

Research Intelligence

Using Advanced Metrics to Identify Unique Institutional Research Strengths, Recruit and Retain Talent, and Drive More Productive University-Industry Partnerships

University-Industry Partnership Workshop 25 September 2017

Detroit

Daniel Calto Global Director of Solution Services Research Intelligence



Agenda

- Data Sources and Technologies: Scopus, SciVal, Elsevier Fingerprint Engine
- Advances in SciVal Metrics Topics of Prominence (TOP)
 - What are TOP, how do they work?
 - Why are they different and better than previous methods?
 - Broadening SciVal use cases
- Research Strengths and SWOT Analysis
 - How can TOP be used to identify specific strengths at a national level?
- Driving University-Industry Partnerships
- Q&A

Elsevier – A Unique Vantage Point on the Global World of Research



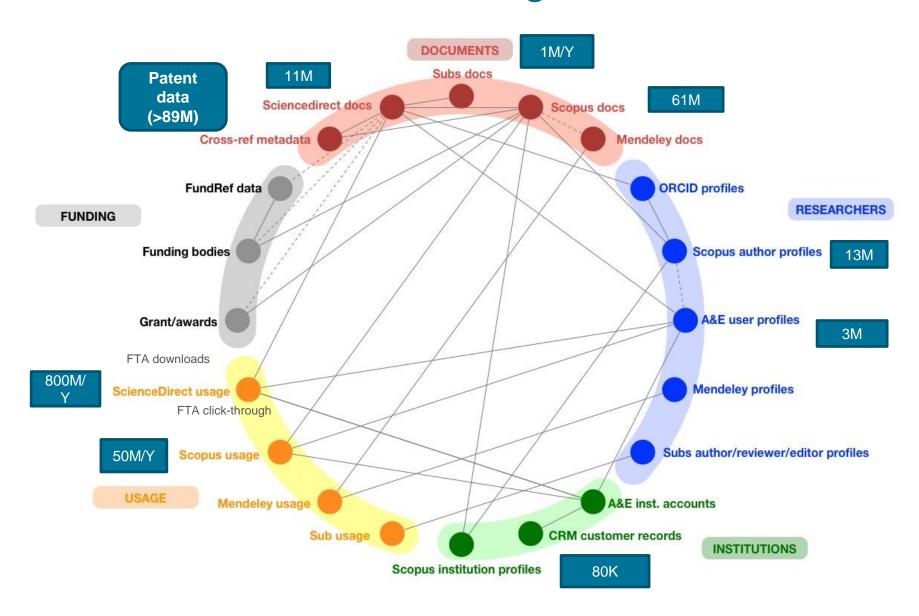
Elsevier – From publisher to solution provider

- Founded 130+ years ago
- Serving 30 million+ scientists, students, health and information professionals in 180+ countries
- 2,500 E-journals, 2,000 E-books published each year, dozens of researchoriented databases
- RELX (Reed Elsevier), the largest digital company in Europe

Each year

- 1.3 million manuscripts submitted to 2,500+ Elsevier journals
- 350,000+ articles published
- 900 million digital article downloads delivered
- 22,000+ journals from 5000+ publishers, >2 million articles per year tracked by Scopus (>69M articles in total)
- Terabytes of data in the Elsevier Research Intelligence suite
- Interactions with every university and government

What Data Do We Bring to the Table?



Scopus Coverage Summary

World's largest Abstract and Citations Database

69M records from **22K** serials, **100K** conferences and **150K** books from more than **5000** publishers and **105** countries

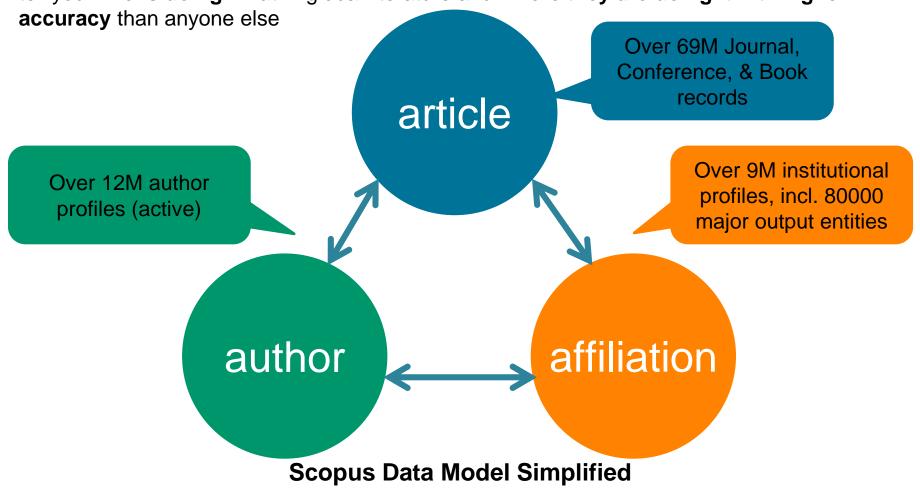
- Updated daily
- Records back to 1823
- "Articles in Press" from > 8,075 titles
- 40 different languages covered
- 3,643 active Gold Open Access journals indexed

CONFERENCES JOURNALS BOOKS PATENTS* Physical Sciences 562 book series 7,441 **27M** patents 21,951 peer-reviewed 100K conference iournals Health events **Sciences** 280 trade journals 8M conference 150K stand-alone 7,133 From 5 major books papers Social patent offices Full metadata, abstracts 1.2M items Sciences - WIPO and cited references (refs 8,698 - EPO post-1970 only) Life - USPTO Funding data from Mainly Engineering Sciences - JPO Focus on Social acknowledgements and Computer - UK IPO Sciences and A&H Citations back to 1970 Sciences

Scopus Data Model

The Scopus data model is designed around the notion that *articles* are written by *authors* that are *affiliated with institutions*. Visually and rather simplistically, this relational model is represented below.

What is the value of this structured data? This relational data model means that Scopus can tell you who is doing what in global literature and where they are doing it with higher

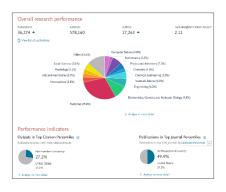


SciVal in a nutshell

SciVal offers quick, easy access to the research performance of 220 nations and 8,500 research institutions worldwide, and groups of institutions



Ready-made-at a glance snapshots of any selected entity





Flexibility to create and compare any research groups



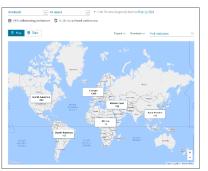


Identify and analyze existing and potential collaboration opportunities

Develop

collaborative

partnerships



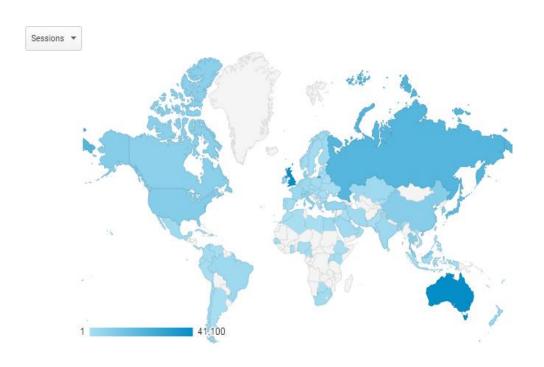


Analyze research trends to discover the top performers and rising stars

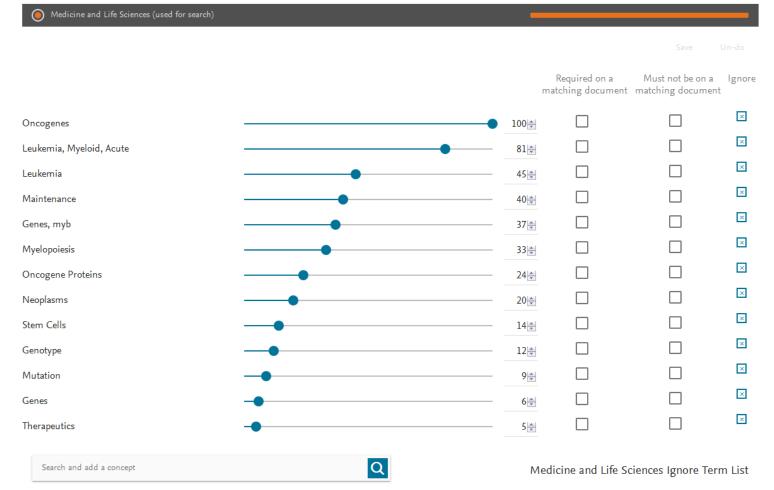


SciVal at this Moment

- To evaluate and demonstrate performance of research teams, institutions, cities, provinces and countries all around the world
- Launched in 2014. More than 500 customers, predominantly academic institutions.
- Corporate customers include: Unilever, Siemens, Boeing, Mercedes
- Several funding organizations and national government bodies also use SciVal for both internal and external analysis



Fingerprints Can Be Created from Any Text, or Group of Texts



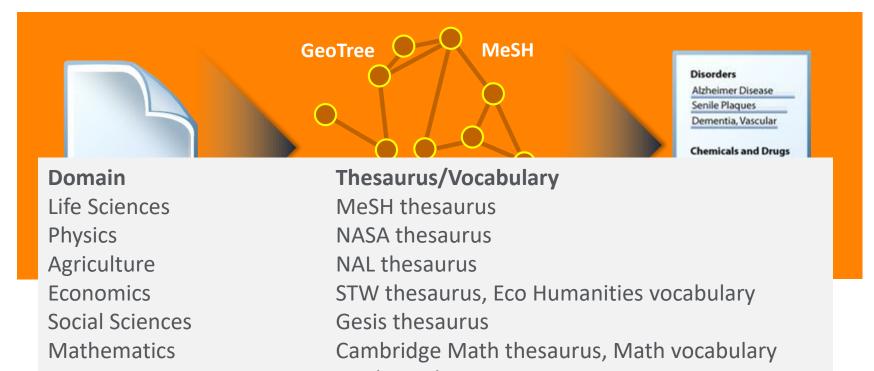
2

Concepts are derived from the text. Each concept is found in an underlying thesaurus suitable for the scientific area of the text. Concepts are weighted to create a precise summary of the text's meaning.



