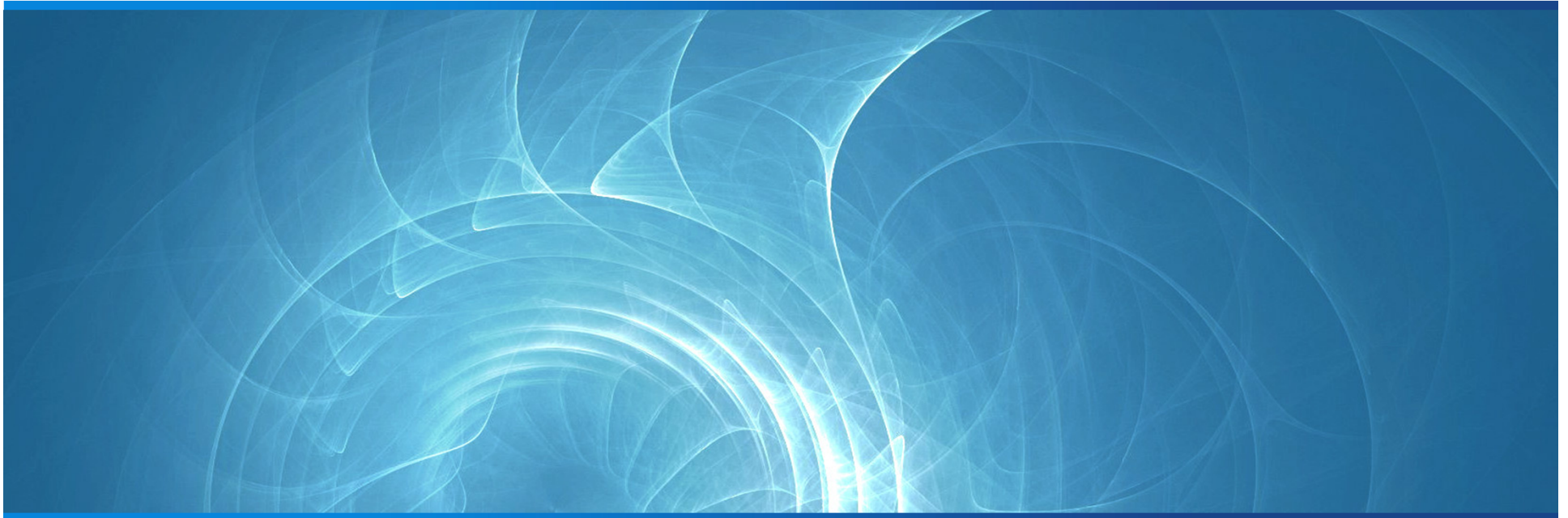# Scale-Free Geographical Mapping of Scientific Collaboration of Metropolitan Statistical Areas

4S Annual Meeting, Cleveland OH
November 3, 2011 | 1:30 to 3:00 PM

Context – Why is collaboration useful?

- Papers written in collaboration are usually more cited

- Collaboration among researchers increases the flow of knowledge

- Researchers who collaborate more frequently have access to more/better equipment

- An high collaboration rate is an indicator of an open innovation culture in an institution, region, state or country.

- Most work has focused on the country or state level.

➢ What about the city level, or Metropolitan Statistical Area?

Measuring using classical methods?

- **Using the number of papers (whole counting) written in collaboration**

Metropolitan Statistical Areas	Collaboration between MSA
Boston-Cambridge-Quincy, MA-NH	292,441
Los Angeles-Long Beach-Santa Ana, CA	227,800
Washington-Arlington-Alexandria, DC-VA-MD-WV	221,013
New York-Northern New Jersey-Long Island, NY-NJ-PA	216,221
San Francisco-Oakland-Fremont, CA	172,181
Chicago-Naperville-Joliet, IL-IN-WI	151,778
Baltimore-Towson, MD	136,932
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	132,454
Houston-Sugar Land-Baytown, TX	128,169
Durham, NC	116,204

- So the biggest universities and institutions collaborate more?

Science-Metrix

Measuring using classical methods?

- Maybe using the proportion of papers written in collaboration?

