

Science Overlay Mapping for Benchmarking, Complementarities, and Possible Collaborations

Alan Porter

Technology Policy & Assessment Ctr (TPAC)

Georgia Tech

alan.porter@isye.gatech.edu

Presentation

- Science overlay mapping
- Research assessment applications

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.

Science Overlay Mapping

- Use the science map as a base upon which to locate given bodies of research
- Plot Subject Category concentrations
 - Node size ~ #
 - Can plot publication SCs, cited SCs, &/or citing SCs
- Convey research field engagement (i.e., locate research activity among the disciplines)
- Show cross-disciplinary research knowledge interchange
- Show change over time (e.g., Before vs. After researchers receive an award)

Tracking multi-generational research knowledge transfer with

- **Interdisciplinarity metrics**
- **Science overlay mapping**

#1: Papers Citing Level #2 Papers
– Citing Paper Overlay Maps
[Knowledge Diffusion]

- Diffusion scores
- **Science Citing Overlay Maps**
- Relative engagement by ISI Subject Categories



#2: Main Level (e.g., research outputs of a target program) – publication overlay maps

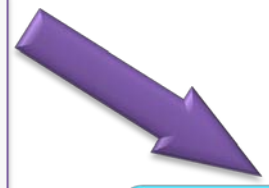
- “Specialization” scores (Diversity of areas of publication)
- **Science overlay maps** (Location of publications among ISI Subject Categories)

- Integration scores (Average diversity of areas of citation)
- **Science citation maps**
- Bibliographic coupling



#3: Papers cited by #2

- Coherence measures (do #3 papers draw upon distinct topics?)
- [“Bibliographic Coupling” measures available – e.g., % shared references]



#4: Papers cited by #3

The Data

- **Rely notably on Web of Science (“WOS”)**
 - Science Citation Index (SCI) & Social Science Citation Index (SSCI)
 - Backward & Forward Citation information
- **WOS Subject Categories (“SCs”)**
 - Thomson Reuters Institute for Scientific Information (ISI) uses to categorize journals
 - ~221 for SCI + SSCI (+ additional for Arts & Humanities Index, which behave differently, so not included now)
 - Have devised our interdisciplinarity metrics based on the SCs
 - Have based our science overlay maps on these – using Leydesdorff’s SC-to-SC cross-citation matrix (basis of similarity among SCs)
- **Macro-Disciplines**
 - Ismael Rafols’ factor analyses provide – statistically based; human assigned names
 - 18 for our 221 SC maps (based on 2007 WOS SC-to-SC cross-citations)
- **Meta-Disciplines**
 - 4 overarching categories

The Data: WOS Abstract Records (partial record)

AU **Guerra-Filho, G; Aloimonos, Y**

TI A sensory-motor language for human activity understanding

SO 2006 6th IEEE-RAS International Conference on Humanoid Robots

DT Proceedings Paper

CT 6th IEEE/RAS International Conference on Humanoid Robots

AB We have empirically discovered that the space of human actions has a linguistic framework. This is a sensory-motor space consisting of the evolution of the joint angles of the human body in movement. The space....

C1 **Univ Maryland**, Dept Comp Sci, Comp Vis Lab, College Pk, MD 20742 **USA**.

CR ATKESON CG, 1997, **P 14 INT C MACH LEAR**, P12

BARBIC J, 2004, **PROC GRAPH INTERF**, P185

FOD A, 2002, **AUTON ROBOT**, V12, P39

GALLESE V, 1996, **BRAIN 2**, V119, P593

NR 22

TC 1

PY 2006

SC **Automation & Control Systems; Computer Science, Artificial Intelligence; Robotics**

Magic in Analyses: Macros

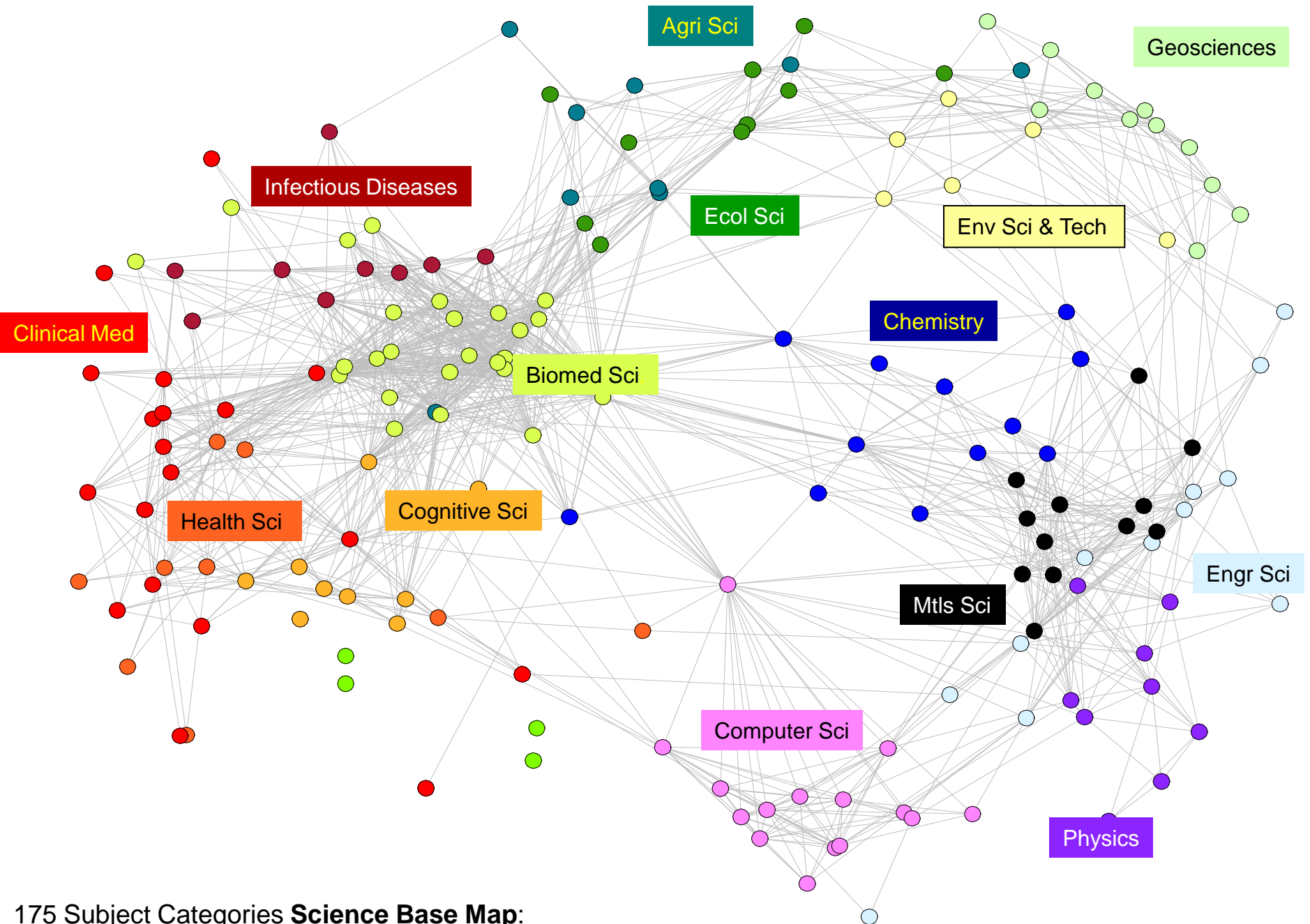
- Retrieving records
- Extracting “Cited SCs”
- Calculating integration scores
- Preparing science overlay maps

[Special thanks to Stephen Carley at GT!]

Text Data Mining

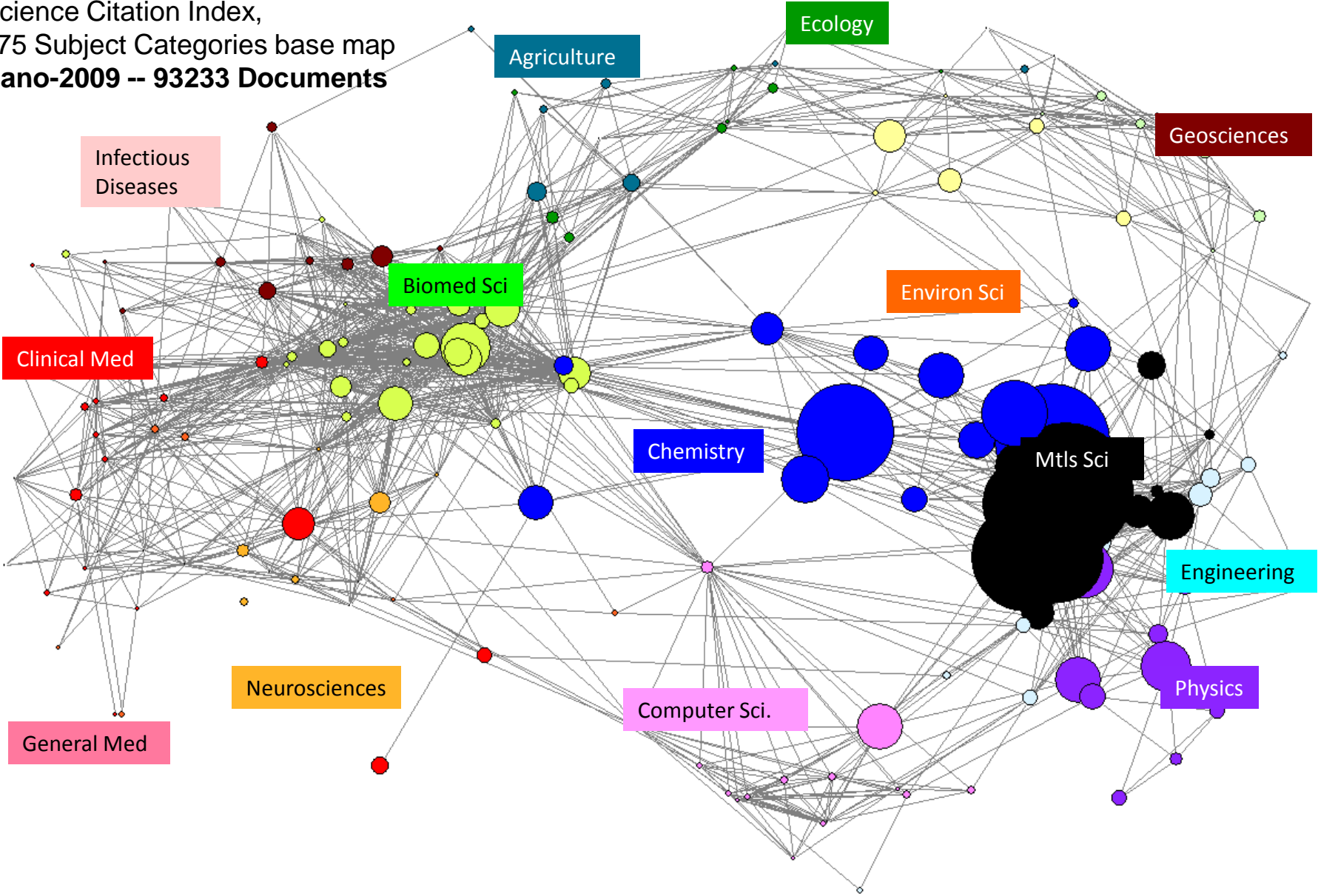
- Extract field information (authors, cited authors, cited journals, topics, etc.)
- Clean (fuzzy matching) & apply thesauri to extract “cited Subject Category,” etc.
- Clean WOS (& other) records (e.g., to identify correctly with our HSD researchers – especially for “After HSD” pubs)
- Clean WOS **citations to** those publications (to associate correctly & extract desired fields)
- Analyze citation records **by** those publications to get at research stream “coherence”

[Using *VantagePoint* (~TechOASIS; Thomson Data Analyzer) – www.theVantagePoint.com]

