

# Science Mapping: Convergence, Consensus, Policy Implications?

Kevin W. Boyack SciTech Strategies, Inc. kboyack@mapofscience.com http://www.mapofscience.com



- Accuracy studies
- Overview of 2 recent papers
  - Convergence or consensus?
    - Comparison of 20 maps of science
    - Submitted to JASIST (under review)
  - Linking science/technology through inventor-authors
    - Using rare names
    - Accepted for publication in Journal of Informetrics
- Simplified circle map and uses



- Use of science maps for policy requires that they be as accurate as possible
  - If metrics or rankings are to be used, this is also true for any science classification structure, regardless of whether it is "mapped" or not
- We have done several accuracy studies
  - Journal maps: explored the local and regional accuracy of different similarity measures using ISI categories as a standard
  - Paper-level maps: explored local accuracy and disciplinary bias of two similarity measures, and two levels of pruning
  - In general, normalized similarity measures show reasonable agreement with category structures



### Accuracy studies have limitations

- Is there a single map of science or classification structure?
  - Some conversation between Börner, Boyack, Leydesdorff, Rosvall, Small, and others several months ago ended up with a "NO"
  - Multi-dimensional system, any single map might miss important facets
  - Disagreement on whether we should even try to come to some "convergence" or not
- So, where do we go next, given that
  - There is no single accepted map
  - There is no fully accepted standard



- We decided to compare/contrast all of the "comprehensive" maps of science that we could find
- We found 20 in the following categories
  - Hand-drawn (4)
    - 3 by experts
    - 1 based on course pre-requisites
  - Electronic (16)
    - 6 reference paper maps
    - 7 journal maps
    - 3 journal category maps



### Maps and references

Researcher(s) & Reference	Name	Method	Elements	# Clust	Database & Year	Form
(Bernal, 1939)	Bernal	Expert		14, 110		Hierarchical
(Ellingham, 1948)	Ellingham	Expert		13, 51,		Hierarchical &
		-		130		Non-centric
(Balaban & Klein, 2006)	Balaban-I	Expert	16 fields	16		Hierarchical &
						Centric
(Griffith, Small, Stonehill, &	Small74	Reference papers	1,150 pap	41	SC, 1972 Q1	Centric
Dey, 1974)						
(Small & Garfield, 1985)	Small85	Reference papers	~11,000 pap	51	SC+SS, 1983	Hierarchical &
						Centric
(Small, 1999)	Small99	Reference papers	36,720 pap	35	SC+SS, 1995	Hierarchical
(Klavans & Boyack, 2008)	KB-Para	Reference papers	800k pap	776	SC+SS, 2003	Non-centric
(Klavans & Boyack, 2007)	KB06-TS	Reference papers	1.9M pap	283	SC+SS, 2004	Non-centric
(Klavans & Boyack, 2007)	KB06-SC	Reference papers	2.1M pap	554	Scopus, 2004	Non-centric
(Bassecoulard & Zitt, 1999)	B-Z	Journals	~2,000 jnl	29	SC/JCR, 1993	Hierarchical &
						Centric
Klavans, unpublished, 2002	K02	Journals	5,647 jnl	69	SC+SS+AH, 2000	Non-centric
		-				
(Boyack, Klavans, & Börner,	Backbone	Journals	7,121 jnl	205	SC+SS, 2000	Non-centric
2003) (Deveels at al. 2009)	DDV02 S	Torrala	7 227 1	671	80188.2002	Non contria
(Boyack et al., 2008)	BBK02-5		7,227 <u>ј</u> ш	0/1	5C+55, 2002	Non-centric
(Boyack, 2008)	B03-51	Journais	8,00/ JII	832	SC+SS+PK, 2005	Non-centric
(Klavans, Patek, & Boyack,	UCSD	Journals	16,235 jni	334	SC/SS/AH +	Non-centric
2007)	D11	T1_	6 116 -1	07	Scopus, 2001-05	NT
(Rosvall & Bergstrom, 2007)	Kosvall	Journals	6,116 jni	8/	SC+SS, 2004	Non-centric
(Moya-Anegon et al., 2004)	Scimago-I	Journal categories	25 categ	25	SC+SS+AH, 2000	Non-centric
	а. н	<b>T</b>	210	210	Spanish papers	0.1
(Moya-Anegon et al., 2007)	Scimago-II	Journal categories	219 categ	219	SC+SS+AH, 2002	Centric
(Leydesdorff & Katols,	Г-К	Journal categories	0,104 jni;	172	SC, 2006	Mixed
2007)	<b>D</b> 1 1		172 categ			<b>a</b> . •
(Balaban & Klein, 2006)	Balaban-II	Course prerequisites		11	Texas A&M	Centric
					undergraduate	



