Delving Deeper into Topic Prominence in Science
Advanced user webinar

Kevin Boyack :: President SciTech Strategies
Martin Edling Andersson :: Senior Product Manager, SciVal
Aileen Christensen :: Global Marketing Manager, SciVal
Agenda

Introduction & Demonstration

Methodology

• Portfolio analysis » Supply and demand
• Technical background
• Design choices
• Model characteristics » Topics, Prominence and Funding

Topic Prominence Roadmap

Q&A
Going beyond evaluation and benchmarking

- The **Trends Module** was launched in 2015 for **in-depth analysis of any research area created by the user**

- But what if we could **help the user find their topics of interest**?
So that we could…

…Help Research managers

• Identify pockets of well funded research in the research portfolio.

• Find the top performers and rising stars in those areas for recruitment, tenure and collaboration.

• Showcase that their institution is active in topics with high momentum

• Identify which topics other universities are active in that have high momentum.
Solution

- We have identified ~97,000 global research topics by clustering all of Scopus and ranked them by Prominence.
- Prominence is a new indicator that shows the current momentum of a topic by looking at very recent citations, views and CiteScore values.
- Prominence = momentum (not the same as importance!).
- Prominence predicts funding – helps researchers and research managers identify topics which are likely to be well funded.
- Going way beyond what the competition can do…
Demo:
Topics and Prominence in action
First of its kind

The first truly global detailed research portfolio analysis – this has never been done before – we use all of Scopus 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 are therefore truly multidisciplinary.
Research portfolio analysis

• Funding bodies allocate resources among topics
• Administrators choose which topics to support internally and who to hire
• Researchers choose which topics to work and submit proposals on

• Portfolio analysis definition: An analysis of elements of a product mix to determine the optimum allocation of resources.
  - If research is the product … then
  - Agency – optimizes a mix of funded topics
  - University – optimizes a mix of research topics
  - Researcher – optimizes a mix of research topics
Supply and demand

- Research is the product
- Topics (groups of publications) represent the supply
- Funding (purchasing of research) represents the demand

- Research portfolio analysis can be framed in terms of the supply of and demand for research

- Stakeholders who have information about the supply of and demand for topics will be able to make more information portfolio decisions
Current gap

• To optimize topic portfolios, stakeholders must know what the topics are and what value they bring to the portfolio

• GAP: There is currently no comprehensive and accepted model or list of topics (and their relative values) in science

• This is the gap that we address by
  o Creating a detailed model of the topics in science
  o Creating an indicator of demand (value) for topics that correlates with funding
Background
What is a “Topic”? 

• A topic is a collection of documents with a common intellectual interest – a “research problem”
• Topics can be large or small, new or old, growing or declining
• Topics are dynamic and can evolve
• New topics can be born
• Many topics are inherently multidisciplinary
• Old topics may be dormant, but still exist
• Researchers have mobility and can contribute to multiple topics
How can we identify topics

• The general approach is to take a set of documents and divide it into smaller groups

• The key is to do this in the most transparent and accurate manner possible
Background

- To divide documents into groups, we need to estimate relatedness between pairs of articles

- There are many ways to do this
  - Keyword similarity – not very discriminatory
  - Textual similarity – extremely computationally expensive
  - Citation links
    - Direct citation (DC) – simple and accurate
    - Indirect methods (BC, CC, LC) – computationally intensive
**Background**

- Direct citation has variants

- Indexed items only
  - 32 million articles
  - 440 million links
  - But ... only 29 million linked

- Including cited, non-indexed items
  - 32 million articles
  - 31 million non-indexed items
  - 830 million links
  - Far more signal with which to cluster the 32 million source articles
Background

• Clustering is done using algorithms that
  o Divide the documents into groups
  o Have a resolution parameter where increasing the resolution increases the number of clusters and reduces cluster sizes
  o Maximize the links within clusters and minimize the links between clusters
Design choices
Design needs and choices

• NEEDS
  o Coverage: A complete list of topics in science
  o Granularity: Topics that are of the appropriate size and number
  o Accuracy: Accurate topics that contain the right papers
  o Stability: Topics with realistic dynamics
Topic coverage

