

Open Tool Development and Physics Career Trajectories across Geo and Science Space

Robert P. Light

*Senior Systems Analyst, Database Administrator
Cyberinfrastructure for Network Science Center
Indiana University*

June 16, 2014

Research Team at IU

- Dr. Katy Börner, Principal Investigator
- Robert Light, Senior Systems Analyst, Database Administrator
- Ted Polley, Research Assistant
- Michael Ginda, Research Assistant



Research Team at UC Davis

- Dr. Jim Crutchfield, Co-Principal Investigator
- Dr. Dowman Varn, Postdoc



Research Goals

- Develop and demonstrate that free and open tools can be used to visualize career trajectory in geographic, network, and topic space.
- Establish the value and uses of publication data in analyzing and modelling career trajectories.

Today, we present results published in

Börner, Katy, and David E. Polley. 2014. "Replicable Science of Science Studies". In *Measuring Scholarly Impact - Methods and Practice*, edited by Ding, Ying, Ronald Rousseau, and Dietmar Wolfram, Ch. 12. Springer Verlag.

Plus recent work.

Dataset Details

The core dataset was provided by Vincent Larivière from the University of Montreal.

Source – Web of Science

Coverage –Top 10,000 physicists in terms of productivity from 1980-1987

Timeframe covered – 1988-2010

	A	B	C	D	E	F	G
1	Name	Year	Institution	Country	Journal	Papers	Citations
2465	AGAKISHIEV-GN	1992	DUBNA-JOINT-NUCL-R	USSR	SOVIET JOURNAL OF NUCLEAR PH	1	3
2466	AGAKISHIEV-GN	1993	JOINT-INST-NUCL-RES	RUSSIA	PHYSICS OF ATOMIC NUCLEI	1	6
2467	AGARWAL-GS	1988	UNIV-HYDERABAD	INDIA	PHYSICAL REVIEW A	7	209
2468	AGARWAL-GS	1988	UNIV-HYDERABAD	INDIA	IEEE JOURNAL OF QUANTUM ELE	1	9
2469	AGARWAL-GS	1988	UNIV-HYDERABAD	INDIA	JOURNAL OF THE OPTICAL SOCI	3	250
2470	AGARWAL-GS	1988	UNIV-HYDERABAD	INDIA	OPTICS LETTERS	1	15
2471	AGARWAL-GS	1989	UNIV-HYDERABAD	INDIA	OPTICS COMMUNICATIONS	3	53
2472	AGARWAL-GS	1989	UNIV-HYDERABAD	INDIA	PHYSICAL REVIEW A	6	206
2473	AGARWAL-GS	1990	UNIV-HYDERABAD	INDIA	OPTICS COMMUNICATIONS	1	28
2474	AGARWAL-GS	1990	UNIV-HYDERABAD	INDIA	PHYSICAL REVIEW A	7	376
2475	AGARWAL-GS	1991	UNIV-HYDERABAD	INDIA	PHYSICAL REVIEW LETTERS	2	177
2476	AGARWAL-GS	1991	UNIV-HYDERABAD	INDIA	PHYSICAL REVIEW A	6	305
2477	AGARWAL-GS	1991	UNIV-HYDERABAD	INDIA	ADVANCES IN ATOMIC MOLECU	1	3
2478	AGARWAL-GS	1991	UNIV-HYDERABAD	INDIA	JOURNAL OF THE OPTICAL SOCI	1	79
2479	AGARWAL-GS	1991	UNIV-HYDERABAD	INDIA	OPTICS COMMUNICATIONS	2	28
2480	AGARWAL-GS	1991	UNIV-NEW-MEXICO	USA	PHYSICAL REVIEW A	1	4
2481	AGARWAL-GS	1992	UNIV-HYDERABAD	INDIA	OPTICS COMMUNICATIONS	3	31
2482	AGARWAL-GS	1992	UNIV-HYDERABAD	INDIA	MODERN PHYSICS LETTERS A	1	5
2483	AGARWAL-GS	1992	UNIV-HYDERABAD	INDIA	PHYSICAL REVIEW A	5	200
2484	AGARWAL-GS	1992	UNIV-HYDERABAD	INDIA	PHYSICS LETTERS A	1	18
2485	AGARWAL-GS	1993	MAX-PLANCK-INST-QU	GERMANY	PHYSICAL REVIEW A	3	46
2486	AGARWAL-GS	1993	MAX-PLANCK-INST-QU	GERMANY	PHYSICAL REVIEW LETTERS	1	25
2487	AGARWAL-GS	1993	MAX-PLANCK-INST-QU	GERMANY	OPTICS COMMUNICATIONS	2	28
2488	AGARWAL-GS	1993	MAX-PLANCK-INST-QU	GERMANY	PHYSICA SCRIPTA	1	7
2489	AGARWAL-GS	1993	UNIV-HYDERABAD	INDIA	JOURNAL OF MODERN OPTICS	1	9
2490	AGARWAL-GS	1993	UNIV-HYDERABAD	INDIA	OPTICS COMMUNICATIONS	2	8
2491	AGARWAL-GS	1993	UNIV-HYDERABAD	INDIA	PHYSICAL REVIEW LETTERS	1	83
2492	AGARWAL-GS	1993	UNIV-HYDERABAD	INDIA	PHYSICAL REVIEW A	2	35
2493	AGARWAL-GS	1993	UNIV-MUNICH	GERMANY	PHYSICAL REVIEW A	1	65
2494	AGARWAL-GS	1994	MAX-PLANCK-INST-QU	GERMANY	PHYSICS LETTERS A	1	15
2495	AGARWAL-GS	1994	MAX-PLANCK-INST-QU	GERMANY	PHYSICAL REVIEW A	1	20
2496	AGARWAL-GS	1994	MAX-PLANCK-INST-QU	GERMANY	OPTICS COMMUNICATIONS	1	7
2497	AGARWAL-GS	1994	UNIV-HYDERABAD	INDIA	PHYSICAL REVIEW LETTERS	1	7
2498	AGARWAL-GS	1994	UNIV-HYDERABAD	INDIA	PHYSICAL REVIEW A	4	144