Metropolitan Statistical Areas	Papers	Collabo.	%
Olympia, WA	1,004	647	64%
Anchorage, AK	1,878	1,175	63%
Flagstaff, AZ	2,657	1,620	61%
Oxnard-Thousand Oaks-Ventura, CA	2,836	1,637	58%
Fresno, CA	1,325	761	57%
Killeen-Temple-Fort Hood, TX	1,467	826	56%
Grand Rapids-Wyoming, MI	1,544	869	56%
Springfield, MA	1,253	691	55%
Santa Fe, NM	1,518	832	55%
Hilo, HI	1,927	1,046	54%

- Even with a threshold of 1,000 papers, small MSAs are unfairly advantaged.

More publications, more problems
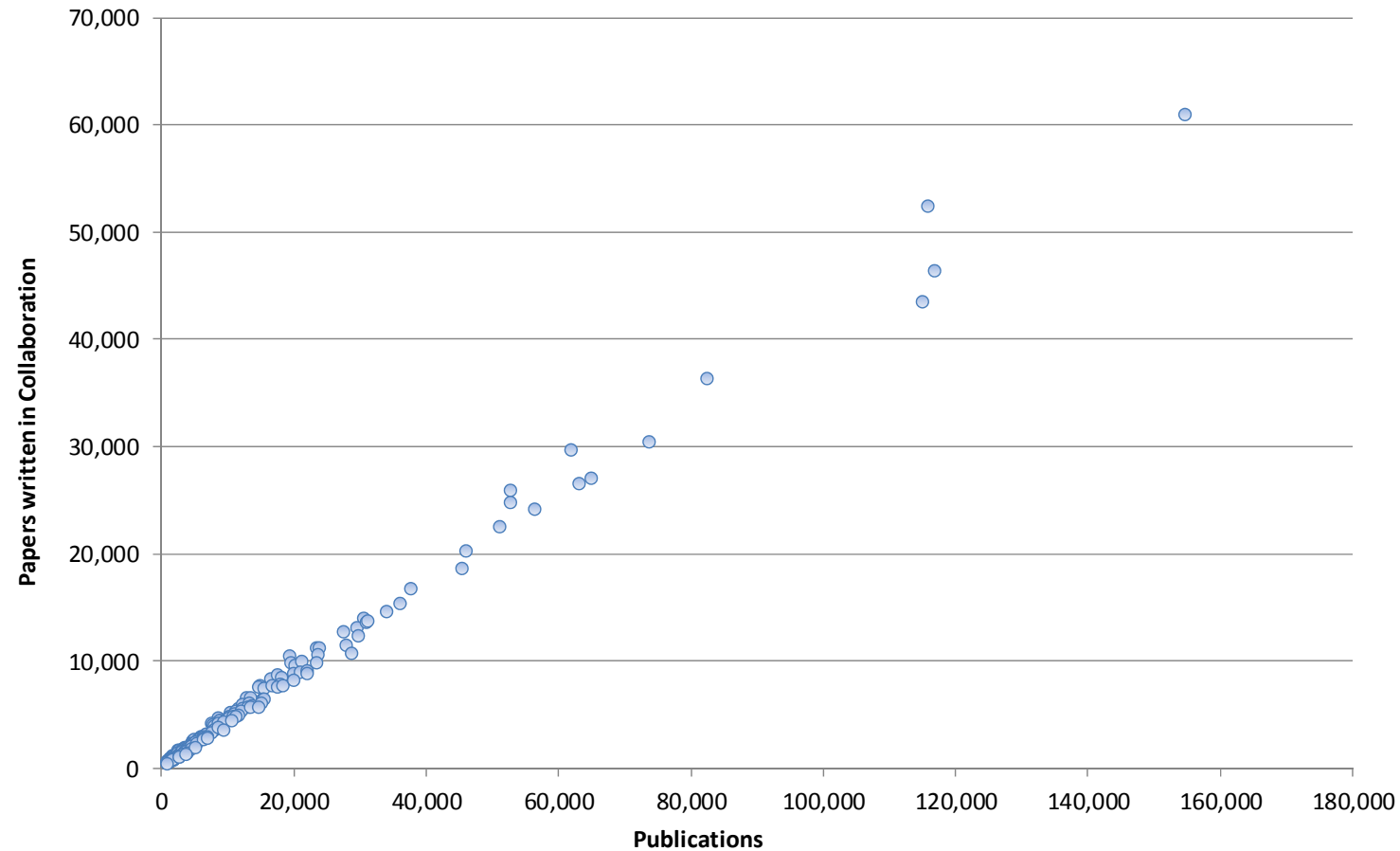
- The higher the publication count, the higher the number of collaborations. Now what?

