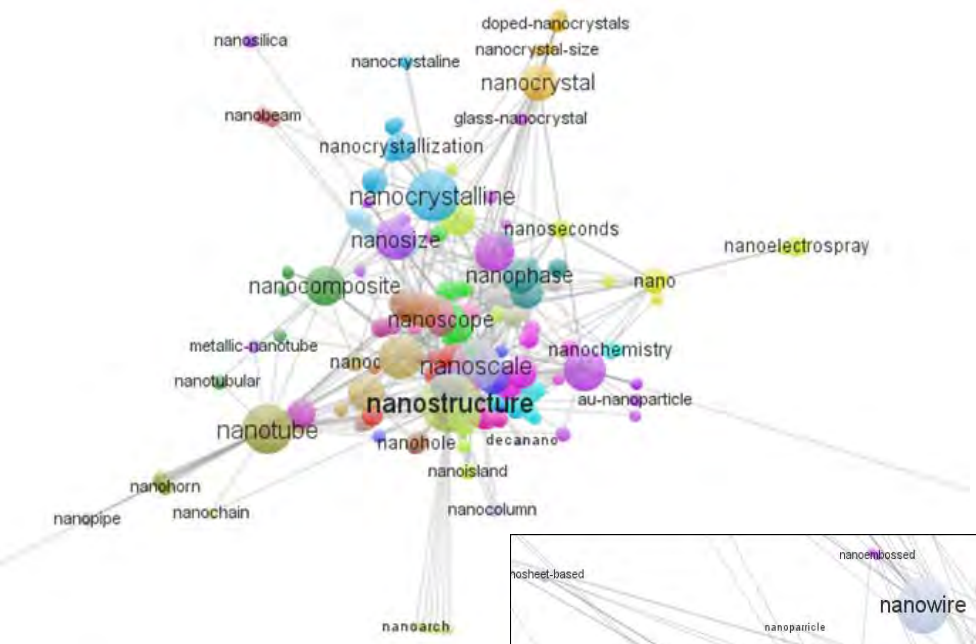
Word 1 ⟺ **Word 2**

Factor map of graphene keywords in patent abstracts

six-factor solution shown

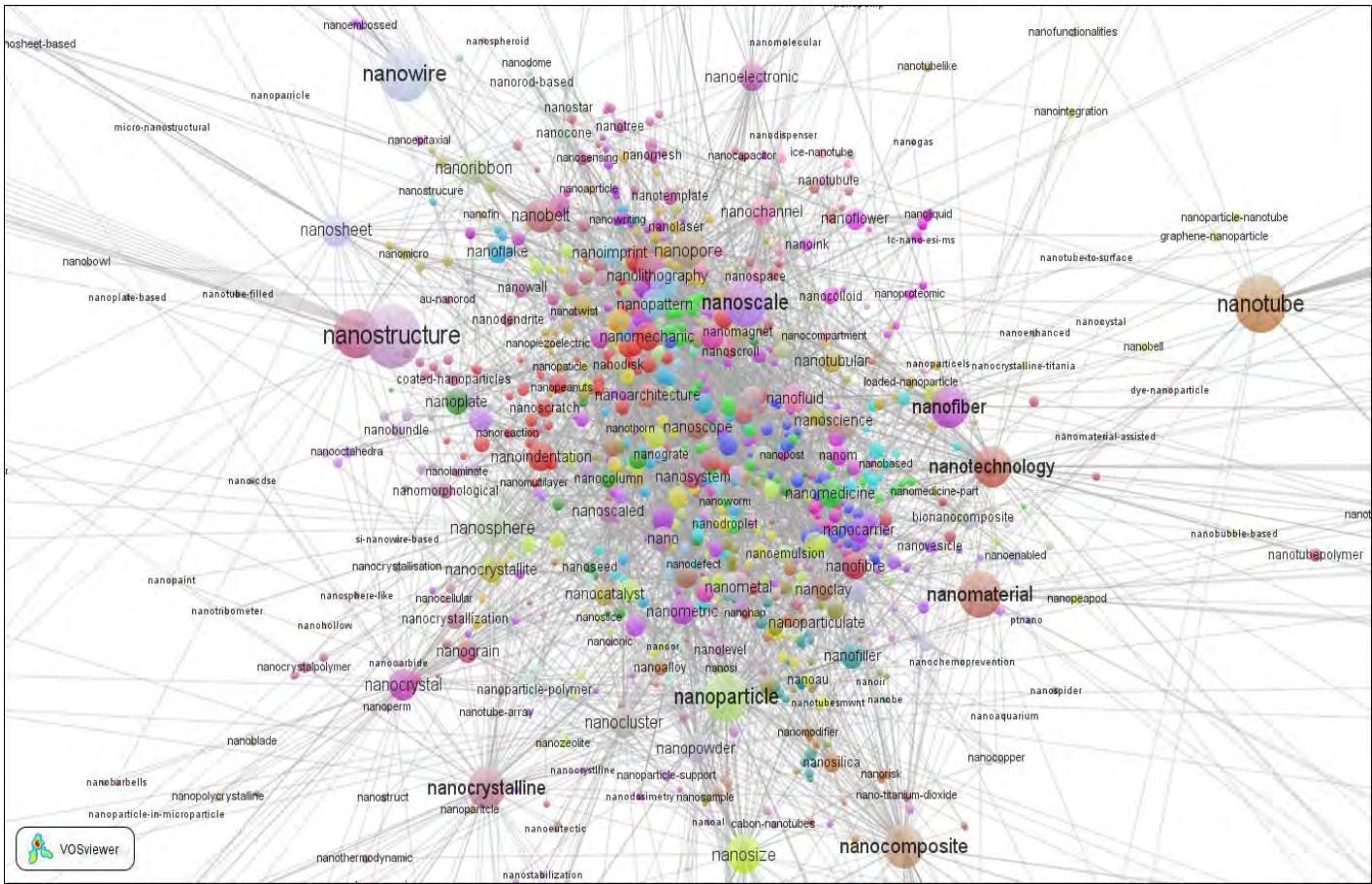


Source: Shapira et al., JNR, 2012



Nano-prefixed co-occurrence maps: 2000 v. 2010

S. Arora, J. Youtie, S. Carley, A.L. Porter, P. Shapira (2014) **Measuring the development of a common scientific lexicon in nanotechnology, JNR**



3. Defining a Technology Domain

- By keyword
  - Center: set of keywords that capture 75% of lead researchers' publication
  - Distinctive domain: graphene in title but not in abstract
  - Complex domain (see next session)
- By citation patterns: core articles – who cites them?
- By ontology: key functions

Defining a Technology Domain

- Domain definition concepts
  - Recall: captures all relevant records
  - Precision: truly relevant records with limited “noise”