Dataset Pros and Cons

Strengths

- Provides Year, Journal and Location for publications over the 1988-2010 timeframe.

Weaknesses

- No data on publications or locations prior to 1988 (just paper counts)
- Restricted to physics journals
- No author disambiguation

Dataset Cleaning

Google Refine - Basic text cleaning

- Case-normalization
- Removal of white space
- Basic group identification

Author Disambiguation

- Li, Y – Associated with over 200 locations.
- Rare Name Identification technique
- If more than half of an author name's publications are associated with a single location, that author name is deemed to be unique.

Boyack, K. W., & Klavans, R. (2008). Measuring Science-Technology Interaction Using Rare Inventor-Author Names. Journal of Infometrics, 2(3), 173-182.

Dataset Cleaning

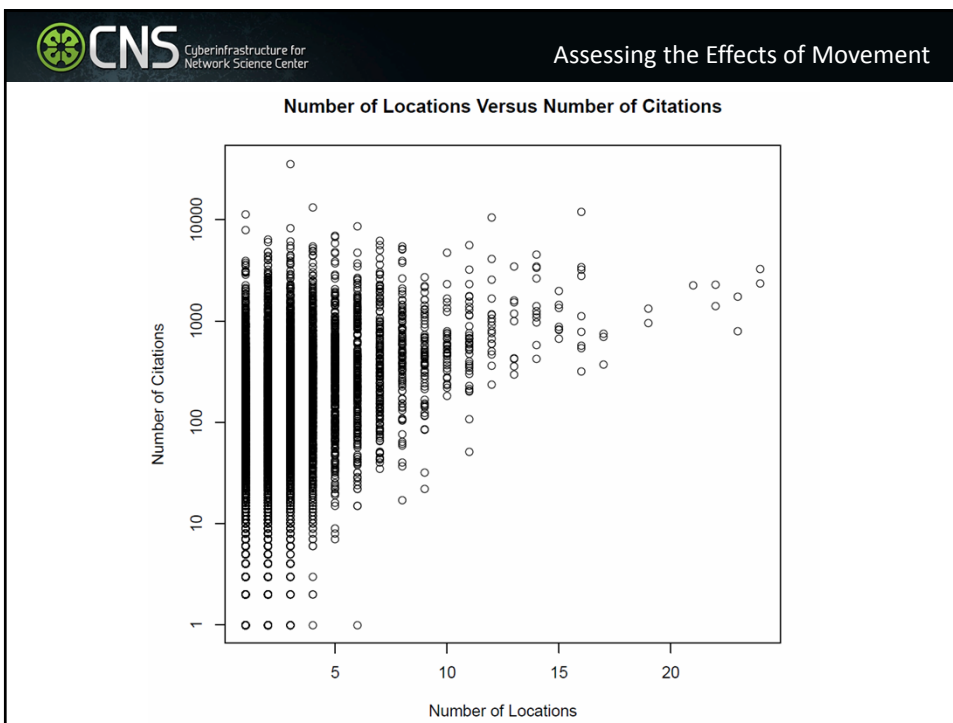
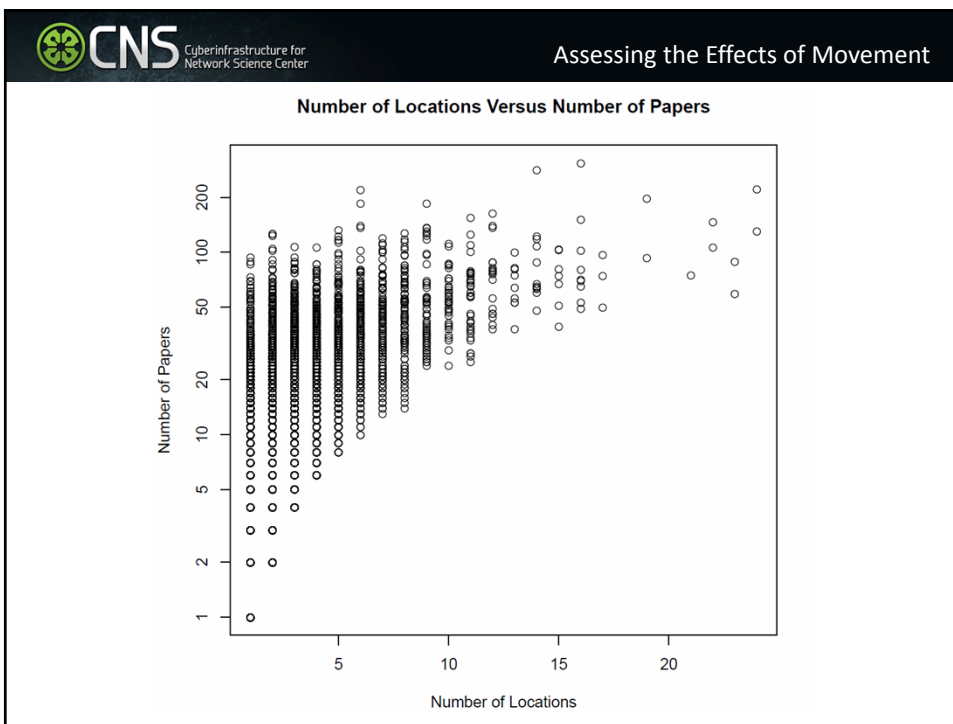
Rare Name Identification