More publications, more problems

- Raw collaboration counts are highly correlated with the number of publications

- Larger MSAs would be seen as keen to collaborate

- The percentage of collaboration sends the opposite message – Smaller MSAs are seen as keen to collaborate

- What's wrong with these indicators?

- ANSWER: **Neither is a reliable performance indicator because they don't adjust for scale**

- Why? Scientific systems are generally **non-linear**.
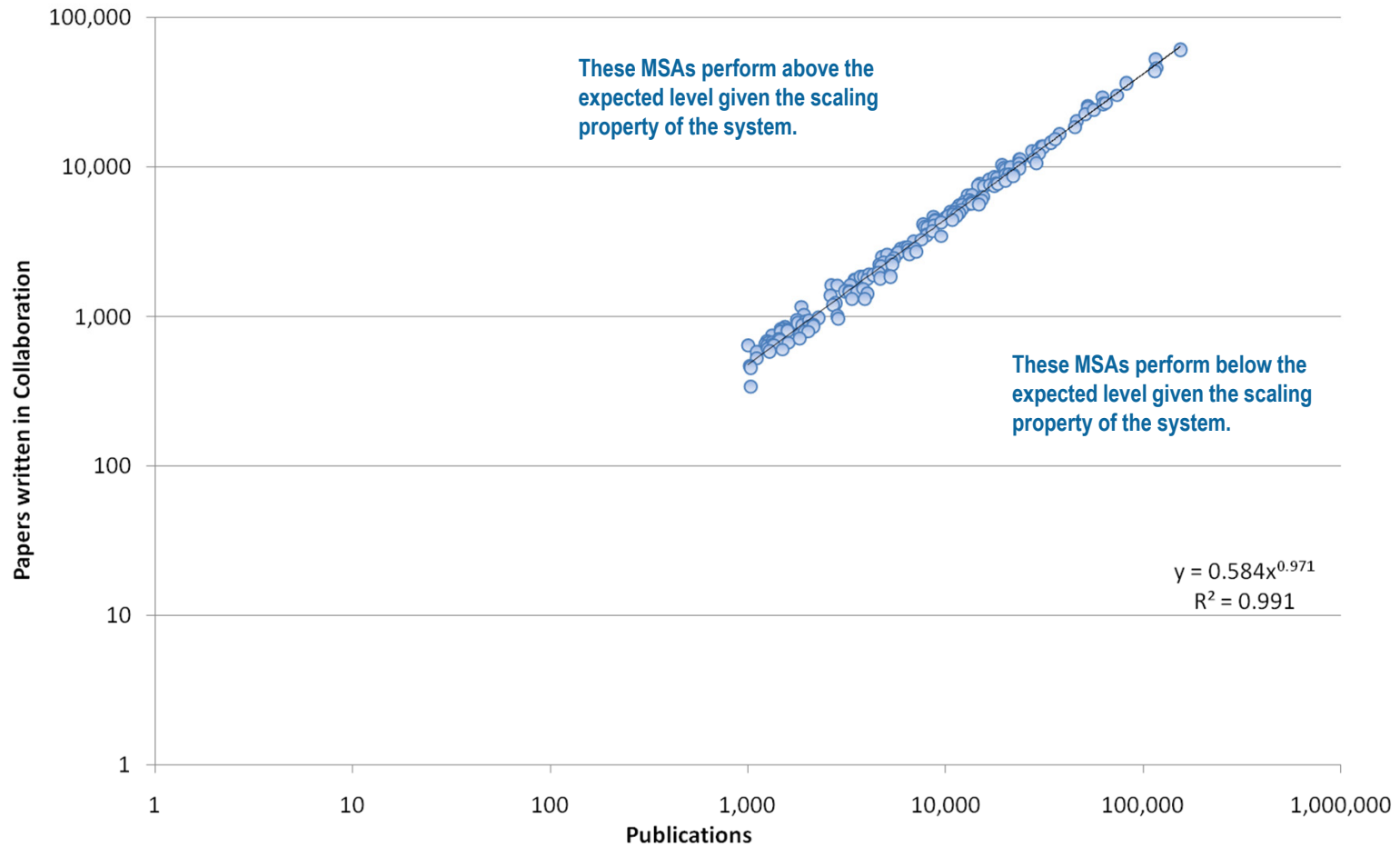
Adjusting for scale