• Hierarchical (linear)

• Centric (hub/spokes)

• Non-centric (ring)



Klavans, R. (2003). Poster at the <u>Sackier</u> Colloquium on Mapping Knowledge Domains, Irvine, CA, May 9-11, 2003.

## Comparing maps of science

- Maps all conform to the following:
  - Division (separation of science into parts)
  - Proximate location (related parts are adjacent)
  - Linkage (additional linkages for non-adjacent parts)
- Need a basis of comparison since all maps are at different levels of detail
- It was quickly determined that convergence was not happening, so we switched to looking for "consensus"



- Develop a framework
- Code all maps using the framework
- Simplify each coded map (eliminate duplicate edges)
- List paired relationships
- Used paired relationships from all maps to determine consensus, and to measure relative accuracies of all maps



- Science was divided into 16 broad areas
  - Fundamental areas (4): Math, Physics, Chemistry, Biology mentioned in all maps
  - Fundamental combinations (2): Physical Chemistry, Biochemistry mentioned in most maps
  - Applied areas related to physics/chemistry (3): Computer Science, Engineering, Earth Sciences
  - Applied areas related to biology (3): Infectious Disease, Medical Specialties, Brain Research
  - Applied areas dealing with social issues (3): Social Sciences, Health Services, Psychology
  - AHCI (1): Humanities



### Coding example



SCImago-II 1 – Map



Moya-Anegón, F., Vargas-Quesada, B., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., Munoz-Fernández, F.J., & Herrero-Solana, V. (2007). Visualizing the marrow of science. Journal of the American Society for Information Science and Technology, 58(14), 2167-2179. SCImago-II 2 - Coding



Moya-Anegón, F., Vargas-Quesada, B., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., Munoz-Fernández, F.J., & Herrero-Solana, V. (2007). Visualizing the marrow of science. Journal of the American Society for Information Science and Technology, 58(14), 2167-2179. SCImago-II 3 - Simplification and edges



All maps and codings available at <u>www.mapofscience.com/history/maps</u>

в

С

MD

#### ELLINGHAM 2 - Coding

G ence



5 HS

KB06-TS 2 - Coding

N

8 HS

Small, H., & Garfield, E. (1985). The geography of science mappings. Journal of Information Science, 11, 147-159.

1



4

В

PS,SS

I H I

BC,I

HS;1;MD

na PC mee

CS T

es E ag



Small, H. (1999). Visualizing science by citation mapping. Journal of the American Society for Information Science, 50(9), 799-813.

KB-PARA 2 - Coding

(ale)

3

MD

(1)

(A)



PC

Griffith, B.C., Small, H.G., <u>Stonehill</u>, J.A., & <u>Dey</u>, S. (1974). Structure of scientific literatures. 2. Toward a macrostructure and microstructure for science. Science Studies, 4(4), 339-386.

Klavans, R., & Boyack, K.W. (2008). Thought leadership: A new indicator for national and institutional comparison. To appear in Scientometrics, 75(2).



Klavans, R., & Boyack, K.W. (2007). Is there a convergent structure of science? A comparison of maps using the ISI and Scopus databases. In D. Torres-Balinas & H. Medgi (Eds.), 11th International Conference of the International Society for Scientometrics and <u>Informetrics</u> (pp. 437-448). Madrid, Spain, June 2007: ISSI. Attories Clance Applied Physics Applied Physics Physical Clenistry Check Medice Cincel Medice Cancel

Dashed edges indicate edges from nodes in the global map to nodes in the NODE1 detail.