- If more than half of the publications associated with an author name come from a single location, that author is determined to be unique.
- 8,783 of 10,000 names marked as unique
- Eliminates all common names.
- May eliminate a single author who has a publication portfolio that is evenly spread among locations.
- May combine an author with a small portfolio in under an author with a much larger body work.

Boyack, K. W., & Klavans, R. (2008). Measuring Science-Technology Interaction Using Rare Inventor-Author Names. Journal of Infometrics, 2(3), 173-182.

Does Movement Matter?

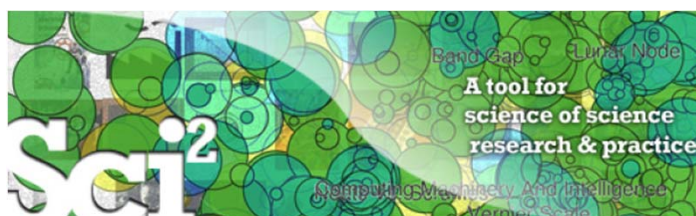
- Do career changes help or hinder researchers?
- Are researchers who move around more productive or more heavily cited?
- Is there a magic number of a career changes that is optimal?



Introduction to Sci2

The Science of Science (Sci2) Tool is an open-source modular toolset originally designed for the study of science. It is widely used to support temporal, geospatial, topical, and network analysis and visualization of scholarly datasets.

The tool can be freely downloaded at <http://sci2.cns.iu.edu>



Macrosopes

Decision making in science, industry, and politics, as well as in daily life, requires that we make sense of the massive amounts of data that result from complex systems.

Rather than making things larger or smaller, **macrosopes let us observe what is too great, slow, or complex for us to comprehend or sometimes even notice.**



Microscopes














Telescopes



Macrosopes

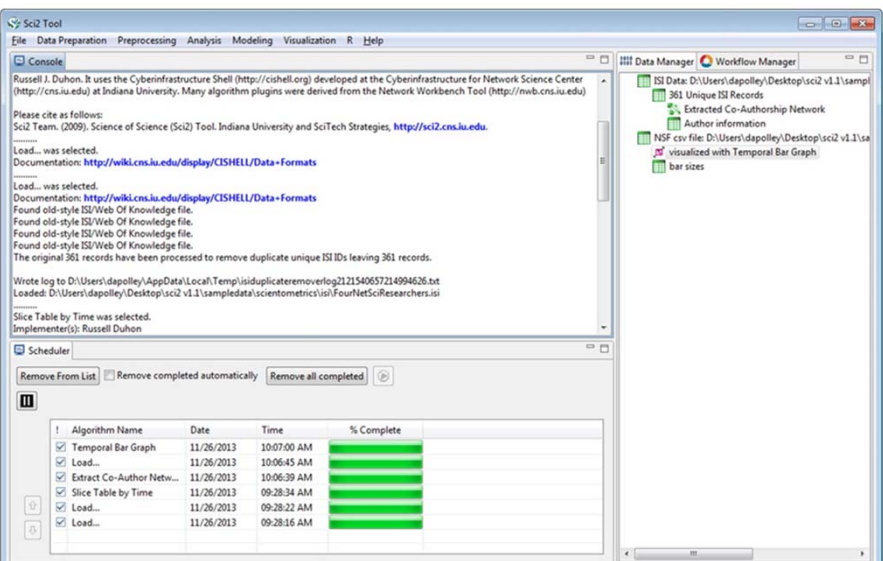
CNS Cyberinfrastructure for Network Science Center Tool Development

Visualization Design

	Micro/Individual (1-100 records)	Meso/Local (101–10,000 records)	Macro/Global (10,000 < records)
Statistical Analysis/Profiling	Individual person and their expertise profiles	Larger labs, centers, universities, research domains, or states	All of NSF, all of science 
Temporal Analysis (When)	Funding portfolio of one individual	Topic bursts of PNAS 	113 Years of P Research 
Geospatial Analysis (Where)	Career trajectory of one individual	Mapping a state intellectual landscape 	PNAS publications 
Topical Analysis (What)		flows in research 	VxOrd/Topic n NIH funding 
Network Analysis (With Whom?)	NSF Co-Authorship network of one individual 	NSF Co-Authorship network 	NIH's core co-authorship network 