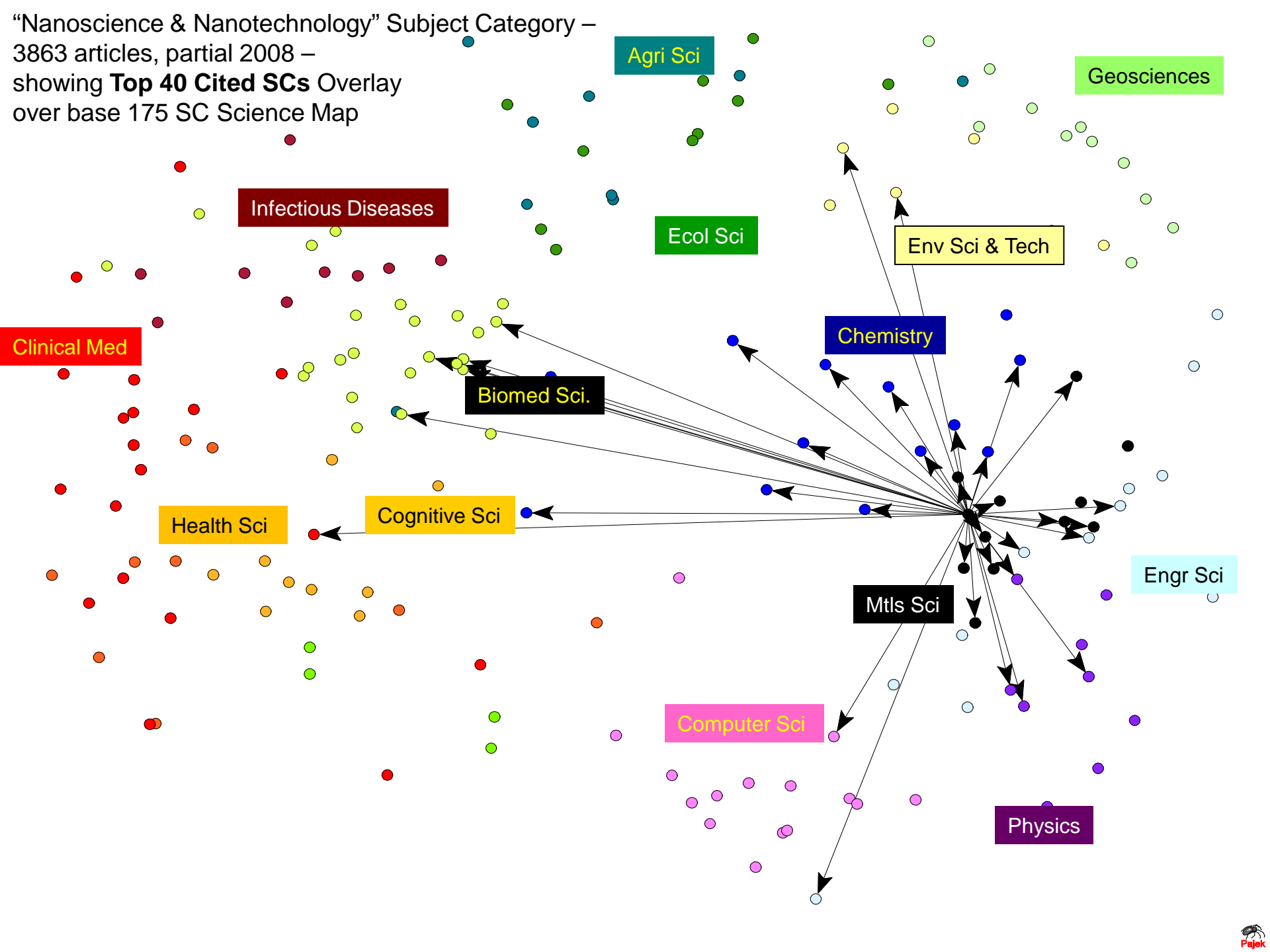


175 Subject Categories **Science Base Map:**
See References

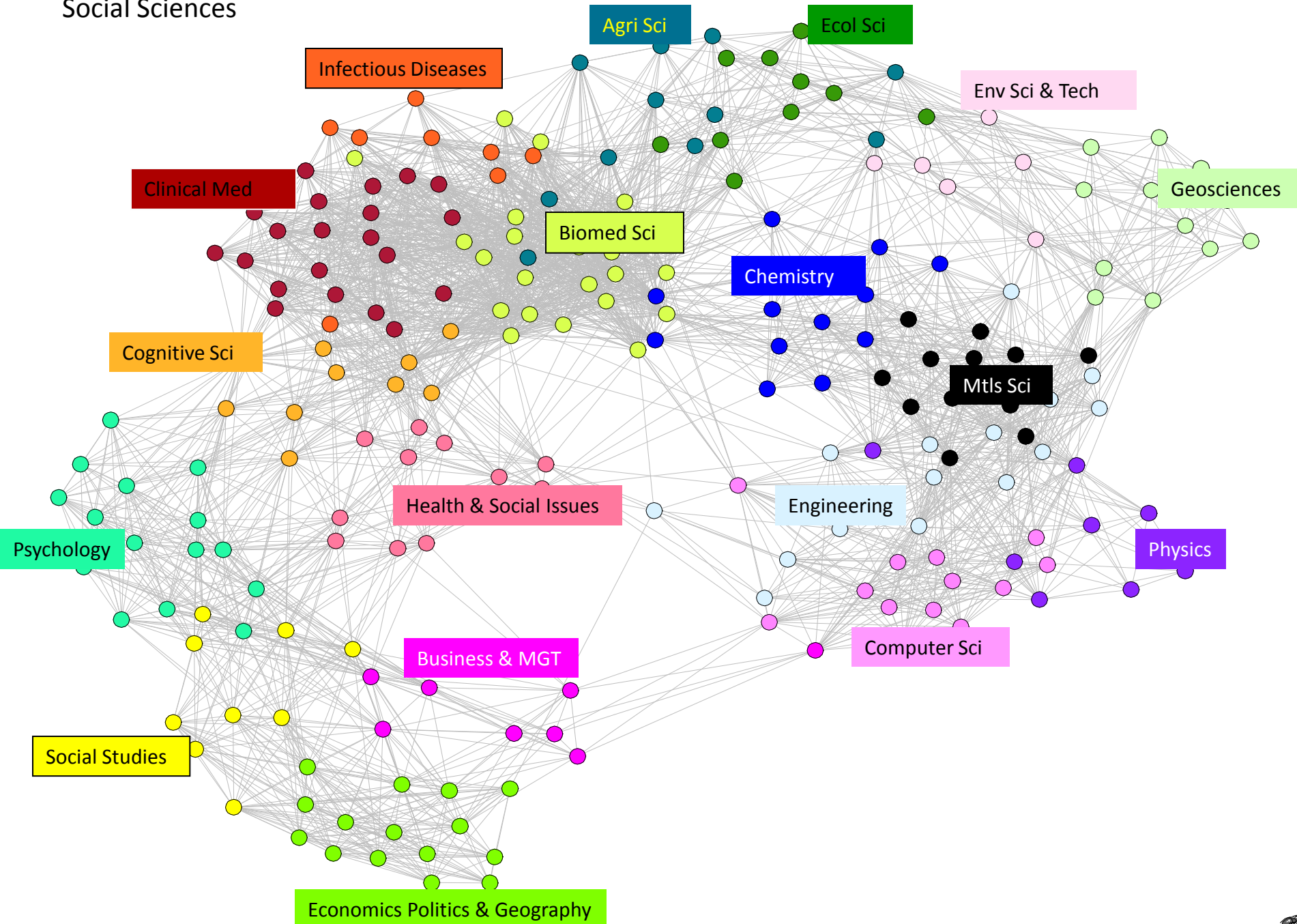
Web of Science –
Science Citation Index,
175 Subject Categories base map
Nano-2009 -- 93233 Documents

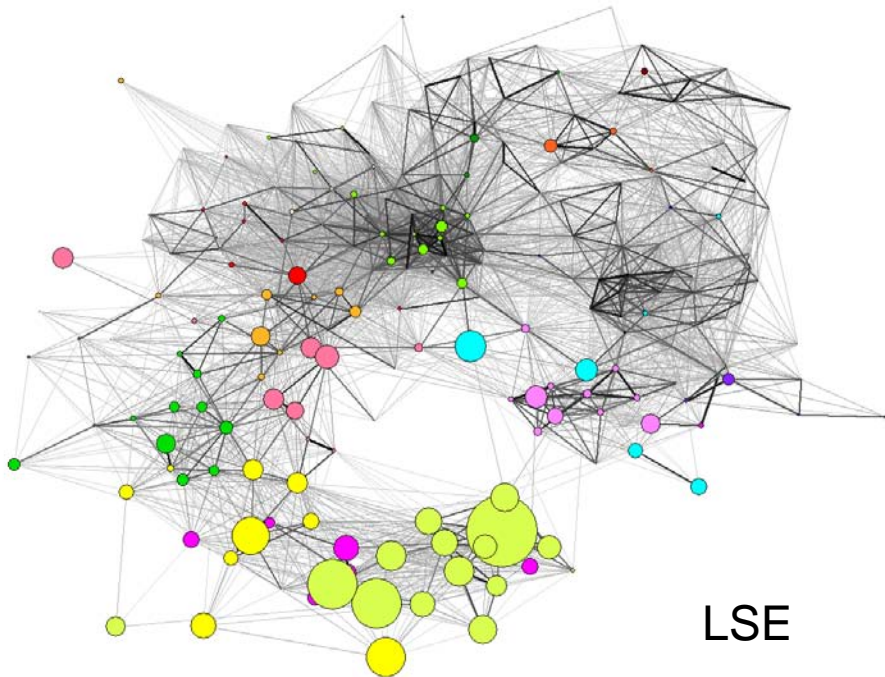
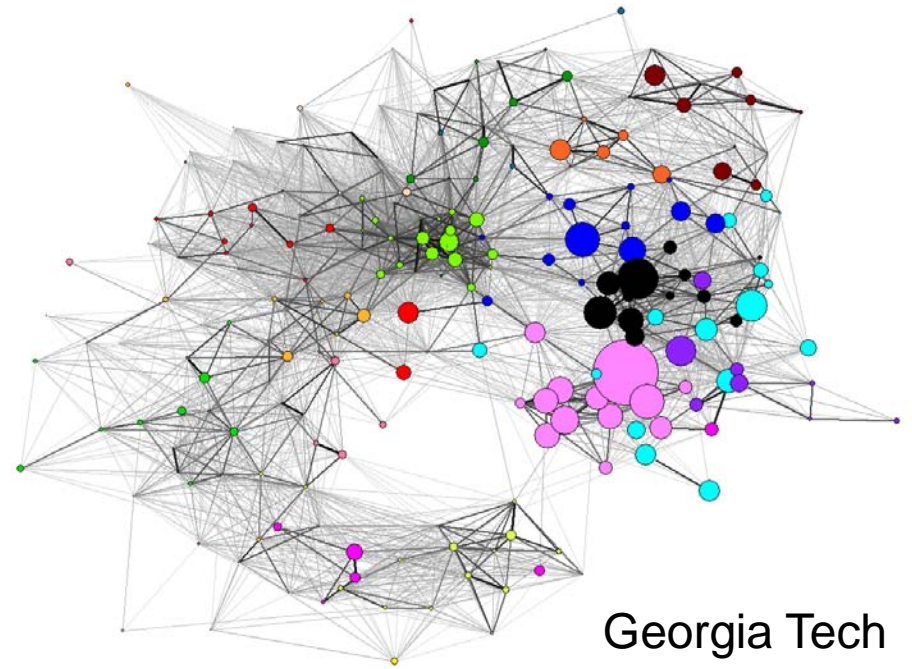
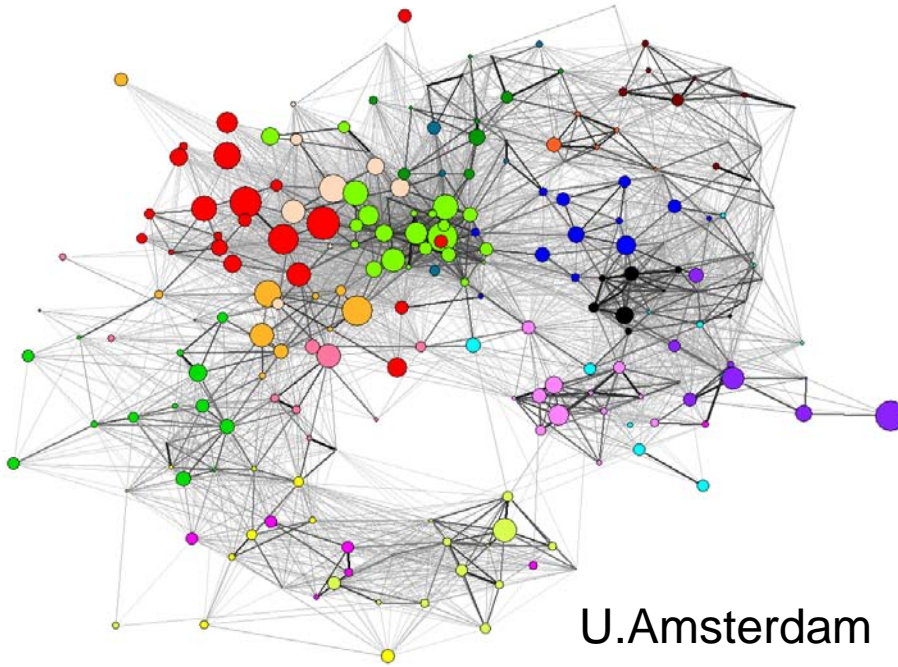