- The relationship of papers and collaboration follows a power-law. We can model this system and make predictions.

- Sylvan Katz (@SPRU), suggested a scale-free method. By calculating the ratio between the expected (or predicted) number of collaborations and the observed number of collaboration, we get a performance indicator that accounts for the scaling dynamic of the system.

- Archambault et al. (2011) applied this method to countries, states, provinces and Canadian institutions.

- Let's do the same thing with Metropolitan Statistical Areas!

Adjusting for scale

These MSAs perform above the expected level given the scaling property of the system.

These MSAs perform below the expected level given the scaling property of the system.

$$y = 0.584x^{0.971}$$
$$R^2 = 0.991$$

- Very interesting, but what about that scale adjusted indicator?

Scale adjusted collaboration indicator

Log of the ratio of observed/expected collaboration of the 30 most publishing MSAs
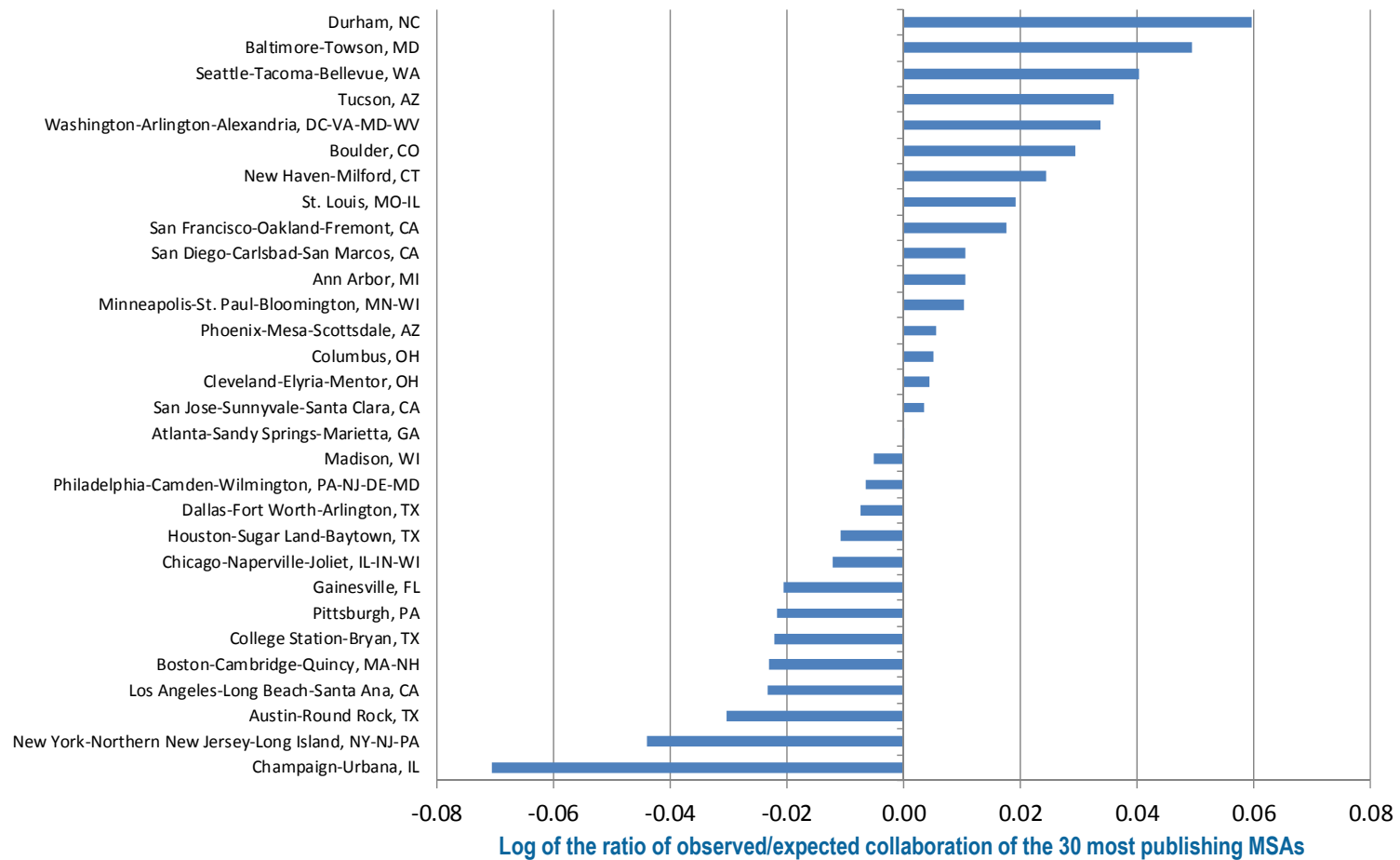
- There must be a prettier, and more intuitive, way to present and explore this data…
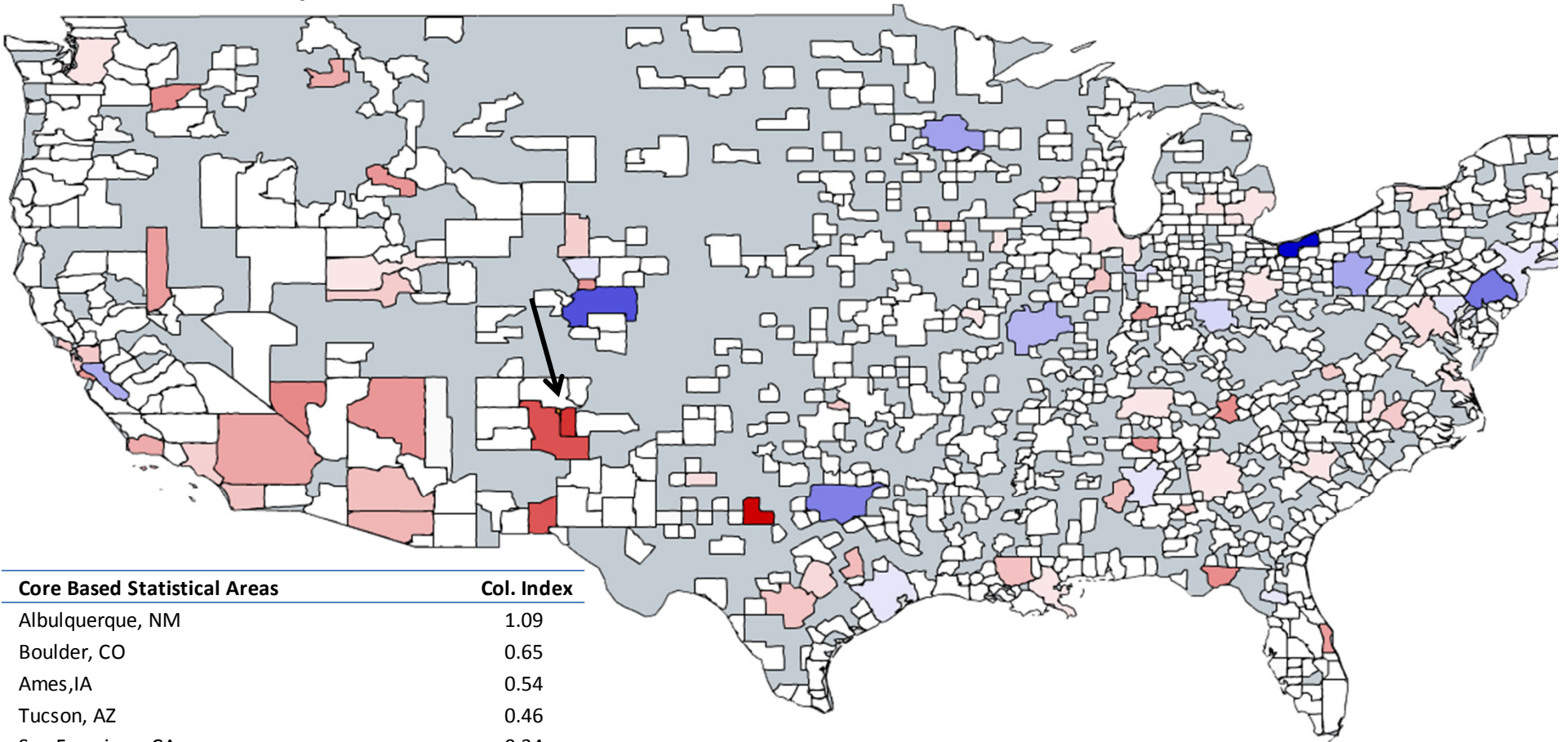
MSAExplorer