- Any text can be Fingerprinted, from grant applications to publications
- Fingerprints are generated from the title and abstract
- Natural Language Processing techniques are applied

How Does the Elsevier Fingerprint Engine Work?

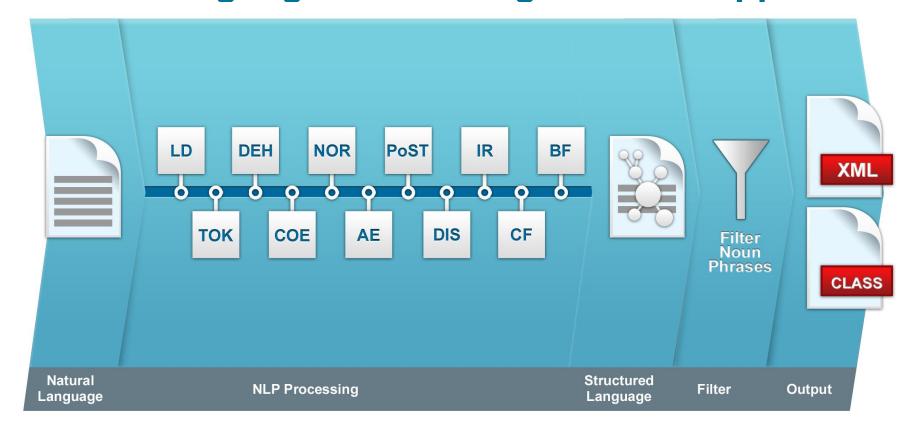


Geosciences Geobase thesaurus

Engineering Compendex thesaurus Humanities Humanities vocabulary

Compounds (Chemistry) Compendex thesaurus, MeSH thesaurus

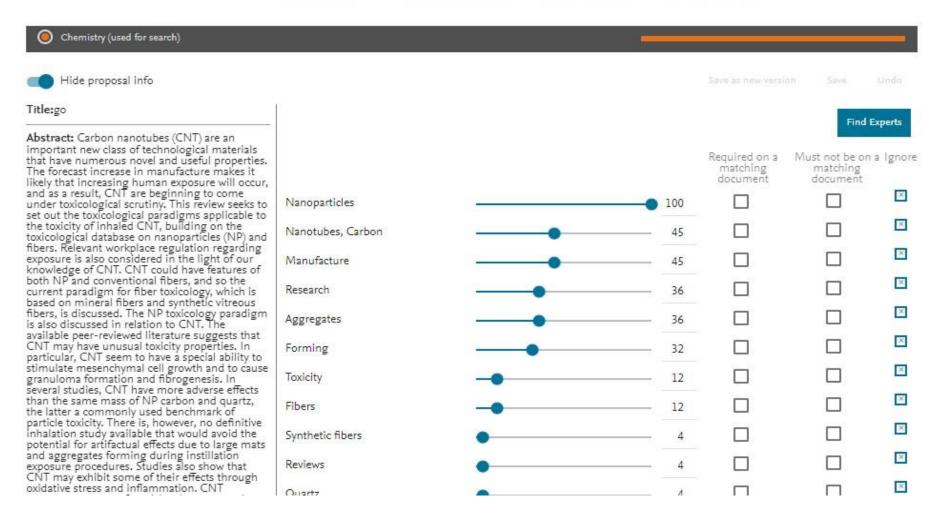
Natural Language Processing Modules Applied to Text



Natural Language via NLP Processing to Structured Semantic Machine-Readable Text

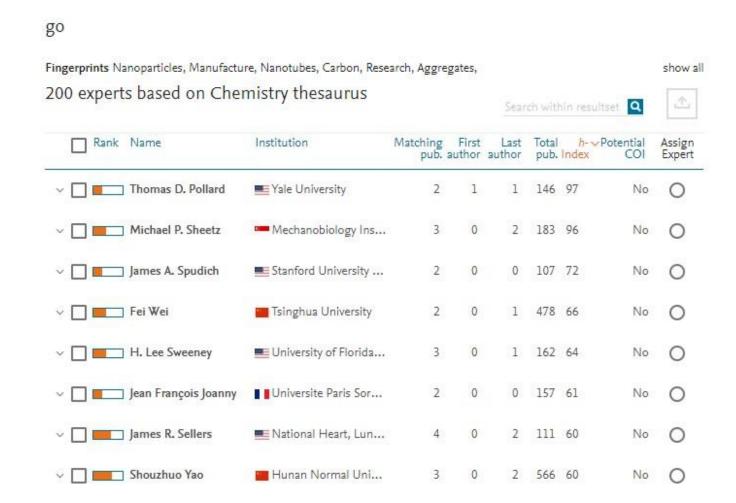
FPE In Action--Expert Lookup—Global Expertise Search





Expert Lookup—Global Expertise Search—Top Scholars

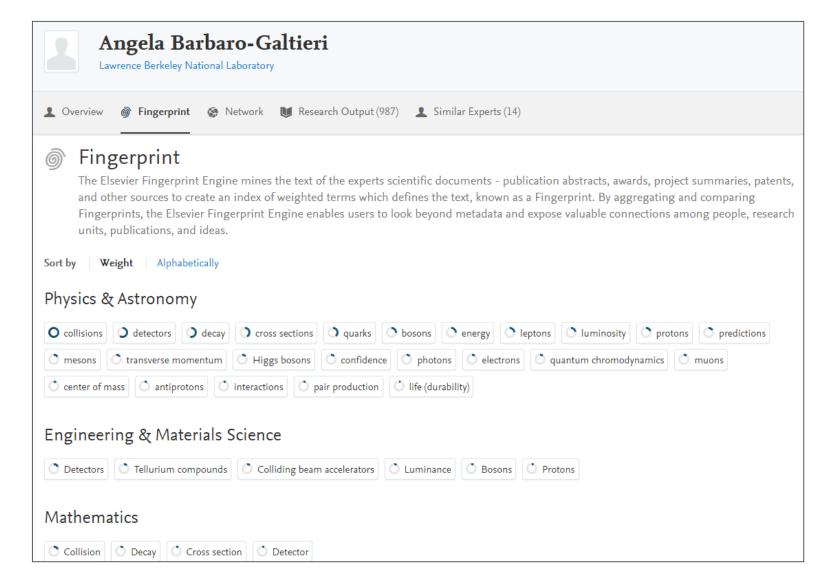




FPE in Action: SciVal

Keyphrase analysis Top 50 keyphrases by relevance, based on 1,030 publications | Learn about keyphrase calculations > Cobalt compounds Solid electrolytes Durability Electrochemical impedance spectroscopy Electrooxidation Nanoparticles lonomers Catalyst supports Precious metals Alkaline fuel cells Protons Electrocatalysis Oxygen Cathodes Membranes Hydrogen fuels Chemical stability Solid oxide fuel cells (SOFC) Electrodes Diffusion in gases Platinum alloys Fuel storage Catalysts Fuel cells Ionic conductivity **Platinum** Fuels Cerium compounds Electrolytic cells Direct methanol fuel cells (DMFC) Yttria stabilized zirconia **Electrolytic reduction** Rotating disks Anodes Methanol fuels Negative ions Gas fuel purification Electrocatalysts Polyelectrolytes **Energy conversion** Electrolytes Microbial fuel cells Gadolinium Hydrogen storage Ion exchange membranes Materials AAA relevance of keyphrase | declining growing (2011-2015)