• NEEDS
  o Coverage: A complete list of topics in science
  o Keyword/journal based datasets
    o Contain many papers with a low fraction of links within the set; these papers belong to topics that are not core to the local dataset (low precision)
    o Exclude many papers that are strongly linked to the dataset (low recall)
  o By definition, a full database contains most relevant papers
  o We use the full Scopus database
Number of topics

• NEEDS
  o Granularity: Topics that are of the appropriate size and number
    o Most analysis is done using (~250-350) journal subject categories
      o While this is OK for some tasks, it is not sufficient for portfolio analysis
    o Researchers and funding officers can differentiate between 100,000 topics
  o Early work on scientific specialties suggested that they were comprised on average of about 100 authors
    o 10,000,000 Scopus authorIDs have published in the last 4 years
      o This suggests around 100,000 topics in science
  o For portfolio analysis, use around 100,000 topics
Design needs and choices

• 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

Design needs and choices

• NEEDS
  - Stability: Topics with realistic dynamics
  - Topics created using bibliographic coupling or co-citation are inherently unstable; most new topics disappear after one or two years
  - Topics created using direct citation have realistic dynamics; low birth and death rates, s-curve histories

Design needs and choices

• NEEDS
  o Coverage: A complete list of topics in science
    o We use the full Scopus database to get complete coverage
  o Granularity: Topics that are of the appropriate size and number
    o 100,000 topics
  o Accuracy: Accurate topics that contain the right papers
    o Topics are based on extended direct citation
  o Stability: Topics with realistic dynamics
    o Papers don’t move around; S-curve-like dynamics
Creating topics in practice
Model of topics – process

- Create list of citing-cited (paper-reference) pairs
  - This forms a large graph
  - 80M docs, 900M pairs

- Calculate relatedness value for each pair, based on number of links

- Using the full list of links and relatedness values, cluster the documents using the VOS code to create topics
Model of topics – process

- Create list of citing-cited (paper-reference) pairs
  - This forms a large graph
  - 80M docs, 900M pairs

- Calculate relatedness value for each pair, based on number of links

- Using the full list of links and relatedness values, cluster the documents using the VOS code to create topics
Model of topics – process

- Create list of citing-cited (paper-reference) pairs
  - This forms a large graph
  - 80M docs, 900M pairs

- Calculate relatedness value for each pair, based on number of links

- Using the full list of links and relatedness values, cluster the documents using the VOS code to create topics
SciVal topics

- Using source data 1996-present
  - ~900 million citing-cited pairs
  - ~40 million source docs
  - ~35 million cited non-indexed docs
- Calculated relatedness for 900 million pairs
- VOS code used for clustering
- Result – ~96,000 topics
Model can be extended in time

• New documents are added to the existing model using their references and text
  - Over 90% accuracy

• A new model is not needed each year
• This stability is good for users

• A few emerging topics can be split from existing large topics each year
Topic characteristics
Topic characteristics – size distribution

- Numbers of documents and authors per topic are distributed in a way that is typical of complex networks
- This correlates with the notions of big science and little science

What does a topic look like?

**RNA, Long Untranslated; Neoplasms; cancer tissues T.115**

*Summary*
- Institutions: 5
- Countries: 3
- Authors: 28
- Scopus Sources: 2
- Keyphrases: 10

**Overall research performance**
- Scholarly Output: 3,329
- Field-Weighted Citation Impact: 4.05
- International Collaboration: 544

**Views Count**: 39,837
**Citation Count**: 103,471
**Topic Prominence percentile**: 99.983

Source: Scopus

**Keyphrase analysis**
Top 50 keyphrases by relevance, based on 3,329 publications

---

**Organic light emitting diodes (OLED); Phosphorescence; host materials T.158**

*Summary*
- Institutions: 5
- Countries: 3
- Authors: 28
- Scopus Sources: 2
- Keyphrases: 10

**Overall research performance**
- Scholarly Output: 2,323
- Field-Weighted Citation Impact: 2.09
- International Collaboration: 328