  - Noise: irrelevant records
- Strategies
  - High recall can only be attained at the expense of lower precision
  - Optimize between the extremes of high recall and high precision
  - Avoid keywords that produce too much noise

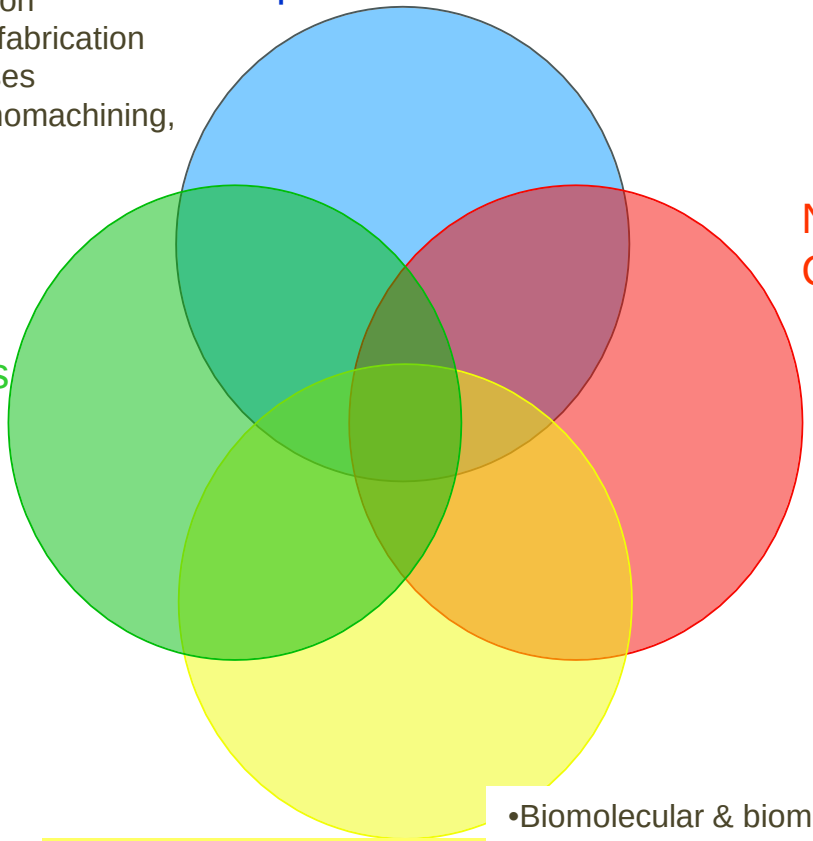
Simple Precision Ratio

Selected Terms	Nano Search	Random Web of Science	Precision
biosensor*	818	29	8.30%
photonic*	841	47	13.0%
fluidic*	603	43	16.6%
self-assembl*	4600	117	5.9%
epitaxial*	997	56	13.1%
epitaxy*	912	52	13.3%
scanning electron microscop*	6638	161	5.7%
transmission electron microscop*	5163	145	6.5%
atom* force microscop*	3445	93	6.3%

Nanotechnology Research Foci & Key Concepts

•Microscopy
- Scanning probe microscopy
- Electron microscopy
•Self assembly; Directed assembly
•Nanomechanics
•Molecular simulation
•Scanning probe writing & fabrication
•Top-down processes
(nano-lithography, laser nanomachining, etc.)