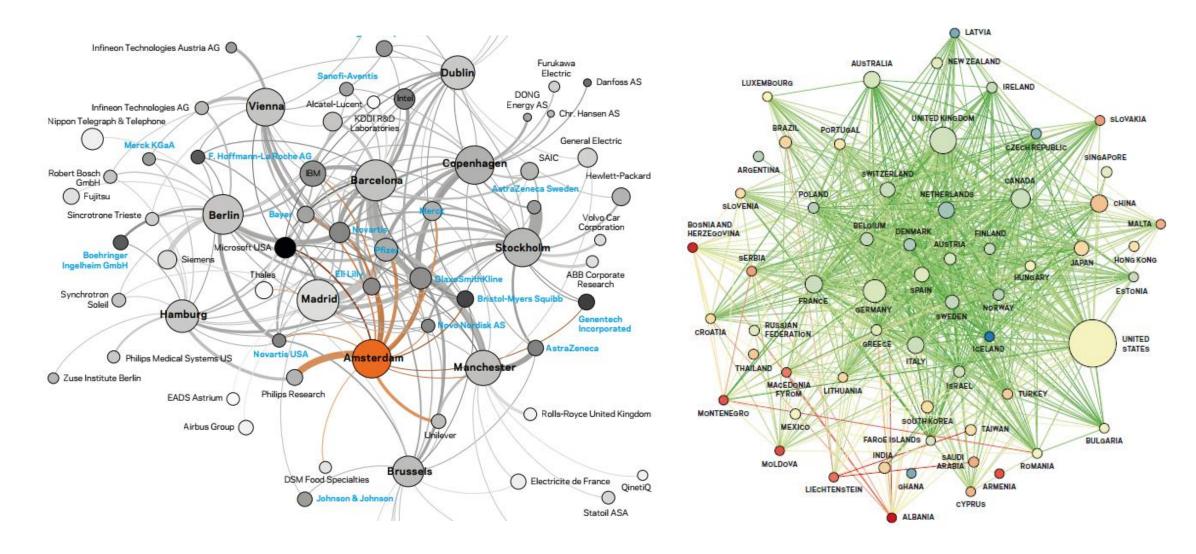
FPE in Action: Pure



Agenda

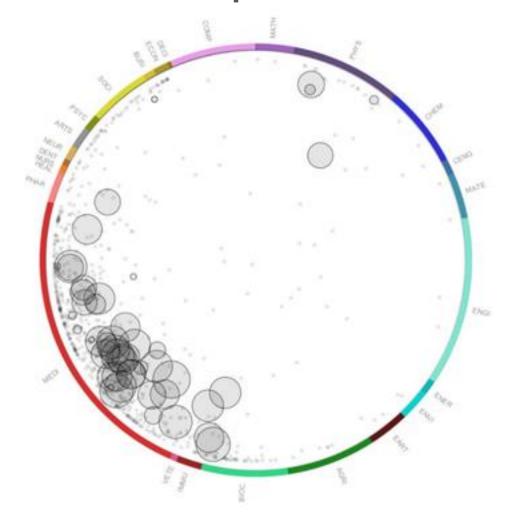
- Data Sources and Technologies: Scopus, SciVal, Elsevier Fingerprint Engine
- Advances in SciVal Metrics Topics of Prominence (TOP)
 - What are TOP, how do they work?
 - Why are they different and better than previous methods?
 - Broadening SciVal use cases
- Research Strengths and SWOT Analysis
 - How can TOP be used to identify specific strengths at a national level?
- Driving University-Industry Partnerships
- Q&A

Network Graphs – Production, Collaboration, Impact, Visualization..



Topics of Prominence

Recent advances in clustering and modelling science allow us to partition science into Topics

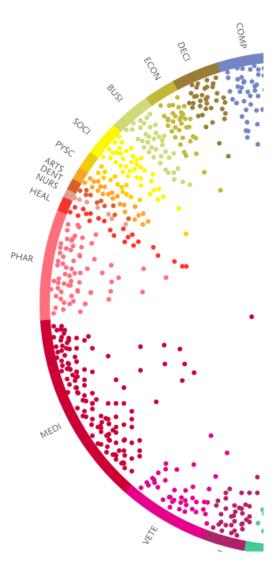


Looking at a specific institution to collaborate with..

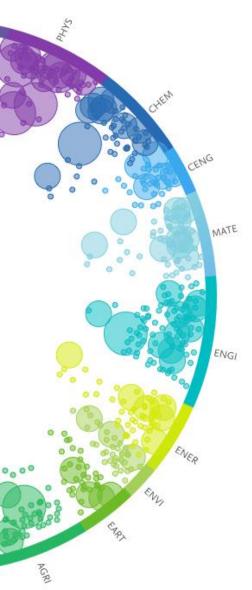
- What are the most prominent topics this institution is active in?
- What is this institution's presence in topics of prominence across the subject fields?
- Show me the dynamics within this topic, the top performers and rising stars?
- Are there any other topics I should keep an eye on?

Topics of Prominence—A Planning Solution

- We have identified ~100.000 global research topics and ranked them by Prominence.
- Prominence is a new indicator that shows the current momentum of a topic by looking at citations, views and CiteScore values.
- Prominence highly correlated with funding –
 helps researchers and research managers identify
 topics in which funding will increase.
- Going way beyond what the competition can do...



First Solution of its Kind



The first truly global detailed research portfolio analysis – this has never been done before – we use <u>all of Scopus</u> to form topics.

- Who's leading the way we can identify emergent topics with high momentum to understand who is currently leading the way.
- What's related We can tell you how the topics are related to your research portfolio.
- A better reflection of reality topics are an excellent reflection of reality since they are based on citation patterns and not journal categories and therefore truly multidisciplinary.

A groundbreaking concept

- Researchers in topics with high prominence receive more funding on average We have evidence that researchers in prominent topics receive more funding (per researcher) than their peers in other topics.
- Help improve grant applications we can truly help researchers to increase their grant success rate by focusing on high prominence topics.
- Topics resonate with researchers researchers recognize them intuitively and agree with the level of granularity.



Mapping Research Topics--History and Competition

- 1985 ISI (now Clarivate) develops Research Fronts
 - A bibliometric way to identify research opportunities
- 1988 CRP (now SciTech) develops Research Communities
 - Same algorithms and lower thresholds to increase coverage
- 2007 SciTech develops Distinctive Competencies
 - Clusters research communities using University strengths
- 2015 SciTech develops Topics
 - Significantly increases coverage and accuracy
- 2017 SciTech develops Topic Prominence indicator
 - Uses citations, downloads and journal impact
 - First time a bibliometric indicator is used to predict funding patterns

Mapping Research Topics--History and competition

Research Fronts (1985)

- 2% coverage
- 10,000 clusters

- (Clarivate is still using this!)
- Research Communities (1988) 4% coverage
- 35,000 clusters
- Distinctive Competencies (2007)15% coverage 200,000 clusters
- Topics (2015)

95% coverage 100,000 clusters

Topic Prominence (2017)

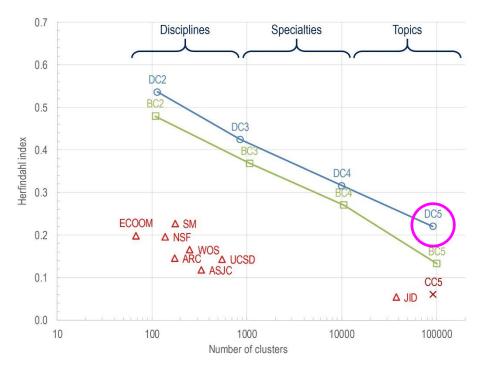
- 95% coverage Predicts funding
- Full coverage, accurately models supply/demand for science

Topics of Prominence—Accuracy

NEEDS

- Accuracy: Accurate topics that contain the right papers
- Comprehensive analysis at scale shows that topics based on direct citation are far more accurate than those based on bibliographic coupling or co-citation
- Also, they are much more accurate than journal categories

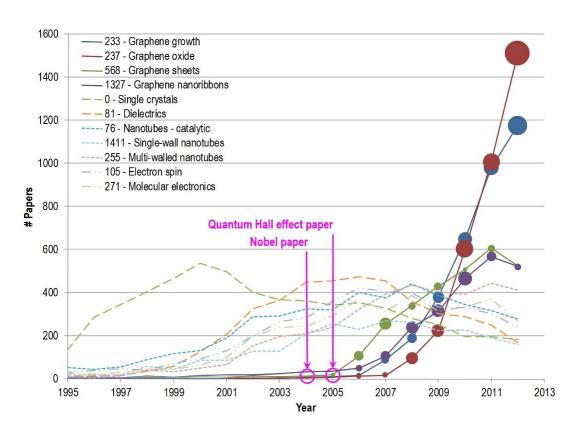
Use topics identified using direct citation



Klavans, R. and K.W. Boyack, Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge? JASIST, 2017. 68(4): p. 984-998.

Topics of Prominence – Variance and Dynamics

- Stability: Topics with realistic dynamics
- Topics can be new or old, large or small, growing, emerging, declining, interdisciplinary, etc., and have varied histories
- Topics have persistent dynamics; low birth and death rates, s-curve histories



Boyack, K.W. and R. Klavans, R. Creation and analysis of large-scale bibliometric networks. Springer Handbook of Science and Technology Indicators, 2018 (to appear).

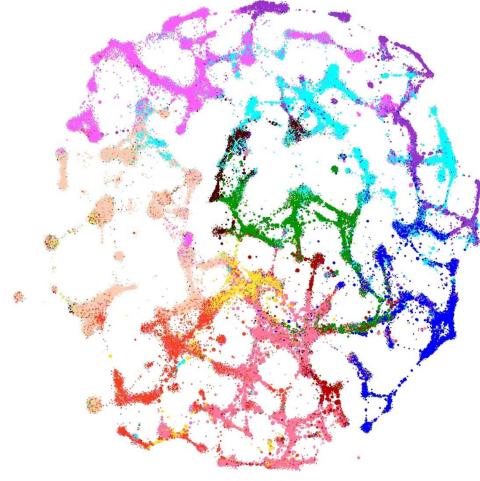
Topic Prominence—How is It Calculated?

Table 2. Factor loadings and scoring coefficients used to calculate topic prominence.