CS

KB06-SC 2 - Coding

B 13, 14

Disciplinary and national

Klavans, R., & Boyack, K.W. (2007). Is there a convergent structure of science? A comparison of maps using the ISI and Scopus databases. In D. Torres-Salinas & H. Model (Eds.), 11th International Conference of the International Society for Scientometrics and <u>Informetrics</u> (pp. 437-448). Madrid, Spain, June 2007: ISSI.

Bassecoulard, E., & Zitt, M. (1999). Indicators in a research institute: A multi-level classification of journals. Scientometrics, 44(3), 323-345.



B-Z 2 - Coding

BACKBONE 2 - Coding

BBK02-S 2 - Coding

B03-ST 2 - Coding



Klavans, R. (2003). Poster at the <u>Sackler</u> Colloquium on Mapping Knowledge Domains, Irvine, CA, May 9-11, 2003.

UCSD 2 - Coding



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

SCImago-I 2 - Coding



Boyack, K.W., <u>Börner</u>, K., & Klavans, R. (2008). Mapping the structure and evolution of chemistry research. Scientometrics, accepted. OR related conference paper: In D. Torres-Salinas & H.Moge (disk). 11th International Conference of the International Society for Scientometrics and <u>Informetrics</u> (pp. 112-123). Madrid, Spain.

SCImago-II 2 - Coding



Boyack, K.W. (2008). Using detailed maps of science to identify potential collaborations. Scientometrics, accepted. OR related conference paper: In D. Torres-Salins at H. Mogel (Eds.). 11th International Conference of the International Society for Scientometrics and <u>Informetrics</u> (pp. 124-135). Madrid, Spain.



Note: this is the Mercator projection of a map that was originally on a sphere. The visual break between SS and CS was necessitated by use of the Mercator projection. There is an edge between CS and SS.

Klavans, R., Patek, M.D., & Boyack, K.W. (2007). Maps of Science: Forecasting Large Trends in Science: Places & Spaces (<u>http://www.scimaps.org/dev/map\_detail.php?map\_id=164</u>). Detail on this map is also available at <u>www.mapofscience.com</u>.



Moya-Anegón, F., Vargas-Quesada, B., Herrero-Solana, V., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., & Munoz-Fernández, F.J. (2004). A new technique for building maps of large scientific domains based on the cocitation of classes and categories. Scientometrics, 61(1), 123-145.

ROSVALL 2 - Coding





Moya-Anegón, F., Vargas-Quesada, B., Chinchilla-Rodriguez, Z., Corera-Álvarez, E., Munoz-Fernández, F.J., & Herrero-Solana, V. (2007). Visualizing the marrow of science. Journal of the American Society for Information Science and Technology, Sel (14), 2167-2179.

BALABAN-II 2 - Coding



Leydesdorff, L., & <u>Rafols</u>, I. (2007). A global map of science based on the ISI subject categories, from <u>http://users.fmg.uva.nl/lleydesdorff/map06/texts/index.htm</u> Balaban, A.T., & Klein, D.J. (2006). Is chemistry 'The Central Science'? How are different sciences related? Co-citations, reductionism, emergence, and posets. Scientometrics, 69(3), 616-637.



## Consensus pairs

Rank	Pair	N	N-poss	%	Rank	Pair	Ν	N-poss	%
1	B-BC	20	20	100.0	15	BC-MD	14	20	70.0
2	I-MD	20	20	100.0	16	BC-C	14	20	70.0
3	H-SS	8	8	100.0	17	E-P	12	18	66.7
4	C-PC	19	20	95.0	18	B-I	13	20	65.0
5	HS-MD	16	17	94.1	19	CS-SS	10	16	62.5
6	PS-SS	16	17	94.1	20	H-PS	5	8	62.5
7	P-PC	18	20	90.0	21	M-P	11	19	57.9
8	MD-N	16	18	88.9	22	С-Е	10	18	55.6
9	E-G	16	18	88.9	23	C-P	11	20	55.0
10	B-G	17	20	85.0	24	HS-N	8	15	53.3
11	BC-I	16	20	80.0	25	CS-E	9	17	52.9
12	E-PC	14	18	77.8	26	C-G	10	20	50.0
13	N-PS	14	18	77.8	27	HS-PS	8	16	50.0
14	CS-M	13	18	72.2					