Metrology & Nanoprocesses



Nanostructure Chemistry & Materials

• Nanoscale chemical structures
• Nanocomposites
• Sol-gels; quasi-crystals
• Growth methods
(epitaxy – MBE, CBE,MOCVD)
• 0D – Quantum dots
• 1D – Nano/quantum tubes, rods or fibers; nanopolymers
• 2D – graphite layers
• 3D -; fullerenes; nanocrystals

Nanodevices & Nanoelectronics

•Nanocomputing devices
•Nanotransistors
•NEMS; PEBBLES
•Molecular electronics
•Nanoscale magnetics

Nanomedicine & Nano-biotechnology

•Biomolecular & biomemetic devices
•Biosensors
•Molecular motors
•Biomolecular fabrics
•Engineered enzymes & proteins
•Drug discovery and delivery

Search	Contingency	Terms*
1. Nano*	No	TS=(nano*)
2. Quantum	No	TS=((“quantum dot*” OR “quantum well*” OR “quantum wire*”) NOT nano*)
3. Self-assembly	Yes, MolEnv-I	TS=((“self assembl*” OR “self organiz*” OR “directed assembl*”) AND MolEnv-I)
4. Nano-related	No	TS=(("molecul* motor*" OR "molecul* ruler*" OR "molecul* wir*" OR "molecul* devic*" OR "molecular engineering" OR "molecular electronic*" OR "single molecul*" OR fullerene* OR buckyball OR buckminsterfullerene OR C60 OR "C-60" methanofullerene OR metallofullerene OR SWCNT OR MWCNT OR "coulomb blockad*" OR bionano* OR "langmuir-blodgett" OR Coulombstaircase* OR "PDMS stamp*" OR graphene OR "dye-sensitized solar cell" OR DSSC OR ferrofluid* OR "core-shell") NOT nano*)
5. Microscopy and spectroscopy	Yes, MolEnv-R	TS=((((TEM or STM or EDX or AFM or HRTEM or SEM or EELS or SERS or MFM) OR "atom* force microscop*" OR "tunnel* microscop*" OR "scanning probe microscop*" OR "transmission electron microscop*" OR "scanning electron microscop*" OR "energy dispersive X-ray" OR "xray photoelectron*" OR "x-ray photoelectron" OR "electron energy loss spectroscop*" OR "enhanced raman-scattering" OR "surface enhanced raman scattering" OR "single molecule microscopy" OR "focused ion beam" OR "ellipsometry" OR "magnetic force microscopy") AND MolEnv-R) NOT nano*)
6. Nano-pertinent	Yes, MolEnv-I	TS=(((NEMS OR Quasicrystal* OR “quasi-crystal*” OR “quantum size effect” OR “quantum device”) AND MoleEnv-I) NOT nano*)
7. Nano-pertinent	Yes, MolEnv-R	TS=(((biosensor* OR NEMS OR (“sol gel*” OR solgel*) OR dendrimer* OR CNT OR “soft lithograph*” OR “electron beam lithography” OR “e-beam lithography” OR “molecular simul*” OR “molecular machin*” OR “molecular imprinting” OR “quantum effect*” OR “surface energy” OR “molecular sieve*” OR “mesoporous material*” OR “mesoporous silica” OR “porous silicon” OR “zeta potential” OR “epitax*”) AND MolEnv-R) NOT nano*)
8. Nano journals	No	SO=((Fullerene* OR IEEE Transactions on Nano* OR Journal of Nano* OR Nano* OR Materials Science Engineering C* OR ACS Nano OR Current Nanoscience OR Digest Journal of Nanomaterials and Biostructures OR IEE Proceedings Nanobiotechnology OR IET Nanobiotechnology OR International Journal of Nanomedicine OR International Journal of Nanotechnology OR Journal of Biomedical Nanotechnology OR Journal of Computational and Theoretical Nanoscience OR Journal of Experimental Nanoscience OR Nature Nanotechnology OR Photonics and Nanostructures* OR Wiley Interdisciplinary Reviews Nano*) NOT nano*)

* Monolayer, molecul*, polymer, and others

Records containing these terms are removed from the nano* dataset	Exclude any nano* records containing only one of these terms and no other nano terms
plankton*	nanometer*
n*plankton	nano-metre
m*plankton	nano-meter
b*plankton	nano-metre
p*plankton	nanosecond*
z*plankton	nano-second
nanoflagel*	nanomolar*
nanoalga*	nano-molar
nanoprotist*	nanomole(s)
nanofauna*	nanogram*
nano*aryote*	nano-gram
nanoheterotroph*	nanoliter*
nanophtalm*	nanolitre*
nanomeli*	nano-liter
nanophyto*	nano-litre*
nanobacteri*	
⁂ ~270 organism names beginning with nano*	
nano2	
nano3	
nanos	
nanog	
nanor	
nanoa	
nano-	
nanog-	
nanoa-	
nanor-	
nanosatellite*	

Learning Objectives

- Link bibliometric methods to research questions
- Provide introduction into practical bibliometric analytical techniques
- Examine how to define emerging technology in bibliometric terms