	Factor 1	Factor 2	Normalized Score
L:Citations	0.837	- 0.244	0.495
L:Views	0.812	- 0.262	0.391
L:CiteScore	0.653	0.154	0.114
L:Authors	0.593	0.334	(not used)
Vitality	0.441	0.269	(not used)

- Factor 1 has an eigenvalue of 2.33 (very high), suggesting that the composite indicator should include Citations, Views, Citescore
- Other formulations with more features were tested, but they did not have greater explanatory power than the 3-feature indicator
- $P_j = 0.495 (C_j mean(C_j))/stdev(C_j) + 0.391 (V_j mean(V_j))/stdev(V_j) + 0.114 (CS_j mean(CS_j))/stdev(CS_j),$

- Using 2013-10 datacut (source data 1996-2012)
- 582 million citing-cited pairs, 24.6 million source EID, 23.8 million cited non-indexed EID
- Calculated relatedness for 582 million pairs
- Ran SLM using resolution of 3 x 10⁻⁵
- A few clusters with <50 items were merged with larger clusters
- Result 91,726 clusters (topics)



Klavans, R. and K.W. Boyack, Research portfolio analysis and topic prominence. Journal of Informetrics, 2017 (under review).

Single Topic Characterization for 92,000 Topics

DC5 7909

TOP PHRASES (2011-2015)	score
1 anode material	20
2 anode materials	20
3 batteries LIBs	20
4 capacity retention	20
5 cycling stability	20
6 discharge capacity	20
7 electrochemical performances	20
8 electrode materials	20
9 electron microscopy	20
10 graphene oxide	20

FOP CATEGORIES (2011-2015)	score
1 Nanoscience & Nanotechnology	0.98
0,	
2 Energy	0.78
3 Materials	0.27
4 General Chemistry	0.05
5 Unclassified	0.04
6 Physical Chemistry	0.03
7 Inorganic & Nuclear Chemistry	0.03
8 Organic Chemistry	0.01
9 Chemical Physics	0.01
10 Applied Physics	0.01

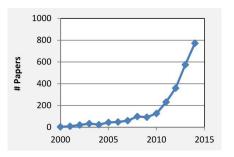
OP AUTHORS (2011-2015)	score
1 Ni S. (China Three Gorges University)	29
2 Qian Y. (University of Science and Techn	44
3 Yang X. (China Three Gorges University)	29
4 Ma J. (China Three Gorges University)	14
5 Lv X. (China Three Gorges University)	14
6 Pereira N. (Rutgers University)	14
7 Amatucci G.G. (Rutgers University)	19
8 Xiong Q.Q. ()	16
9 Zhang J. (China Three Gorges University	10
10 Xiong S. (Shandong University)	19

FOM: 2.9852 (98.07%); CPP: 21.069

IDIOSYNCRATIC PHRASES (2011-2015)	score
1 mA g ^{-12 batteries LIBs 3 superior electrochemical}	60.97 40.07 30.01
4 lithium storage	22.27
5 anode materials6 anode material	16.07 15.77
7 mAh g ^{-18 reversible capacity}	15.65 15.13
9 metal oxides 10 conversion reaction	13.96 12.85

OP SOURCES (2011-2015)	score
1 electrochim acta	2.96 2.85
2 j mater chem a 3 j power sources	1.78
4 nano energy	1.13
5 acs appl mater interfaces	0.78
6 rsc adv 7 nanoscale	0.59 0.45
8 j mater chem	0.43
9 mater lett	0.41
10 j alloys compd	0.26

ENGNG; DC4:20; DC3:269; DC2:23; REG:105



TOP INSTITUTIONS (2011-2015)	count
1 Nanyang Technological University	130
2 University of Science and Technology of	108
3 Shandong University	115
4 XiangTan University	37
5 CAS - Changchun Institute of Applied Ch	40
6 China Three Gorges University	30
7 University of Wollongong	49
8 Anhui University of Technology	24
9 Zhejiang Normal University	26
10 CAS - Shanghai Institute of Ceramics	24

REPRESENTATIVE PAPERS (2011-2014)

1 Reddy M.V. (2013) Metal oxides and oxysalts as anode materials for Li ion batteries. Chemical Reviews	530
2 Zhu X. (2011) Nanostructured reduced graphene oxide/Fe2O3 composite as a high-performance anode mater	514
3 Ji L. (2011) Recent developments in nanostructured anode materials for rechargeable lithium-ion batteries. En	576
4 Wang Z. (2012) Assembling carbon-coated î±-Fe2O3 hollow nanohorns on the CNT backbone for superior lith	270
5 Wang JZ. (2011) Graphene-encapsulated fe3O4 nanoparticles with 3d laminated structure as superior anode	230
6 Wang B. (2011) Quasiemulsion-templated formation of α-Fe2O3 hollow spheres with enhanced lithium storac	350
7 Sun B. (2011) MnO/C core-shell nanorods as high capacity anode materials for lithium-ion batteries. Journal o	118
8 Deng Y. (2011) One-pot synthesis of ZnFe2O4/C hollow spheres as superior anode materials for lithium ion base	106
9 Jin S. (2011) Facile synthesis of hierarchically structured Fe3O4/carbon micro-flowers and their application to	127
0 Wu H.B. (2012) Nanostructured metal oxide-based materials as advanced anodes for lithium-ion batteries. Na	324

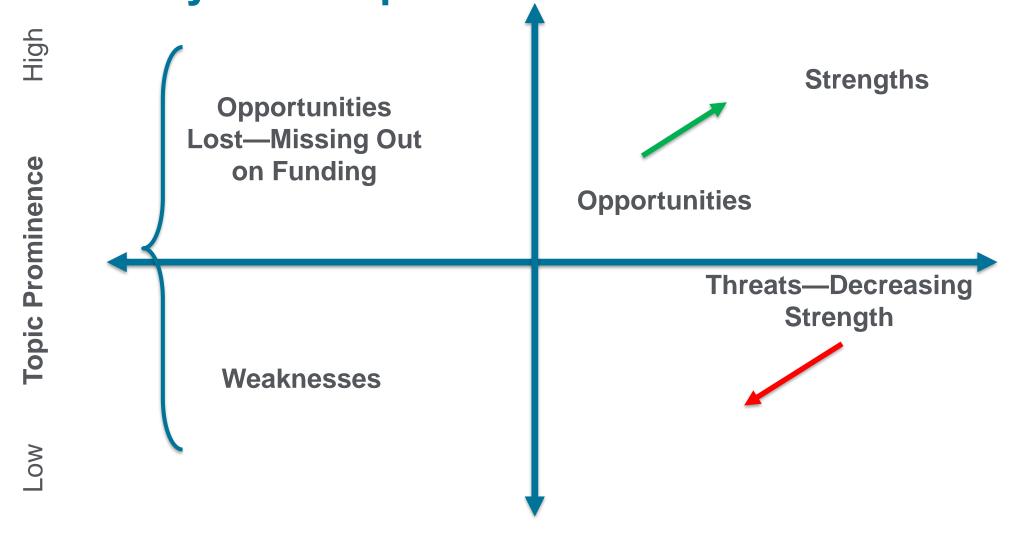
Topics of Prominence—Highly Correlated with Funding

- Funding divided into two time periods (2008-10, 2011-13)
- Prominence was calculated as of 2010

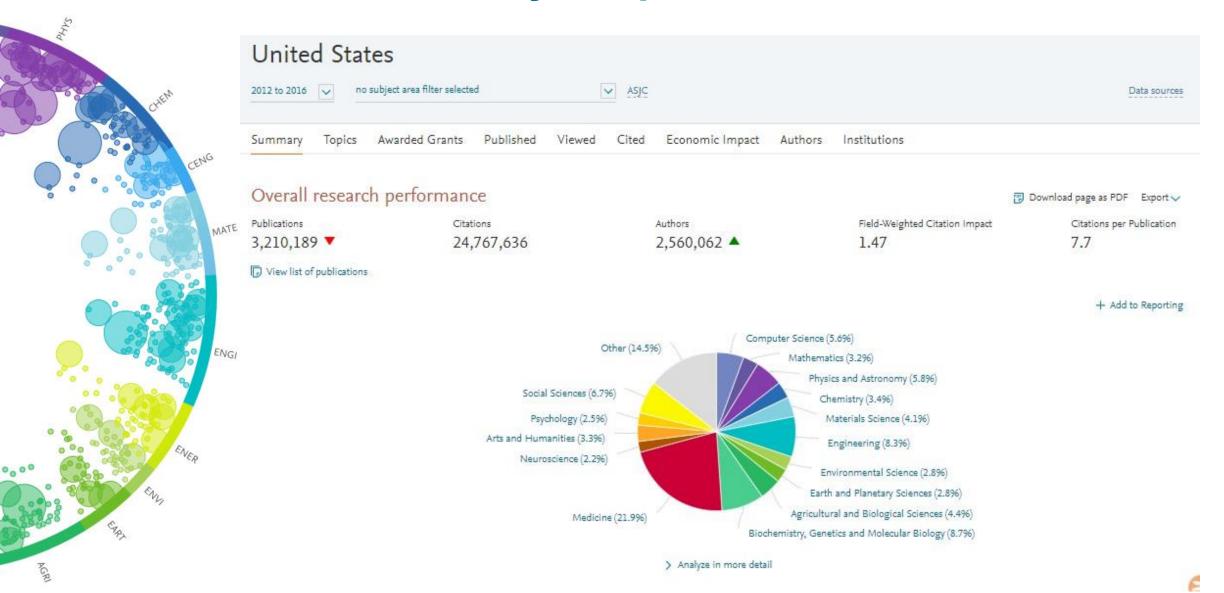
Table 4. Correlation matrix for variables considered in the funding prediction analysis.