### Consensus map





### Consensus map (w/o KBB maps)





- Map is NON-CENTRIC if all consensus (>50% edges) are used
- Map is HIERARCHICAL if only top 15/16 edges used



### 20 maps on the consensus map

Number of hops	Count	Percent	Accuracy value
1	345	78.4	1.0
2	73	16.6	0.5
3	18	4.1	0
4	4	0.9	0
Total	440	100	0.867



### Accuracies of 20 maps

Source map	Year	Туре	Local Acc	Local Acc	Regional	Figure of	# Areas	Multi-
			Type1	Type2	Acc	Merit		nodes
KB06-SC	2006	Paper	95.8	94.0	100.0	96.6	15	0
Backbone	2004	Jnl	97.6	88.0	100.0	95.2	15	0
UCSD	2007	Jnl	95.7	88.9	100.0	94.8	16	0
Ellingham	1948	Expert	90.0	92.1	100.0	94.0	12	1
KB-Para	2005	Paper	92.3	94.4	93.8	93.5	16	1
Bernal	1939	Expert	85.7	94.0	100.0	93.2	15	2
Scimago-I	2004	Categ	90.9	87.5	100.0	92.8	15	2
KB06-TS	2006	Paper	91.7	90.7	93.8	92.1	16	1
B03-ST	2005	Jnl	92.5	82.0	100.0	91.5	15	0
BBK02-S	2004	Jnl	92.5	80.0	100.0	90.8	15	0
Rosvall	2007	Jnl	78.3	93.2	100.0	90.5	14	2
Small99	1999	Paper	78.6	89.5	100.0	89.3	13	3
Balaban-II	2007	Pre-req	85.0	82.0	100.0	89.0	15	4
K02	2002	Jnl	84.2	81.8	100.0	88.7	15	1
L-R	2007	Categ	86.1	73.9	100.0	86.7	14	0
Balaban-I	2007	Expert	73.9	79.6	100.0	84.5	16	3
Small85	1985	Paper	84.2	76.0	86.7	82.3	15	2
Small74	1974	Paper	69.2	76.5	100.0	81.9	13	2
B-Z	1999	Jnl	80.6	71.7	93.3	81.9	14	1
Scimago-II	2007	Categ	90.0	75.9	75.0	80.3	16	1



## Reasons to favor NON-CENTRIC

• Riemannian (curved) space is inherently more accurate than Euclidean (x,y,z) space



- Hierarchy and centric maps both imply favored status
- Non-centric maps
  - Do not impose artificial boundaries
  - Can show interdisciplinarity and new discoveries in an exciting way

Consensus map summary

- Consensus map generated from 20 maps
  - Robust
  - NON-CENTRIC map if all consensus edges are used
- Simplified map (circular) may be an effective map for policy purposes
  - Show interdisciplinary
  - Show technology (patent) profiles
  - Etc.



## Science / technology interaction



- Variety of methods possible
  - Text (titles, abstracts, etc.)
    - Patent abstracts tend to have different language than article abstracts
  - Non-patent references
    - Roughly 2/3 are to articles/proceedings papers
    - Data cleaning is an issue; no standard format
  - Inventor-authors
    - Name disambiguation is an issue
    - Relatively small overlap between authors / inventors



- Data: Scopus papers, USPTO patents, 2002-2006
- To circumvent the name disambiguation issue we focus on uncommon names



- Assumes that patents belong to the same discipline or paradigm as the papers
  - Both come from the same intellectual space (the inventorauthor's output)



### Articles

- Author-Org pairs for all papers 2002-2006
- Find fraction of author for each Author-Org pair
- Limit set to those with frac>0.5 (assures each name appears only once)

### Patents

- Inventor-Assg pairs for all patents 2002-2006
- Find fraction of inventor for each Inv-Assg pair
- Limit set to those with frac>0.5

