Tech Mining with TDA (Thomson Data Analyzer)

Alan Porter

Director, R&D

Search Technology

&

Technology Policy & Assessment Center

Georgia Tech

aporter@searchtech.com

ISTIC, December, 2010

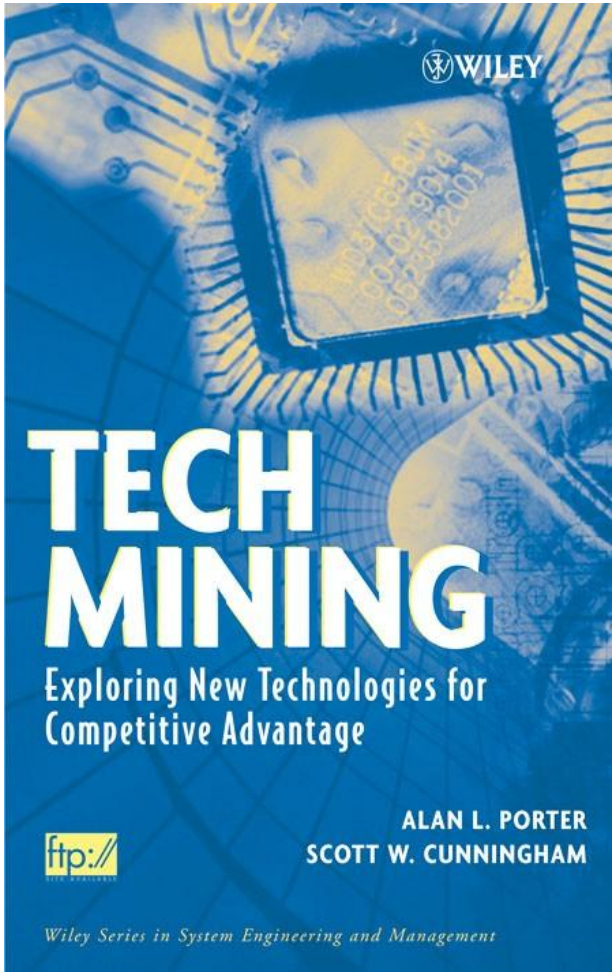
Topics

1. Tech Mining – to generate "innovation indicators" from Science, Technology &Innovation **("ST&I")** information resources
   - Indicators of interdisciplinarity
   - Science overlay maps to locate research activity
2. Illustrating Web of Science analyses:
   - Research assessments for US NSF
   - CAS nano research profiling
     [thanks to Ruimin Pei, CAS]
3. Illustrating Patent analyses
   - Biomaterials Technology Opportunities Analysis
   - Patent Life Cycle Analysis
4. TDA -- Dye Sensitized Solar Cells
   [thanks to Ying Guo, BIT]

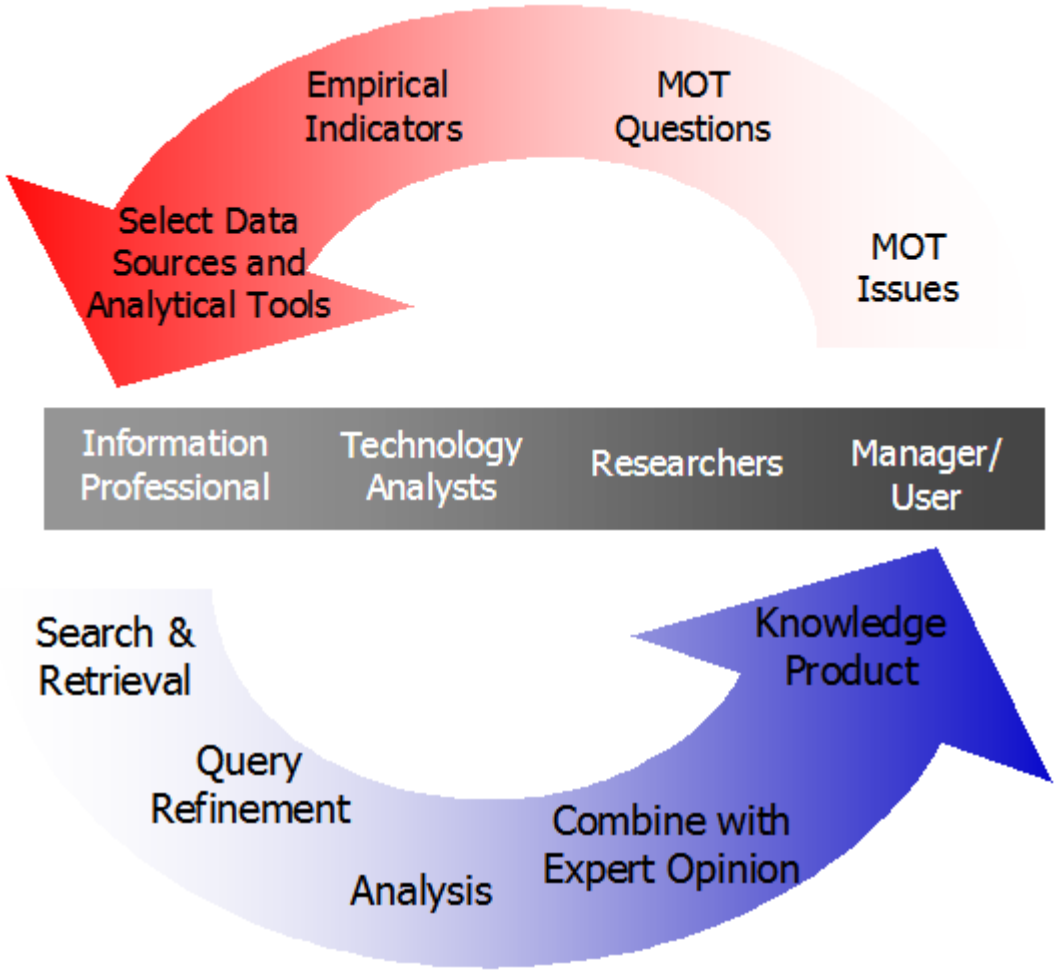
How do you extract effective intelligence from all that Science, Technology & Innovation ("ST&I")information?

Tech Mining

Alan L. Porter and Scott W. Cunningham
John Wiley & Sons Inc., 2005



The Tech Mining Process



How to do Tech Mining: 8-steps

1. Spell out the questions and how to answer them
2. Get suitable data
3. Search (iterate)
4. Import into text mining software (e.g., Thomson Data Analyzer)
5. Clean the data
6. Analyze & interpret
7. Represent the information well – **communicate!**
8. Standardize and semi-automate where possible

13 MOT Issues

- R&D Portfolio Mgt
- R&D Project Initiation
- Engr Project Initiation
- New Product Development
- Strategic Planning
- Track/forecast emerging or breakthrough technologies
- etc.

39 MOT Questions

What?

- What's hot?
- Fit into tech landscape?
- New frontiers at fringe?
- Drivers?
- Competing technologies?
- Likely development paths?

Who?

- Who are available experts?
- Which universities or labs lead?

~200 Innovation Indicators

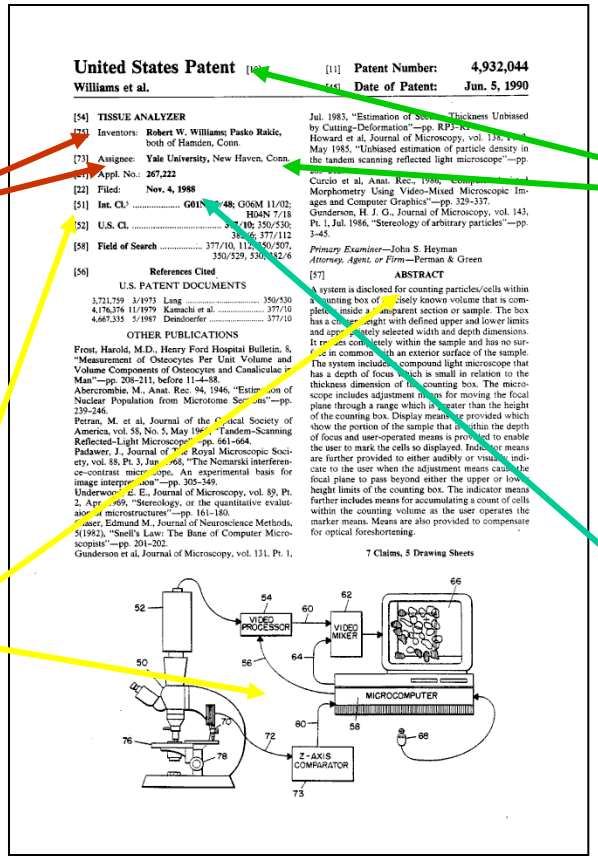
- Mapping of topic clusters within the technology
- 3-D trend charts for topic clusters
- Ratio of conference to journal papers (benchmarked)
- Scorecard rate-of-change metrics for topic clusters
- Time slices to show evolution of topical emphases
- Topic growth modeling (S-curve) fit & extrapolation
- Profile table of main players
- Pie chart: Company vs. Academic vs. Government publishing
- Spreading (or constricting) # of players by topic

MOT Issues, Questions, and Indicators

What kind of questions can we answer by Tech Mining?

Who?

Where?



United States Patent [19]
Williams et al.

[11] Patent Number: 4,932,044
[45] Date of Patent: Jun. 5, 1990

[54] TISSUE ANALYZER

[75] Inventors: Robert W. Williams; Pasko Rakic, both of Hamden, Conn.

[73] Assignee: Yale University, New Haven, Conn.

[21] Appl. No.: 267,222

[22] Filed: Nov. 4, 1988

[51] Int. Cl.[5] G01N 7/48; G06M 11/02; H04N 7/18

[52] U.S. Cl. 377/10; 350/530; 377/112

[58] Field of Search 377/10, 112; 350/507, 350/529, 530; 382/6

[56] References Cited

U.S. PATENT DOCUMENTS

3,721,759 3/1973 Lang 350/530
4,176,376 11/1979 Kamachi et al. 377/10
4,667,335 5/1987 Deindoerfer 377/10

OTHER PUBLICATIONS

Frost, Harold, M.D., Henry Ford Hospital Bulletin, 8, "Measurement of Osteocytes Per Unit Volume and Volume Components of Osteocytes and Canaliculae in Man"—pp. 208–211, before 11-4-88.

Abercrombie, M., Anat. Rec. 94, 1946, "Estimation of Nuclear Population from Microtome Sections"—pp. 239–246.

Petran, M. et al, Journal of the Optical Society of America, vol. 58, No. 5, May 1968, "Tandem-Scanning Reflected-Light Microscope"—pp. 661–664.

Padawer, J., Journal of the Royal Microscopic Society, vol. 88, Pt. 3, Jun. 1968, "The Nomarski interference-contrast microscope, An experimental basis for image interpretation"—pp. 305–349.

Underwood, E. E., Journal of Microscopy, vol. 89, Pt. 2, Apr. 1969, "Stereology, or the quantitative evaluation of microstructures"—pp. 161–180.

Baker, Edmund M., Journal of Neuroscience Methods, 5(1982), "Snell's Law: The Bane of Computer Microscopists"—pp. 201–202.

Gunderson et al, Journal of Microscopy, vol. 131, Pt. 1, Jul. 1983, "Estimation of Section Thickness Unbiased by Cutting-Deformation"—pp. RP3–R4.

Howard et al, Journal of Microscopy, vol. 138, Pt. 2, May 1985, "Unbiased estimation of particle density in the tandem scanning reflected light microscope"—pp. 203–213.

Curcio et al, Anat. Rec., 1986, "Computer Morphometry Using Video-Mixed Microscopic Images and Computer Graphics"—pp. 329–337.

Gunderson, H. J. G., Journal of Microscopy, vol. 143, Pt. 1, Jul. 1986, "Stereology of arbitrary particles"—pp. 3–45.

Primary Examiner—John S. Heyman
Attorney, Agent, or Firm—Perman & Green

[57] ABSTRACT

A system is disclosed for counting particles/cells within a counting box of precisely known volume that is completely inside a transparent section or sample. The box has a chosen height with defined upper and lower limits and appropriately selected width and depth dimensions. It resides completely within the sample and has no surface in common with an exterior surface of the sample. The system includes a compound light microscope that has a depth of focus which is small in relation to the thickness dimension of the counting box. The microscope includes adjustment means for moving the focal plane through a range which is greater than the height of the counting box. Display means are provided which show the portion of the sample that is within the depth of focus and user-operated means is provided to enable the user to mark the cells so displayed. Indicator means are further provided to either audibly or visually indicate to the user when the adjustment means causes the focal plane to pass beyond either the upper or lower height limits of the counting box. The indicator means further includes means for accumulating a count of cells within the counting volume as the user operates the marker means. Means are also provided to compensate for optical foreshortening.

7 Claims, 5 Drawing Sheets

What?

When?

How? & Why? – Need human analyst to interpret the data

Some Web of Science based Innovation Indicators

- ~200+ Innovation Indicators (“Tech Mining”)
- % of R&D publications by industry vs. academia vs. government
- Research network maps
- Interdisciplinarity indicators
- Science overlay maps

Some Patent-based Innovation Indicators

- Technological Maturity: model R&D activity on a growth curve
- Competitive Entry: Slope of the trend in new companies initiating patenting recently
- Competitive Exit: % of top assignees that have left the domain
- Diffusion: Change in number of new IPC codes in recent years
- Patent Quality (next slide)

Indicator System for Core Patent Documents Evaluation

Xian Zhang, Shu Fang, Chuan Tang, Guo-hua Xiao, Zheng-yin Hu, and Li-dan Gao,
Chengdu Library of the Chinese Academy of Sciences

Technology	Technology Scope (TS)	Number of IPC (A_{11})
		Number of Claims (CLN) (A_{12})
	Technology Impact (TI)	Technology Dependence (A_{21})
		Technology Cycle Time (TCT) (A_{22})
	Science Strength	Science Linkage (SL) (A_{31})
		Science Cycle Time (SCT) (A_{32})
		Science Impact (SI) (A_{33})
	Standardization Activity	Standardization Impact (A_{41})
		Standardization Scope (A_{42})
Market	Patent Family	Patent Family Size (B_{11})
		Share of the Triad (US, EP and JP) Patents in a Family (B_{12})
	Input Strength	Human Resource Input (B_{21})
		Collaboration Intensity (B_{22})
	Technology Commercialization	Self-Commercializing or Licensing (B_{31})
		Patent Impawning or Collateral Loan (B_{32})
		Patent Assignment (B_{33})
Legal	Number of years of a patent is renewed	Patent Validity Years (C_{11})
		Extended (C_{12})
		Survived from Patent Opposition Claim(s) (C_{13})
		Survived from Patent Annulment Claim(s) (C_{14})
	Current Legal Status	Granted (C_{21})
		Stability of Legal Status (C_{22})

National Academies Keck *Futures* Initiative [*Facilitating Interdisciplinary Research*]

*www.**keckfutures**.org*

Interdisciplinary research (IDR) is a mode of research by teams or individuals that *integrates*

- **perspectives/concepts/theories and/or**
- **tools/techniques and/or**
- **information/data**

from two or more bodies of specialized knowledge or research practice.

Examples of *bodies of specialized knowledge or research practice* include: low temperature physics, molecular biology, developmental psychology,...

The 221 Web of Science "Subject Categories" [science & social science], are a good fit for this.