	L:Fund1113	L:Fund0810	Prominence	Vitality	L:Authors
L:Fund1113	1.000				
L:Fund0810	0.837	1.000			
Prominence	0.606	0.616	1.000		
Vitality	0.166	0.162	0.314	1.000	
L:Authors	0.160	0.171	0.242	0.202	1.000

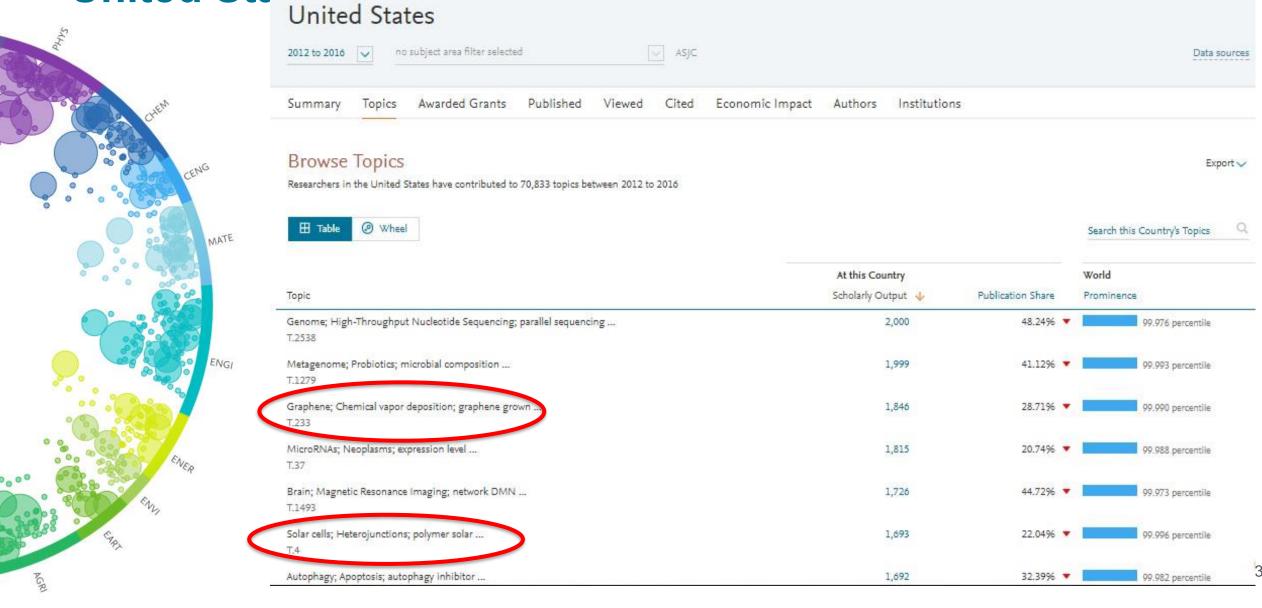
- Funding in two time periods is extremely highly correlated
- Prominence is highly correlated with funding in both time periods



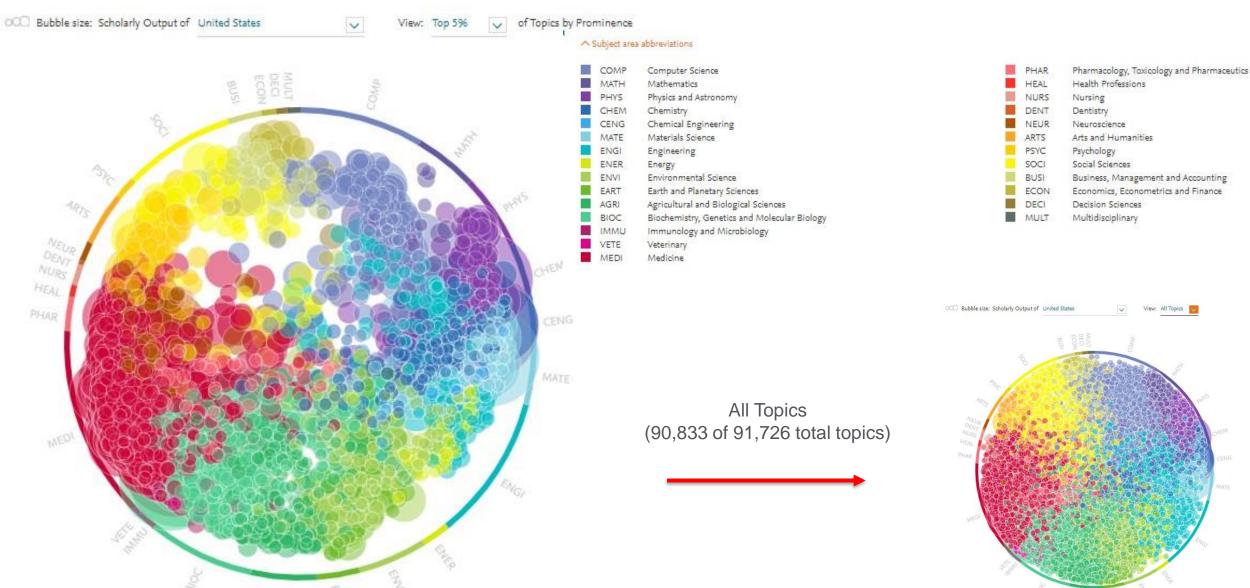
United States—Country Output



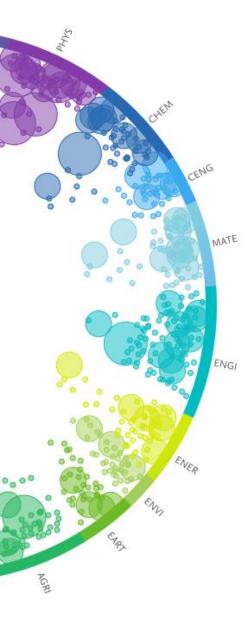
United States—Topics of Prominence United States

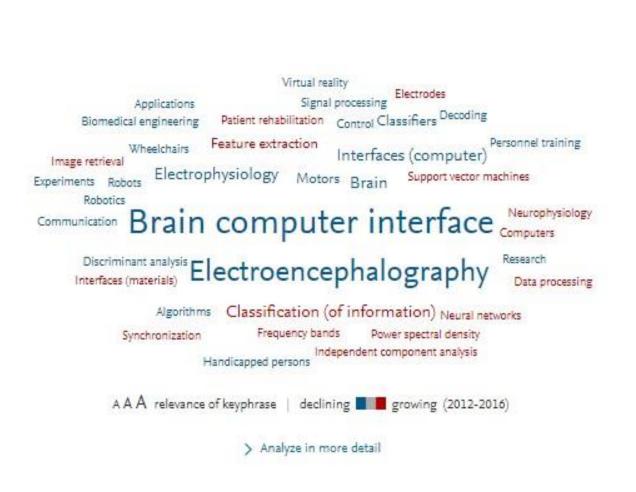


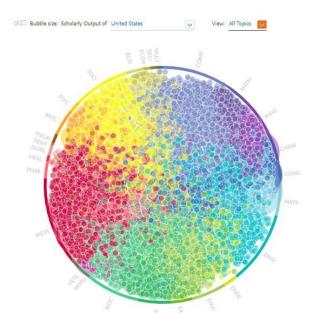
United States—Topics of Prominence—Top 5%