CNS Cyberinfrastructure for Network Science Center Tool Development

Science of Science (Sci2) Tool



The screenshot shows the Sci2 Tool interface. The console window displays the following text:

```

Russell J. Duhon. It uses the Cyberinfrastructure Shell (http://cishell.org) developed at the Cyberinfrastructure for Network Science Center (http://cns.iu.edu) at Indiana University. Many algorithm plugins were derived from the Network Workbench Tool (http://nwb.cns.iu.edu)

Please cite as follows:
Sci2 Team. (2009). Science of Science (Sci2) Tool. Indiana University and SciTech Strategies, http://sci2.cns.iu.edu.

Load... was selected.
Documentation: http://wiki.cns.iu.edu/display/CISHELL/Data+Formats


Load... was selected.
Documentation: http://wiki.cns.iu.edu/display/CISHELL/Data+Formats
Found old-style ISI/Web Of Knowledge file.
Found old-style ISI/Web Of Knowledge file.
Found old-style ISI/Web Of Knowledge file.
Found old-style ISI/Web Of Knowledge file.
The original 361 records have been processed to remove duplicate unique ISI IDs leaving 361 records.

Wrote log to D:\Users\dapolley\AppData\Local\Temp\isiduplicateremoverlog2121540657214994626.txt
Loaded: D:\Users\dapolley\Desktop\sci2 v.1.1\sampdata\scientometrics\isi\FourNetSciResearchers.isi

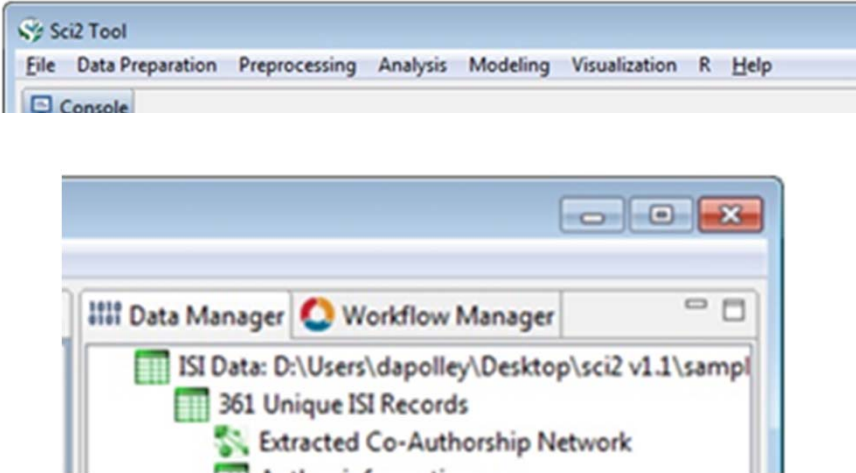
Slice Table by Time was selected.
Implementer(s): Russell Duhon
    
```


The Scheduler window shows a list of tasks with the following columns: Algorithm Name, Date, Time, and % Complete.

Algorithm Name	Date	Time	% Complete
Temporal Bar Graph	11/26/2013	10:07:00 AM	100%
Load...	11/26/2013	10:06:45 AM	100%
Extract Co-Author Net...	11/26/2013	10:06:39 AM	100%
Slice Table by Time	11/26/2013	09:28:34 AM	100%
Load...	11/26/2013	09:28:22 AM	100%
Load...	11/26/2013	09:28:16 AM	100%