Science Overlay Mapping

- Use the science map as a base
- Locate research activity in Subject Categories (SCs):
  - Node size ~ #
  - Can plot publication SCs, cited SCs, &/or citing SCs
- Show cross-disciplinary interchanges
- Show change over time

Based on Leydesdorff and Rafols (2009) and Rafols and Meyer (2010)

Macro- and Meta-Disciplines

1. SC relatedness based on one year's data – WOS Journal X Journal **cross-citation** matrix
2. Loet Leydesdorff transforms to SC X SC matrix.
3. **Macro-Disciplines** come from Ismael Rafols' factor analyses:
   - **175 SC science base map (14 factors)**
   - **221 SC science + social science base map (18 factors)**
4. **Meta-Disciplines** – further group to 4 overarching categories

Topics

1. Tech Mining
2. **Illustrating Web of Science analyses:**
   - **Research assessments for US NSF**
   - **CAS nano research profiling [thanks to Ruimin Pei, CAS]**
3. Illustrating Patent analyses
4. TDA -- Dye Sensitized Solar Cells

NSF Research & Research Assessments

- **Measuring & Mapping Research Knowledge Integration**
  - Science overlay mapping
- **Research on Learning & Education Program**
  - Can proposal references describe research communities engaged (and aid review processes)?
  - Track interplay of Cognitive Science & Education research
- **Research Cooperation Networks (RCN) Program**
  - Can we see network enrichment, Before to After?
- **Human & Social Dynamics (HSD) Program**
  - Influence beyond the Social Sciences

HSD Assessment

- **Locate HSD research among the disciplines**
- **Explore HSD research knowledge diffusion**
- Key on Year 2004 HSD awards (28 Projects with papers in WOS)
- Study research outputs (**publications**)
  - **Before vs. After**
  - **Deriving from the awards**
- Study **citations** too

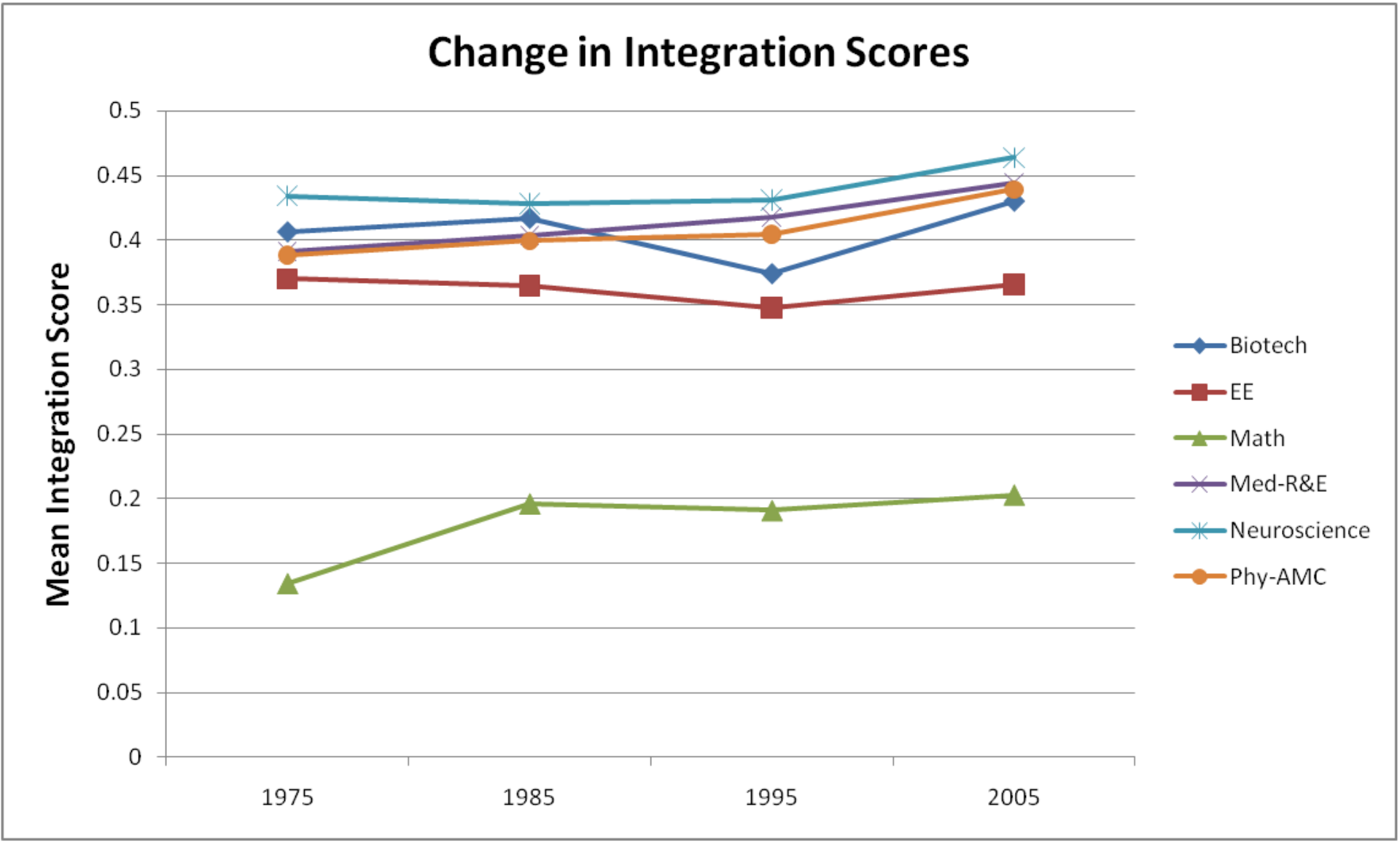
"Tech Mining"

- Publication & citation data gathering
- Data cleaning in [*Thomson Data Analyzer*] software
- Analyses and visualizations, also drawing upon
  - MS Excel
  - Pajek
  - MS Powerpoint
- Here are selected examples from several research assessments

HSD-derived Publication Characteristics

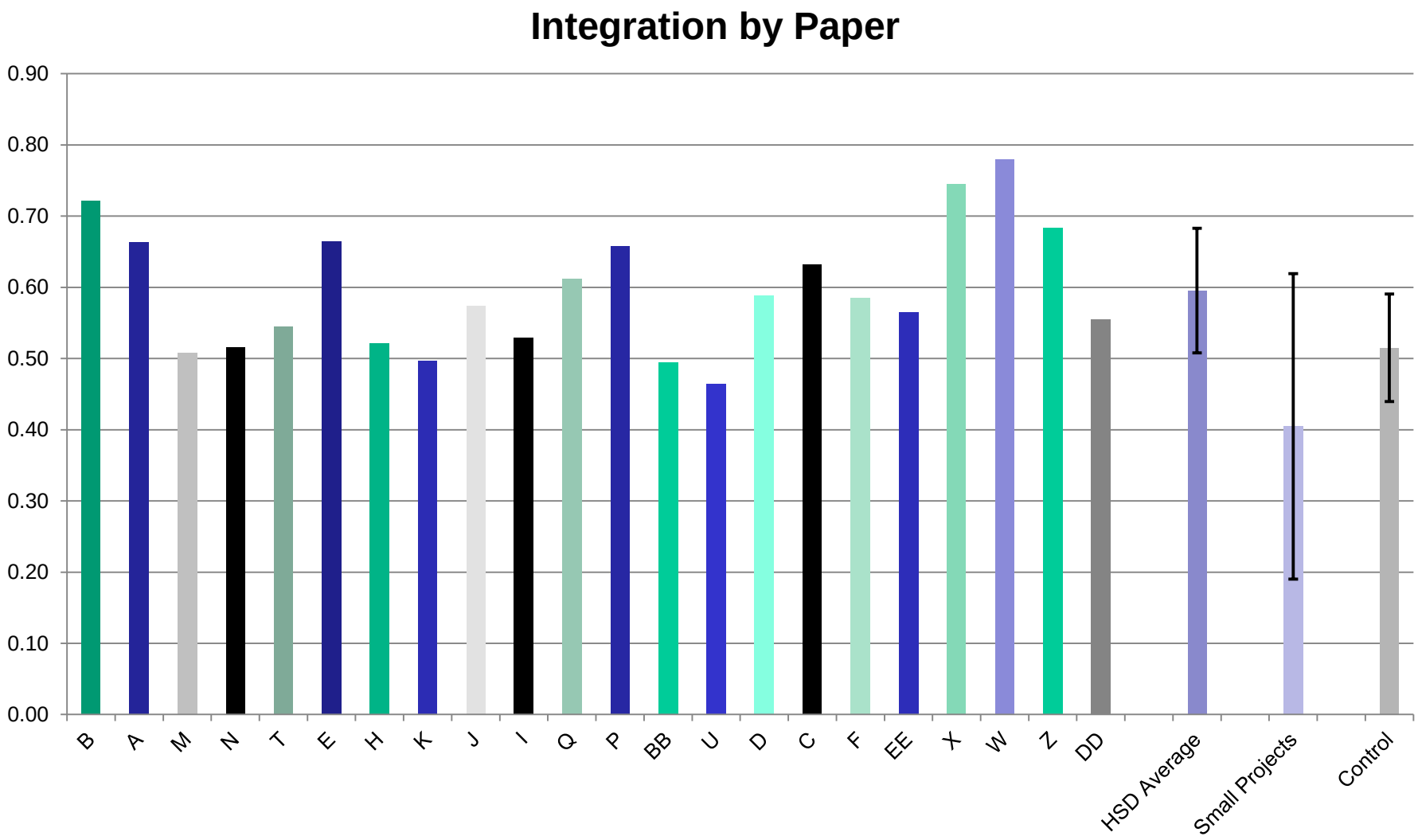
	Overall	Project B	Project A	Project H
			Project	
Authors/Paper	2.79	2.42	3.00	2.27
Author Affiliation	2.26	1.92	2.69	2.09
Cited Reference Count	42.44	38.63	31.00	54.73
Number of Countries	1.44	1.25	2.19	1.55
Integration by Article	0.58	0.72	0.66	0.52
Journal Impact Factor	3.89	2.36	3.14	4.24
Times Cited	7.48	6.46	4.31	3.27

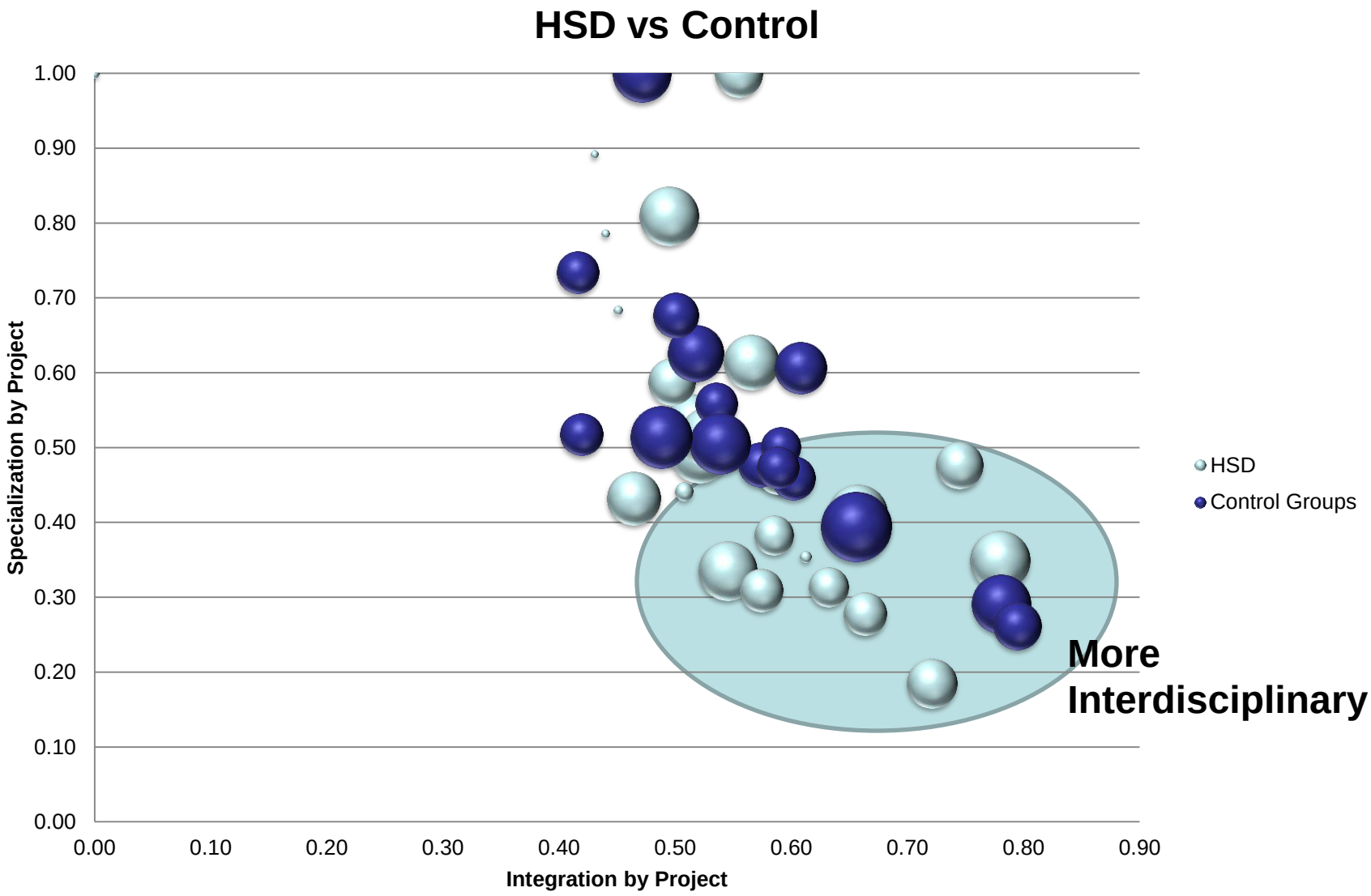
Benchmarking Integration Scores



Porter and Rafols (2009)

HSD Integration Scores



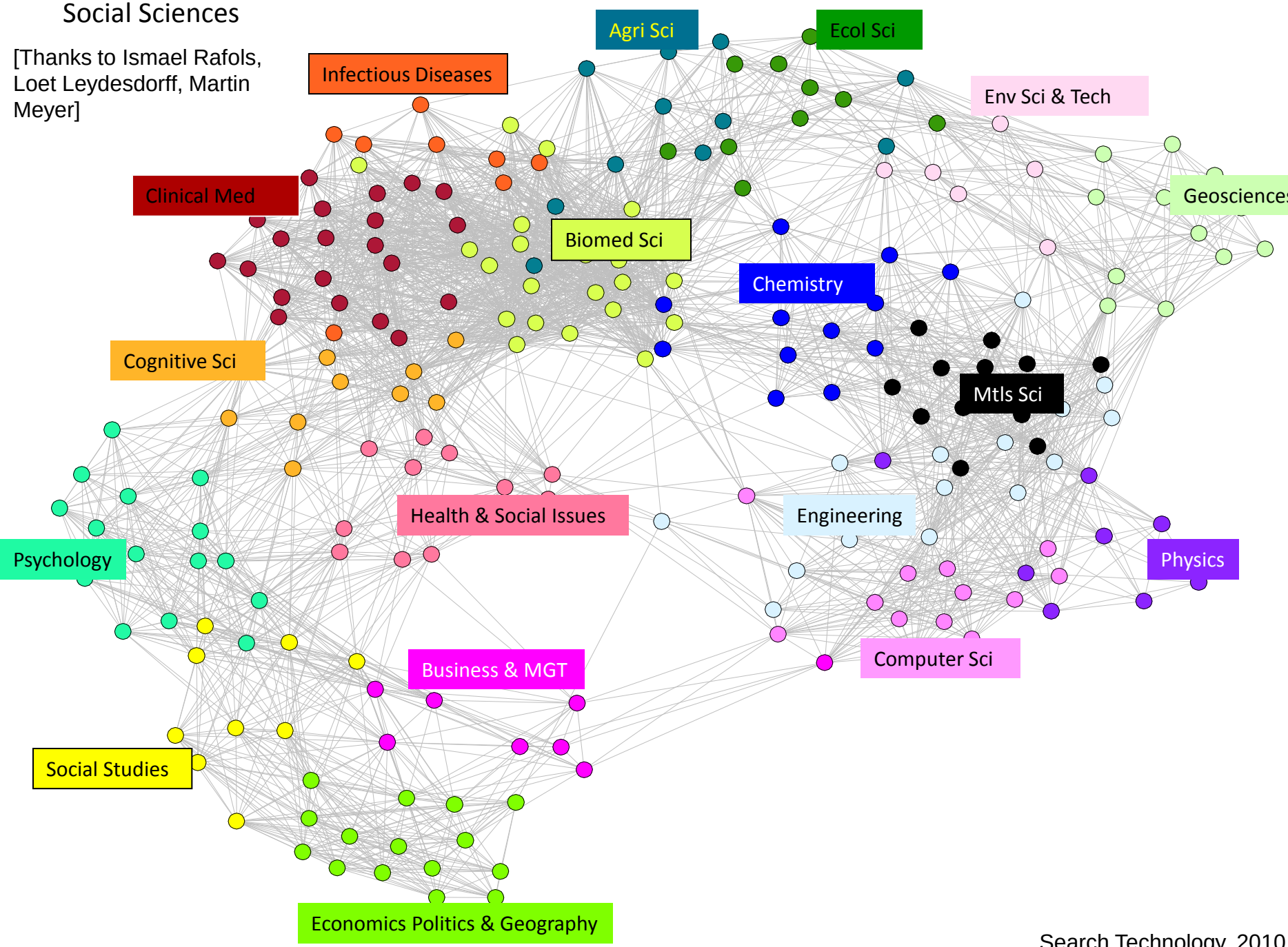

HSD vs Control
Specialization by Project
1.00
0.90
0.80
0.70
0.60
0.50
0.40
0.30
0.20
0.10
0.00
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
Integration by Project
HSD
Control Groups
More
Interdisciplinary