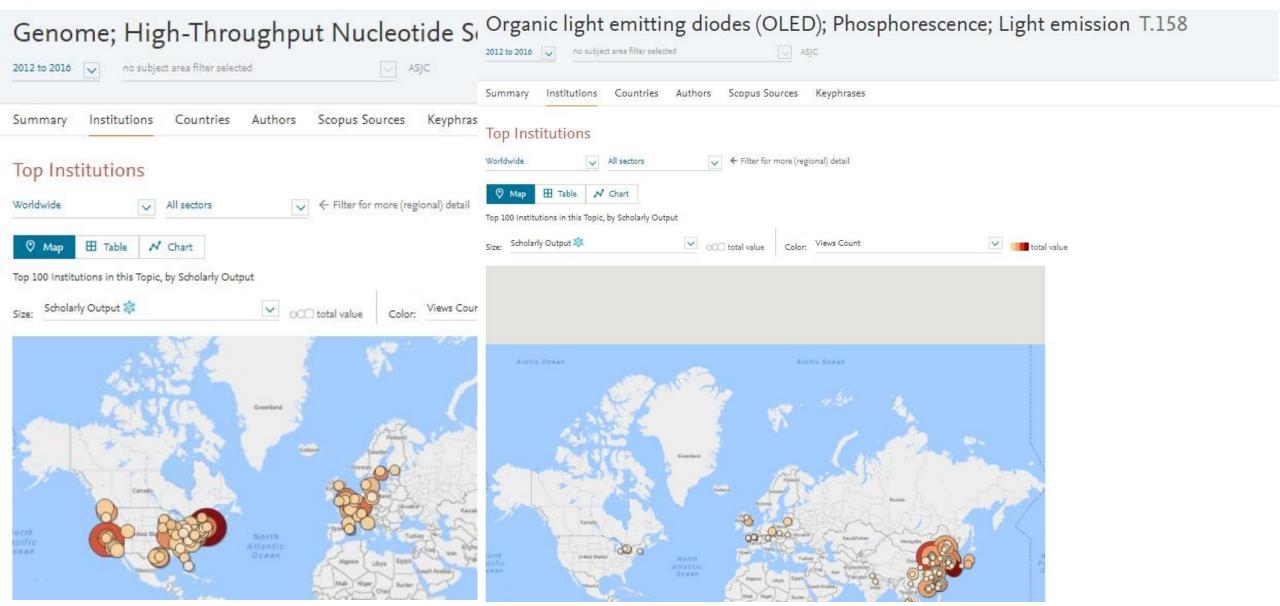
Semantic Word Cloud by Topic







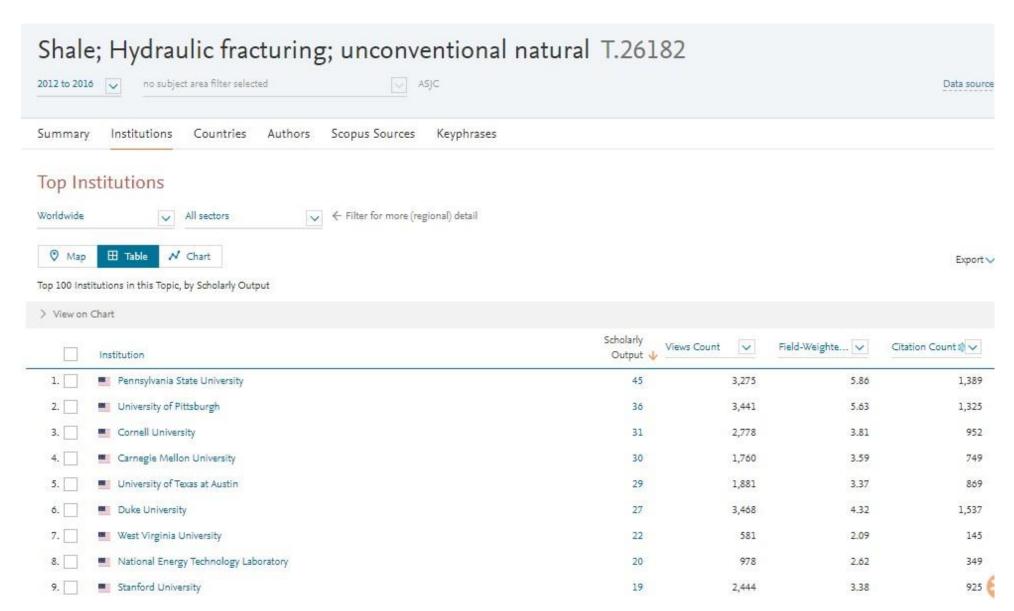
Leading Institutions in High-Throughput Genetic Sequencing



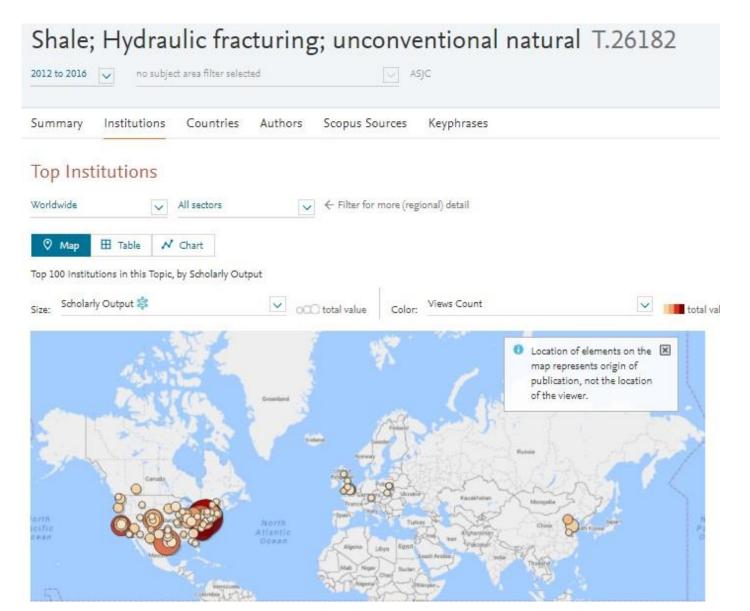
Leading Institutions in High-Throughput Genetic Sequencing

Genome; High-Throughput Nucleotide Sequencing; parallel sequencing T.2538 2012 to 2016 🗸 √ ASJC no subject area filter selected Data source: Scopus Sources Summary Institutions Countries Authors Keyphrases Top Institutions Worldwide All sectors Filter for more (regional) detail O Map **⊞** Table N Chart Export V Top 100 Institutions in this Topic, by Scholarly Output > View on Chart Scholarly Views Count Field-Weighte... V Citation Count 10 V Institution Output 4 1. Harvard University 152 4,231 9.00 15,135 Stanford University 123 3.69 2,873 4,530 National Institutes of Health 88 1,533 3.82 3,672 Washington University St. Louis 85 1,710 4.58 4,692 Chinese Academy of Sciences 76 1,442 1.21 758 Johns Hopkins University 74 2,030 12.03 12,538 Wellcome Trust Sanger Institute 66 1,940 5.42 4,355 Broad Institute 62 1,991 12.77 10,492 University of Washington 62 1,430 4.60 2,951

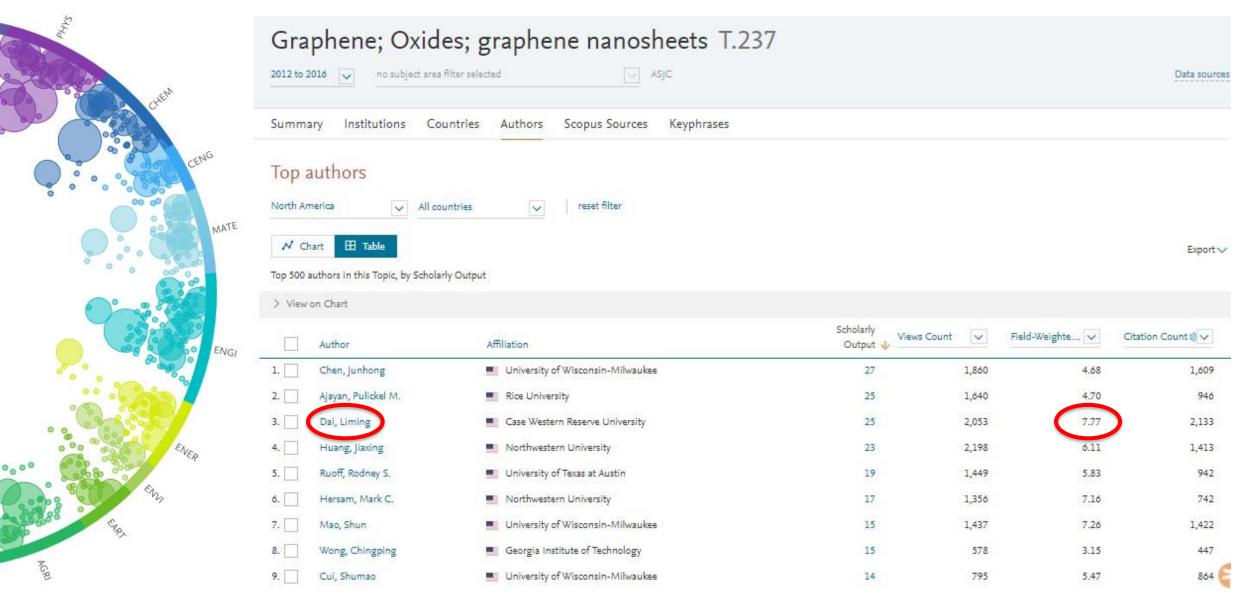
US Topics of Prominence—Shale, Hydraulic Fracturing



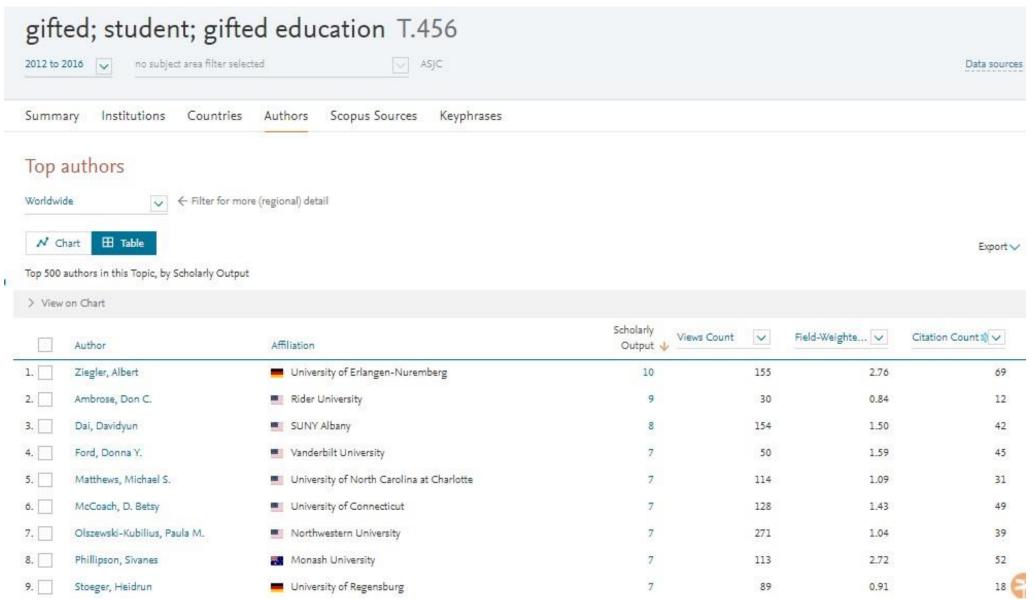
US Topics of Prominence—Shale, Hydraulic Fracturing