“Nanoscience & Nanotechnology” Subject Category –
3863 articles, partial 2008 –
showing **Top 40 Cited SCs** Overlay
over base 175 SC Science Map



221 SC Base Map – Sciences + Social Sciences





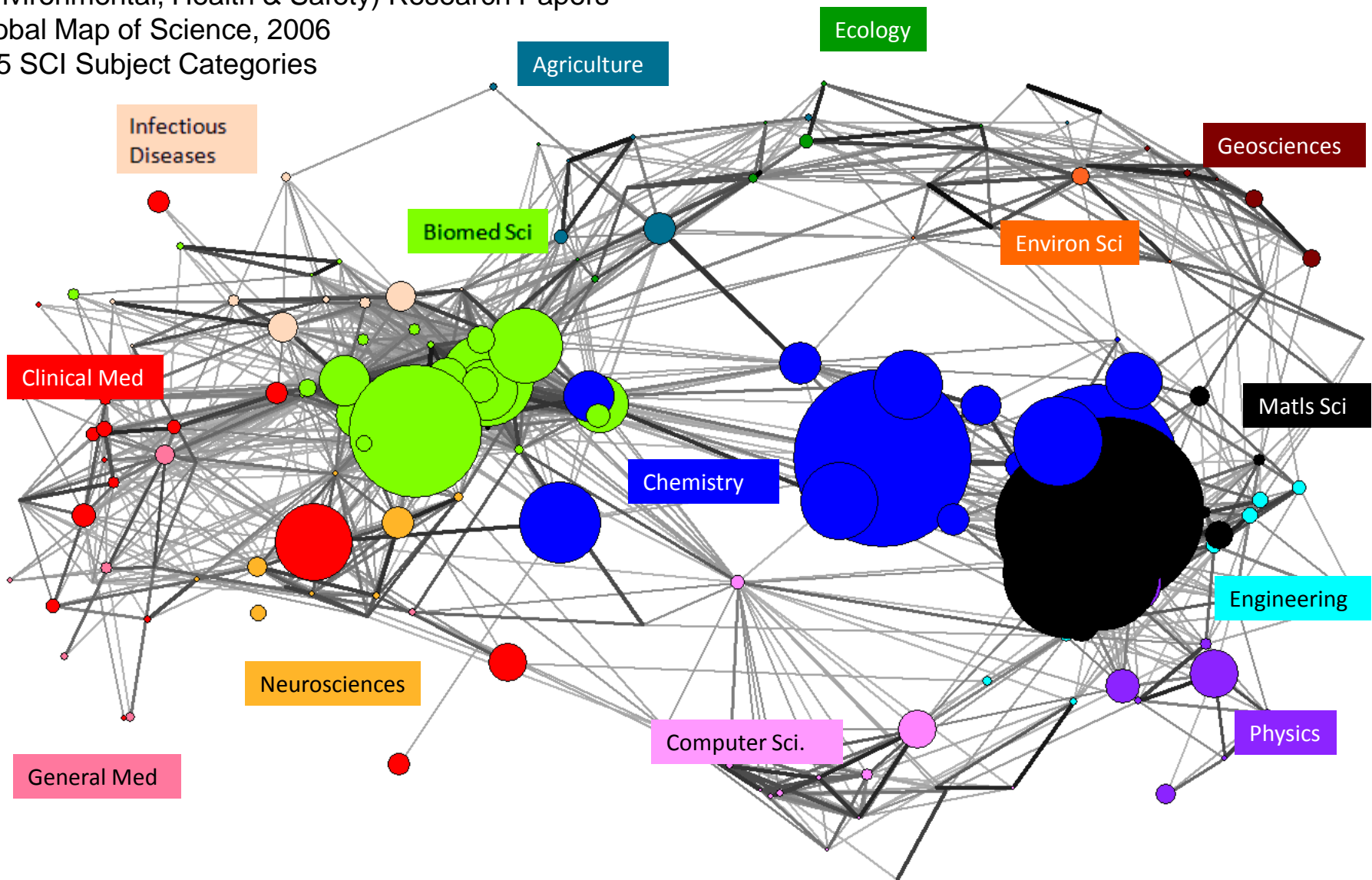
Ismael Rafols, 2010

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

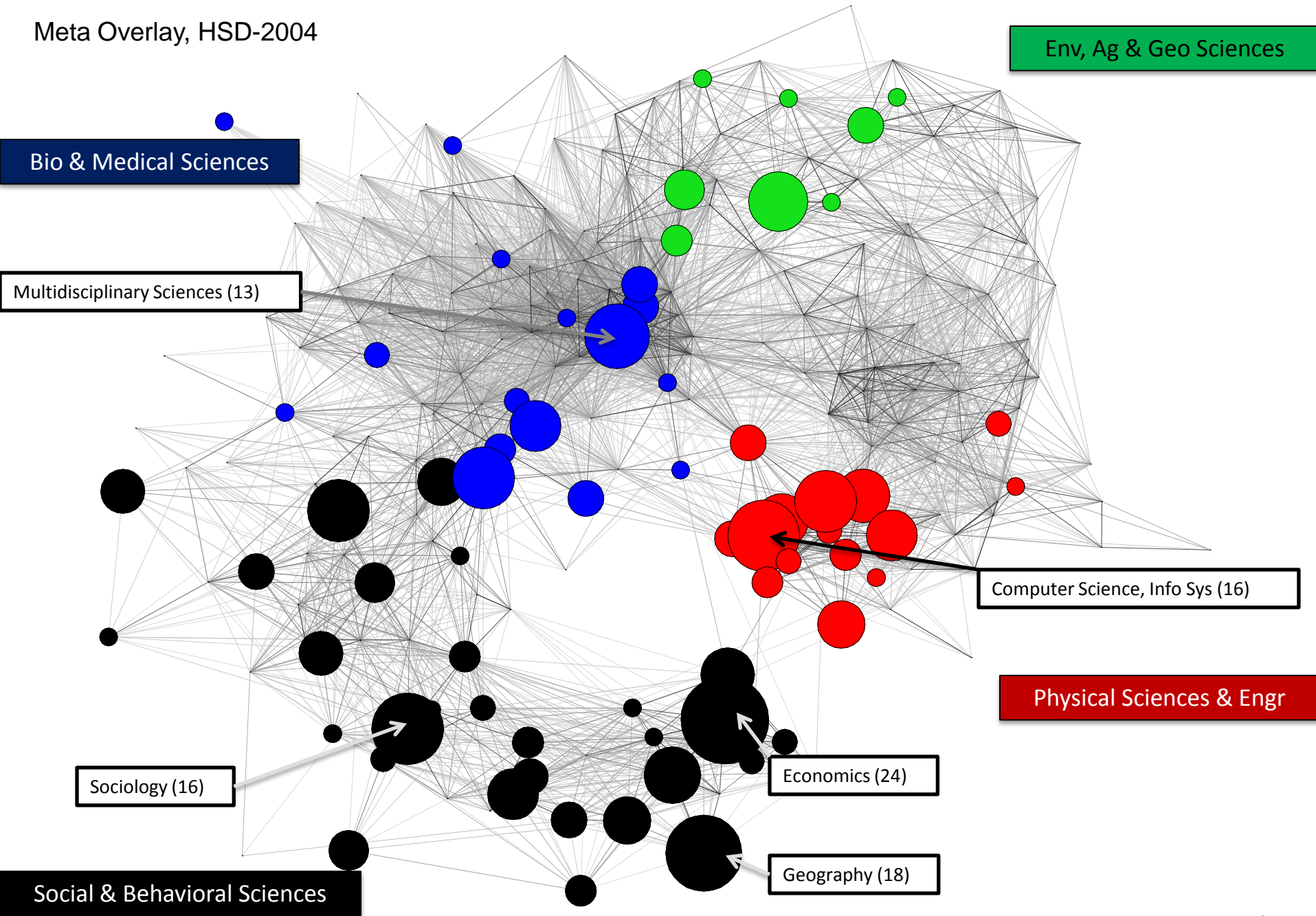
NSF Research Assessments

- **Nano Sci & Engr Ctrs Assessment (Juan Rogers, PI) (current)**
 - Interdisciplinarity, etc.
- **ROLE/REESE Programs (2008; current)**
 - Can proposal references describe research communities engaged (and aid review processes)?
 - Track networking via publications, their referencing, and citations to them?
 - (New) Track interplay with Cognitive Science
- **Research Cooperation Networks (RCN) Program (current)**
 - How interdisciplinary is the research?
 - Can we see network enrichment, Before to After?
- **Human & Social Dynamics (HSD) Program (current)**
 - Interdisciplinarity
 - Influence beyond the Social Sciences

Nano Journal Publications **Citing Nano-EHS**
(Environmental, Health & Safety) Research Papers
Global Map of Science, 2006
175 SCI Subject Categories