- Nationwide collaboration affinities
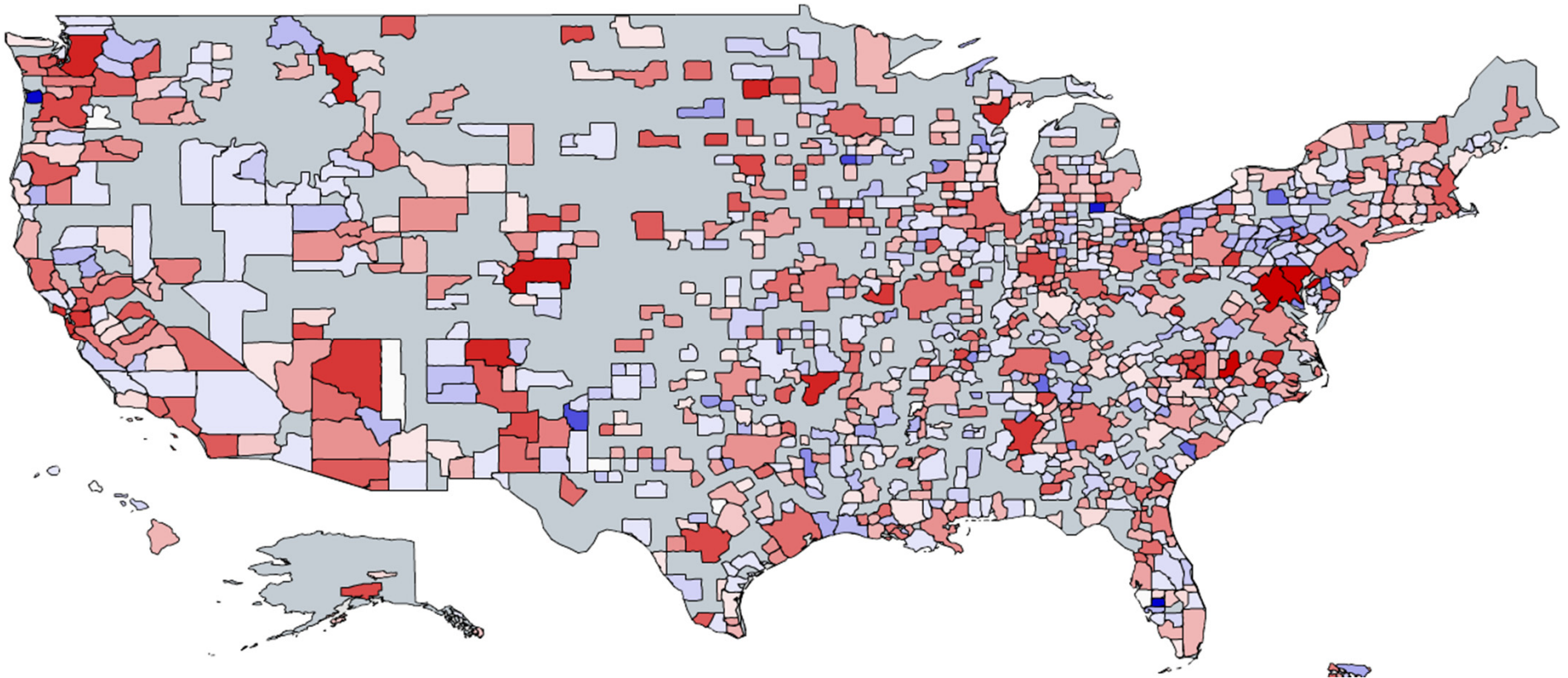
Blue: Higher affinity to collaborate considering their scientific output