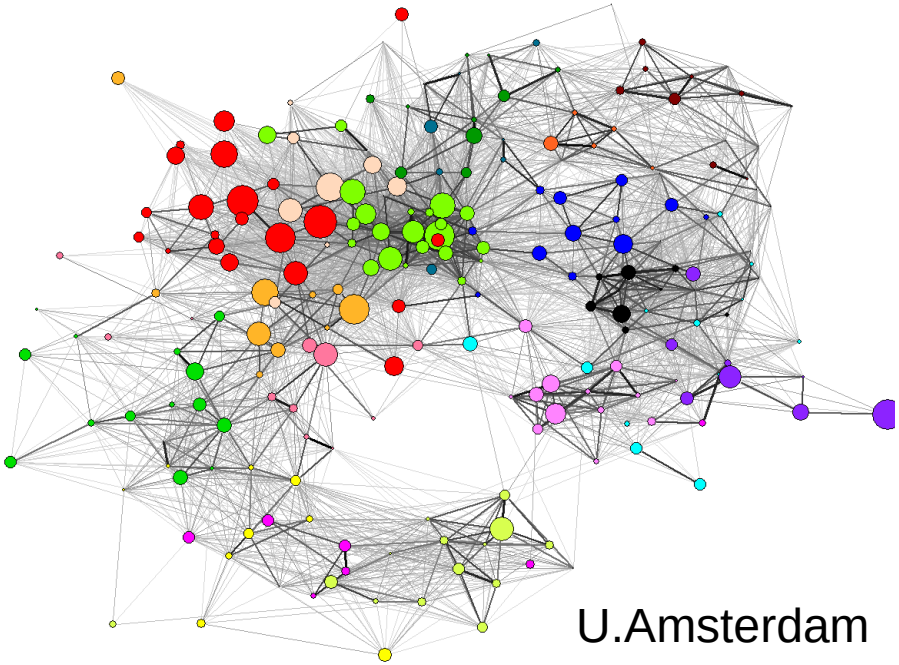
Maps for Research Assessment

- Global – science overlay maps
- Local – social network maps

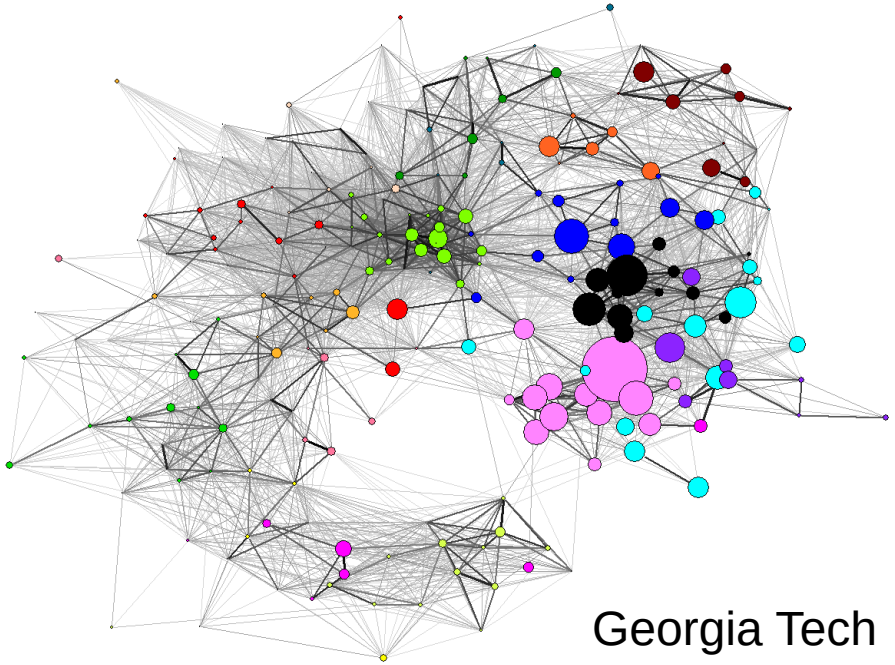
221 SC Base Map – Sciences + Social Sciences

[Thanks to Ismael Rafols, Loet Leydesdorff, Martin Meyer]

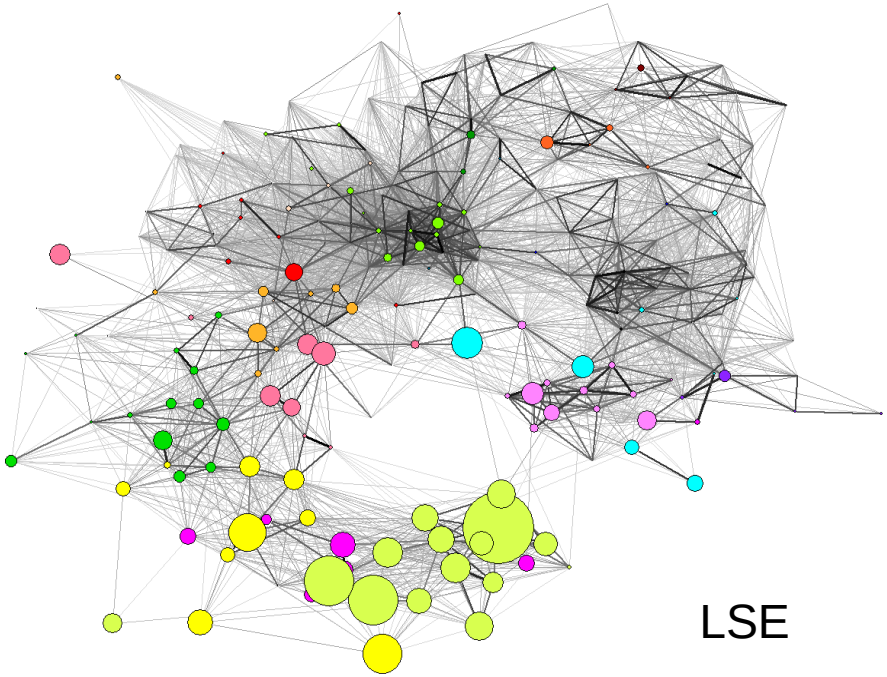




U.Amsterdam



Georgia Tech



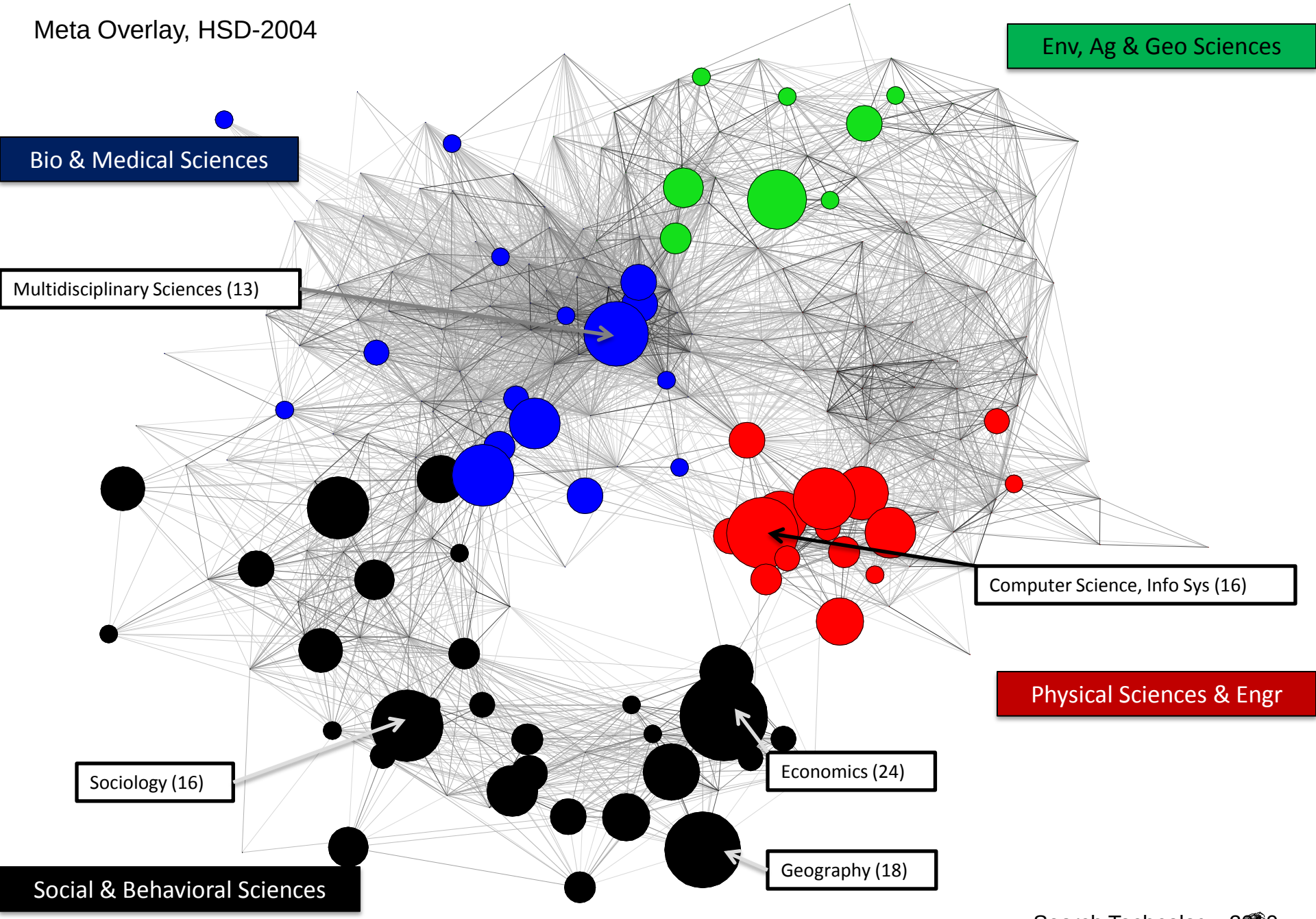
LSE

Ismael Rafols, 2010

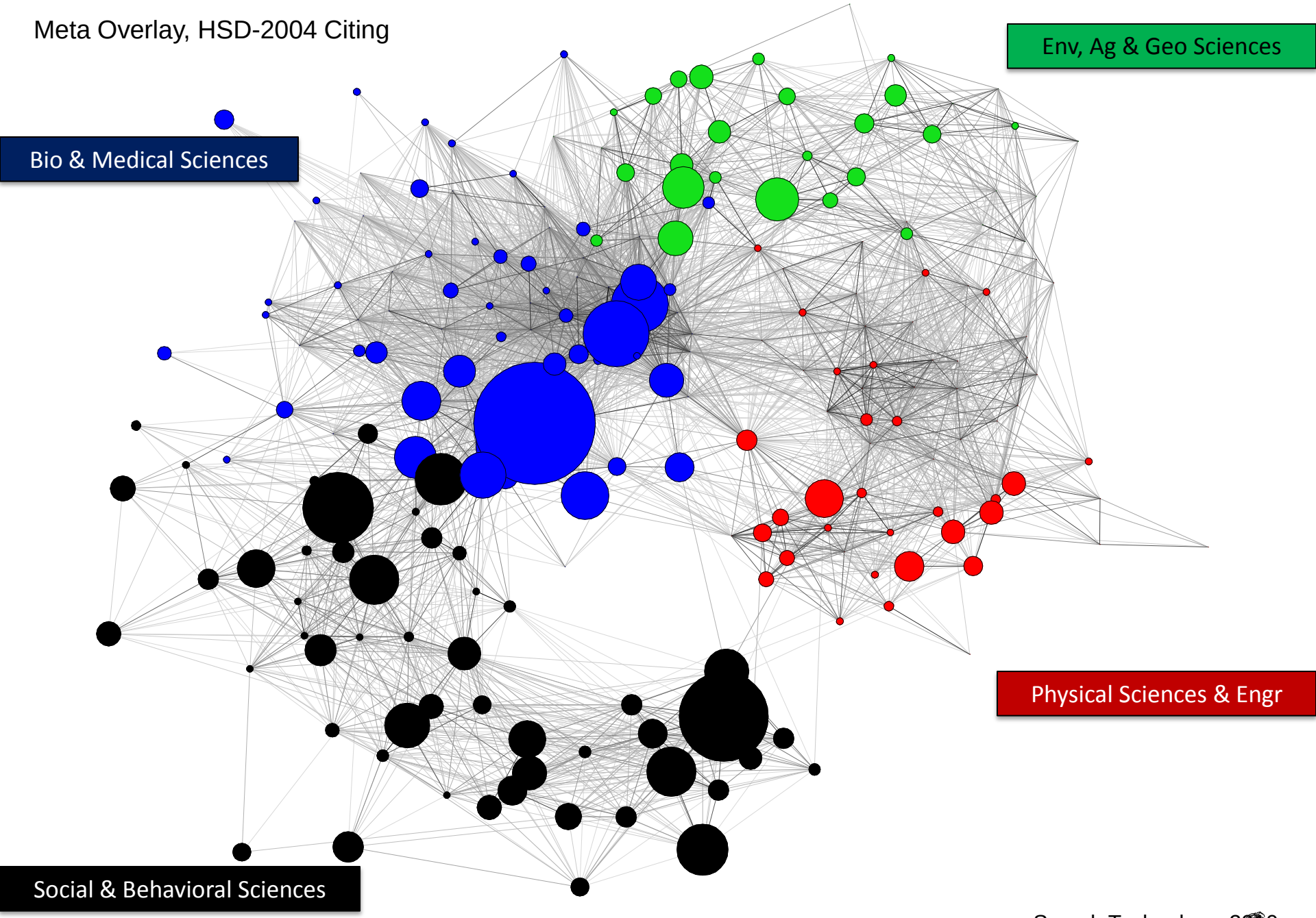
Also, a set of **company** maps @ idr.gatech.edu

HSD-derived Publications

Subject Category	#
Economics	24
Geography	18
Computer Science, Information Systems	16
Sociology	16
Multidisciplinary Sciences	13
Mathematics, Interdisciplinary Applications	12
Psychology, Experimental	12
Public, Environmental & Occupational Health	12
Environmental Sciences	11
Environmental Studies	10
Computer Science, Artificial Intelligence	9
Engineering, Electrical & Electronic	9
Social Sciences, Mathematical Methods	9
Neurosciences	8
Political Science	8
Telecommunications	8
Area Studies	7
Psychology	7
Robotics	7
Linguistics	6
Psychology, Social	6


Meta Overlay, HSD-2004
Env, Ag & Geo Sciences
Bio & Medical Sciences
Multidisciplinary Sciences (13)
Computer Science, Info Sys (16)
Physical Sciences & Engr
Sociology (16)
Economics (24)
Geography (18)
Social & Behavioral Sciences