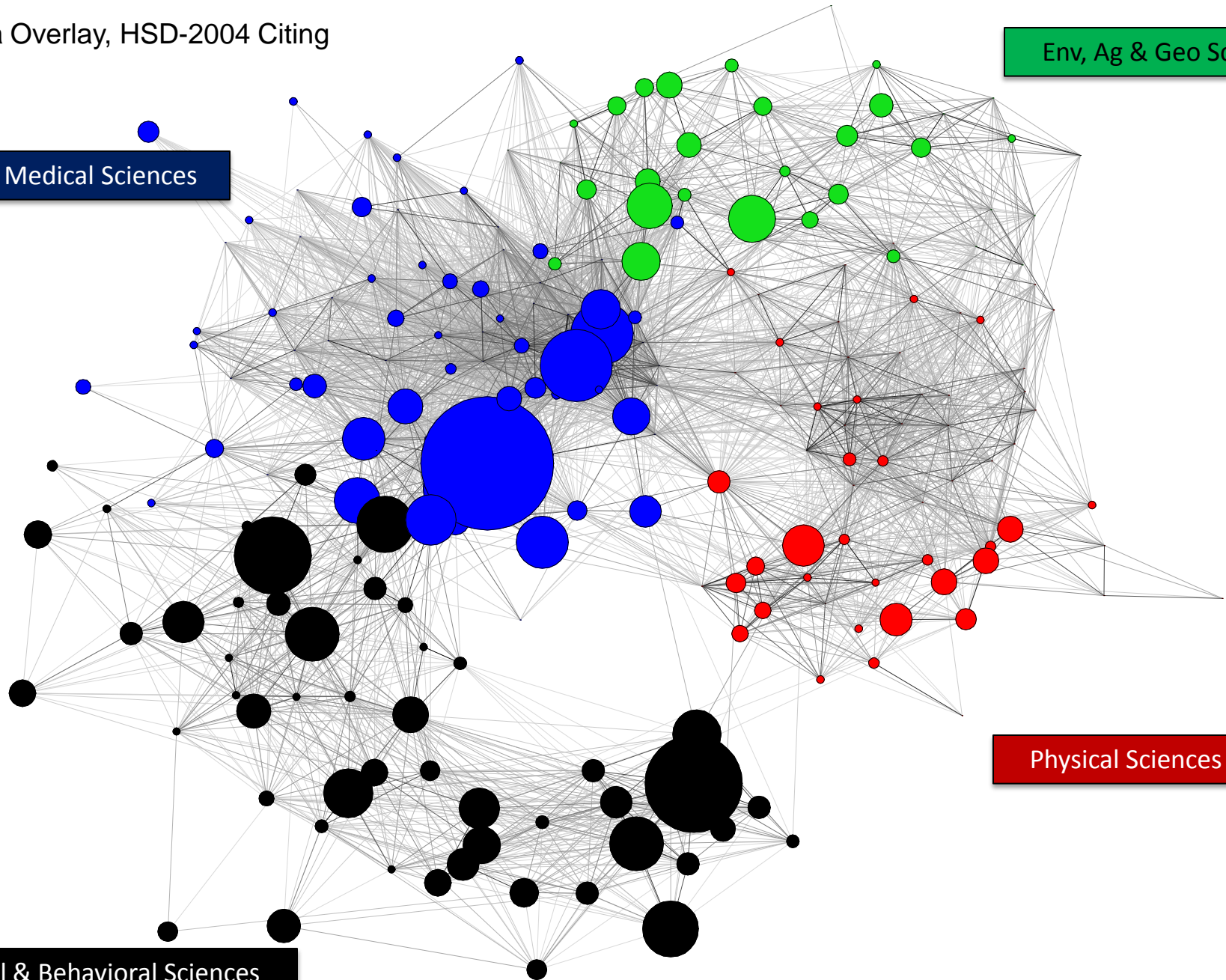
Meta Overlay, HSD-2004



Meta Overlay, HSD-2004 Citing

Env, Ag & Geo Sciences

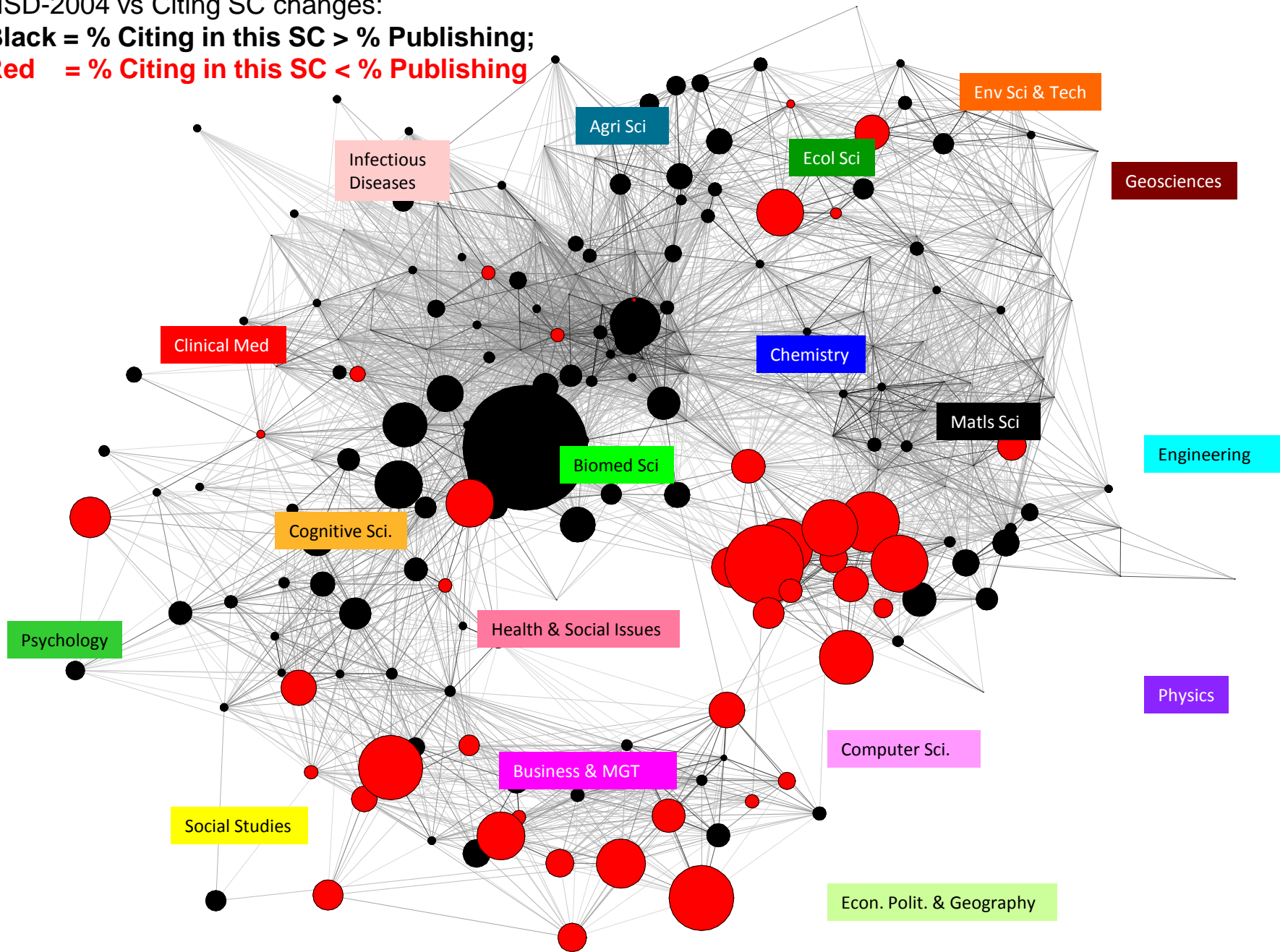
Bio & Medical Sciences



Physical Sciences & Engr

Social & Behavioral Sciences

HSD-2004 vs Citing SC changes:
Black = % Citing in this SC > % Publishing;
Red = % Citing in this SC < % Publishing



Ongoing Research

1. Study research **knowledge diffusion processes**
2. Explore our **Diffusion Indicator** (~inverse of Integration score)
3. Develop **Patent Mapping** (to complement Science Overlay Mapping)
4. Explore MeSH terms and **PubMed** mapping
5. Do in-depth **lab studies** to validate the metrics & maps
6. Case study on **more interdisciplinary** papers' influence?
7. Combine clustering & factoring in a **nanotaxonomy**; use to refine our nano search algorithms
8. Expand Multi-database "**Cross-charting**" to enrich "Tech Mining" (and R&D Profiling)
9. Devise more systematic "**Forecasting Innovation Pathways**"

Science Mapping References

Science Maps

- Chen, C. (2003) *Mapping Scientific Frontiers: The Quest for Knowledge Visualization*, Springer, London
- 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

<http://www.idr.gatech.edu/>

- 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*,
- 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

Research Assessment References

- Porter, A.L., Newman, N.C., Myers, W., and Schoeneck, D., Projects and Publications: Interesting Patterns in U.S. Environmental Protection Agency Research, *Research Evaluation*, Vol. 12, No. 3, 171-182, 2003.
- Porter, A.L., Schoeneck, D.J., Roessner, D., and Garner, J., Practical research proposal and publication profiling, *Research Evaluation*, under submission.

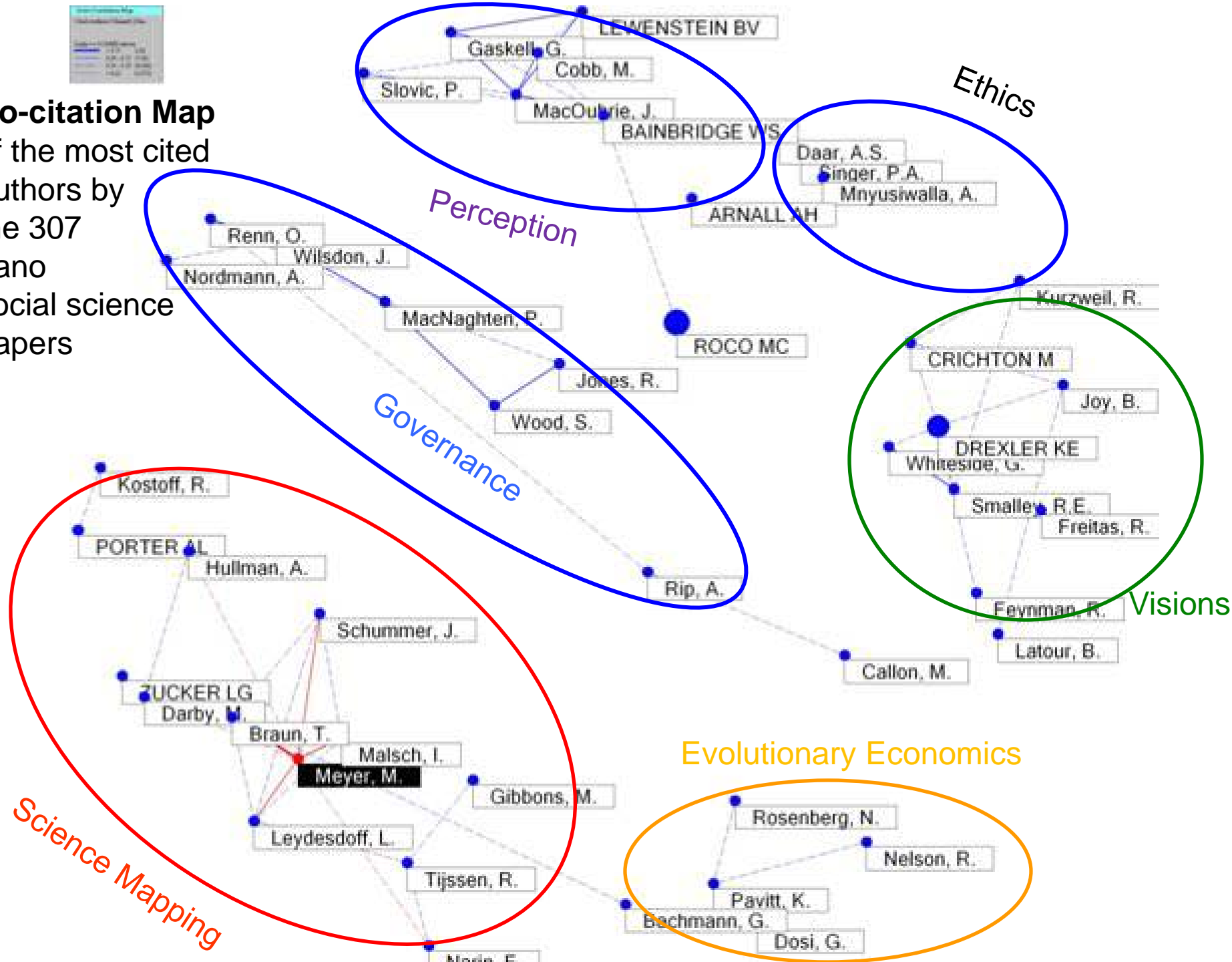
OUTTAKES beyond here

Resources

- **science overlay maps (can make your own there too)**
[//idr.gatech.edu/test/](http://idr.gatech.edu/test/)
- **The text mining software used:**
www.theVantagePoint.com
- **Tech Mining Workshop + Atlanta Conference on Science & Innovation Policy**
Atlanta, Sep 13-17, 2011



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



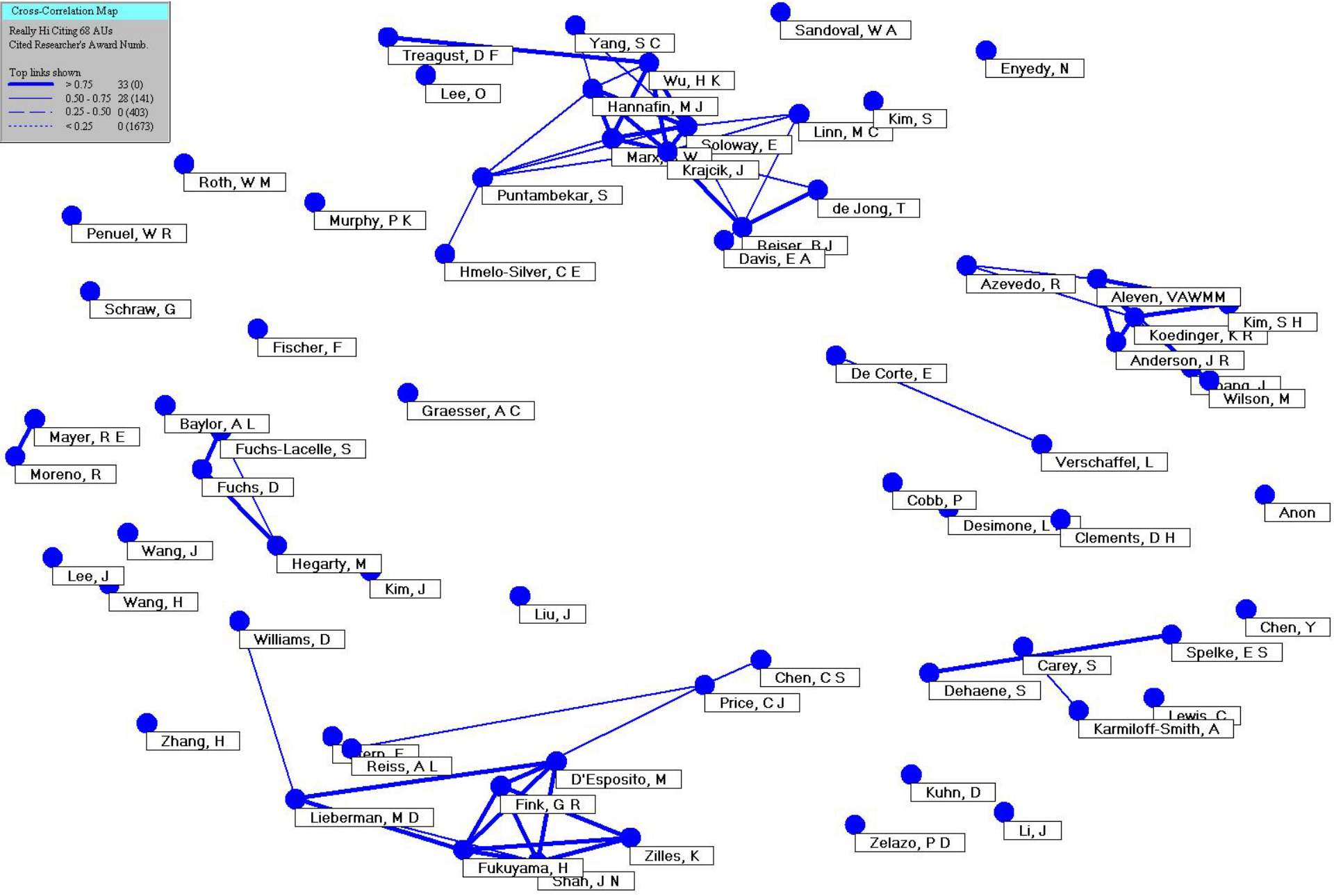
Highly Citing Authors (based on cited ROLE project coupling)