 **CNS** Cyberinfrastructure for Network Science Center Tool Development

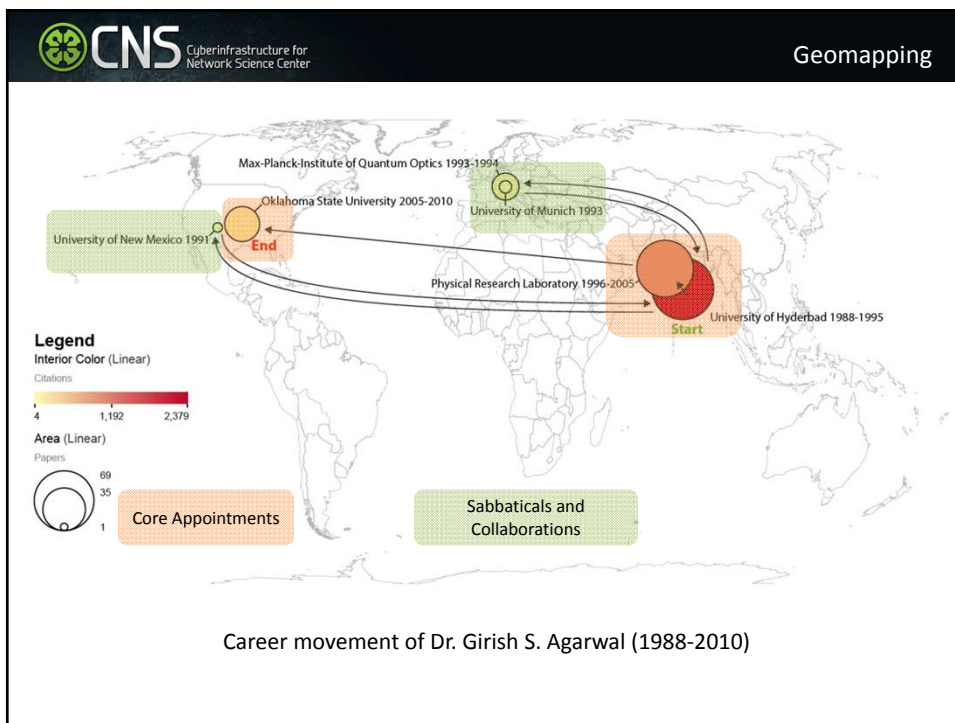
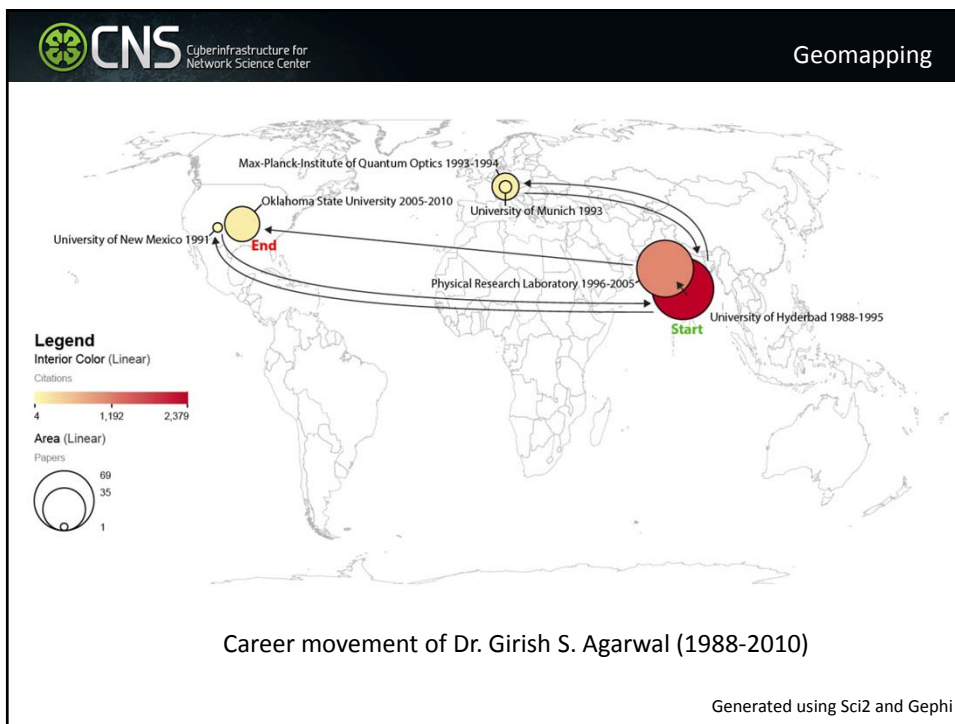
Science of Science (Sci2) Tool




 **CNS** Cyberinfrastructure for Network Science Center Geomapping

Geomapping Methodology

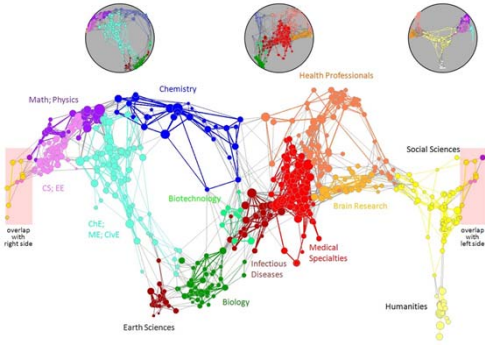
- The Top 10 authors by number of papers were selected for special consideration.
- The institutions associated with these authors were identified and their addresses gathered manually.
- The Bing Geocoder was then used to translate these addresses into coordinates.
- This information was then used to create a Proportional Symbol Map.





Science Mapping

UCSD Map of Science

- Visual representation of science via 13 disciplines and 554 subdisciplines
- Generated via clusters of journals
- First created in 2006 and updated in 2011
- Indexes over 25,000 journals and conference proceedings

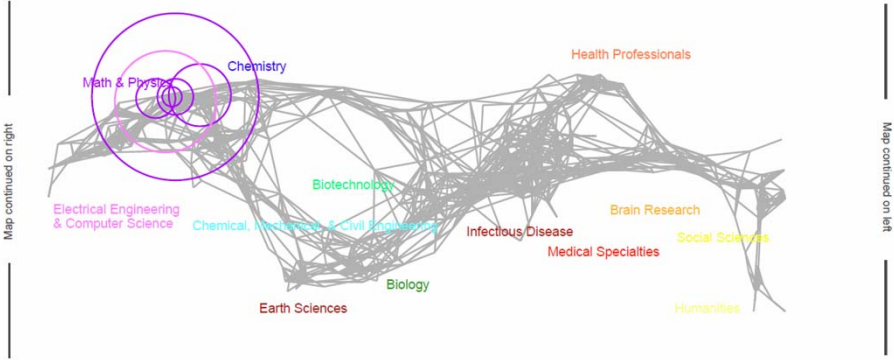


Börner, K., Klavans, R., Patek, M., Zoss, A. M., Biberstine, J. R., Light, R. P., Larivière, V., & Boyack, K. W. (2012). Design and Update of a Classification System: The UCSD Map of Science. PLOS ONE, 7(7), e39464-39464.