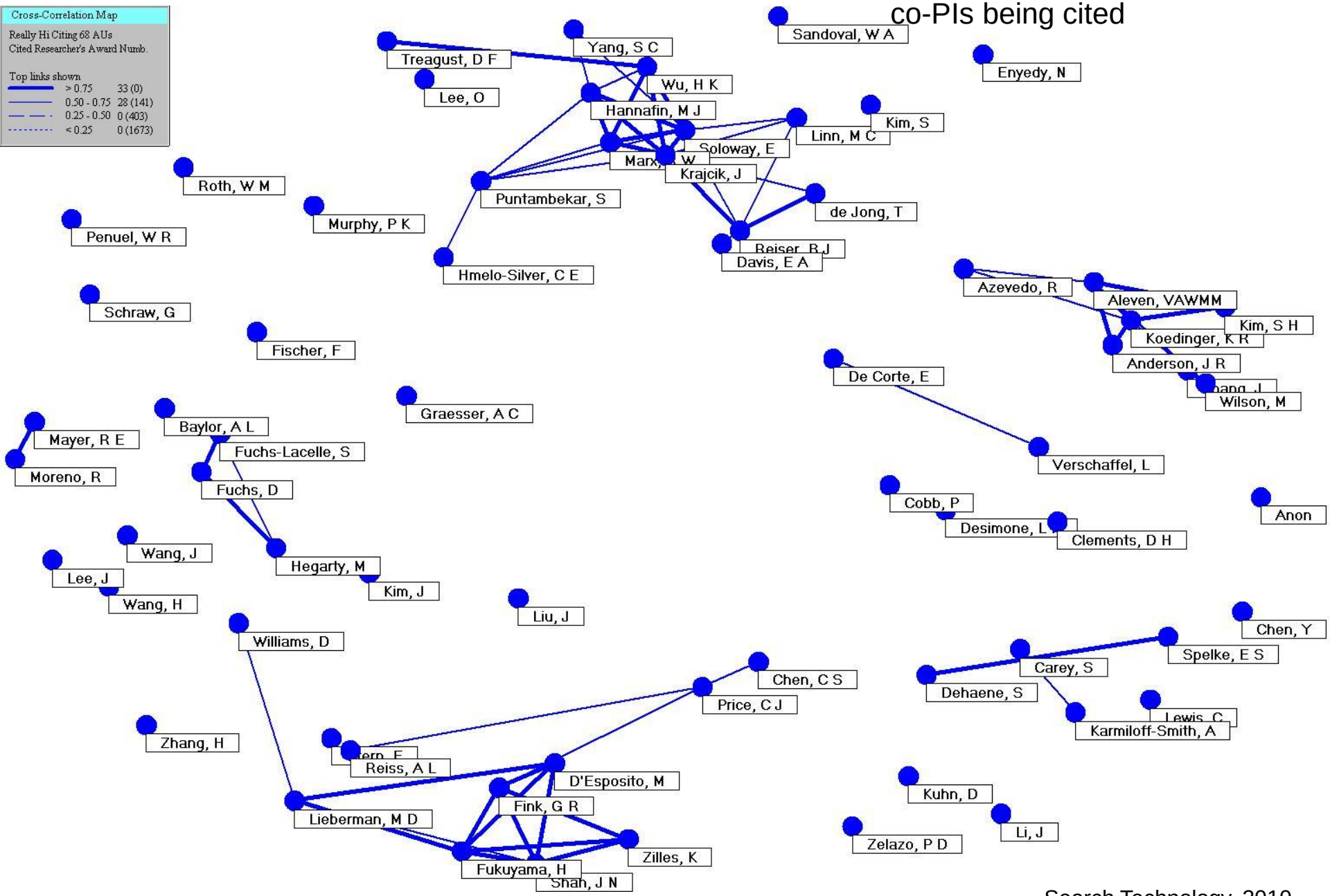
Meta Overlay, HSD-2004 Citing



Thomson Data Analyzer Map Principles

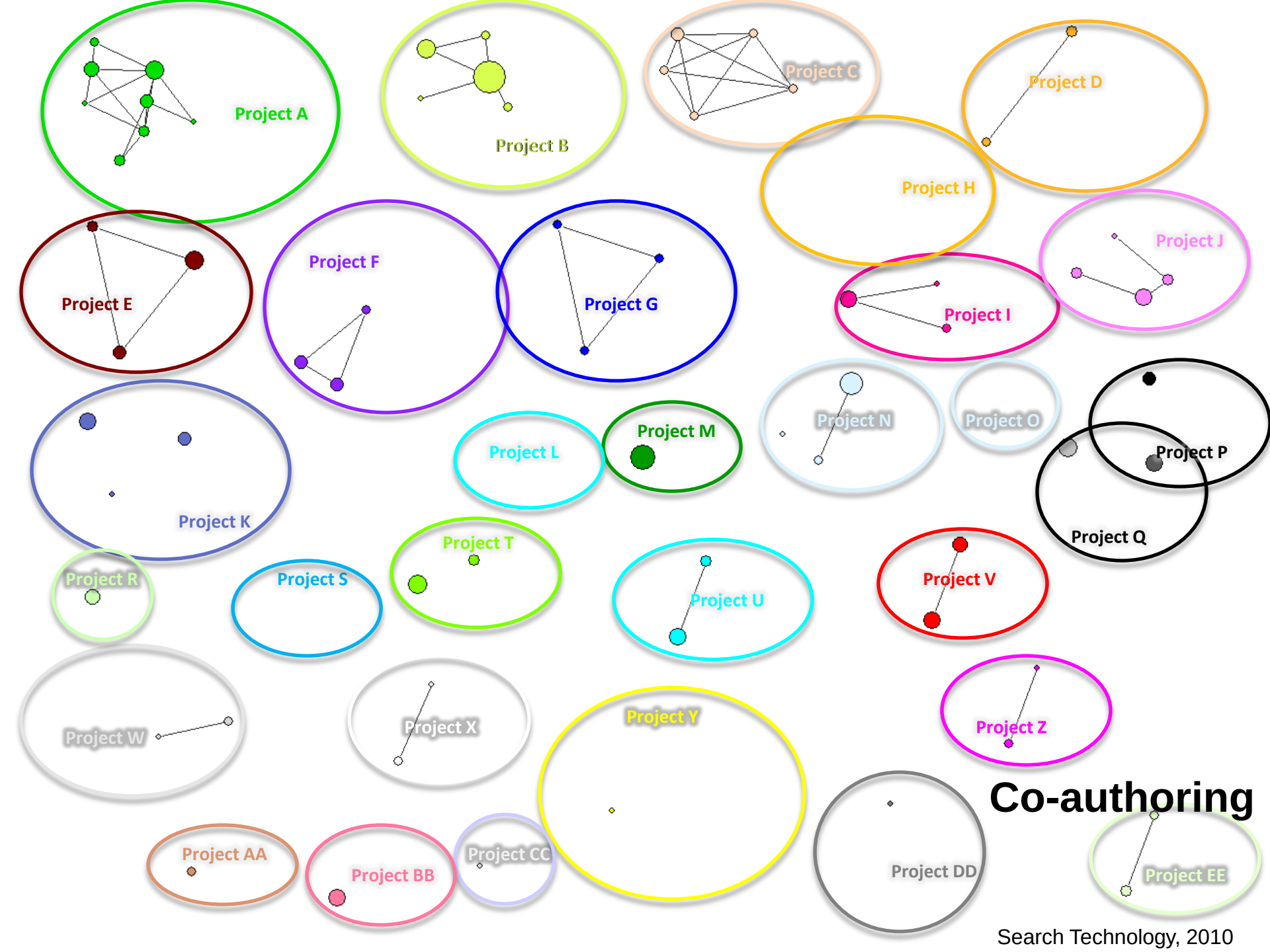
- Nodes = larger is more activity
- Multi-Dimensional Scaling ("**MDS**") representations
  - Proximity suggests relationship
  - Position on X & Y axes has no inherent meaning
- Path-erasing Algorithm added to indicate relationship
  - Heavier links (lines) = stronger relationship
  - Absence of a link only means relationship < threshold selected

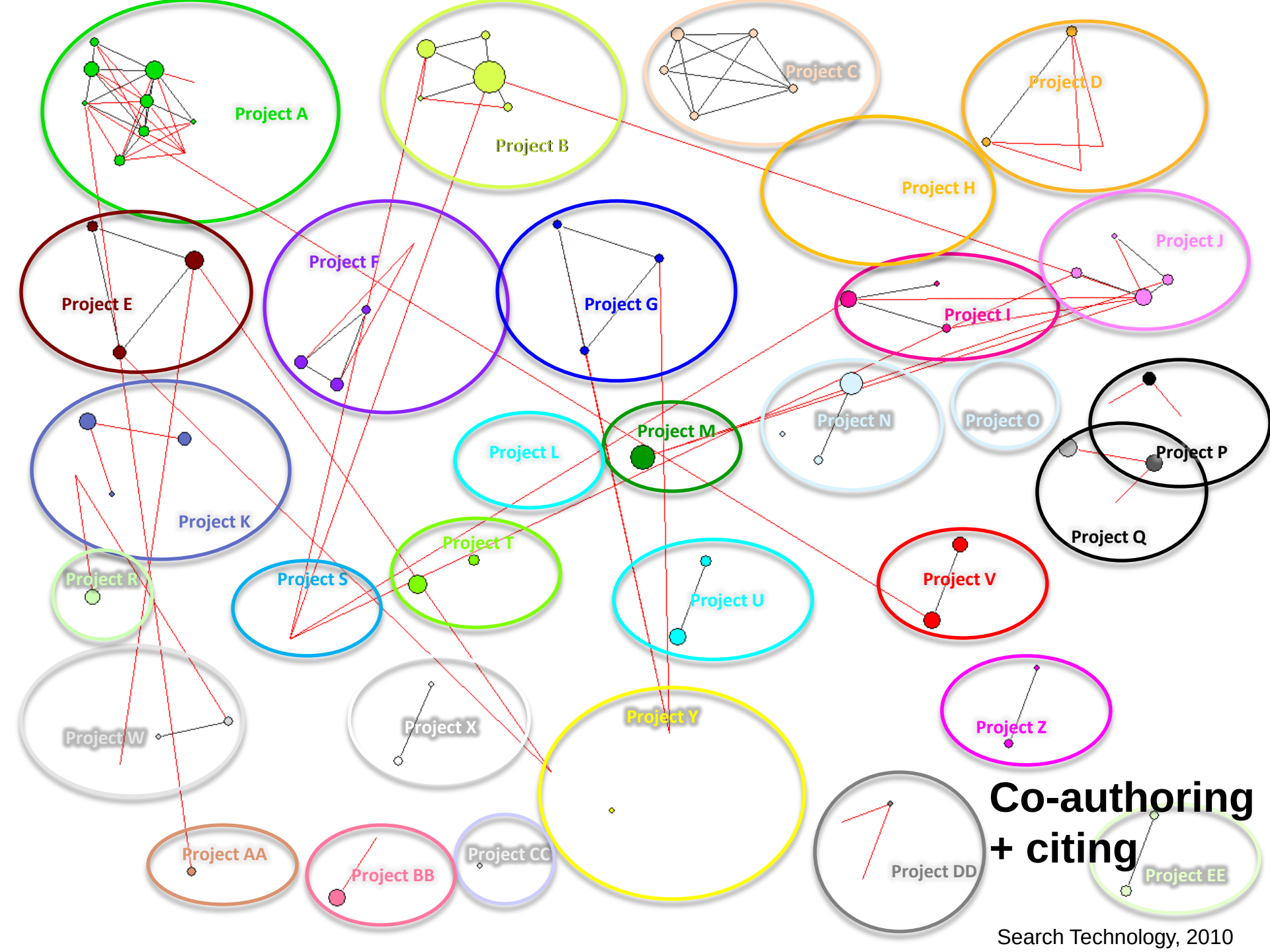
68 Highly **Citing** Authors, based on shared ROLE/REESE PIs & co-PIs being cited



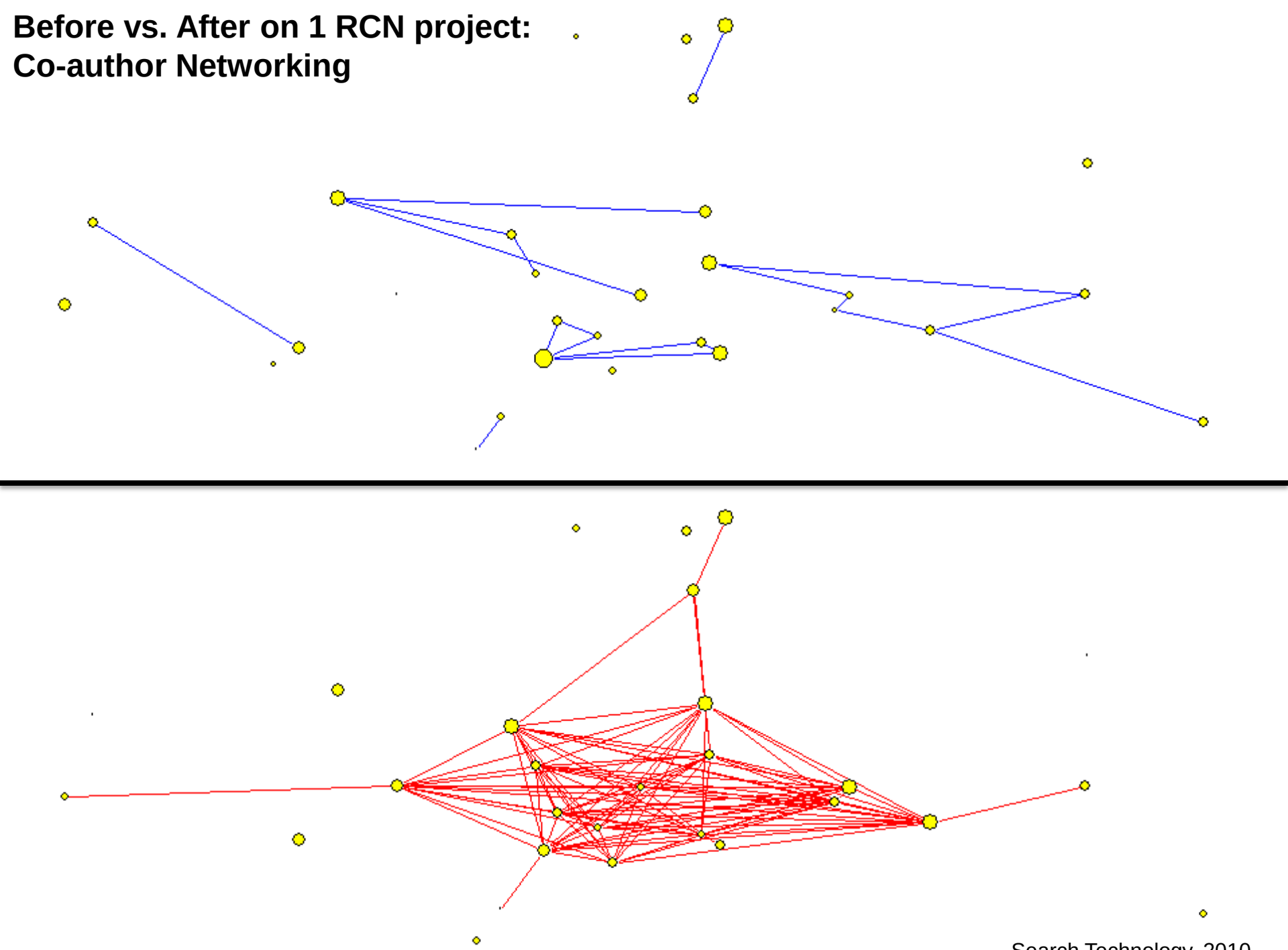
Search Technology, 2010

Co-citation Map of the most cited authors by the 307 nano social science papers


Project A
Project B
Project C
Project D
Project E
Project F
Project G
Project H
Project I
Project J
Project K
Project L
Project M
Project N
Project O
Project P
Project Q
Project R
Project S
Project T
Project U
Project V
Project W
Project X
Project Y
Project Z
Project AA
Project BB
Project CC
Project DD
Project EE
Co-authoring
Search Technology, 2010


Project A
Project B
Project C
Project D
Project E
Project F
Project G
Project H
Project I
Project J
Project K
Project L
Project M
Project N
Project O
Project P
Project Q
Project R
Project S
Project T
Project U
Project V
Project W
Project X
Project Y
Project Z
Project AA
Project BB
Project CC
Project DD
Project EE
Co-authoring
+ citing
Search Technology, 2010

Before vs. After on 1 RCN project: Co-author Networking

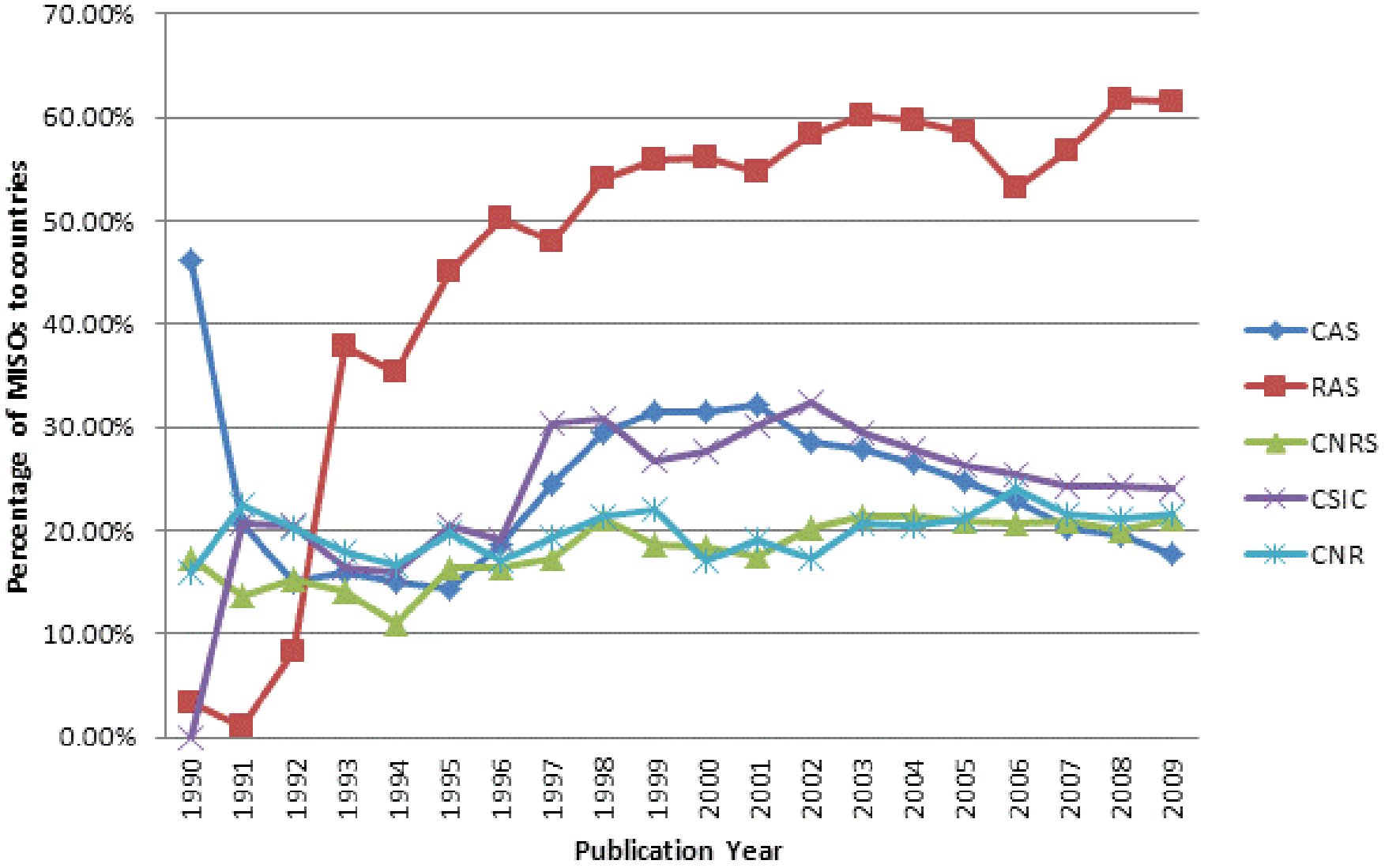


Search Technology, 2010

Nano – MISO/CAS Analyses

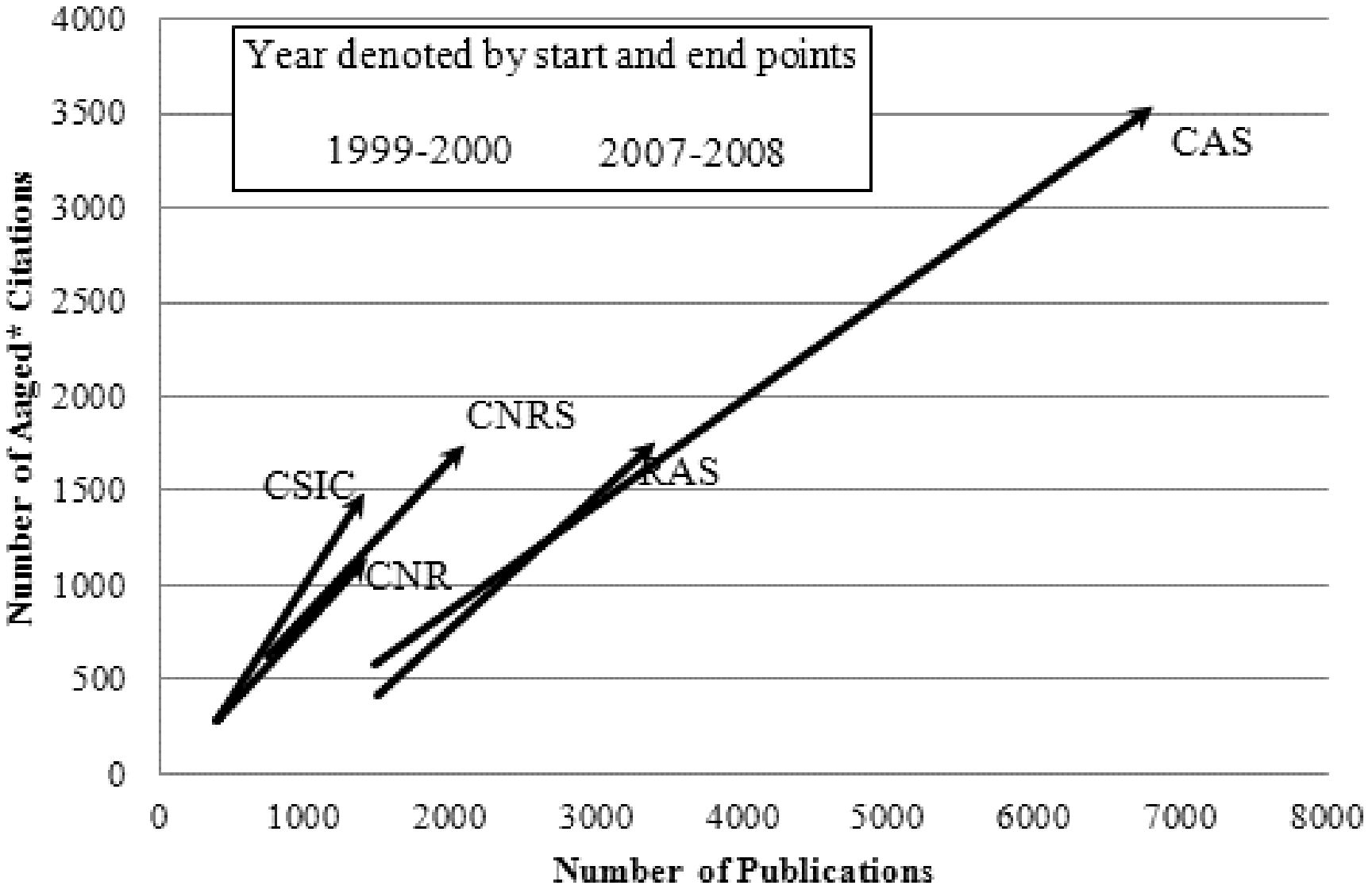
- by Ruimin Pei, CAS
- Using Georgia Tech Web of Science (SCI) nano dataset
- Compare Multi-Institute Scientific Organizations **("MISOs"):**
  - **CAS** (China)
  - RAS (Russian Academy of Sciences)
  - CNRS (France)
  - CNR (Italy)
  - CSIC (Spain)