**Views Count**: 58,329
**Citation Count**: 35,161
**Topic Prominence percentile**: 99.955

Source: Scopus

**Keyphrase analysis**
Top 50 keyphrases by relevance, based on 2,323 publications
What does a topic look like?

Supply chain management; Supply chains; GSCM practices
T.2569

2012 to 2016

Overall research performance

<table>
<thead>
<tr>
<th>Scholarly Output</th>
<th>Field-Weighted Citation Impact</th>
<th>International Collaboration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,266</td>
<td>2.66</td>
<td>350</td>
</tr>
</tbody>
</table>

Views Count: 87,596 Citation Count: 15,561 Topic Prominence percentile: 99.803

Keyphrase analysis

Top 50 keyphrases by relevance, based on 1,266 publications | Learn about keyphrase calculations

Neural networks; Convolution; convolutional layers T.4338

2012 to 2016

Overall research performance

<table>
<thead>
<tr>
<th>Scholarly Output</th>
<th>Field-Weighted Citation Impact</th>
<th>International Collaboration</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,435</td>
<td>8.42</td>
<td>292</td>
</tr>
</tbody>
</table>

Views Count: 10,550 Citation Count: 35,406 Topic Prominence percentile: 99.909

Keyphrase analysis

Top 50 keyphrases by relevance, based on 1,435 publications | Learn about keyphrase calculations
Topic features

• Conceptually simple
  - Articles that cite each other are mostly in the same topic

• Accurate problem-level subdivisions of science
  - We use the most accuracy clustering methods available

• Nearly complete coverage
  - Papers without references or abstracts are missing

• Topics are stable – papers don’t jump around

• Topics reflect known dynamics

• Topics enable direct comparison of institutions
Topic dynamics – graphene
Topic characteristics – prominence
Topic prominence

• Composite indicator
• Features (metrics) considered
  - Citation Count in year n to papers published in n and n-1
  - Scopus Views Count in year n to papers published in n and n-1
  - Average CiteScore for year n
  - Average number of authors per article for year n
  - Vitality – inverse reference age, similar to “state-of-the-art” from Competencies

- Patent reference counts and fraction industry authorship were also originally considered, but discarded to avoid features related to economic motives
Topic prominence

- Citations, Views, CiteScore and Author features were all highly skewed, and thus log-transformed for analysis to reduce skewness.

<table>
<thead>
<tr>
<th></th>
<th>L:Citations</th>
<th>L:Views</th>
<th>L:CiteScore</th>
<th>L:Authors</th>
<th>Vitality</th>
</tr>
</thead>
<tbody>
<tr>
<td>L:Citations</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L:Views</td>
<td>0.810</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L:CiteScore</td>
<td>0.533</td>
<td>0.483</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L:Authors</td>
<td>0.395</td>
<td>0.395</td>
<td>0.509</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Vitality</td>
<td>0.313</td>
<td>0.288</td>
<td>0.290</td>
<td>0.425</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Topic prominence

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

<table>
<thead>
<tr>
<th></th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Normalized Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>L:Citations</td>
<td>0.837</td>
<td>-0.244</td>
<td>0.495</td>
</tr>
<tr>
<td>L:Views</td>
<td>0.812</td>
<td>-0.262</td>
<td>0.391</td>
</tr>
<tr>
<td>L:CiteScore</td>
<td>0.653</td>
<td>0.154</td>
<td>0.114</td>
</tr>
<tr>
<td>L:Authors</td>
<td>0.593</td>
<td>0.334</td>
<td>(not used)</td>
</tr>
<tr>
<td>Vitality</td>
<td>0.441</td>
<td>0.269</td>
<td>(not used)</td>
</tr>
</tbody>
</table>

- 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
- \[ P_j = 0.495 \left( C_j - \text{mean}(C_j) \right)/\text{stdev}(C_j) + 0.391 \left( V_j - \text{mean}(V_j) \right)/\text{stdev}(V_j) + 0.114 \left( CS_j - \text{mean}(CS_j) \right)/\text{stdev}(CS_j), \]
Topic prominence

• Why call it “Prominence”
• Prominence ≠ Importance (Topics can be important but not prominent)
• Prominence ~ Visibility or Momentum
Topic characteristics – correlation with funding
Correlation with funding

• Assign individual grants to topics
• Correlate funding per topic with prominence
Assigning grants to topics