Science Mapping

UCSD Map of Science

Dr. Girish S. Agarwal – All Locations



Legend

Circle area: Fractional record count

Unclassified = 1

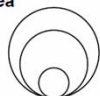
Minimum = 1

Maximum = 102

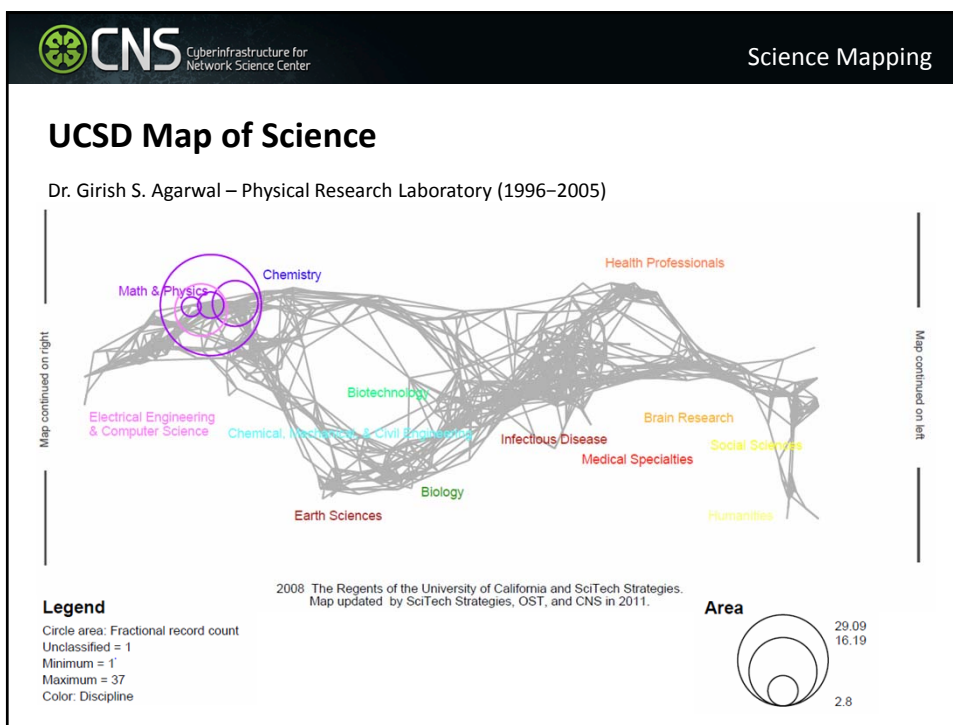
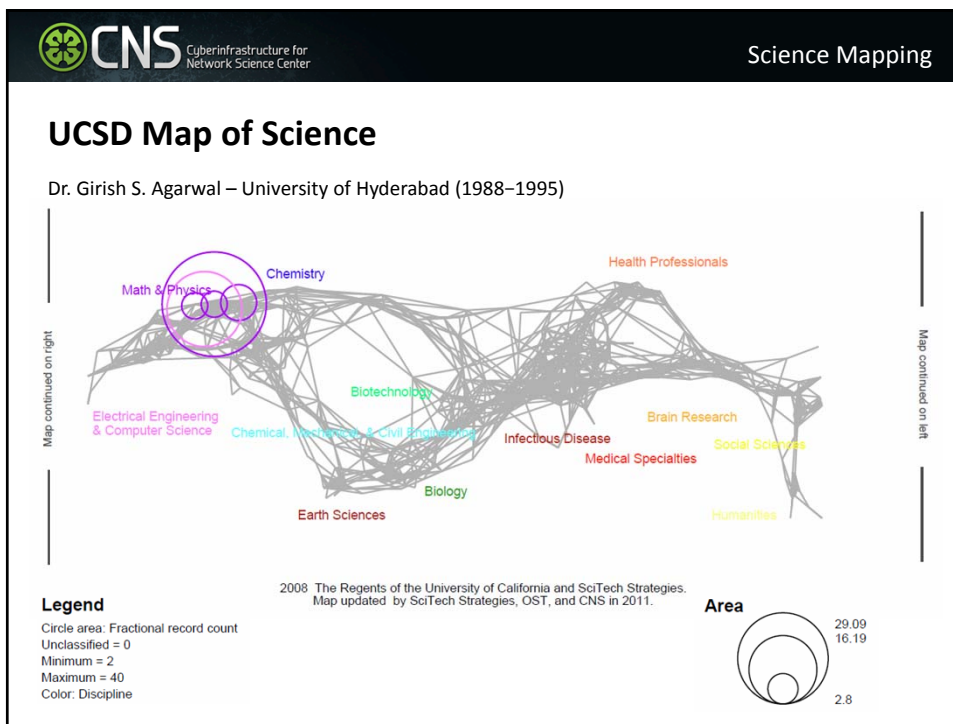
Color: Discipline