**Blue: Lower affinity to collaborate considering their scientific output**

- An example: Los Alamos, NM

| Core Based Statistical Areas	Col. Index
Albulquerque, NM	1.09
Boulder, CO	0.65
Ames,IA	0.54
Tucson, AZ	0.46
San Francisco, CA	0.34
Boston-Cambridge-Quincy, MA-NH	0.14
Atlanta-Sandy Springs-Marietta, GA	0.03
Philadelphia-Camden-Wilmington, PA-NJ-DE-MD	-0.20

- What are Los Alamos researchers collaborating on with their colleagues from Albuquerque?

- Explore @ http://www.science-metrix.com/MSAExplorer
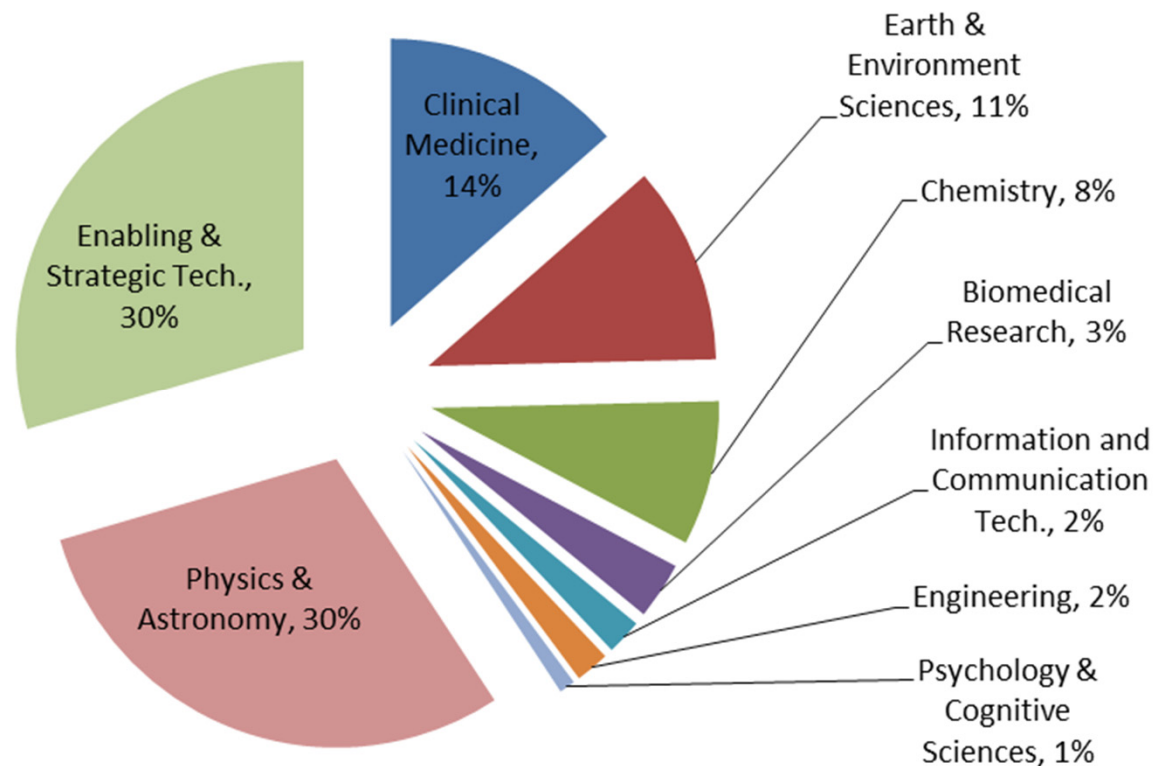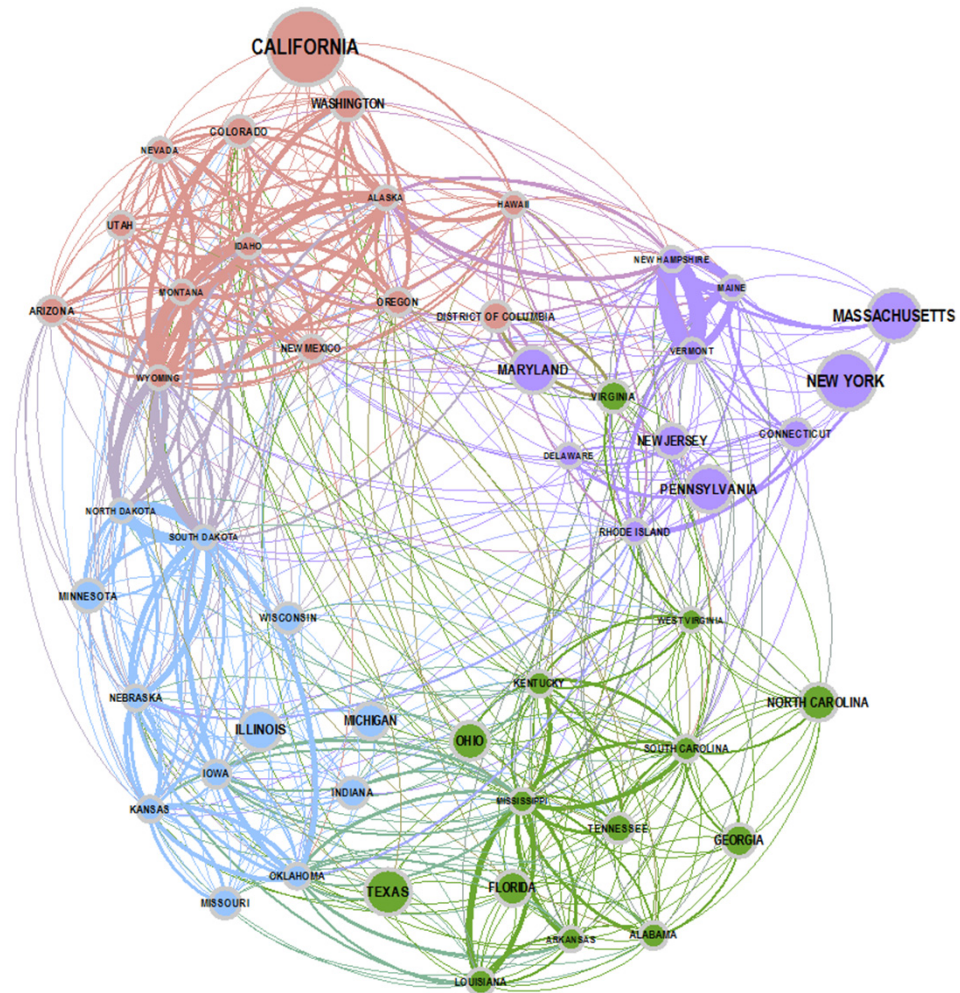
Source: Archambault et al. (2011)

Problems

- Bibliometric data tends to be noisy (even more so at low publication counts)

- A small change in the modeling (regression) can have a big influence on the ratios

- It is somewhat more complex to calculate (and to explain!) scale-free indicators compared to classical indicators

  - Everyone understand what percentages are, scale-free indicators require a tad more involvement

- More research and input from other fields is needed to address these issues

Contact Information

CONTACT INFO

Olivier H. Beauchesne, M.Sc.

Research Analyst| Science-Metrix

olivier.hbeauchesne@science-metrix.com

Éric Archambault, Ph.D.

President and CEO| Science-Metrix

eric.archambault@science-metrix.com

1335, Mont-Royal E.

Montreal, Quebec H2J 1Y6

Telephone: 514-495-6505

Fax: 514-495-6523

E-mail: info@science-metrix.com

www.science-metrix.com