- Join data using Auth=Inv
- Check Org=Assg to assure match



Rareness	# auth	# inv	Inv-auth	Inv-auth +	Fraction	# patents	NULL
fraction	names	names	matches	inst	valid		assignees
range				matches	matches		
f = 1.00	1,106,404	278,146	35,360	7,843	0.222	18,816	3,708
$1.0 \ge f \ge 0.9$	1,138,340	281,214	38,842	9,068	0.233	25,370	3,973
$1.0 \ge f \ge 0.8$	1,222,530	292,594	47,454	11,362	0.239	34,653	4,653
$1.0 \ge f \ge 0.7$	1,305,848	304,527	56,774	13,440	0.237	42,129	5,391
$1.0 \ge f \ge 0.6$	1,462,269	330,886	76,027	17,077	0.225	52,106	6,948
$1.0 \ge f > 0.5$	1,512,207	335,987	84,402	18,251	0.216	55,820	7,703
$1.0 \ge f \ge 0.5$	1,971,180	425,546	148,532				
UNIQUE	2,182,303	436,521					
(authfi)							
UNIQUE	8,712,536	1,049,650					
(authfi+inst)							



- 84,402 potential matches (rare author name matches) were manually inspected. 18,251 of these matches had an institutional match as well
- 55,387 patents (6.7% of US patents over the time frame) were invented by these matched authors
  - Patents were fractionally assigned to disciplines and paradigms through the "patent-inventor/author-paper-paradigm" linkage chain
  - 132,600 papers were authored by these 18,251 inventor-authors



### Patent IPC distribution





### Patent IPC Distribution



- Sample is not representative of actual distribution by class
- Should it be? NO
  - Representative distribution assumes that all patents and classes are equally likely to be inherently linked to science
  - We believe this is a faulty assumption
    - Some classes intuitively do not link to science (e.g., A4 – Personal articles, E0 – Building)
    - Others are intuitively science based (e.g. C0 Chemistry, A6 – Health)
    - Different industries have different publishing habits



### Patent map (IPC subclasses – 3 char)





### Using maps as templates



### Disciplinary map – paper counts





### Circle map



1

SOCIAL SCIENCE		COMPUTER SCIENCE	E	NGINEERING	EARTH SCIENCES	BI	OCHEMIST	RY S	MEDICAL SPECIALTIES		HEALTH SERVICES
H	IUMANITIE	S	MATH & PHYSICS	CHEMISTRY	1	BIOLOGY		NFECTIOUS	5	BRA IN RESEARCH	

**DEGREES OF INTEGRATION** These diagrams of scientific institutions were constructed by sorting over 16,000 academic journals into 554 different clusters, which were then grouped into 13 major fields. These major fields are shown as the 13 colored arcs comprising each circular map. Arc lengths represent the number of journals in the corresponding fields. These circular maps can be used to display the disciplinary makeup of institutions. Take, for example, the Rensselaer Polytechnic Institute Center for Biotechnology and Interdisciplinary Studies. First, the papers authored by this center are mapped to their corresponding disciplines on the circle. Then the average position of these papers is calculated. Colored rays are drawn from this point (the institutional node) to each of the papers on the circle to show disciplinary makeup. The position of the institutional node and distribution of the colored rays give a measure of the interdisciplinarity of the institution. The closer the institutional node is to the center of the circle, and the greater number of colors it incorporates, the more interdisciplinary the institution.



#### SEED · January/February 2008



#### Locating Patents on the Simplified Map of Science

Science-Technology Linkage: Patents were linked to the 554 scientific disciplines on the map of science. These links were based on a set of 18,250 people who were both inventors (on 55,400 patents) and authors (of 132,600 scientific publications) from 2002-2006. Additional information about the method for linking inventors and authors is available in Boyack & Klavans, "Measuring science-technology interaction using rare inventor-author names," Journal of Informetrics, 2008.



### Places & Spaces Iteration 4

CS; EE Math; Physics Chemistry Engineering Earth Sciences Biology Biotech Infect Disease Med Specialties Health Services Brain Research Humanities Social Sciences



- Consensus map generated from 20 maps
  - Robust
  - NON-CENTRIC map if all consensus edges are used
- Simplified map (circular) may be an effective map for policy purposes
  - Show interdisciplinary
  - Show technology (patent) profiles
  - Etc.