Nano – MISO Analyses



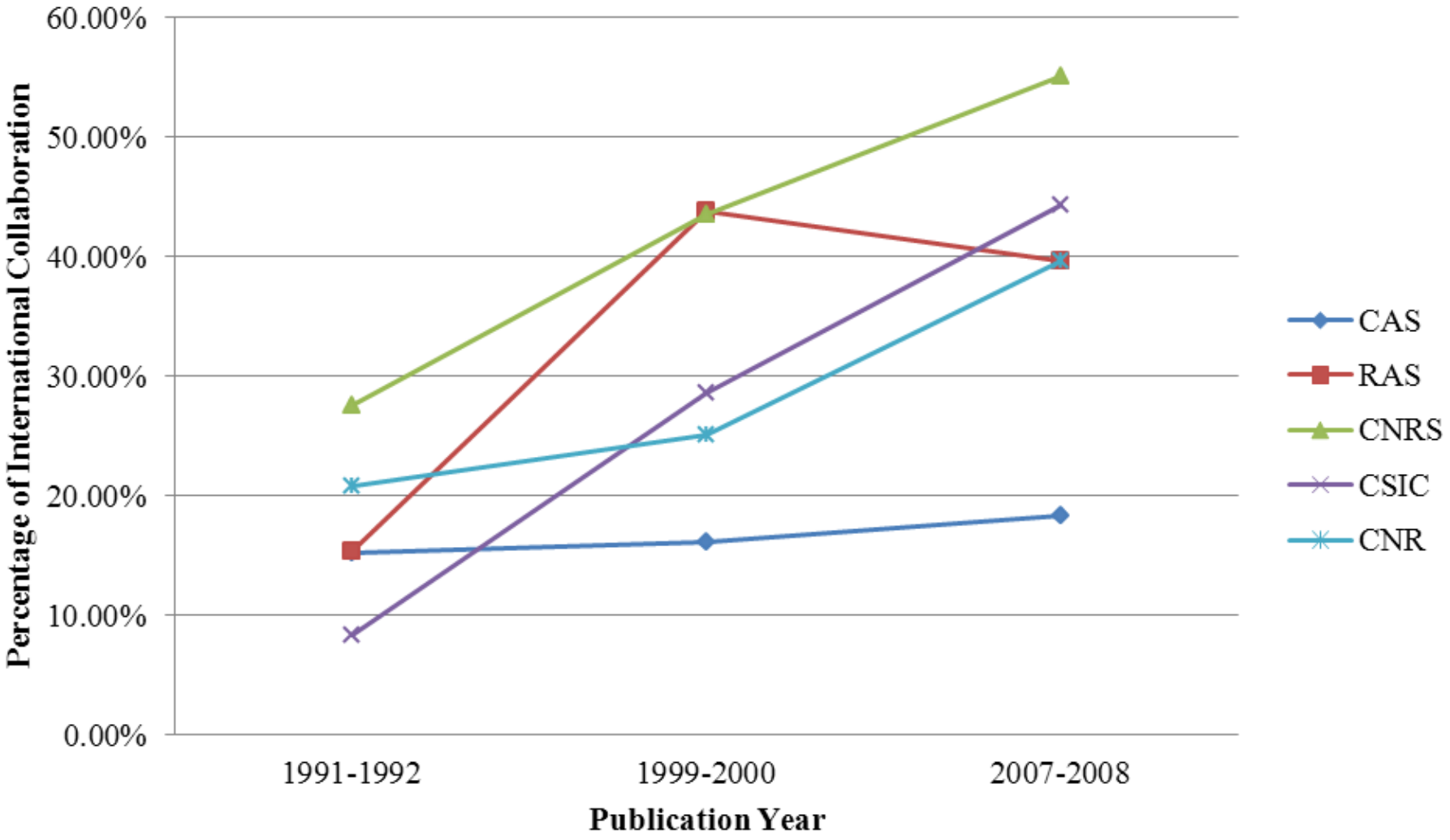
% of Country's nano SCI publications

Nano – MISO Analyses



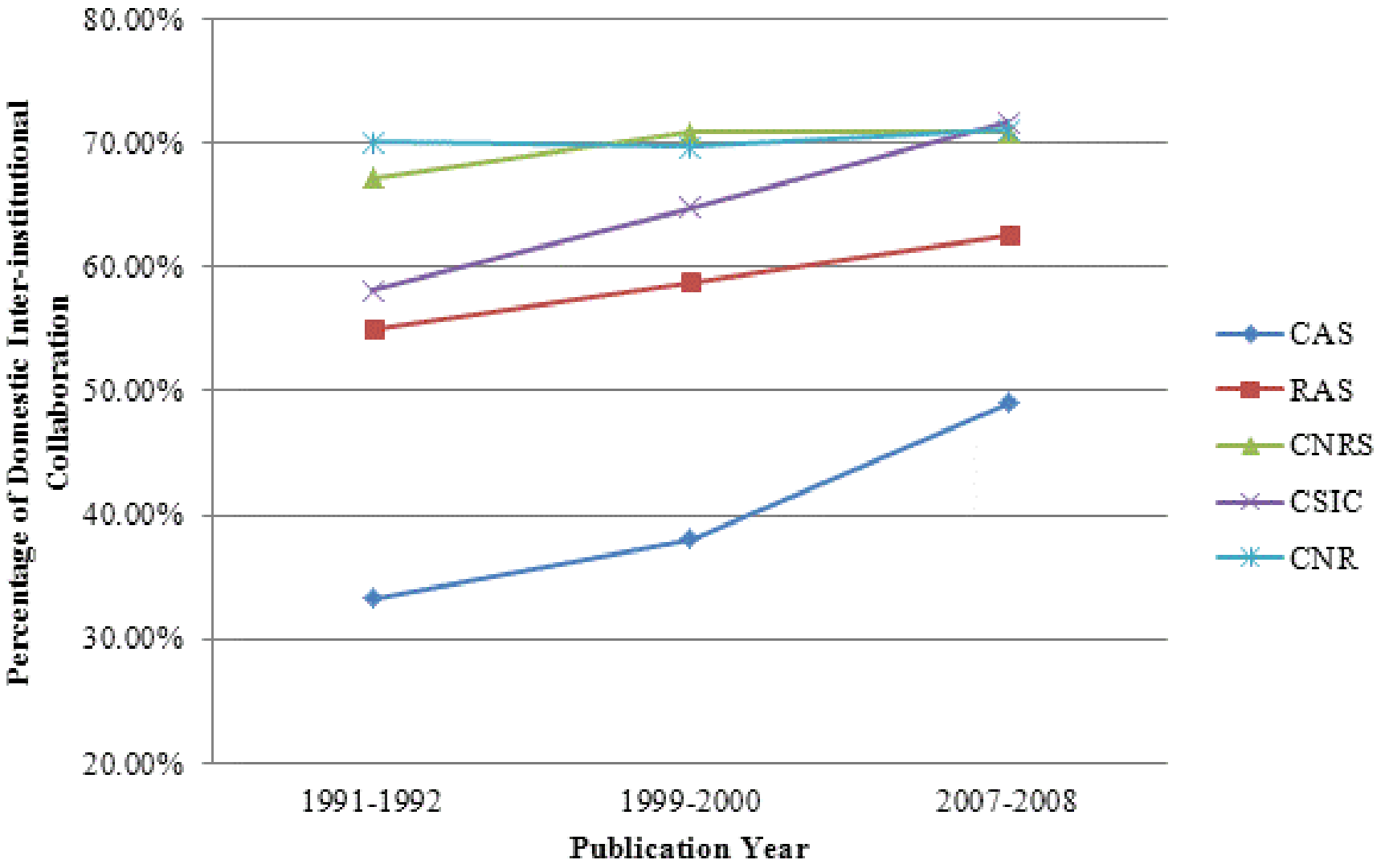
Growth in Publications & Citations/Year since publication

Nano – MISO Analyses

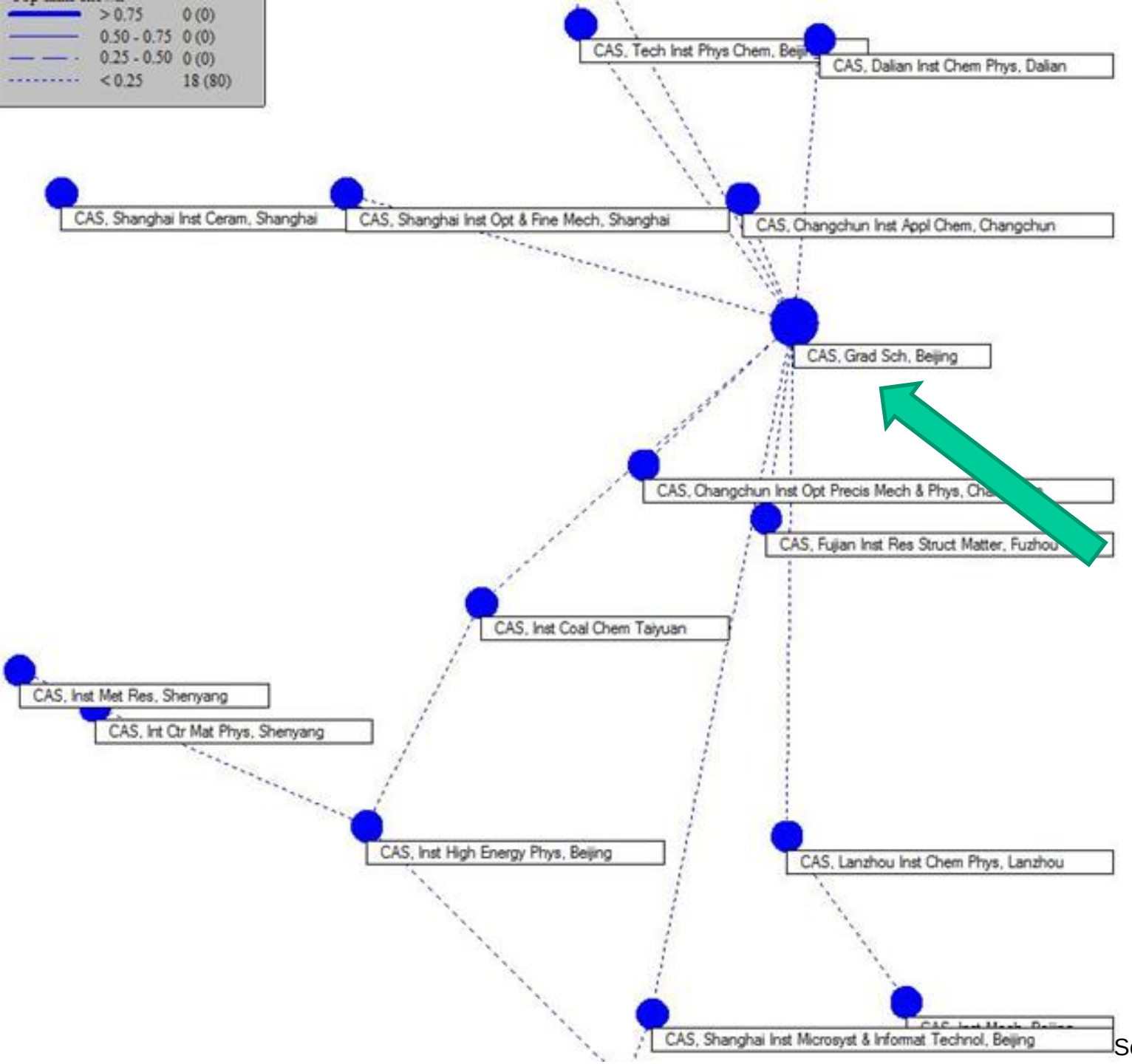


% of nano publications with international collaborators

Nano – MISO Analyses



% of nano publications with domestic collaborators



Co-authoring among CAS institutes on nano [partial network map]

CAS Grad School shows hi centrality

Search Technology, 2010

Topics

1. Tech Mining
2. Illustrating Web of Science analyses:
3. **Illustrating Patent analyses**
   - **Biomaterials Technology Opportunities Analysis**
   - **Patent Life Cycle Analysis**
4. TDA -- Dye Sensitized Solar Cells

Patent Analyses: Biomaterials Case

- Search in MicroPatents database yielded some 10,000 polymer biomaterials patents.
- 2-Dimensional Focusing
  - Type of material: fibrous structural proteins
  - Application area: skin treatments
- 640 patents
- Use patent abstract fields (e.g., assignee) +
- Entity extraction: key term phrases
- Explore various: Who? What? When? Where? Questions
- Identify “blackspaces”

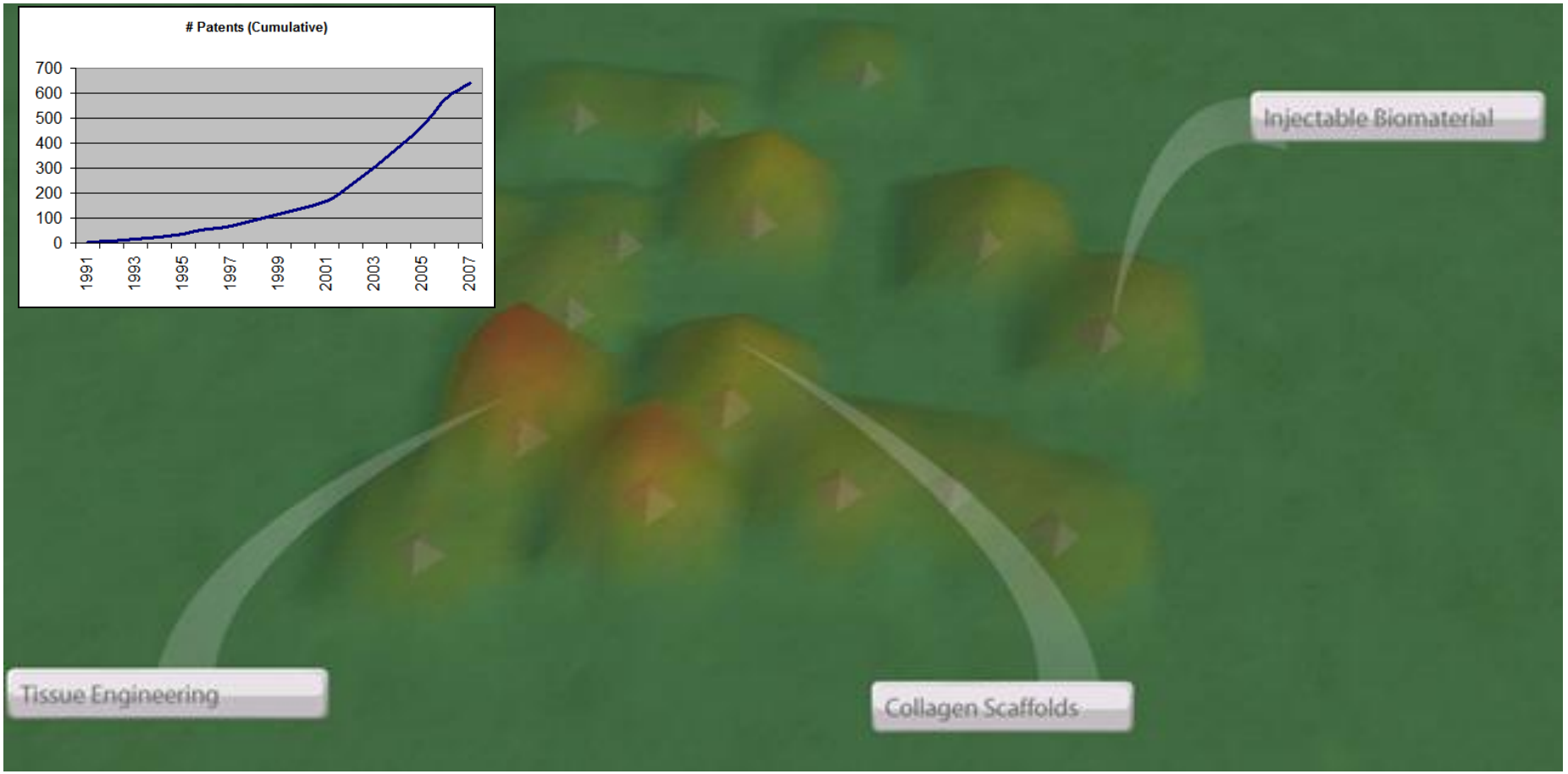
Porter, A.L., and Kayat, M., *International Chemical Information Conference* (ICIC), Barcelona, 2007