• Work done with a previous version of topics
• U.S. Star Metrics data (2008-2014)
  o Primarily from NIH, NSF
  o 364,000 individual grants, $253 Billion
• We have word profiles for each of the 91,726 topics in the model
• Titles and abstracts are available for most of the grants
• Grants were fractionally assigned to the five topics to which they had the greatest textual similarity
  o Some grants had insufficient text
  o 314,000 grants, $203 Billion were assigned
Correlating funding with prominence

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

<table>
<thead>
<tr>
<th></th>
<th>L:Fund1113</th>
<th>L:Fund0810</th>
<th>Prominence</th>
</tr>
</thead>
<tbody>
<tr>
<td>L:Fund1113</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L:Fund0810</td>
<td>0.837</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Prominence</td>
<td>0.606</td>
<td>0.616</td>
<td>1.000</td>
</tr>
</tbody>
</table>

- Funding in two time periods is extremely highly correlated
- Prominence is highly correlated with funding in both time periods
- Future funding is well predicted by past funding, but adding prominence improves this slightly
- However, overall, relatively little project-level funding data
- Thus, in the absence of comprehensive funding data, prominence is an extremely valuable indicator because it acts as a proxy for funding
Is this just a size effect?

- Prominence correlates with funding by topic
- Perhaps even more importantly, funding per author increases with prominence

Prominence examples

• High prominence – high funding (HH)
• High prominence – no funding (HN)
• Low prominence – high funding (LH)
## High prominence – high funding

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2538</td>
<td>674.2</td>
<td>99.8</td>
<td>420.6</td>
<td>807</td>
<td>2379</td>
<td>next-generation DNA sequencing</td>
<td>Cell Biology</td>
</tr>
<tr>
<td>73</td>
<td>364.5</td>
<td>99.3</td>
<td>305.2</td>
<td>1597</td>
<td>1960</td>
<td>T-lymphocytes</td>
<td>Immunology</td>
</tr>
<tr>
<td>1544</td>
<td>181.7</td>
<td>98.3</td>
<td>189.4</td>
<td>747</td>
<td>1297</td>
<td>orbitofrontal cortex and reward</td>
<td>Neurodeg Diseases</td>
</tr>
<tr>
<td>1493</td>
<td>223.7</td>
<td>99.0</td>
<td>180.1</td>
<td>742</td>
<td>2070</td>
<td>default mode network (brain)</td>
<td>Brain,Vision,Hearing</td>
</tr>
<tr>
<td>2771</td>
<td>246.1</td>
<td>98.4</td>
<td>150.2</td>
<td>753</td>
<td>1204</td>
<td>inflammation and obesity</td>
<td>Diabetes</td>
</tr>
<tr>
<td>5042</td>
<td>209.1</td>
<td>95.1</td>
<td>143.4</td>
<td>396</td>
<td>518</td>
<td>autism phenotype</td>
<td>Psychiatry</td>
</tr>
<tr>
<td>236</td>
<td>338.7</td>
<td>99.1</td>
<td>80.1</td>
<td>1215</td>
<td>1675</td>
<td>peptide identification in proteomics</td>
<td>Analytical Chemistry</td>
</tr>
<tr>
<td>205</td>
<td>216.5</td>
<td>98.2</td>
<td>41.6</td>
<td>1053</td>
<td>1523</td>
<td>amyloid function in Alzheimer’s</td>
<td>Neurodeg Diseases</td>
</tr>
<tr>
<td>2646</td>
<td>215.1</td>
<td>98.1</td>
<td>37.3</td>
<td>677</td>
<td>945</td>
<td>solid-state nanopores</td>
<td>Nanochemistry</td>
</tr>
<tr>
<td>2877</td>
<td>128.3</td>
<td>96.4</td>
<td>33.0</td>
<td>472</td>
<td>912</td>
<td>BPA and endocrine disruption</td>
<td>Environ Chemistry</td>
</tr>
</tbody>
</table>