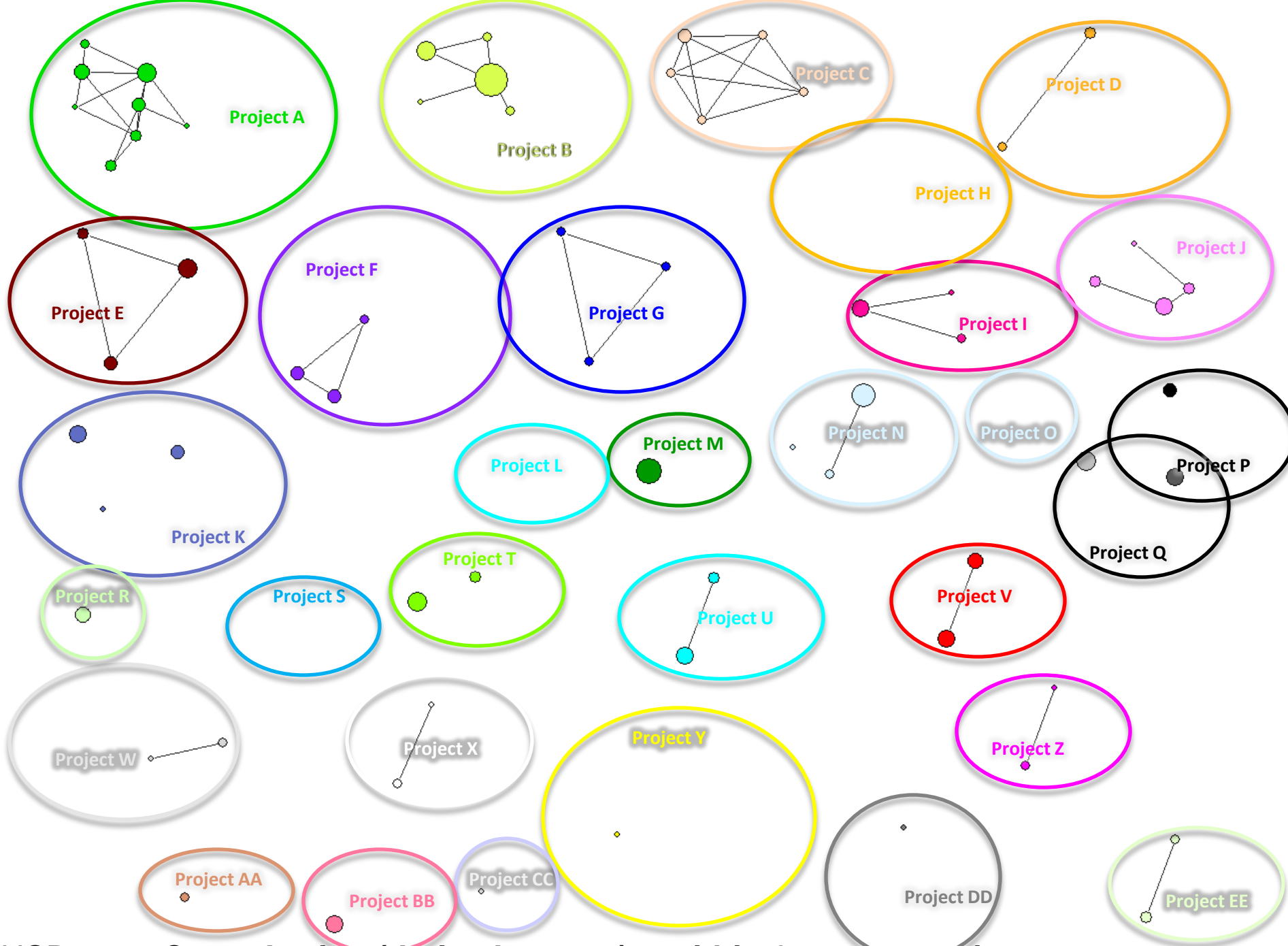
Cross-Correlation Map

Really Hi Citing 68 AUs
Cited Researcher's Award Numbr.

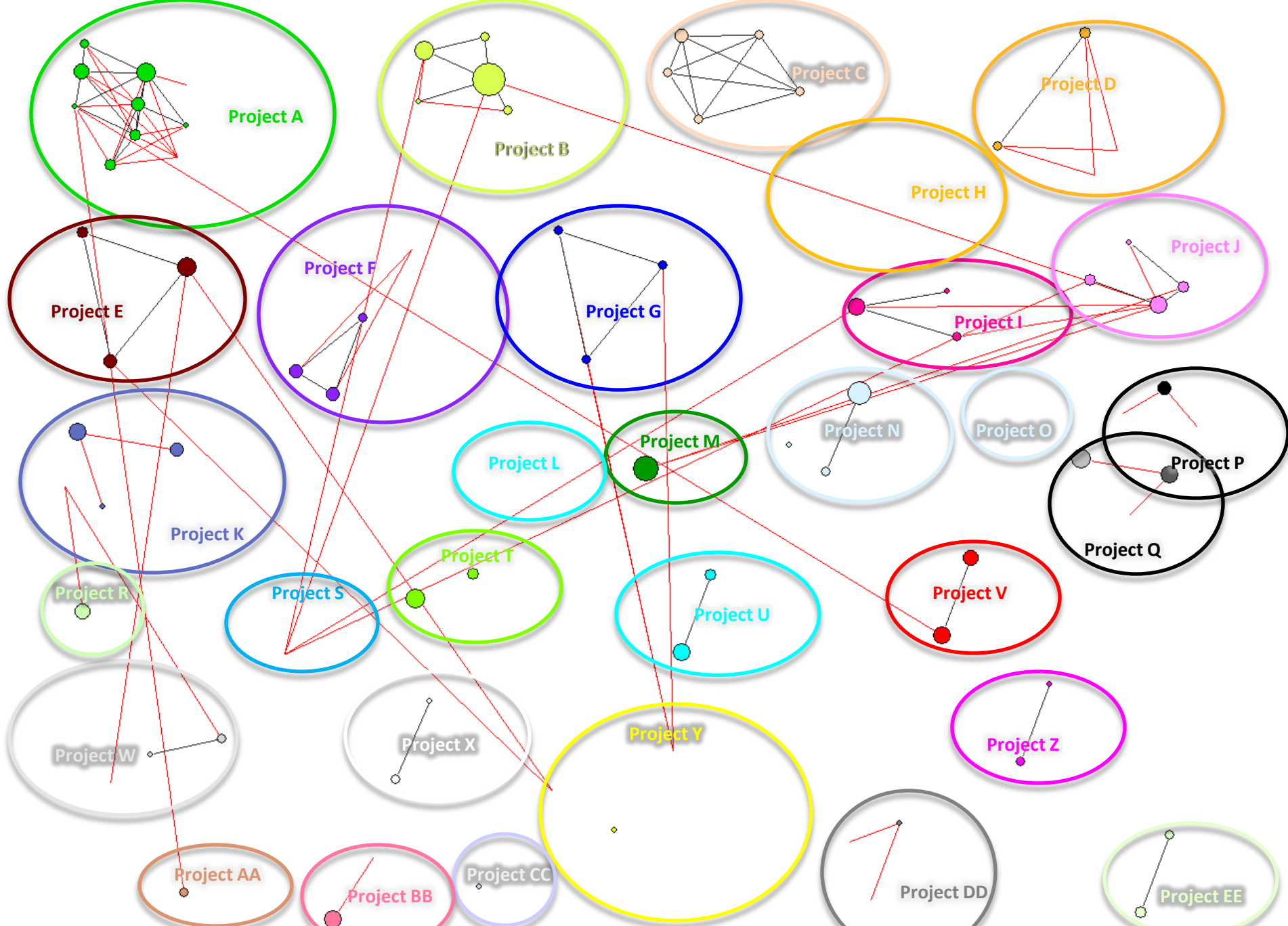
Top links shown

> 0.75	33 (0)
0.50 - 0.75	28 (141)
0.25 - 0.50	0 (403)
< 0.25	0 (1673)



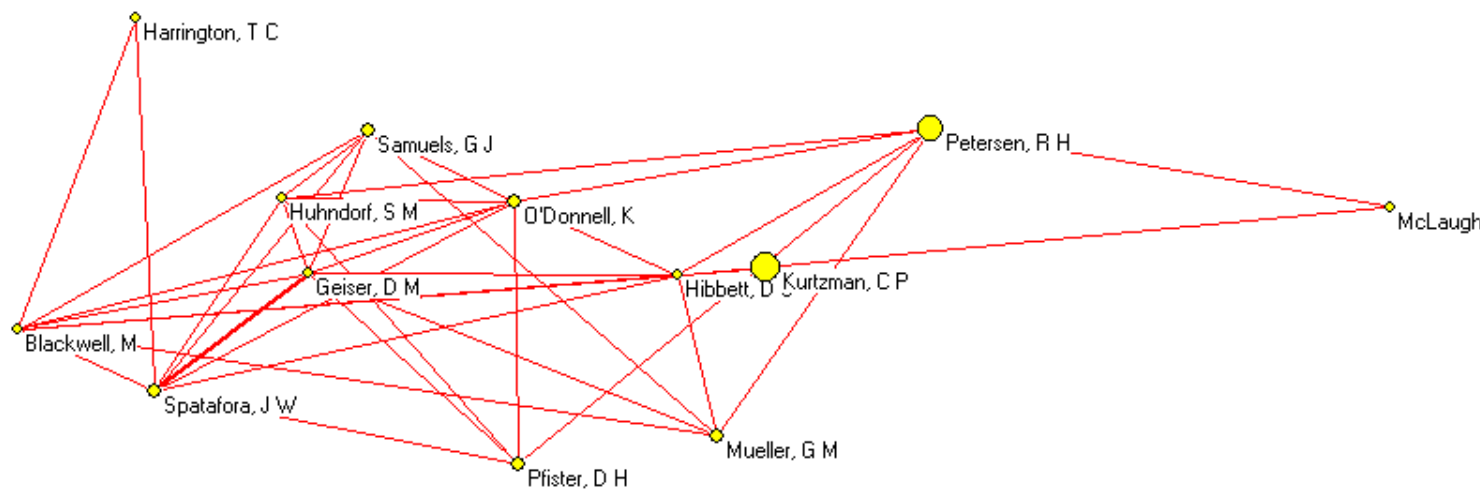
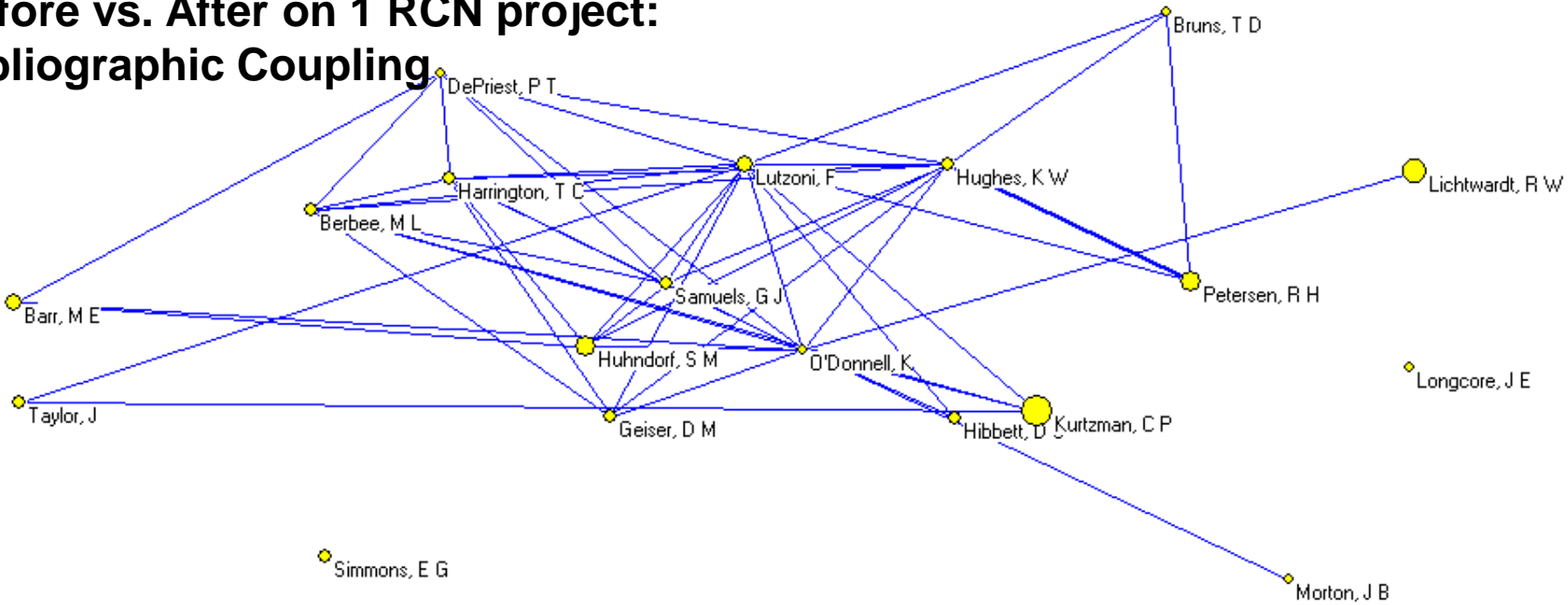