Profiling Particular Patent Assignees' Emphases

Assignee / Applicant	Property Terms in Claims	Publication Year
scaffold	Top 5 Items	% since 2006
CorMatrix Cardiovascular, Inc	scaffold [7]; patch applications [6]; heart [6]; composite/matrix [5]; stem cell [5]	100% of 7
Osteotech, Inc	tumor [6]; adhesive [6]; scaffold [6]; transplant [6]; bone [6]; osteo [6]; stem cell [6]	17% of 6
FIDIA ADVANCED BIOPOLYMERS S.R.L	tumor [6]; adhesive [6]; wound treatment [6]; composite/matrix [6]; scaffold [6]; heart [6]; vascular [6]; radio opaque/therapy [6]; diagnostics [6]; cosmetic [6]; bone [6]; ophthalmic [6]	0% of 6

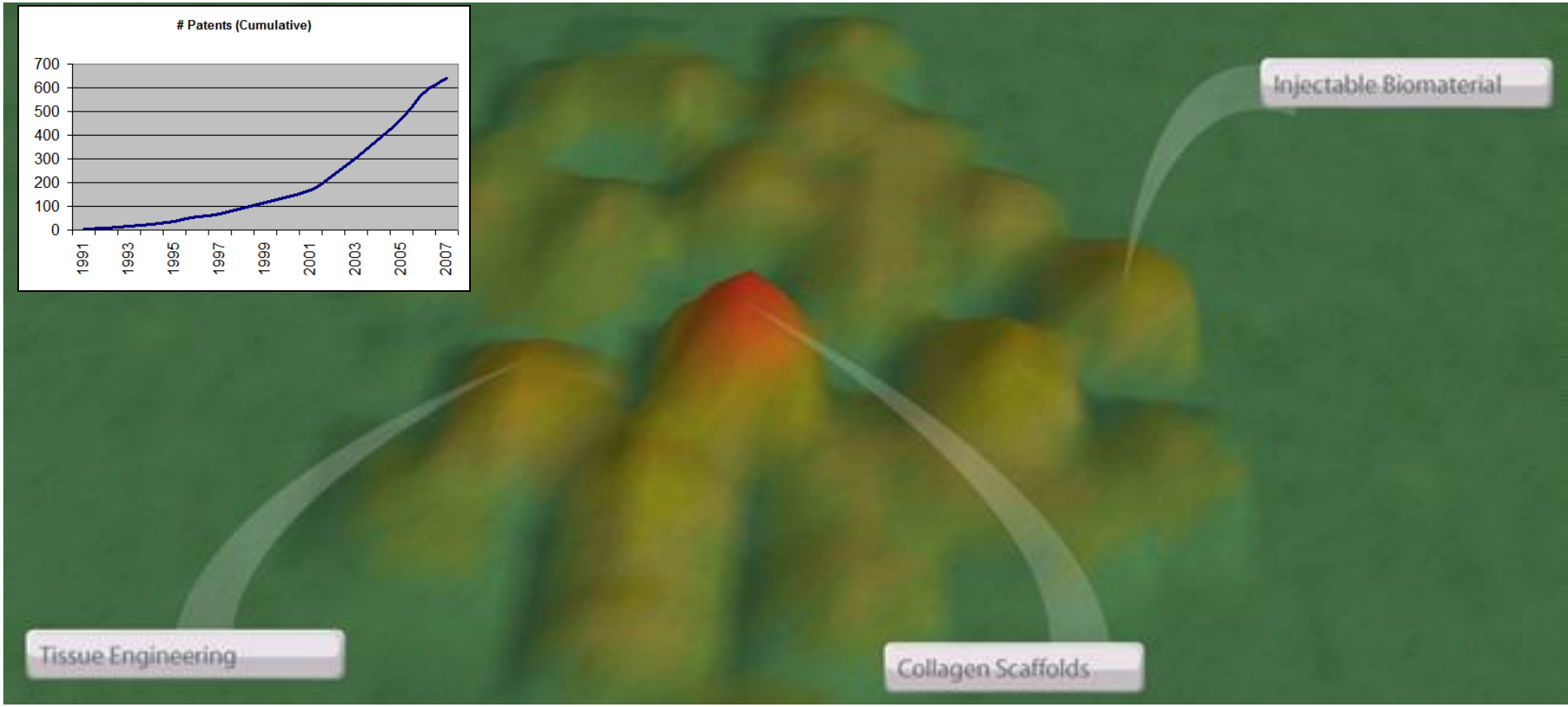
Information Visualizations

Polymer Biomaterials : fibrous structural proteins : skin
1991-1997 (68 patents)



Information Visualization with Animation

Polymer Biomaterials : fibrous structural proteins : skin
1991-2005 (470 patents)



Nanopatenting: Life Cycle Analyses

- **Nanopatenting search (Derwent)**
- **Combine two sources of information (e.g., in TDA)**
  - **Patent sub-classes**
  - **Text mining on "uses"**
- **Categorize technology targets into 3 life stages**
  - **Nano raw materials**
  - **Nano intermediates**
  - **Nano products**
- **Use to analyze company & national patenting strategies**

Alencar, M.S.M., Porter, A.L., and Antunes, A.M.S., Nanopatenting Patterns in Relation to Product Life Cycle, *Technological Forecasting & Social Change,* Vol. 74 (9), 1661-1680, 2007.

Discerning Patent Aims along the Value Chain

Main IPC [# patents]	Main uses description in the nanopatents	Position along the Nano Value Chain
H01L-Semiconductor Devices; Electric Solid State Devices Not Otherwise Provided [2870]	• Electron device • Semiconductor device • Solar cell	• Nanointermediate • Nanointermediate • Nano-products
C01B-Non-Metallic Elements; Compounds Thereof [2716]	• carbon nanotube • fuel cell • catalyst	• Nano-raw material • Nano-products • Nanointermediate
A61K-Preparations For Medical, Dental, Or Toilet Purposes [1863]	• Cancer (treatment, medication) • Cosmetics • drugs	• Nano-products • Nano-products • Nano-products
B82B-Nano-Structures; Manufacture Or Treatment Thereof Chemistry [1615]	• Carbon nanotube • Electron device • catalyst	• Nano-raw material • Nanointermediate • Nanointermediate

Locate along the Patent Value Chain

- Select top assignees for each country
- Cross IPC subclasses with patent uses (from abstract phrases)
- Locate institutions in the nano value chain categories
- Estimate national positions along this nano value chain.

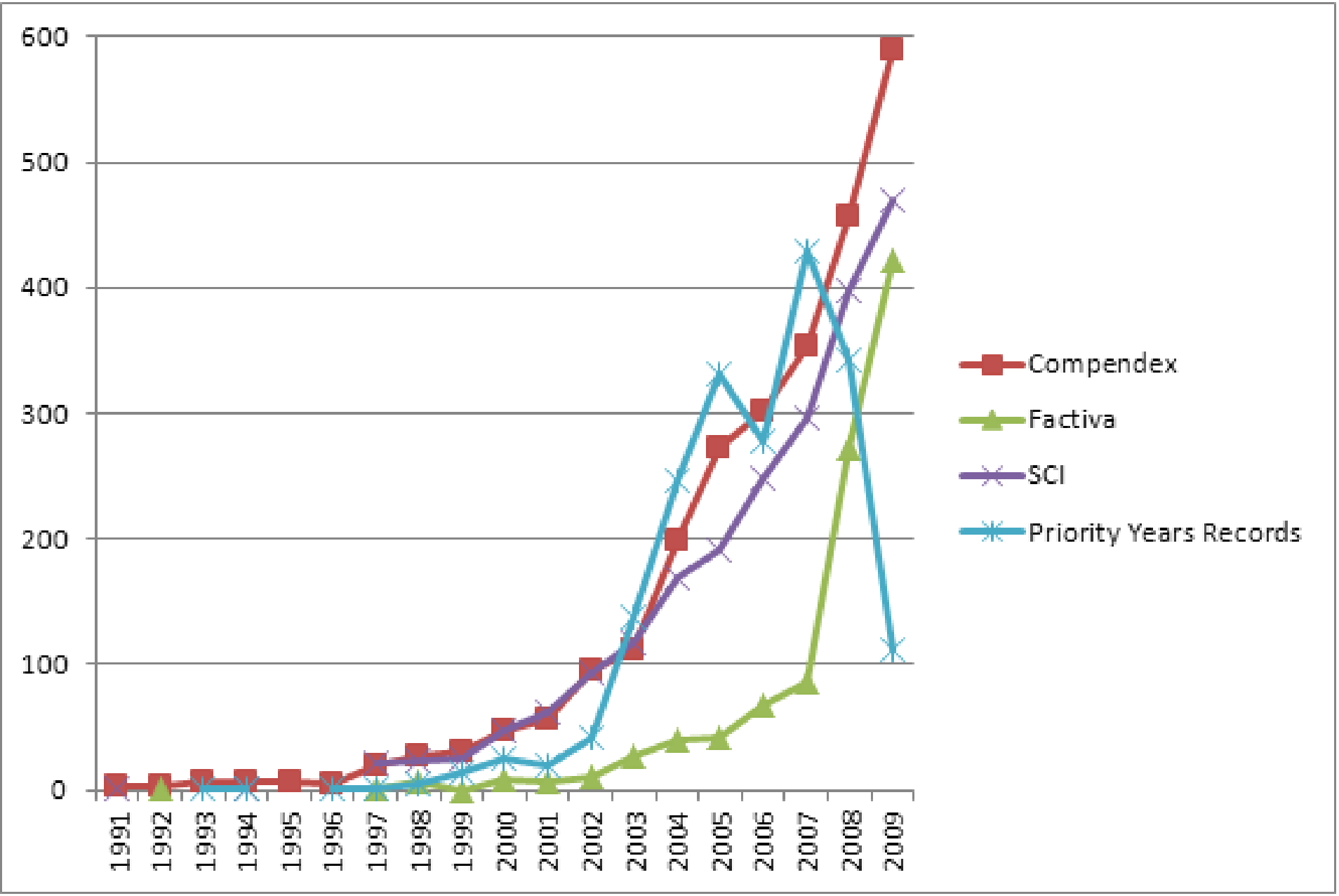
Results: Country Nanopatenting

- US: 6770 patents diffused over the 3 levels
- Japan: 4631 concentrated in the 1st level
- Germany: 1701 concentrate at the 3rd level

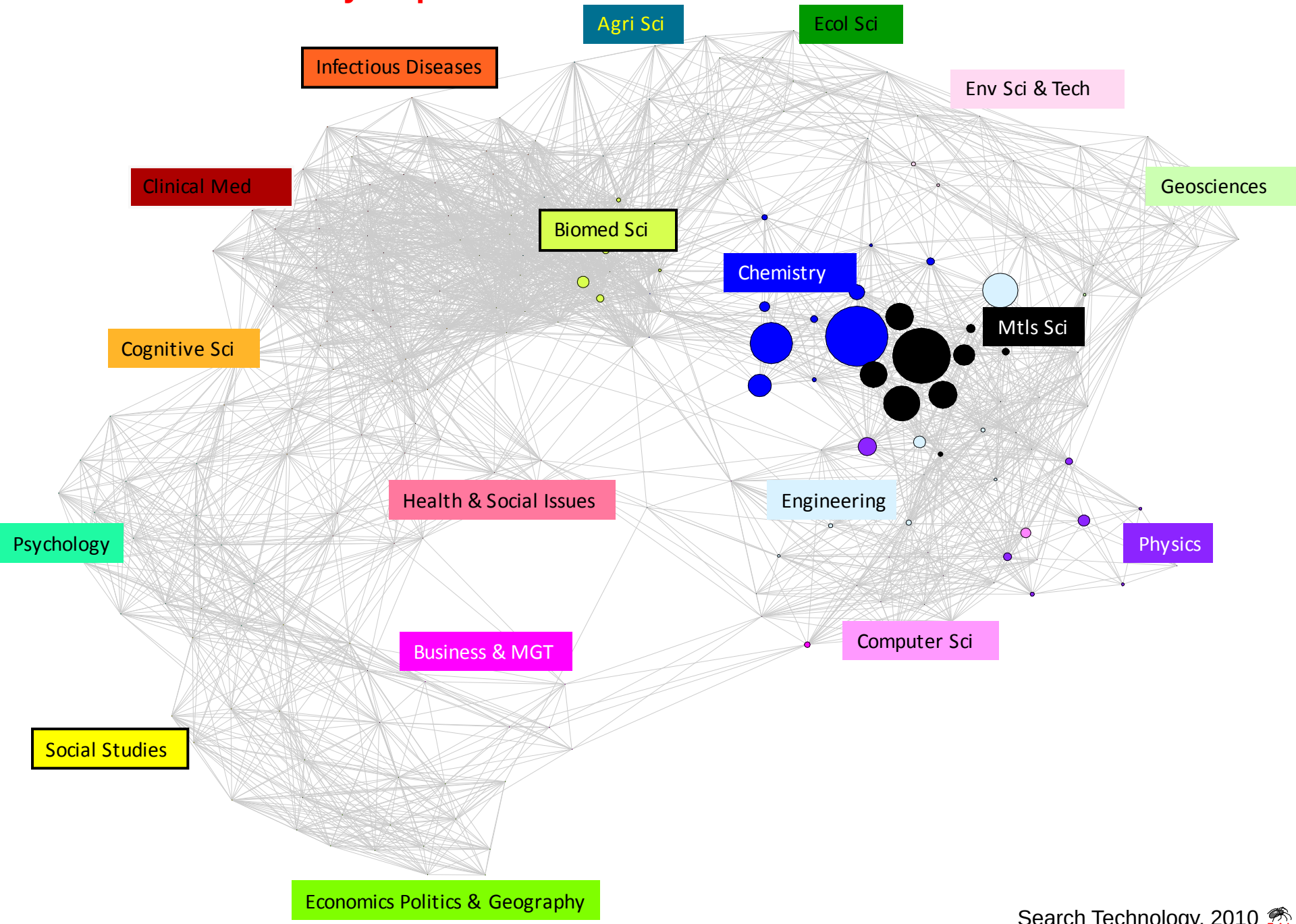
Topics

1. Tech Mining
2. Illustrating Web of Science analyses:
3. Illustrating Patent analyses
4. **TDA -- Dye Sensitized Solar Cells [thanks to Ying Guo with Lu Huang , Beijing Institute of Technology]**
   **** If time & interest, we can work with these data in TDA**

Dye Sensitized Solar Cell Trend Analyses

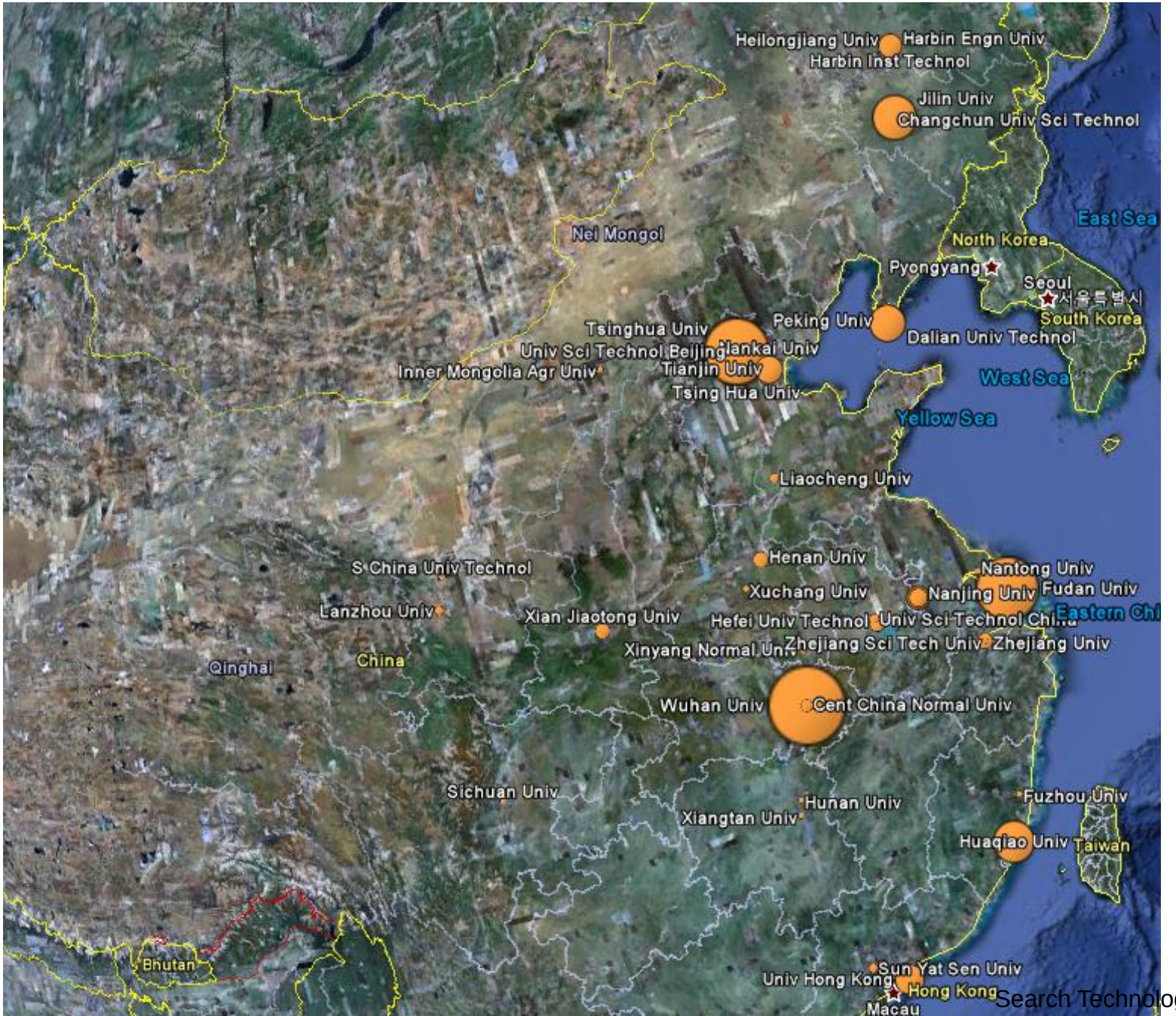


DSSC Science Overlay Map



Search Technology, 2010

Active Chinese Dye Sensitized Solar Cells Research Organizations (SCI)