- Obvious, high profile topics
**High prominence – no funding**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2187</td>
<td>0.5</td>
<td>95.8</td>
<td>0</td>
<td>591</td>
<td>849</td>
<td>electrochem degradation in wastewater</td>
<td>Electrochemistry</td>
</tr>
<tr>
<td>25</td>
<td>1.6</td>
<td>96.5</td>
<td>0</td>
<td>785</td>
<td>1532</td>
<td>corrosion inhibitors (steel)</td>
<td>Materials</td>
</tr>
<tr>
<td>135</td>
<td>2.7</td>
<td>98.2</td>
<td>0</td>
<td>1118</td>
<td>2136</td>
<td>dye remediation in effluents</td>
<td>Electrochemistry</td>
</tr>
<tr>
<td>15</td>
<td>4.2</td>
<td>98.7</td>
<td>0</td>
<td>1263</td>
<td>1961</td>
<td>biosorption of heavy metals</td>
<td>Electrochemistry</td>
</tr>
<tr>
<td>4003</td>
<td>7.8</td>
<td>98.8</td>
<td>0</td>
<td>547</td>
<td>878</td>
<td>dispersive liquid-liquid micro-extraction</td>
<td>Environ Chemistry</td>
</tr>
<tr>
<td>566</td>
<td>9.2</td>
<td>94.7</td>
<td>0</td>
<td>608</td>
<td>753</td>
<td>properties of olive extracts</td>
<td>Animal Science</td>
</tr>
<tr>
<td>4594</td>
<td>14.4</td>
<td>96.9</td>
<td>0</td>
<td>540</td>
<td>841</td>
<td>hollow nanoparticles</td>
<td>Electrochemistry</td>
</tr>
<tr>
<td>580</td>
<td>21.1</td>
<td>96.8</td>
<td>0</td>
<td>1010</td>
<td>1847</td>
<td>phosphors for LEDs</td>
<td>Optical Materials</td>
</tr>
<tr>
<td>644</td>
<td>21.1</td>
<td>95.4</td>
<td>0</td>
<td>725</td>
<td>777</td>
<td>hydrogen energy storage</td>
<td>Electrochemistry</td>
</tr>
<tr>
<td>7</td>
<td>74.3</td>
<td>99.3</td>
<td>0</td>
<td>1963</td>
<td>2407</td>
<td>Zn0 nanostructures</td>
<td>Semicond Physics</td>
</tr>
</tbody>
</table>

- **Good reasons for lack of U.S. funding**
  - Environmental issues for industries outside U.S.
  - Olives not grown in U.S.
  - Materials, energy topics that are dominated by Chinese universities
Low prominence – high funding

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>83249</td>
<td>0.1</td>
<td>0.3</td>
<td>11.9</td>
<td>6</td>
<td>10</td>
<td>loose topic - clinical investigation centers</td>
<td>Patient Care</td>
</tr>
<tr>
<td>25667</td>
<td>2.2</td>
<td>2.2</td>
<td>24.8</td>
<td>6</td>
<td>13</td>
<td>academic medical centers, enthusiasm</td>
<td>Patient Care</td>
</tr>
<tr>
<td>54378</td>
<td>2.3</td>
<td>12.8</td>
<td>89.3</td>
<td>20</td>
<td>20</td>
<td>forestry education</td>
<td>Agricultural Policy</td>
</tr>
<tr>
<td>38569</td>
<td>2.7</td>
<td>14.2</td>
<td>13.7</td>
<td>27</td>
<td>29</td>
<td>agroecology, sustainability</td>
<td>Agricultural Policy</td>
</tr>
<tr>
<td>54158</td>
<td>5.0</td>
<td>4.6</td>
<td>12.0</td>
<td>18</td>
<td>60</td>
<td>loose topic – DMZ, networks, protocols</td>
<td>Computing</td>
</tr>
<tr>
<td>33105</td>
<td>10.9</td>
<td>6.1</td>
<td>29.6</td>
<td>42</td>
<td>60</td>
<td>role of nursing in clinical trials</td>
<td>Patient Care</td>
</tr>
<tr>
<td>18741</td>
<td>30.1</td>
<td>3.5</td>
<td>23.0</td>
<td>71</td>
<td>137</td>
<td>capstone projects, engineering</td>
<td>Learning</td>
</tr>
<tr>
<td>33702</td>
<td>30.4</td>
<td>11.2</td>
<td>18.3</td>
<td>51</td>
<td>77</td>
<td>extension programs and learning</td>
<td>Agricultural Policy</td>
</tr>
<tr>
<td>38645</td>
<td>31.1</td>
<td>8.6</td>
<td>10.9</td>
<td>72</td>
<td>92</td>
<td>systems engineering competency training</td>
<td>Management</td>
</tr>
<tr>
<td>26483</td>
<td>39.8</td>
<td>9.5</td>
<td>44.2</td>
<td>99</td>
<td>176</td>
<td>civil engineering program criteria</td>
<td>Learning</td>
</tr>
</tbody>
</table>