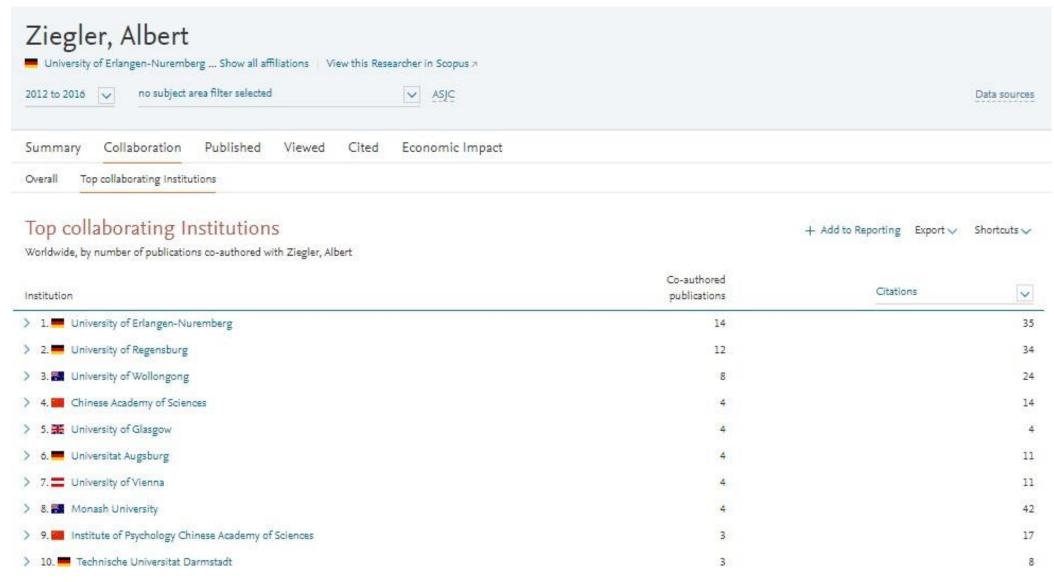
US—Topics of Prominence—Leading Authors



Topics of Prominence—Leading Authors Globally



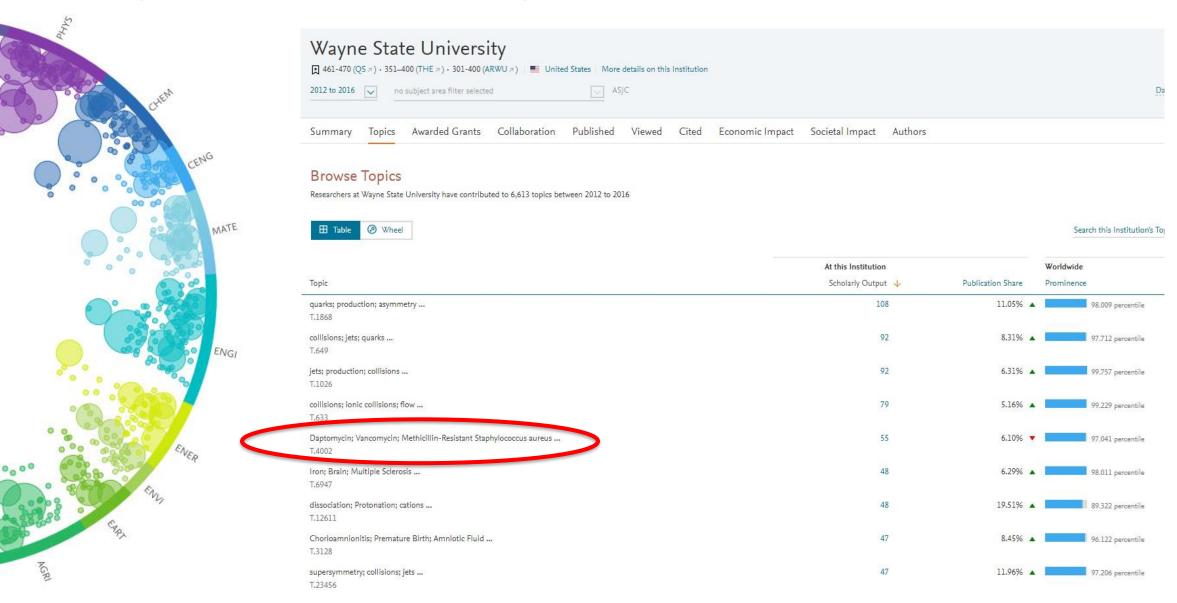
Topics of Prominence—Top Author Collaborations



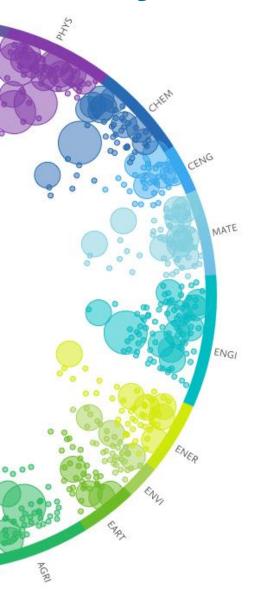
Agenda

- Data Sources and Technologies: Scopus, SciVal, Elsevier Fingerprint Engine
- Advances in SciVal Metrics Topics of Prominence (TOP)
 - What are TOP, how do they work?
 - Why are they different and better than previous methods?
 - Broadening SciVal use cases
- Research Strengths and SWOT Analysis
 - How can TOP be used to identify specific strengths at an institutional or national level?
- Driving University-Industry Partnerships
- Q&A