Search Technology, 2010

Tech Mining+ References

- Roper, A.T., et al., *Forecasting & Management of Technology*, Wiley, 2d edition, due in 2011.
- Porter, A.L., and Cunningham, S.W., *Tech Mining*, Wiley, 2005.
- Robinson, D.K.R., Huang, L., Guo, Y., and Porter, A.L., Forecasting Innovation Pathways for New and Emerging Science & Technologies, *Technological Forecasting & Social Change*, under submission.
- Huang, L., Peng, Z., Guo, Y., and Porter, A.L. (2010), Identifying the emerging roles of nanoparticles in biosensors, *Journal of Business Chemistry,* Vol. 7 (1), 15-30.
- Guo, Y., Huang, L., and Porter, A.L. (2010), The Research Profiling Method Applied to Nano-enhanced, Thin-film Solar Cells, *R&D Management*, Vol. 40 (2), 195-208.
- Porter, A.L., and Rafols, I., Is Science Becoming more Interdisciplinary? Measuring and Mapping Six Research Fields over Time, *Scientometrics*, 81(3), 719-745, 2009 - 10.1007/s11192-008-2197-2
- Porter, A.L., and Youtie, J., Where Does Nanotechnology Belong in the Map of Science?, *Nature-Nanotechnology*, Vol. 4, 534-536, 2009.
- Porter, A.L., Youtie, J., Shapira, P., and Schoeneck, D.J., Refining Search Terms for Nanotechnology, *Journal of Nanoparticle Research*, Vol. 10 (5), 715-728, 2008

Patent Analysis References

- Xian Zhang, Shu Fang, Chuan Tang, Guo-hua Xiao, Zheng-yin Hu and Li-dan Gao. Study on Indicator System for Core Patent Documents Evaluation. *Proceedings, 12th international conference, ISSI - International Society for Scientometrics and Informetrics*, Rio de Janeiro, Brazil. 2009, Vol. 1: 154-164.
- Porter, A.L., and Kayat, M., Innovation Mapping - White space Analysis for Biomaterials in Complex Patent Landscapes, International Chemical Information Conference (ICIC), Barcelona, October, 2007.
- Porter, A.L., and Newman, N.C., "Patent profiling for competitive advantage: Deducing who is doing what, where, and when," in H.F. Moed, W. Glanzel, U. Schmoch (eds.), *Handbook of Quantitative Science and Technology Research*, Kluwer, Dordrecht, The Netherlands, 2005.
- Alencar, M.S.M., Porter, A.L., and Antunes, A.M.S., Nanopatenting Patterns in Relation to Product Life Cycle, *Technological Forecasting & Social Change,* Vol. 74 (9), 1661-1680, 2007.

Science Mapping References

Science Maps

Leydesdorff, L. and 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.

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

Boyack, K. W., Börner, K. & Klavans, R. (2009). Mapping the structure and evolution of chemistry research. Scientometrics, 79(1), 45-60.

Klavans, R. & Boyack, K. W. (2009). Toward a Consensus Map of Science. *Journal of the American Society for Information Science and Technology*, 60(3), 455-476.

Places & Spaces: http://www.scimaps.org/

Science Overlay Maps

Rafols, I., Porter, A.L., and Leydesdorff, L., Science overlay maps: A new tool for research policy and library management, *Journal of the American Society for Information Science & Technology*, 61 (9), 1871-1887, 2010.

Rafols, I. and Meyer, M. (2009) Diversity and Network Coherence as indicators of interdisciplinarity: case studies in bionanoscience. *Scientometrics*, 82(2), 263-287.

The Challenge: ST&I Policy and Technology Management are weak in use of empirical intelligence

1. Derive innovation indicators that answer decision-makers' questions
2. "Tech Mining" of R&D literature and patent literature can provide empirical intelligence (for evidence-based decision making)
3. We are pursuing:
   - R&D Assessment, incorporating patent analyses
   - Tracking Research Knowledge Diffusion
   - Forecasting Innovation Pathways

Resources

- **The text mining software used:**
  //ip.thomsonreuters.com/training/tda/
- **Ongoing Research on Interdisciplinarity & to make your own science overlay maps:**
  //idr.gatech.edu/
- **Future-oriented Technology Analysis Conference,** Seville, May 12-13, 2010 – see jrc-ipts-foresight@ec.europa.eu
- **Tech Mining Workshop + Atlanta Conference on Science, Technology & Innovation Policy**, Atlanta, Sep 13-17, 2011

Outtakes

Heuristics of diversity

(Stirling, 1998; 2007)
(Rafols and Meyer, 2009)

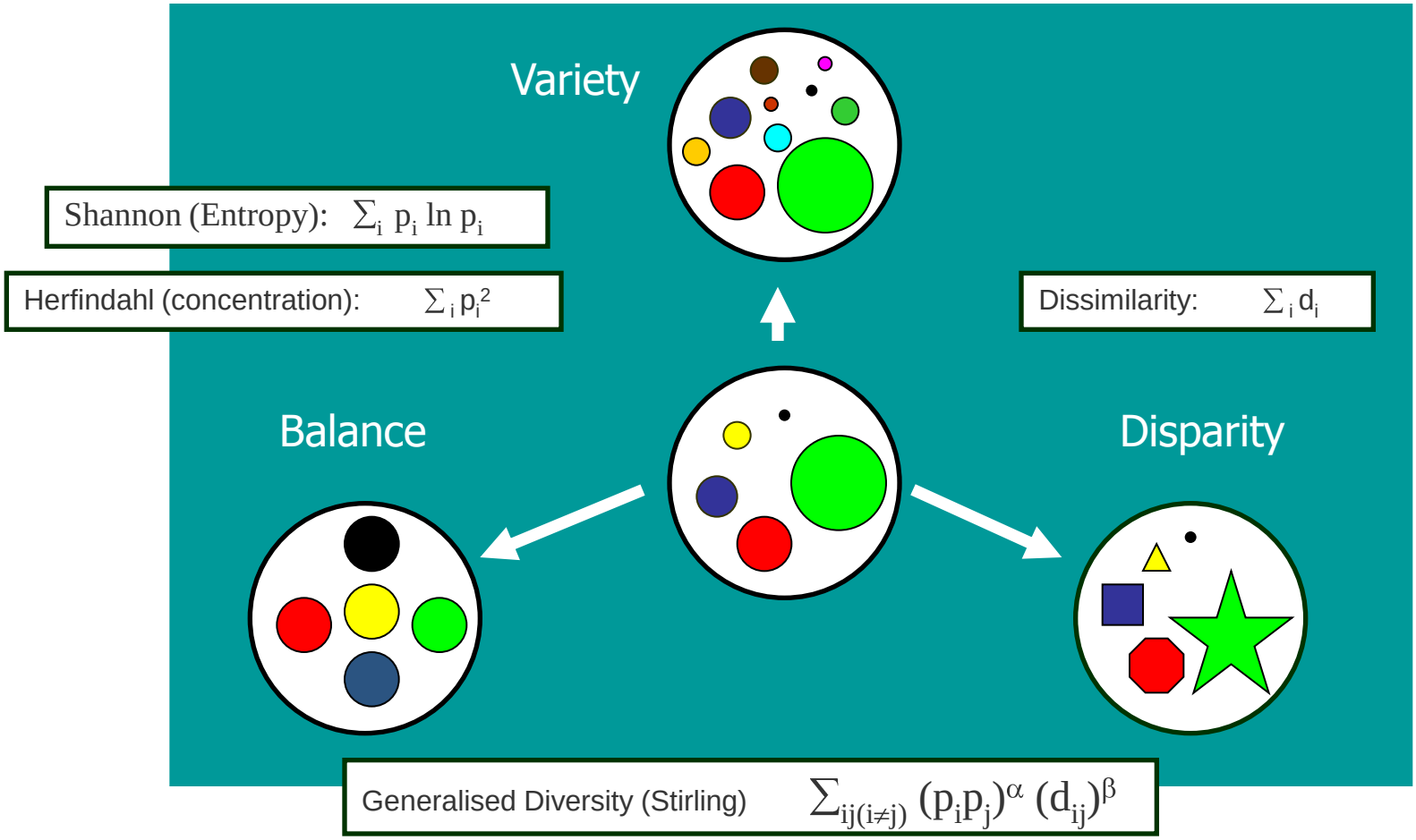
Diversity:

'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]



Integration Score

$$I = 1 - \left[\frac{\sum \left(f_i \times f_j \times COS \left(SC_i - SC_j \right) \right)}{\sum \left(f_i \times f_j \right)} \right]$$

Porter et al. (2007)

where i = *row* ; j = *column* ; f = *frequency*

"cos (SCi – SCj)" measures the association between two SCs, based on a national co-citation sample from Web of Science. It reflects the relative tendency of two particular SCs to be co-cited.

****equivalently,**

$$I = 1 - \sum_{i,j} p_i p_j s_{ij}$$

Rafols and Meyer (2009)

where p_i is the proportion of references citing the SC i in a given paper. The summation is taken over the cells of the SC x SC matrix. s_{ij} is the cosine measure of similarity between SCs i and j

Behavior of our Integration Score

A paper cites N references

Integration increases:

1. As the # of Cited SCs increases (higher **Variety**)

2. As the **Balance** among those Cited SCs increases

3. As the **Disparity** among those Cited SCs increases

Integration ranges from:

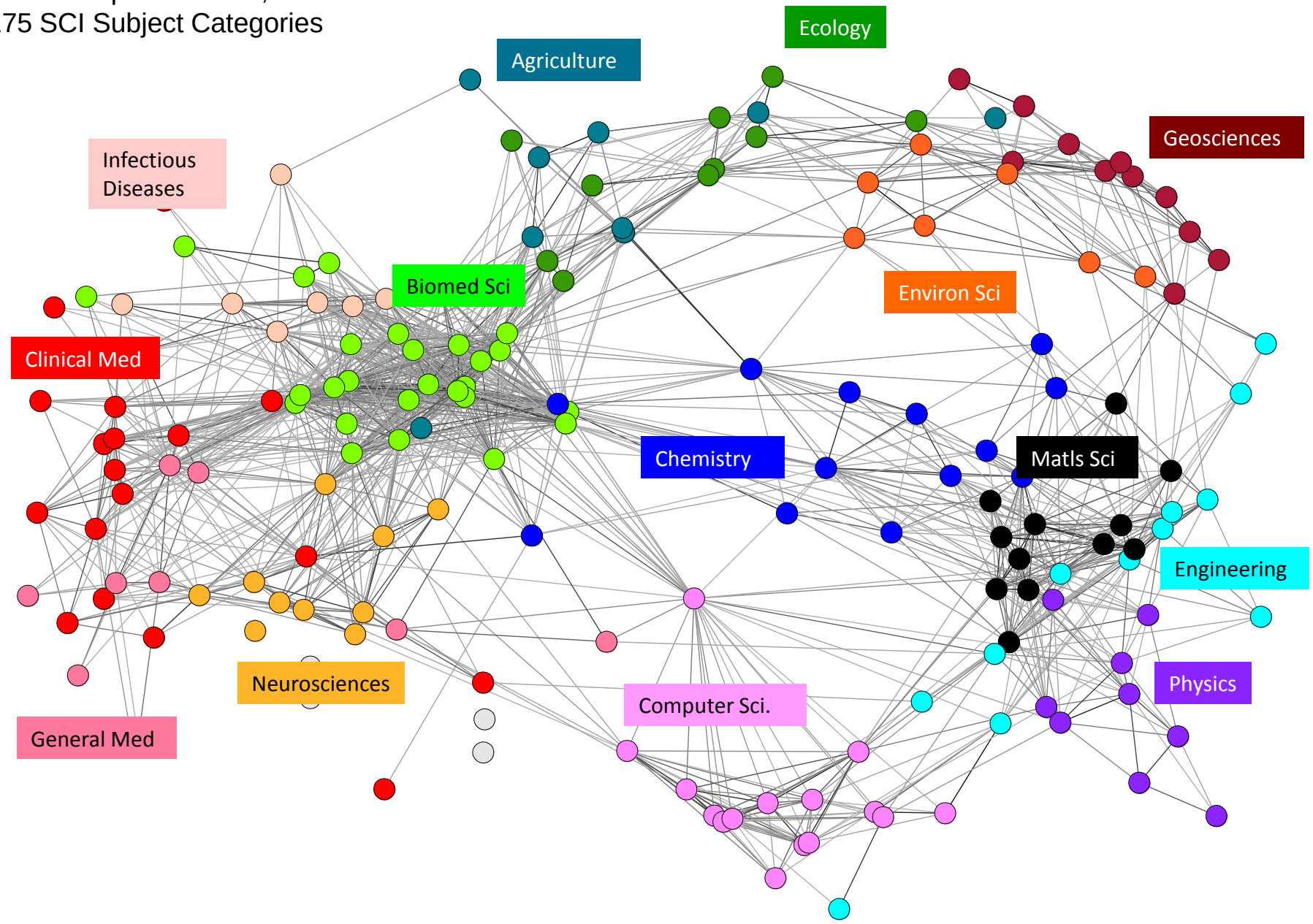
0 (research that cites work from a single SC) to

1 (research drawing from multiple, ~unrelated SCs)

Science Mapping: We're not alone

- Chen, C. (2003) *Mapping Scientific Frontiers: The Quest for Knowledge Visualization*, Springer, London
- Places & Spaces: Mapping Science **http://sci.slis.indiana.edu/**
- Klavans, R. & Boyack, K. W. (2009). Toward a Consensus Map of Science. *Journal of the American Society for Information Science and Technology*, 60(3), 455-476.
- An essential difference
  - **Global** ("all" of science)
  - **Local** (research network analyses)

Global Map of Science, 2006
175 SCI Subject Categories

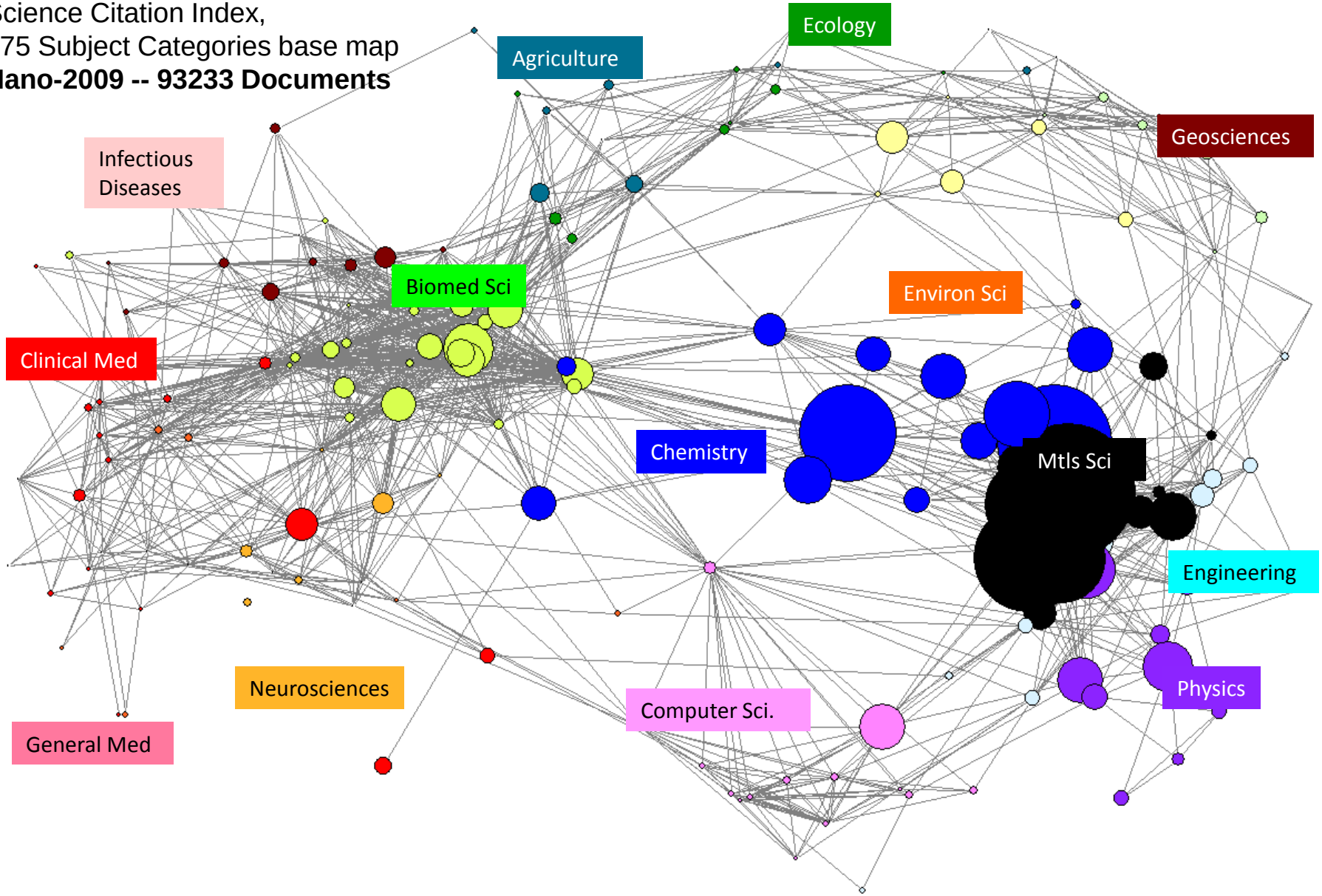


Leydesdorff and Rafols (2009)