- Improper assignment of grants to topics due to common text
  - “clinical trials”, “medical center”
  - Education is addressed as a broader impact in many NIH, NSF grants
  - Small topics with loose text
Publication


• Available at https://authors.elsevier.com/a/1Vyjo6EAijhtjU or https://www.sciencedirect.com/science/article/pii/S1751157717302110
Updating topics
Updating topics

• For purposes of topic stability, we do not create a completely new set of topics each year

• Instead, we add new papers from Scopus to the existing topics
  - This can be done very accurately using the references in each paper

• However, there are new topics that emerge each year, and we need to account for these in SciVal

• Thus, in addition to adding papers to existing topics, we plan to create a few new topics each year
Creation of emerging topics

• Emerging topics have the following properties
  - Relatively small at the time of emergence with a high growth rate
  - An event – e.g., breakthrough paper(s) – that acts as the trigger for emergence
  - Persistence – the topic doesn’t die

• Emerging topics are, in all cases, related to existing topics – they don’t materialize from thin air

• Most emerging topics appear to be embedded in larger existing topics – these are candidates for topic splitting
  - This is based on comparing models from different years
Creation of emerging topics

• Process
  - Create a new set of topics using current information
  - Identify topics that meet our emergence criteria (size, growth, highly cited papers) in the current year and previous year (persistence) – our experiments suggest 30-50 new topics per year
  - Many of these emerging topics are subsets of a larger current topic in SciVal
  - Construct new topics in SciVal using the emerging topics from the new model, and remove their papers from their existing topics

• We may choose to augment this algorithmic identification of emerging topics with some manual identification of emerging topics based on current events in science
Contextual history
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
History and competition

• **Research Fronts (1985)**  
  - Clarivate is still using this!  
  2% coverage  10,000 clusters

• **Research Communities**  
  4% coverage  35,000 clusters

• **Distinctive Competencies**  
  15% coverage  200,000 clusters

• **Topics**  
  95% coverage  100,000 clusters

• **Topic Prominence (2017)**  
  Predicts funding  
  - Full coverage; accurately models supply of and demand for science
Summary

• We have created an accurate model with nearly 100,000 topics that is suitable for portfolio analysis
  o The methodology can be reproduced, but requires a full database

• We have created a topic-level indicator – Prominence – that is strongly correlated with future funding

• Funding per author increases with increasing topic prominence

• Topics and their prominence enable stakeholders in the science system to have the knowledge necessary to make portfolio decisions
Want to know more?

• Visit our info page for further information about the methodology, future iterations, benefits and use cases etc. >> elsevier.com/topic-prominence

• Social media >> following the conversations via Twitter, Facebook and LinkedIn #TopicProminence

• Introductory webinar >> via BrightTALK

• Read Journal article >> Research Portfolio Analysis and Topic Prominence

• The dawn of predictive analytics to measure research performance: SciVal’s Topic Prominence >> Read via Elsevier Connect
What’s Next: Topic Prominence in Science Roadmap

Just released
• Export 500 most prolific Topics via excel/csv

Coming up
• Filter Topics by ASJC subject classification in Overview module
• See representative papers (central papers) in the Trends module
• Get an overview of the related Topics in the Trends module
• Define new research areas based on Topics
• We will be supporting researchers, groups of researchers and groups of institutions
• Analyze any institution’s performance to see which prominent Topics you’re active in
Join the Conversation

SciVal Twitter
Twitter.com/scival

SciVal Facebook
Facebook.com/scival

Research Intelligence YouTube
Youtube.com/channel/UCffqlj7ICiYccxNDPEEjHHg

Research Intelligence LinkedIn
Linkedin.com/showcase/10576088

Research Intelligence BrightTALK
Brighttalk.com/channel/13819/elseviers-research-intelligence
Thank you.
Stay up to date: don’t forget to follow @SciVal on Twitter