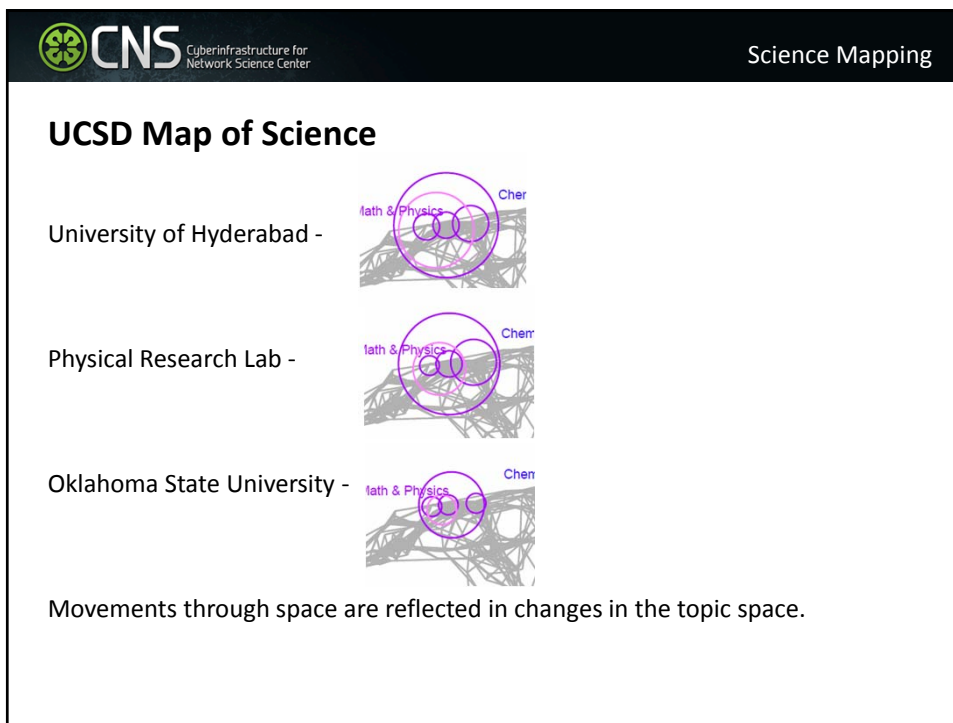
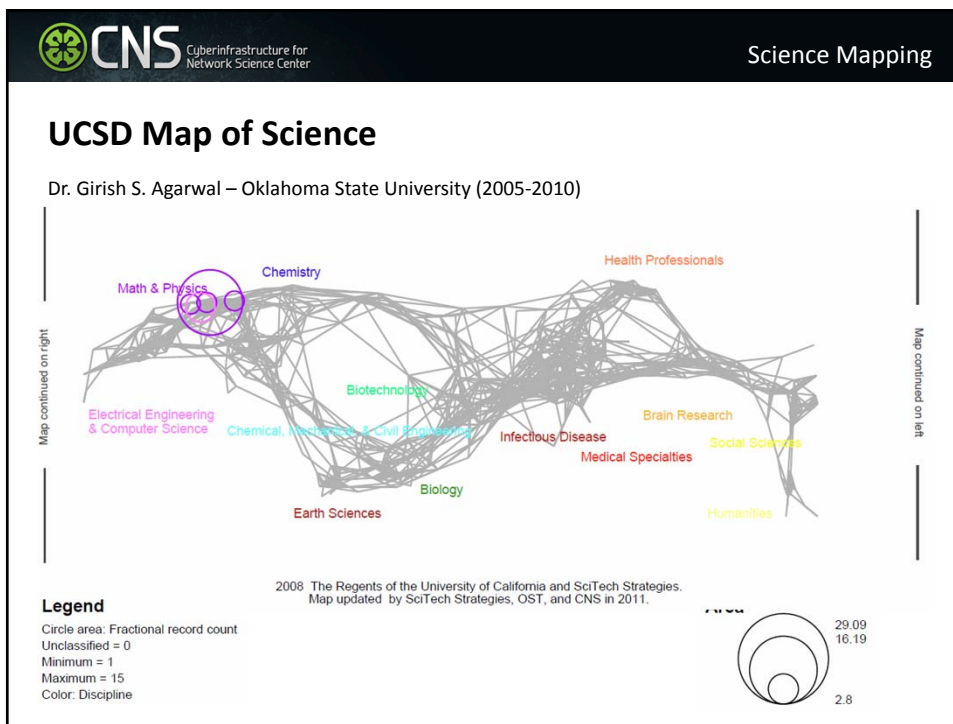
2008 The Regents of the University of California and SciTech Strategies.
Map updated by SciTech Strategies, OST, and CNS in 2011.

Area



29.09
16.19
2.8





UCSD Map of Science

Movements through space are reflected in changes in the topic space.

WHY?

A change in position may reflect a change in career focus.

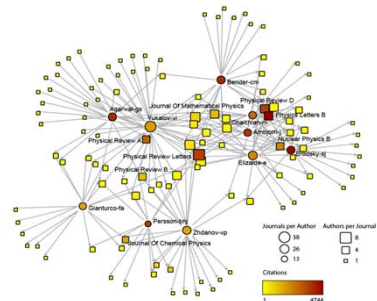
New collaborators may lead to publications in new journals.

A higher position with more funding may lead to publication in higher profile journals.