Pajek

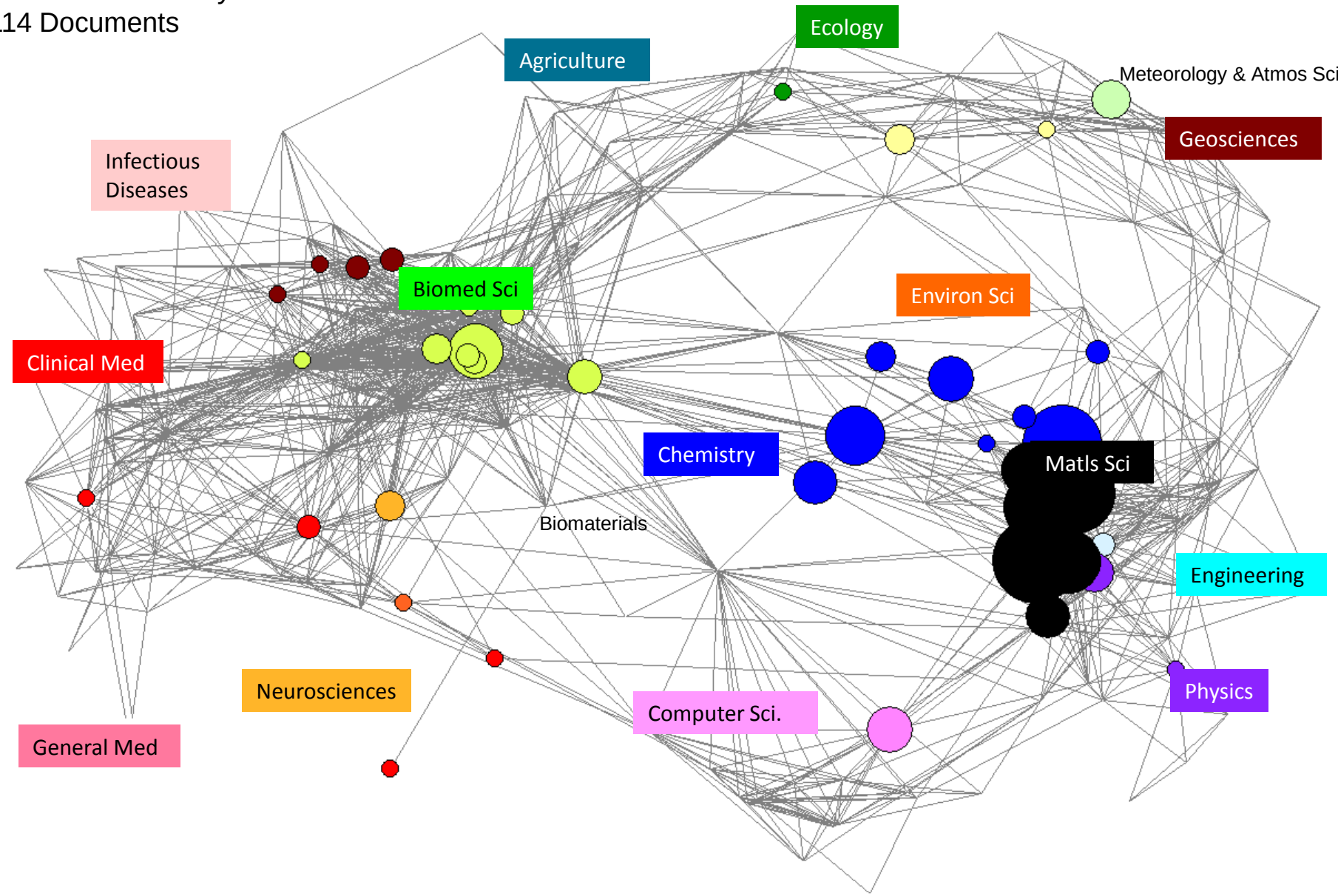
Search Technology, 2010

Web of Science –
Science Citation Index,
175 Subject Categories base map
Nano-2009 -- 93233 Documents

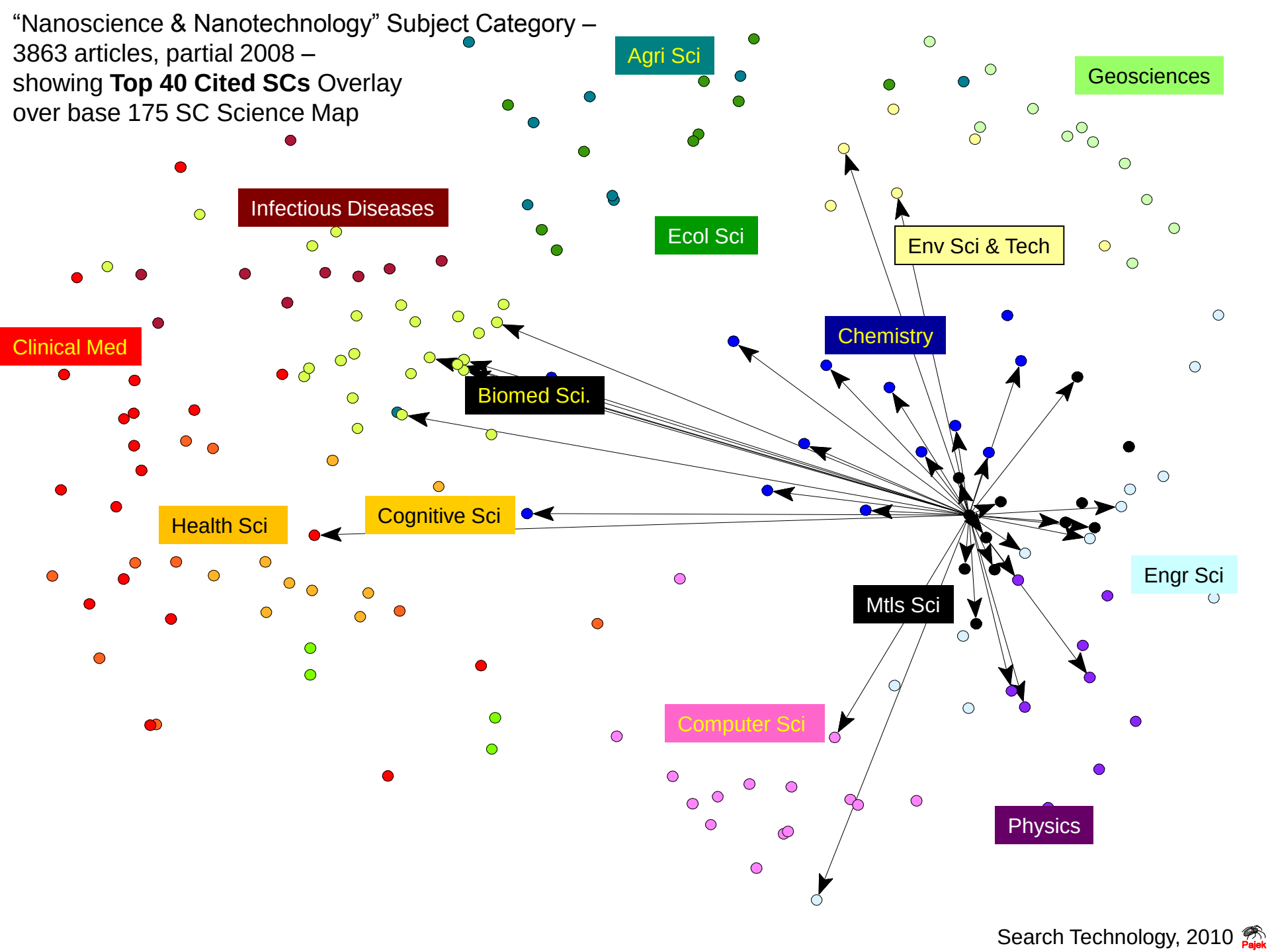


Search Technology, 2010

Nano-SUNY-Albany 2008-09 Web of Science
114 Documents



Search Technology, 2010

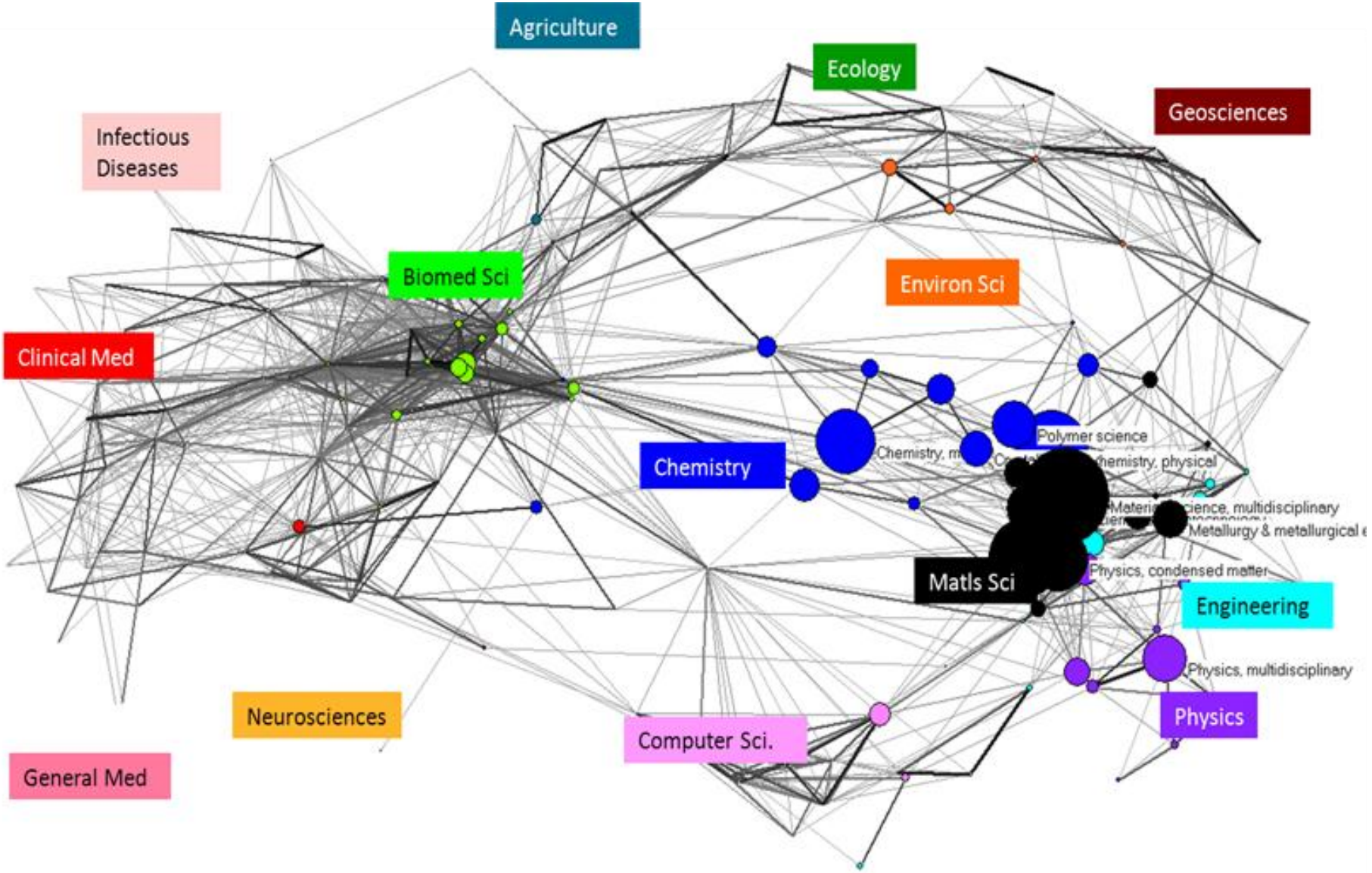

"Nanoscience & Nanotechnology" Subject Category –
3863 articles, partial 2008 –
showing **Top 40 Cited SCs** Overlay
over base 175 SC Science Map
Agri Sci
Geosciences
Infectious Diseases
Ecol Sci
Env Sci & Tech
Clinical Med
Chemistry
Biomed Sci.
Health Sci
Cognitive Sci
Engr Sci
Mtls Sci
Computer Sci
Physics
Search Technology, 2010
Pajek

Nano Global Mapping "Results"

- Does nano engage multiple disciplines?
  - Very much
  - But core is Materials Science

- Does nano research integrate knowledge from multiple disciplines?
  - YES -- nano is not just ~disciplinary research "silos"
  - But citation is heaviest to nearby research areas

Nano – CAS Analyses



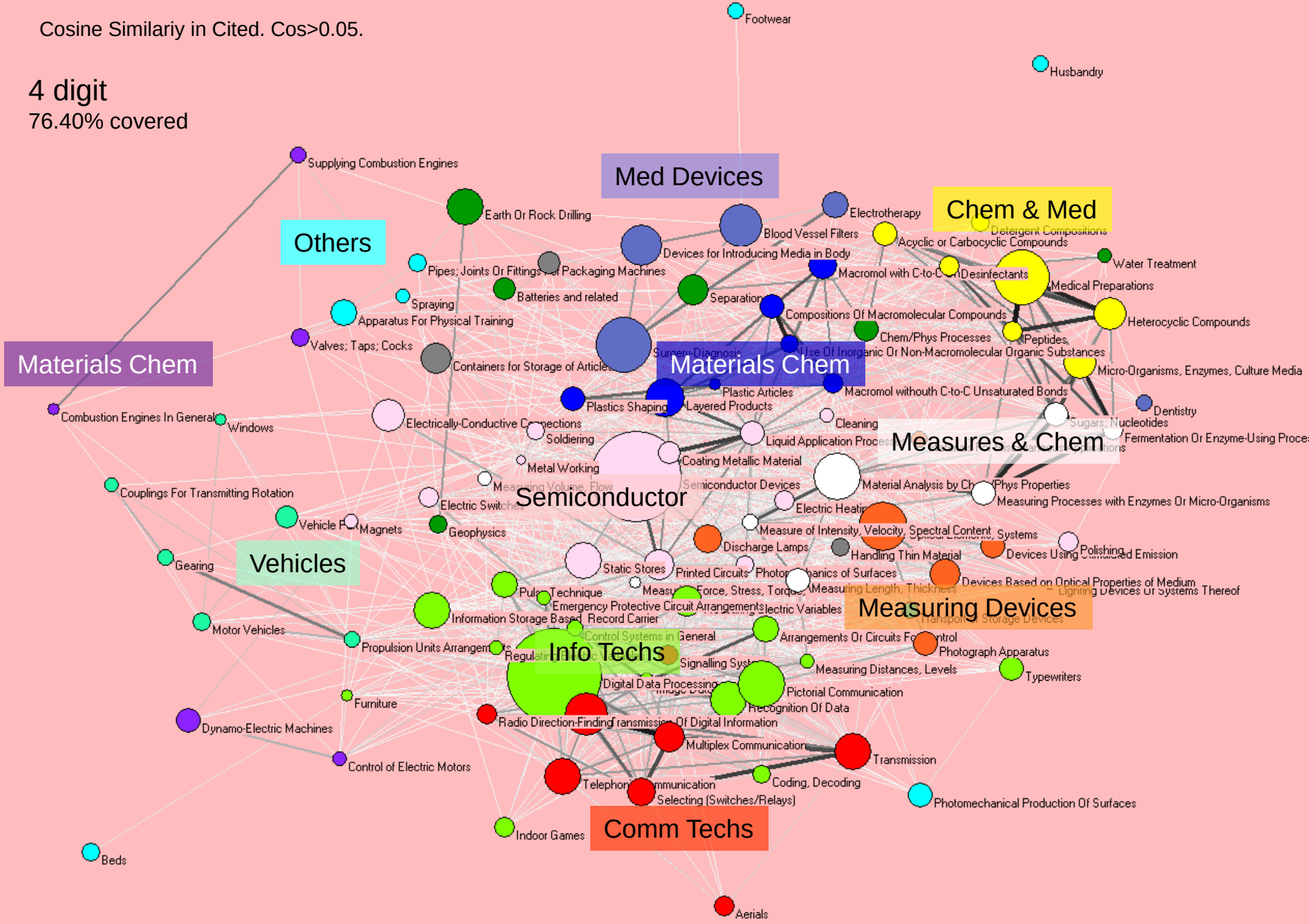
Locating CAS nano publications among the Disciplines

2005 Patent Map

Cosine Similariy in Cited. Cos>0.05.

4 digit

76.40% covered



Search Technology, 2010