HSD-2004 Co-authoring (derived papers) – within & among projects

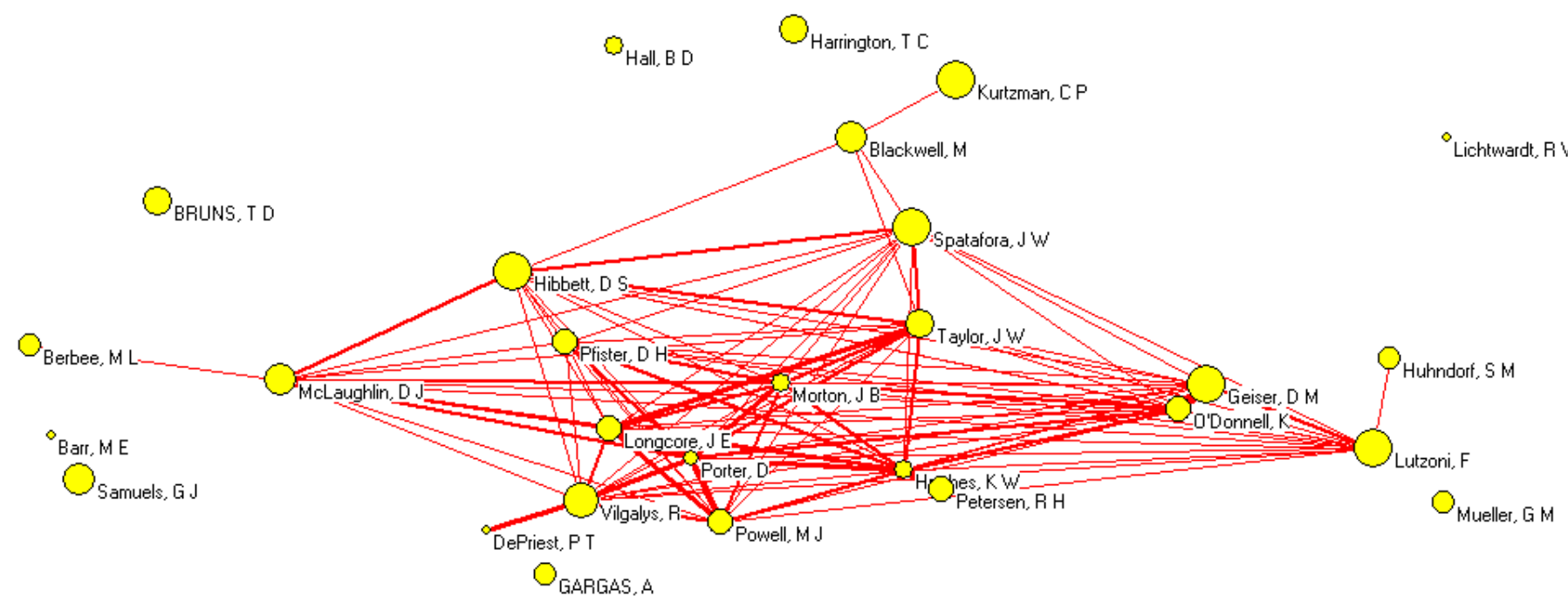
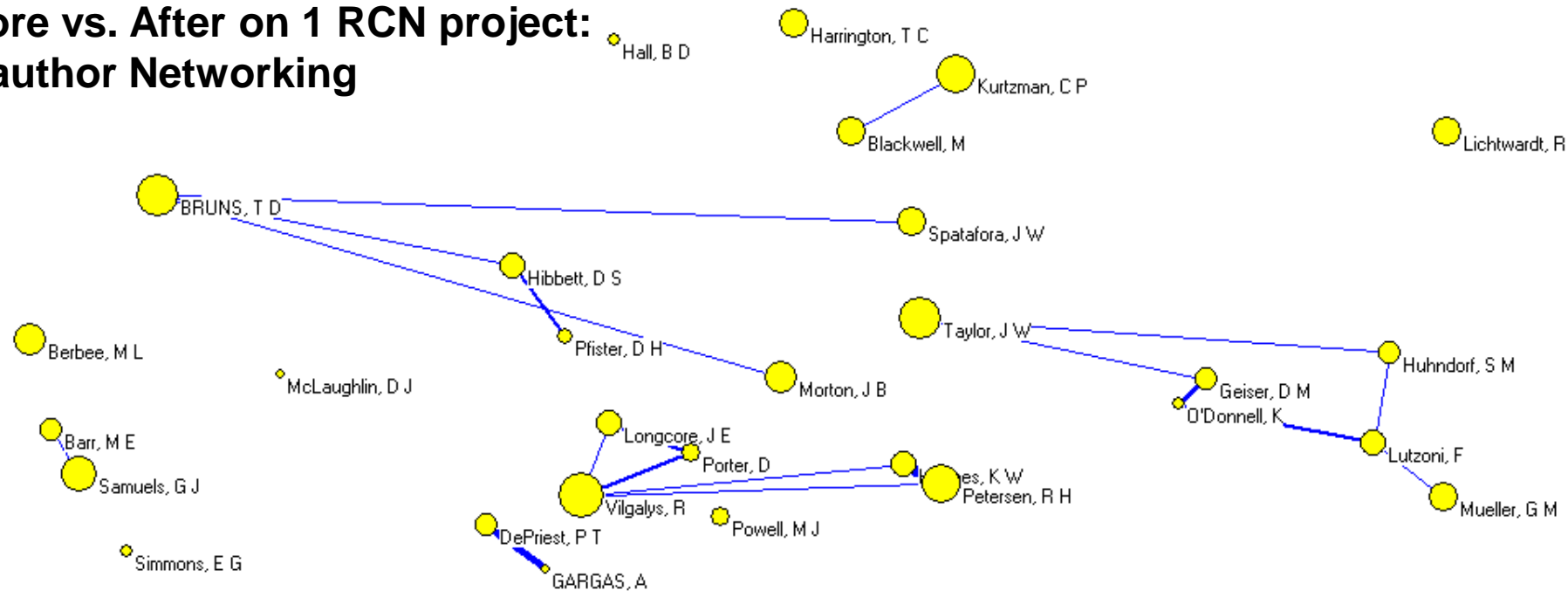


HSD-2004 Co-authoring + cross-citation (derived papers) – within & among projects

Before vs. After on 1 RCN project: Bibliographic Coupling



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



Interdisciplinarity References

National Academies Keck Futures Initiative: www.keckfutures.org

National Academies Committee on Facilitating Interdisciplinary Research, Committee on Science, Engineering and Public Policy (COSEPUP) (2005). *Facilitating interdisciplinary research*. (National Academies Press, Washington, DC).

Klein, J. T. (1996), *Crossing boundaries: Knowledge, disciplinarity, and interdisciplinarity*. (University Press of Virginia, Charlottesville, VA.).

Porter, A.L., Cohen, A.S., Roessner, J.D., and Perreault, M. Measuring Researcher Interdisciplinarity, *Scientometrics*, Vol. 72, No. 1, 2007, p. 117-147.

Porter, A.L., Roessner, J.D., and Heberger, A.E., How Interdisciplinary is a Given Body of Research?, *Research Evaluation*, Vol. 17, No. 4, 273-282, 2008.

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](https://doi.org/10.1007/s11192-008-2197-2)

[\[http://www.springerlink.com/content/c87k7754t4n24385/fulltext.pdf\]](http://www.springerlink.com/content/c87k7754t4n24385/fulltext.pdf)

Behavior of our Integration Score (“I”)

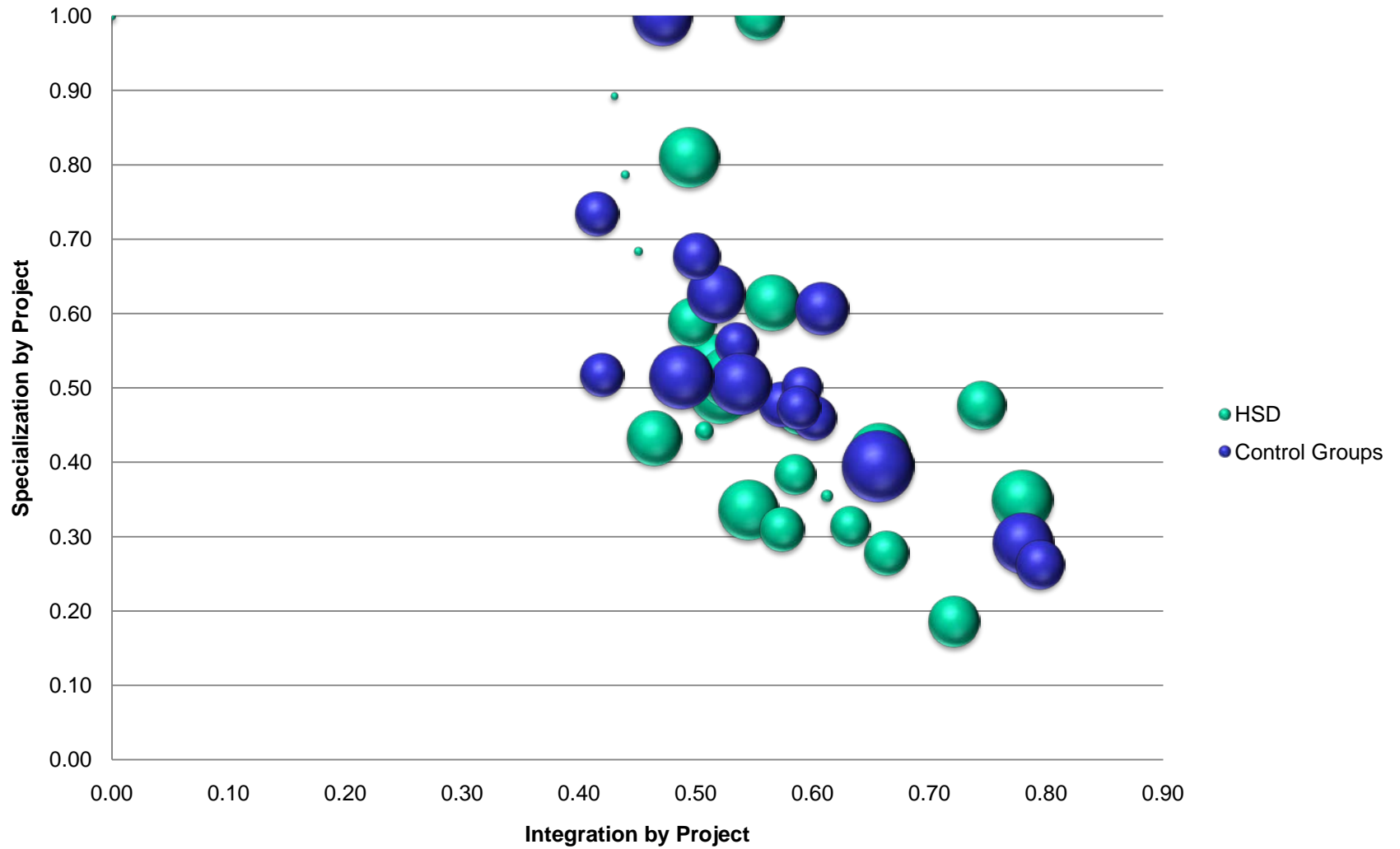
A paper cites N references [considering only those to journals indexed by WOS), and associated with journals associated with 1 or more Subject Categories (SCs)]. Integration varies:

1. As the # of Cited SCs increases (higher **Variety**), I increases
2. As the **Balance** among those Cited SCs increases, I increases
3. As the **Disparity** among those Cited SCs increases [less tendency for them to be co-cited by national samples of articles], I increases

Ranges from:

0 (stand alone research that cites work from a single SC) to
1 (highly Integrative research drawing from multiple,
~unrelated SCs)

HSD vs Control

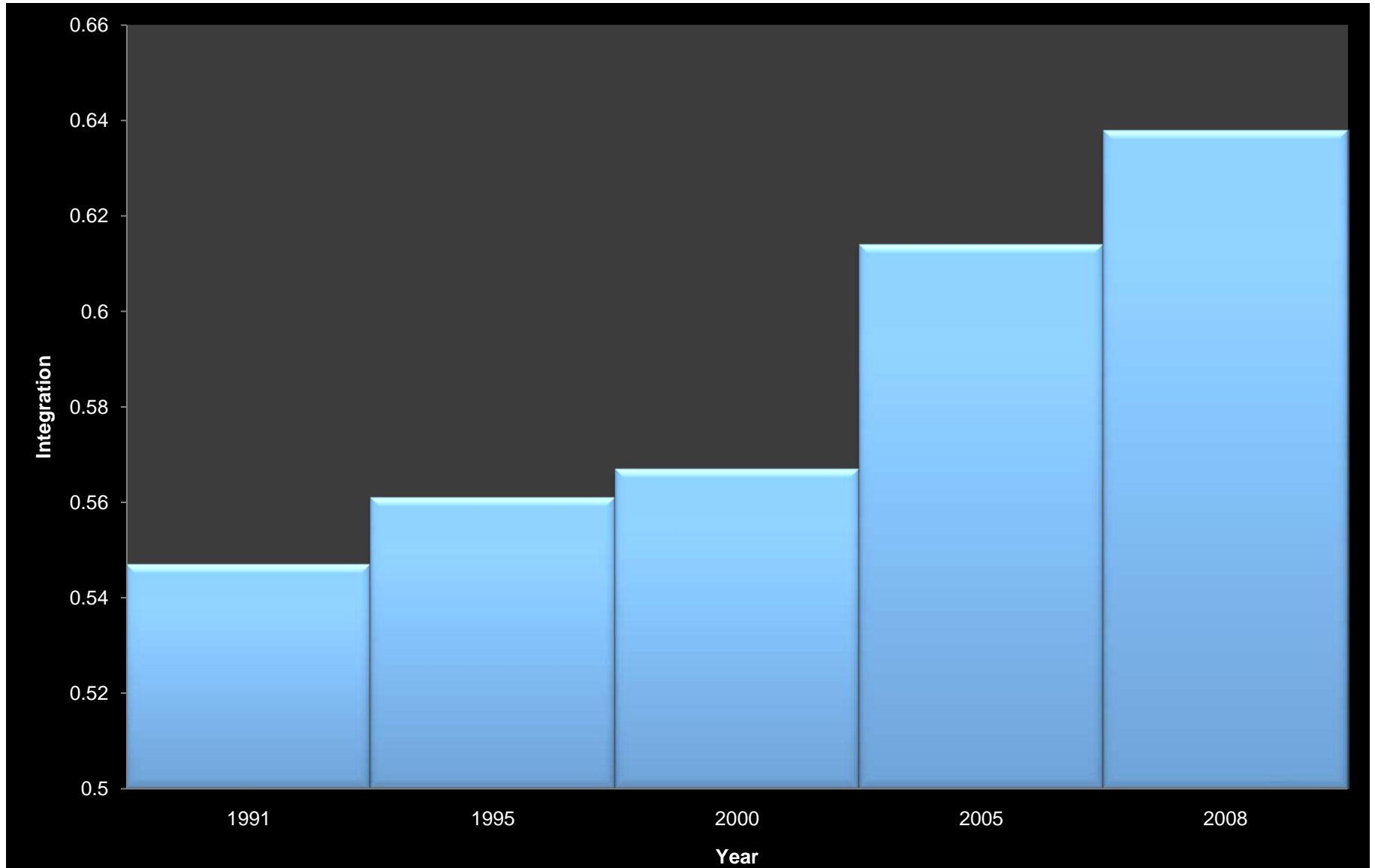


Percentage of Articles in each Macro-Discipline (rows) Citing at least one source in a given Macro-Discipline (columns)

# Pubs	Macro-Disciplines: Publications \ Cited	MtIs Sci	Chemistry	Physics	Biomed Sci	Engr Sci	Clinical Med	Computer Sci	Agri Sci	Env Sci & Tech	Ind Engr/Mgt Sci (including Math)
19301	MtIs Sci	98	77	57	58	44	12	22	16	12	36
7020	Chemistry	91	96	53	77	33	19	13	20	13	29
2989	Physics	89	68	90	56	29	8	28	5	6	22
2647	Biomed Sci	51	83	24	94	19	47	6	34	18	12
2503	Engr Sci	95	74	48	54	81	11	19	24	28	32
973	Clinical Med	65	69	20	81	24	93	7	22	12	35
872	Computer Sci	89	47	63	36	34	5	75	3	3	21
793	Agri Sci	77	82	22	71	52	18	5	81	38	26
630	Env Sci & Tech	71	75	26	62	64	20	5	46	92	19
158	Ind Engr/Mgt Sci (including Math)	86	41	38	32	67	13	27	9	11	72

Average Integration Scores for Samples of Nano Articles

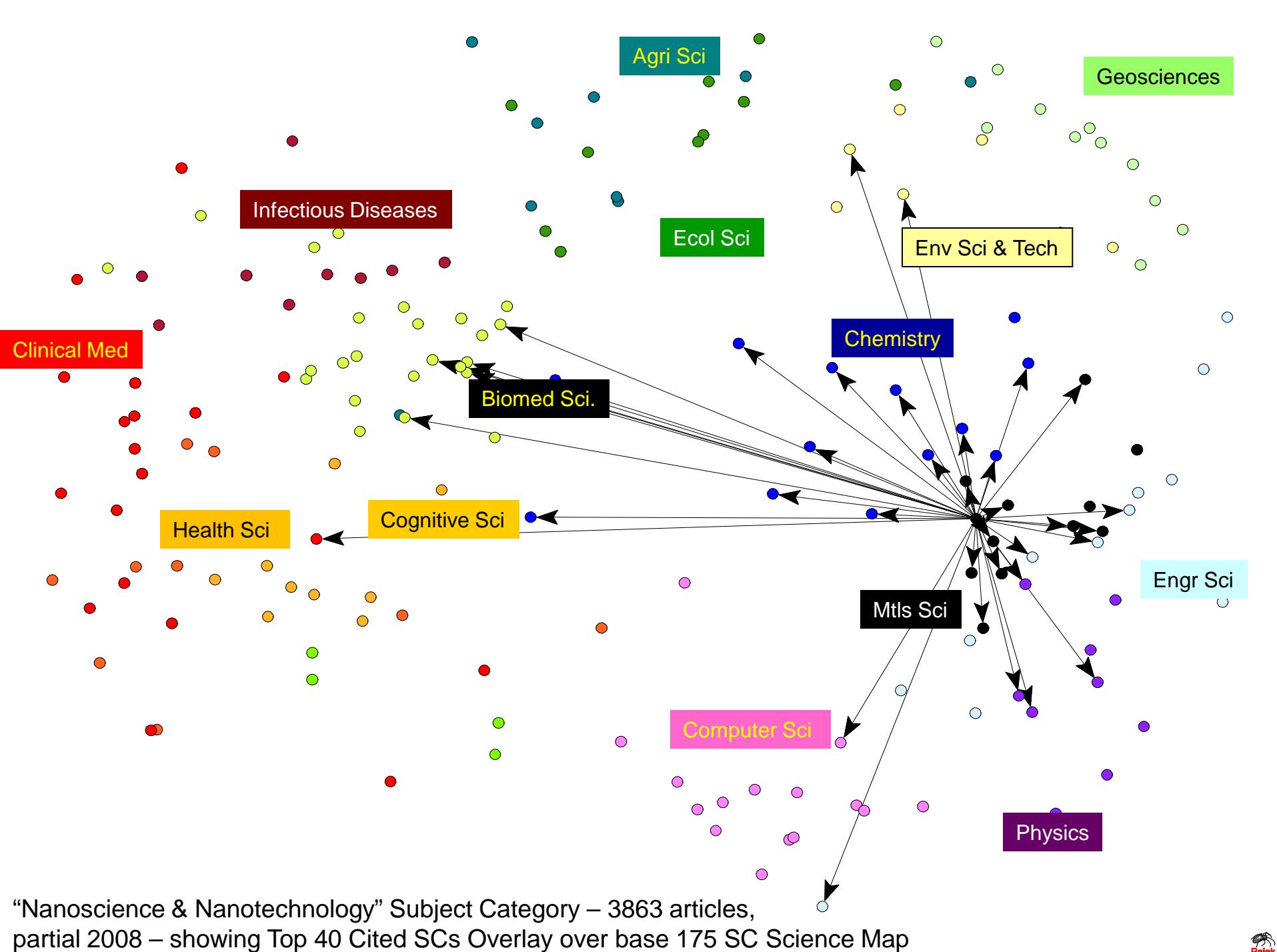
[Benchmark: 1985-2005, 6 SCs increased average of 0.03 – Here 1991-2008, increase of ~0.09] (old version scoring)





Science Mapping

- Based on Bibliometrics
- Since the 1970's
- But usually local – e.g., research networking in a specific research arena
- Recently, mapping “all” of science – Klavans, Boyack, Borner; Chen; Leydesdorff, Rafols, Meyer, Porter
- ~Robust to different data and representations



SC Issues

- Studies indicate ~50% disagreement on journal assignment
 - Researchers' views on what belongs in their field
 - Alternative categorizations
- But ~good news
 - Science maps show general pattern consistency
 - Disagreements are usually within a science neighborhood
 - So, both our Integration measure and maps are not overly sensitive to misclassifications
- For our analyses, recognize that we use SCs, thereby assigning articles to categories based on the journal in which they appear (not upon article content)

Integration Score

$$I = 1 - \left[\frac{\left[\sum (f_i \times f_j \times \cos(SC_i - SC_j)) \right]}{\sum (f_i \times f_j)} \right] \quad \text{Porter et al. (2007)}$$

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

“cos (SC_i – SC_j)” 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,**

$$I = 1 - \sum_{i,j} p_i p_j s_{ij} \quad \text{Rafols and Meyer (2009)}$$

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

[This measure is basically 1 – Stirling D.]

Specialization Formula:

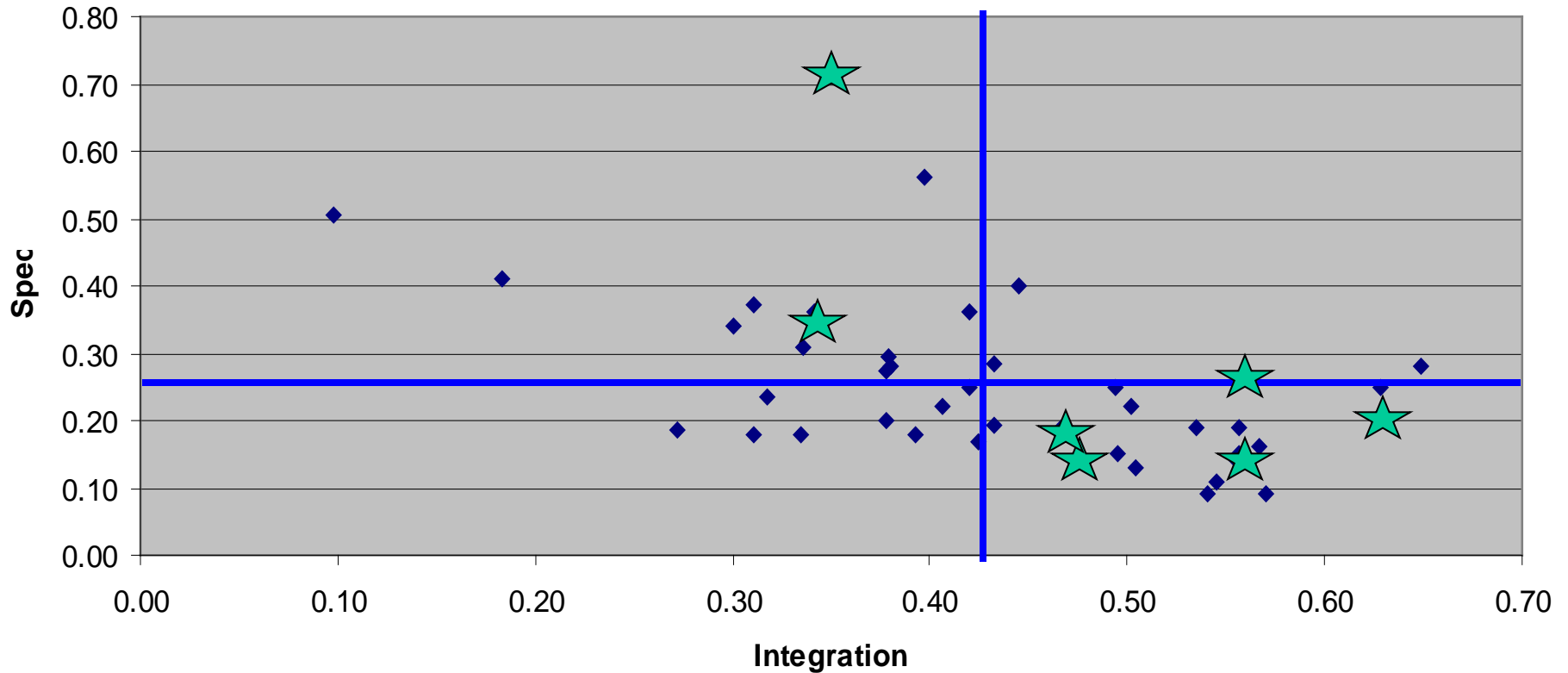
$$S = \left[\frac{\left[\sum (f_i \times f_j \times \cos(SC_i - SC_j)) \right]}{\sum (f_i \times f_j)} \right]$$

Where the **Subject Categories (SCs 1--n)** are those reflecting the journals in which the set of papers was published.