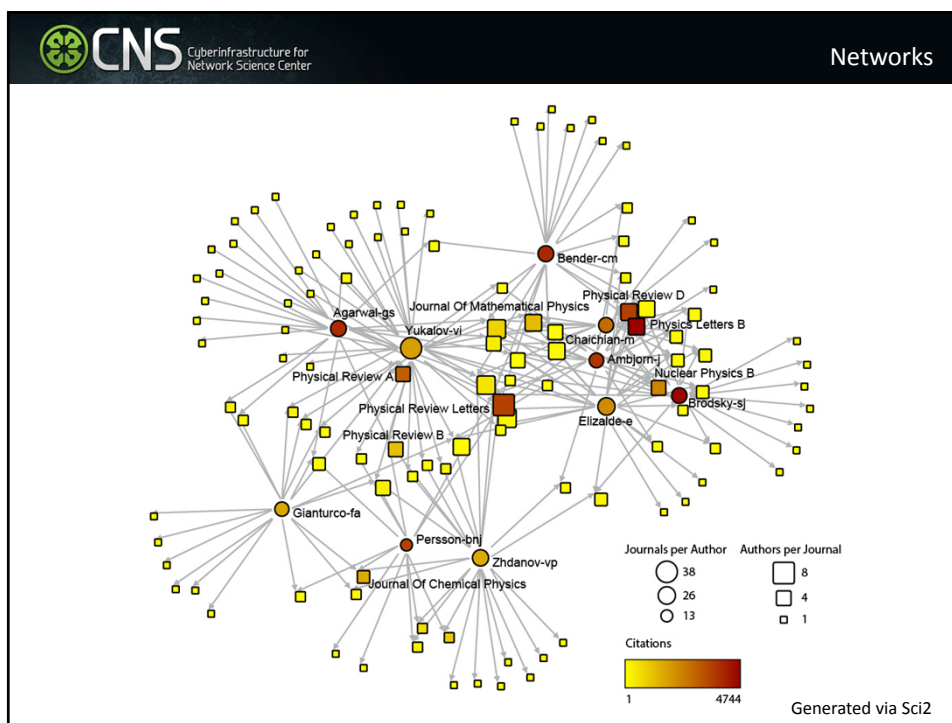
In this case, as he moved from job to job, Dr. Agarwal transitioned from optics journals to more general physics journals like *Physical Review A*.

Author Networking

- A visual way to see who an author works with and what they are collaborating on.
- Based on an ISI set of works by GS Agarwal.
- Circles are authors
- Squares are the journals that published their collaborations
- Larger nodes reflect more papers.
- Darker nodes reflect more citations.
- How would this change as the author moves from location to location?



Generated via Sci2



CNS Cyberinfrastructure for Network Science Center Future Work

Future Work

- How many of the changes shown here today could be predictive?
- Does a new collaboration serve as a possible predictor of an upcoming career change?
- Do bursts of activity (or depressions in activity) serve as predictors of a career change?
- Are authors with certain trajectories (across certain fields) more likely to have higher citation counts?
- Is this data predictive via a Markov model?
- Using SDR Data linked to publication/patent data, see next slide.

Survey of Doctorate Recipients Patent/Publication Matching Project

Respondents to 1993-2010 SDR have been matched to the US Patent and Trademark Offices Patent Record File, and the Thomson-Reuters Web of Science data. By doing so, a series of over 150 bibliometric variables on the patent and publication outcomes of these researchers has been created. Variables cover aggregate measures (e.g. Respondent X produced Y publications in 19XX) and individual patent or publication measures (details about each individual patent or publication produced by the respondent). These new matched variables can be used as a supplement to the regular SDR file.

The data will be available to U.S.-based individuals who obtain a license for confidential data from NCSES.

Recent Publications

- Bollen, Johan, David Crandall, Damion Junk, Ying Ding, and [Katy Börner](#). 2014. "[From funding agencies to scientific agency: Collective allocation of science funding as an alternative to peer review](#)". *EMBO Reports* 15 (1): 1-121.
- Reijnhoudt, Linda, Rodrigo Costas, Ed Noyons, [Katy Börner](#), and Andrea Scharnhorst. 2014. "['Seed+Expand': A general methodology for detecting publication oeuvres of individual researchers](#)". *Scientometrics*.

Available at <http://cns.iu.edu>

Recent Publications

- Mazlounian, Amin, Dirk Helbing, Sergi Lozano, [Robert Light](#), and [Katy Börner](#). 2013. "[Global Multi-Level Analysis of the 'Scientific Food Web'](#)". *Scientific Reports* 3, 1167; DOI:10.1038/srep01167.
- [Light, Robert](#), David E. Polley, and [Katy Börner](#). 2014. "MOOC Visual Analytics: Empowering Teachers, Students, Researchers, and Developers of Massively Open Online Courses". *LAK 2014*.

Available at <http://cns.iu.edu>

Recent Publications

- [Börner, Katy](#), Mike Conlon, Jon Corson-Rikert, and Ying Ding, eds. 2012. [VIVO: A Semantic Approach to Scholarly Networking and Discovery](#). Morgan & Claypool Publishers LLC.
- [Börner, Katy](#), and David E. Polley. 2014. [Visual Insights: A Practical Guide to Making Sense of Data](#). Cambridge, MA: The MIT Press.

Questions?

lightr@indiana.edu

<http://cns.iu.edu>