Wayne State University--Topics

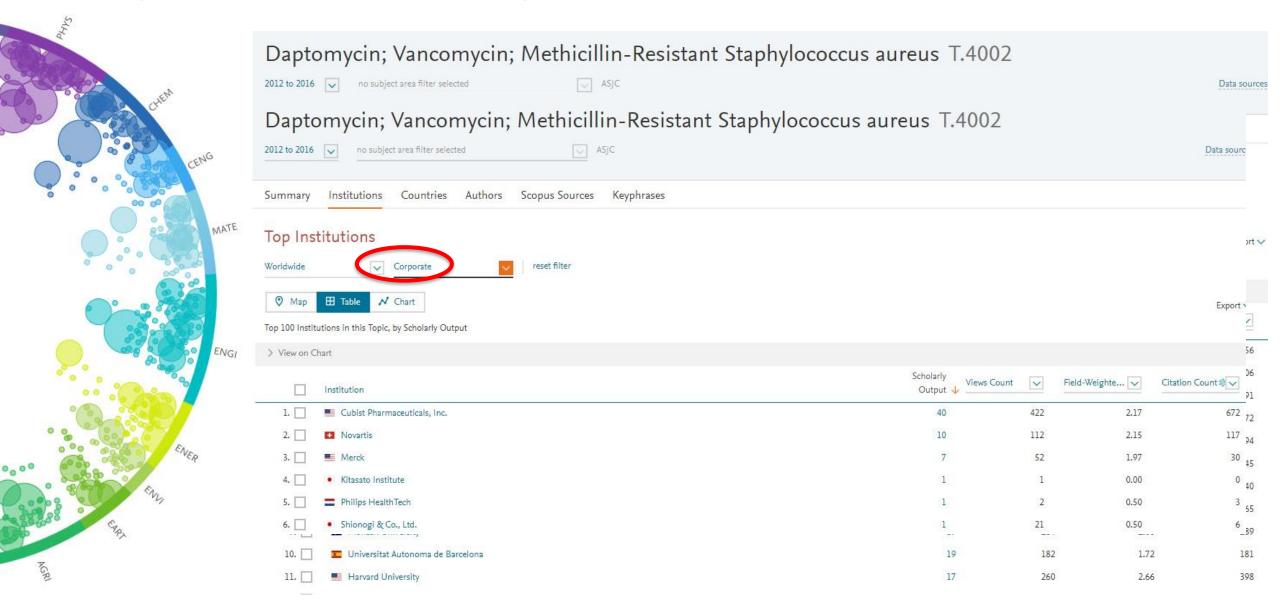


Wayne State University--Topics

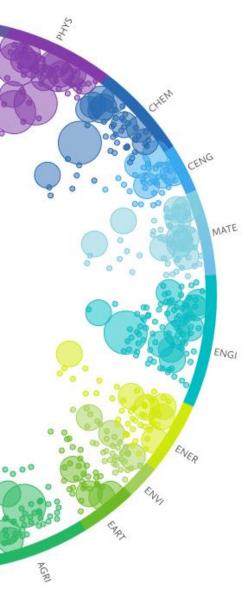


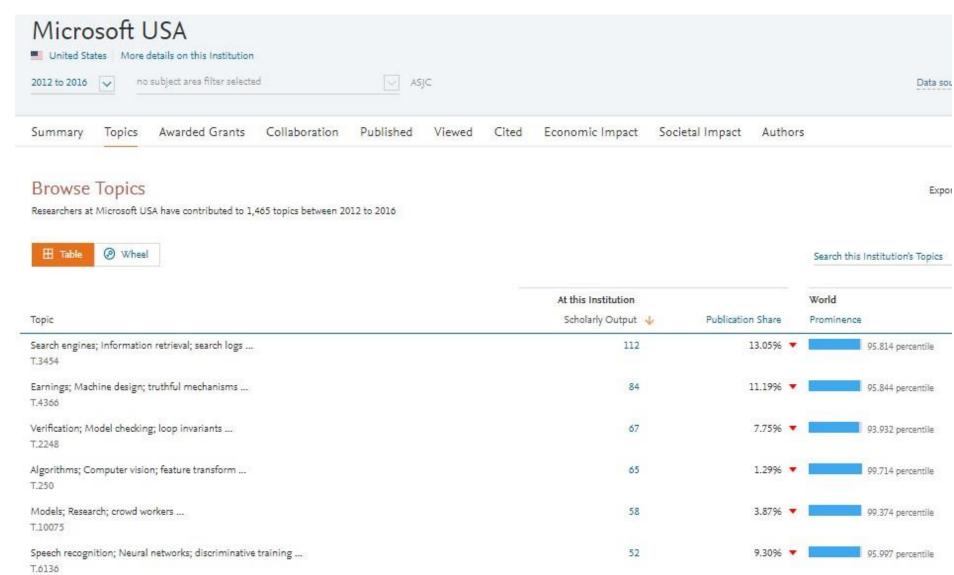


Wayne State University--Topics

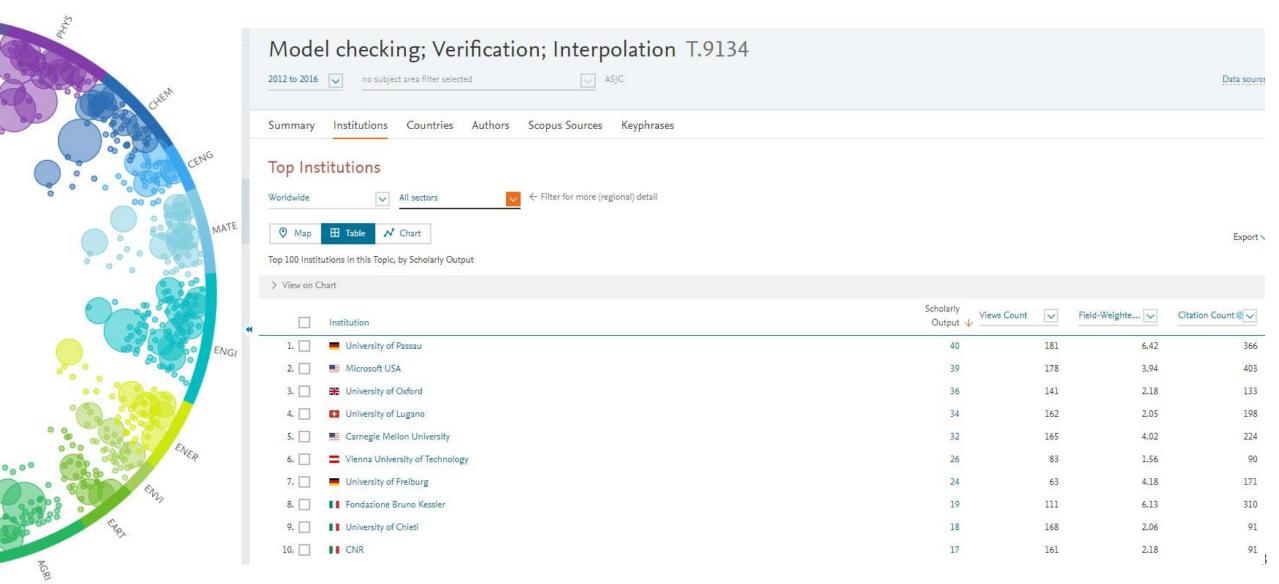


Microsoft USA--Topics

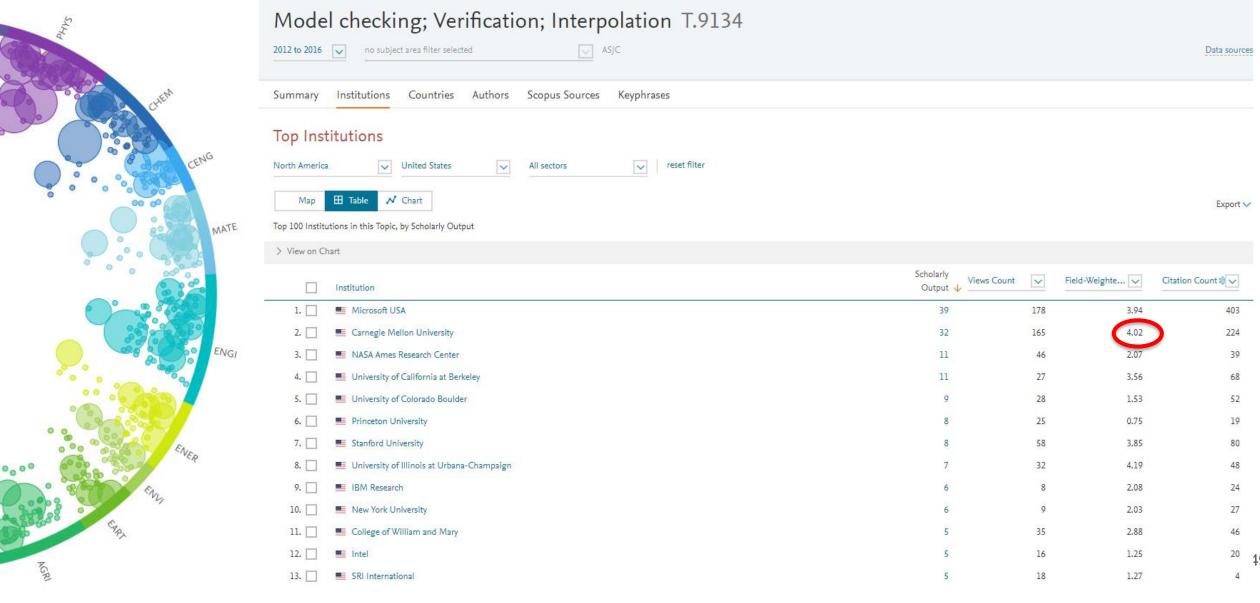




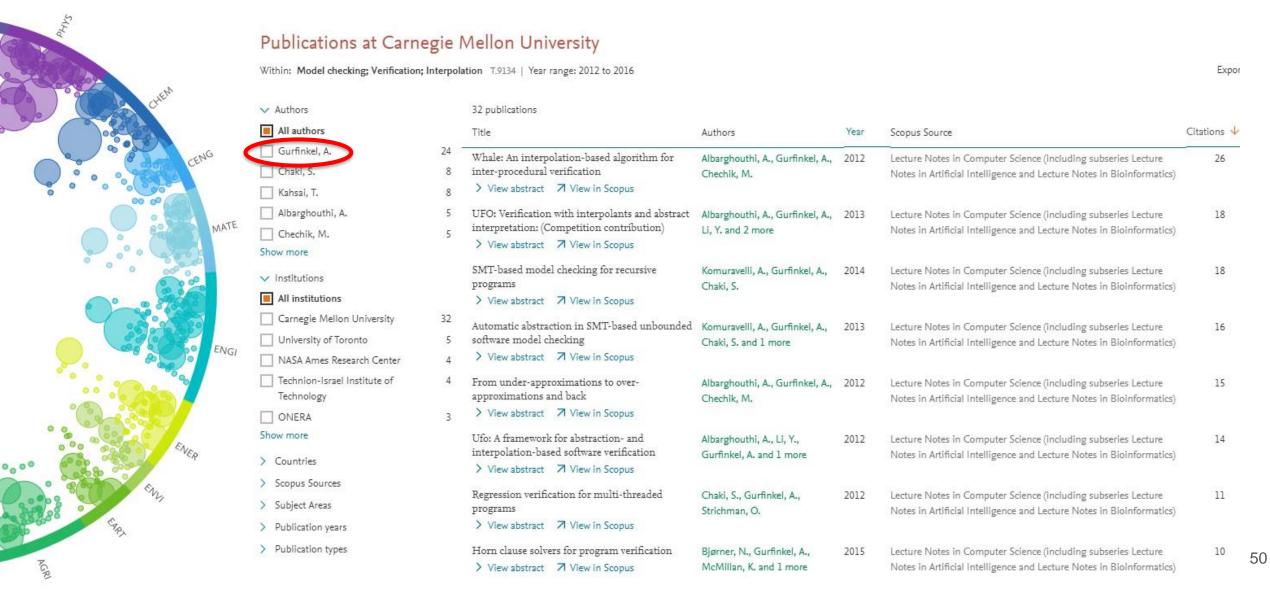
Top Global Institutions—Verification, Model Checking



Top US Institutions—Verification, Model Checking



Top US Institutions—Verification, Model Checking



Topics of Prominence Paper References

Which Type of Citation Analysis Generates the Most Accurate Taxonomy of Scientific and Technical Knowledge? Journal of the Association for Information Science and Technology, 68(4):984–998, 2017

Richard Klavans, SciTech Strategies, Inc., Wayne, PA 19087. E-mail: rklavans@mapofscience.com

Kevin W. Boyack, SciTech Strategies, Inc., Albuquerque, NM 87122. E-mail: kboyack@mapofscience.com

Elsevier Research Intelligence

Q&A Session

Thank You!

Daniel Calto
Global Director of Solution Services
Research Intelligence
Elsevier
d.calto@Elsevier.com
+1-917-455-4788