Note that Specialization only pertains to a set of papers – e.g., a researcher’s work during a given time period.

The cosine values for the SC x SC matrix derive from general samples (not from the body of research under study).

Specialization vs. Integration for the Publication Outputs of 43 Researchers [7 women starred]



Integration Scores for 22 Sample Subject Categories (based on samples of ~100 papers published in 2005 associated with each indicated SC)

Subject Category	Integration (Mean)	Integration (SD)
Engineering, Biomedical	0.69	0.13
Materials Science, Biomaterials	0.68	0.11
Geriatrics & Gerontology	0.68	0.08
		0.11
Clinical Neurology	0.61	
Medicine, Research & Experimental	0.61	0.13
Engineering, Chemical	0.61	0.14
		0.13
Biochemistry & Molecular Biology	0.53	
Statistics & Probability	0.53	0.2
Engineering, Electrical & Electronic	0.52	0.19
Average	0.61	0.13
Standard Deviation	0.05	0.03

~ Macro-Disciplines of the Proposal References

6200 Total Number of Cited References:		848	2514	1249	1589
No. of Records	Cited Macro-Disciplines	Linked Project Year			
		2000	2002	2004	2006
2511	Education & Educational Research	47.1%	40.7%	45.4%	32.9%
1503	Scientific Disciplines Education	27.2%	25.5%	26.7%	18.8%
712	Social Psychology & Related	8.4%	13.6%	10.2%	10.8%
665	Computer Sciences	16.9%	10.1%	7.0%	11.4%
614	Educational Psychology	10.3%	11.7%	8.6%	8.0%
589	Cognitive & Behavioral Sciences	10.3%	10.5%	8.8%	8.1%
188	Engineering Sciences	0.6%	0.2%	2.6%	9.1%
125	Policy Sciences	1.5%	2.0%	2.9%	1.6%
112	Physics	1.1%	1.4%	0.5%	3.9%
103	Biomedical Sciences	1.1%	0.8%	1.2%	3.7%
89	Chemistry	0.9%	1.0%	2.5%	1.6%
84	Special Education	0.2%	2.3%	1.3%	0.5%
80	Ethical & Social Issues	0.4%	0.5%	1.0%	3.3%
79	Clinical Medicine	0.4%	2.4%	0.6%	0.6%
75	Industrial Engineering & Management Sciences	0.4%	1.2%	1.6%	1.4%
67	Ecological Sciences	0.6%	0.5%	0.2%	3.0%
66	Mathematics Across Disciplines	0.4%	1.0%	1.6%	1.1%
56	Environmental Sciences & Technology	0.1%	0.0%	0.1%	3.3%
42	Health Sciences	0.1%	0.7%	0.4%	1.2%

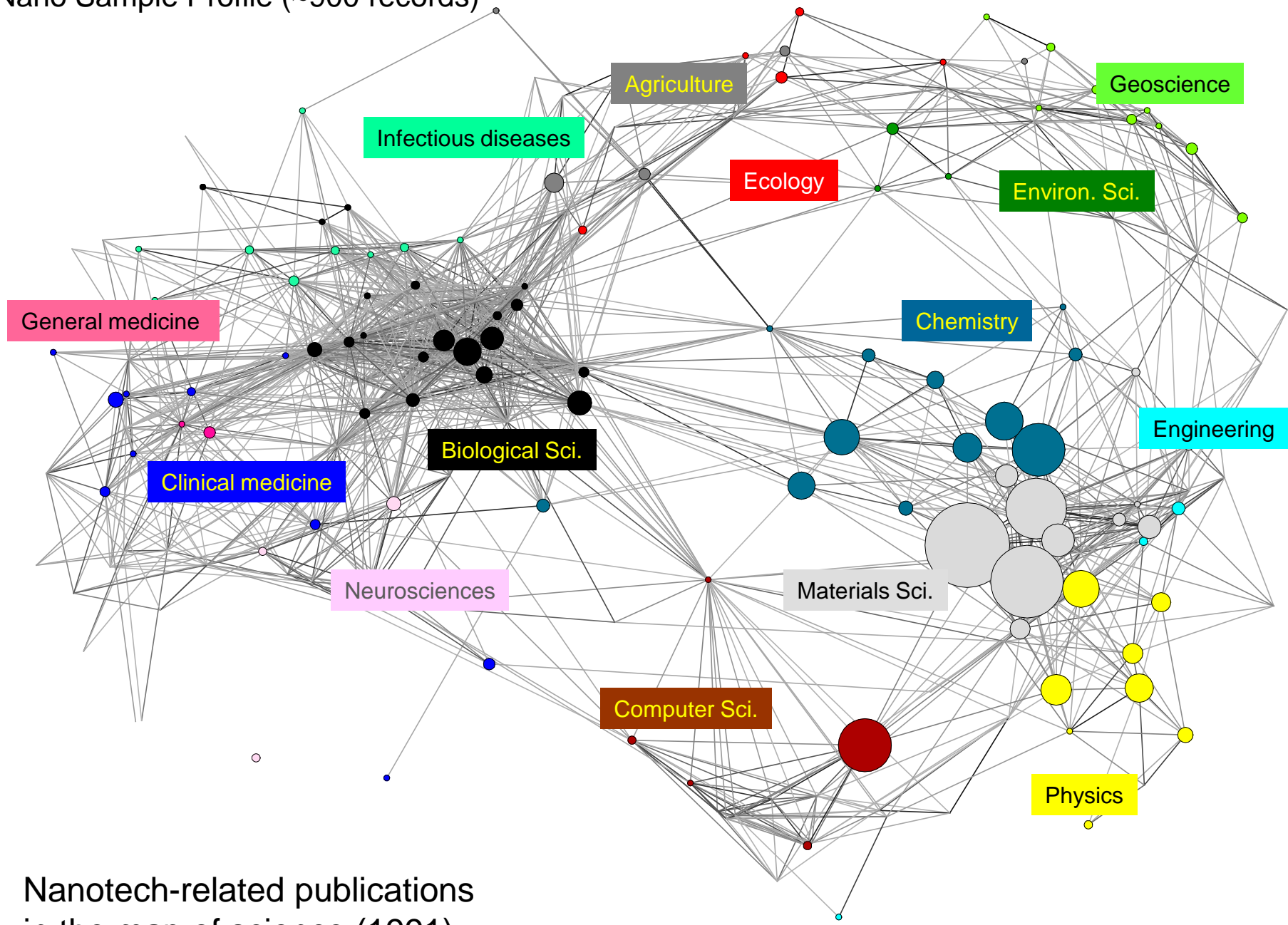
How Interdisciplinary is this research?

- “Integration” score – the more diverse the Subject Categories cited, the higher the integration
- Average for 91 proposals = 0.57
- Average for 32 WOS articles deriving from ROLE support = 0.62
- Benchmark analyses of large sets of WOS articles in 6 Subject Categories
 - Math = 0.29 [“disciplinary”]
 - Other 5 average = 0.62 [quite interdisciplinary]

Mining NSF Proposal Data

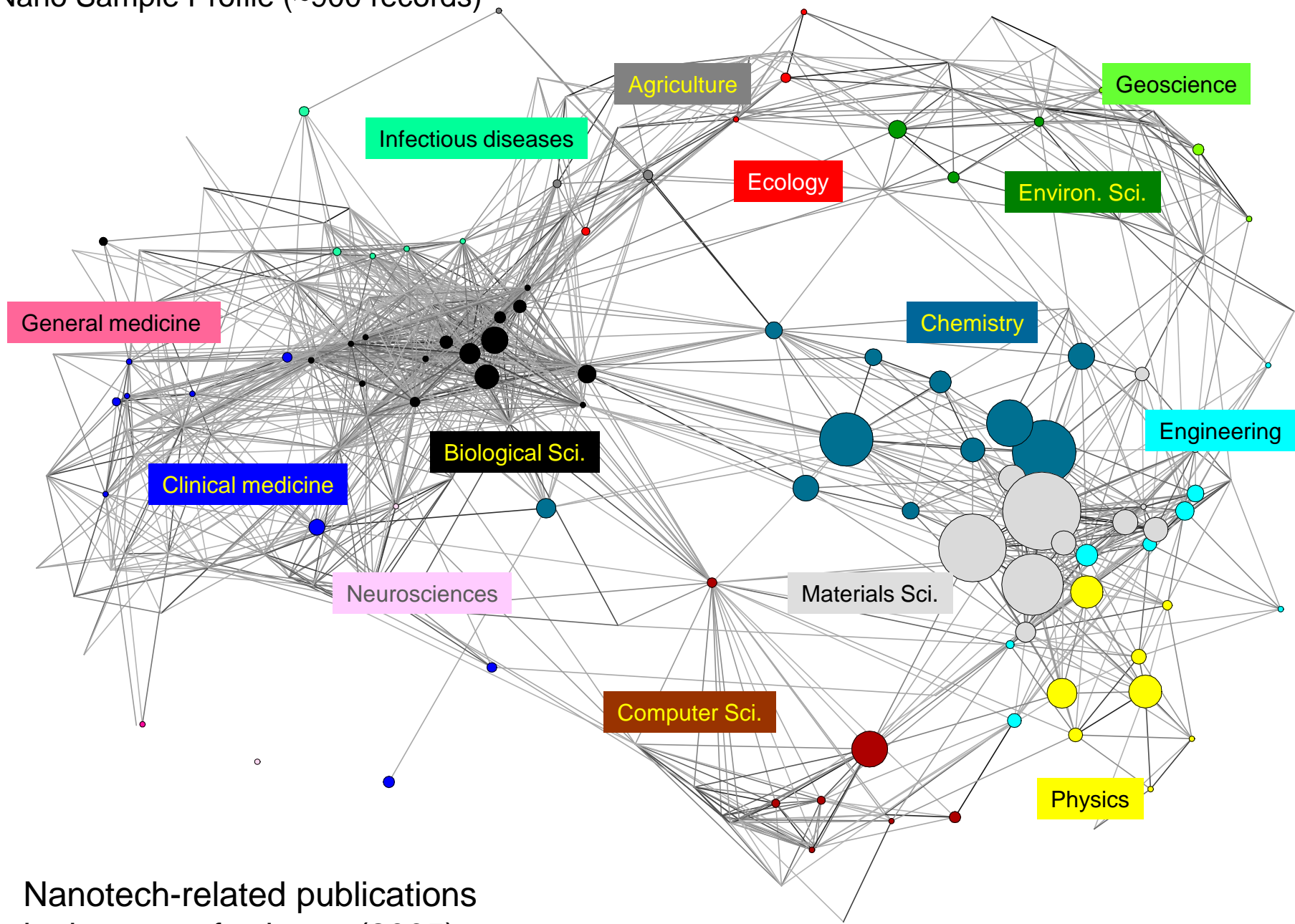
- We keyed on proposal references
 - Identify widely cited researchers
 - Characterize participating research communities
- Also used biosketches to help identify publications and citations in WOS
- Could also do content analyses via text mining on proposal descriptions
- Potential value here
 - To understand research networks & participation patterns
 - For NSF Program Managers?

Nano Sample Profile (~900 records)



Nanotech-related publications
in the map of science (1991)

Nano Sample Profile (~900 records)



Nanotech-related publications in the map of science (2005)

Relationships among Proposals (based on shared macro-disciplines referenced)

