



# Topic Prominence in Science FAQ

*Updated September 2018*



# NOTE

The following extensive FAQ document is based upon customer feedback and questions. It should answer any of the questions which you have, and will be built upon as new questions come in.

Should you have any questions please don't hesitate to reach out to your sales representative, or alternatively you can contact our support team via

<http://scival.com/contact>

# Topic Prominence in Science: FAQ 1/9

## **What is a Topic?**

A Topic is a collection of documents with a common focused intellectual interest, such as the work on a specific research problem. Documents within a Topic cite and are cited by other documents in the Topic.

## **How are Topics created?**

We take the entire citation network – roughly 70 million documents and a billion citation links between documents – and break that network into roughly 96,000 clusters, where the linkages within each cluster are strong and linkages between clusters are weak. Each cluster is a Topic.

## **How is Prominence calculated?**

Prominence is comprised of three metrics – recent citation counts, recent views counts, and journal impact (CiteScore). These three metrics are computed and then normalized using log transforms and standard deviations. The results are then combined as a weighted average.

## **What does Prominence mean?**

Since prominence is based mostly on citations and views of only the most recent papers in a Topic, it reflects current visibility and momentum of the Topic.

## **What is the correlation between Prominence and awarded grants?**

Prominence is strongly correlated with both past funding and future funding. Even more importantly, however, funding per author increases with Topic prominence.

# Topic Prominence in Science: FAQ 2/9

## Which entities can I analyze based on the Topics they are active in?

Institutions and researchers, can all be analyzed based on the Topics they are active in. Further entities will be added in future releases.

## How are Topics assigned to ASJC categories?

Since papers are assigned to ASJC categories independently of Topics, for each Topic we simply count the number of papers per ASJC category, and assign the Topic to its dominant ASJC category.

## How can I filter on Topics?

You can now filter Topics by ASJC subject classifications, with a threshold implemented. Currently the ASJC cut off is based on relative share of articles. The cut off threshold is 40%. It reduces the number of Topics attributed to an institution e.g. when an institution has relatively few papers attributed to a particular Topic.

E.g. If Topic A has ASJC articles as follows, only the highlighted ASJCs (2701, 2703, 1703, and 1503) would be retained.

ASJC	Number of Articles	Relative share percentage
2701	50	$50/50 * 100 = 100$
2703	40	$40 / 50 * 100 = 80$
1703	25	$25 / 50 * 100 = 50$
1503	20	$20 / 50 * 100 = 40$
1705	17	$17 / 50 * 100 = 34$
1301	5	$5 / 50 * 100 = 10$

# Topic Prominence in Science: FAQ 3/9

## **How are new Topics created?**

Although this is a complex process, we create new Topics by identifying emerging Topics (small, high growth, with highly cited papers) using citation linkages, and then splitting the emerging Topics from existing Topics. We use the VOS algorithm developed by CWTS.

## **How often are new Topics created?**

New Topics are split off on a yearly basis. We anticipate 30-50 new Topics each year.

## **When do Topics get updated?**

Topics are updated when we receive the data from Scopus, on a weekly basis.

## **How do new papers get assigned to Topics?**

Two approaches are used for this. Firstly, it is based upon the references within the paper. However if it is ambiguous, a 'similarity algorithm' is employed

## **How is an institution's article share calculated?**

That is calculated with the following formula [scholarly output of institution in topic/scholarly output of entire topic]

## **How are the names of Topics generated?**

The labelling of Topics is done using a combination of Elsevier Fingerprint Technology (EFT) and idiosyncratic phrases. The first two parts are generated by using EFT and provide a high level description of the Topic. The second two parts are idiosyncratic phrases, phrases relatively unique to the Topic, and give a more specific description of the Topic.

---

# Topic Prominence in Science: FAQ 4/9

## How many Topics is an average institution active in?

It depends on the institution size and the threshold. For example, an institution publishing around 5,000 papers per year will be active (at least 1 paper per year) in around 1,000 Topics.

## When will the recalculation of new Topics be done each year?

After initial set of Topics are created, during every weekly run new papers will be assigned to existing Topics.

After that either annually or semi-annually, we will be running algorithm to identify emerging Topics. The exact schedule has not been decided upon yet. New Topics are split out but we actually never recalculate the whole model. That is done once (and therefore stable).

## An institution was just added to SciVal, can I see which Topics they are active in?

It will be included with the processing of the next Scopus snapshot.

## Does Topics replace Competencies?

Yes, Topics has replaced Competencies. If you have Research Areas created with Competencies those will remain usable in SciVal but won't be updated with new publications.

## How are Key Contributors calculated?

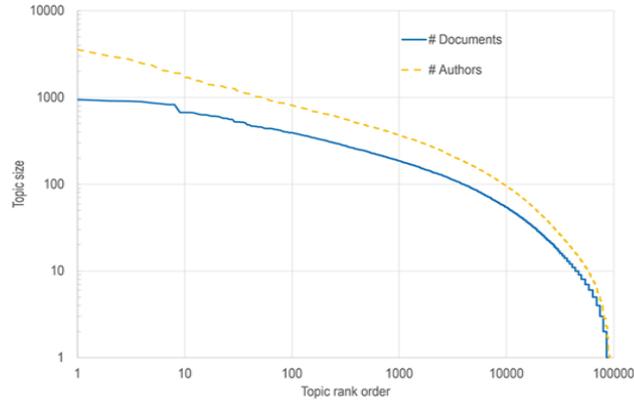
The Key Topics functionality allows you to see only the Topics where the entity is considered to be a key contributor. This allows you to filter out Topics to which the entity has a lower contribution and focus on the Topics where the entity has a higher potential influence. An entity will be a Key Contributor in a Topic if:

- They have at least 1/3 (33.3%) as many papers, as the top publishing entity in a topic AND / OR
- They have at least 1/3 (33.3%) as many citations, as the top cited entity in a topic

# Topic Prominence in Science: FAQ 5/9

## What is the distribution of Topic sizes?

This is an example of the Topic size distribution based on a single publication year.



## Does Prominence equate to Importance?

Due to the nature of certain research fields there are Topics which, will never become "Prominent", however this is not mutually exclusive with the Topic not being important. **Prominence is an indicator of momentum/movement or visibility** of a particular Topic. Comparisons of Prominence are best done with Topics in similar disciplines for fair and meaningful comparisons.

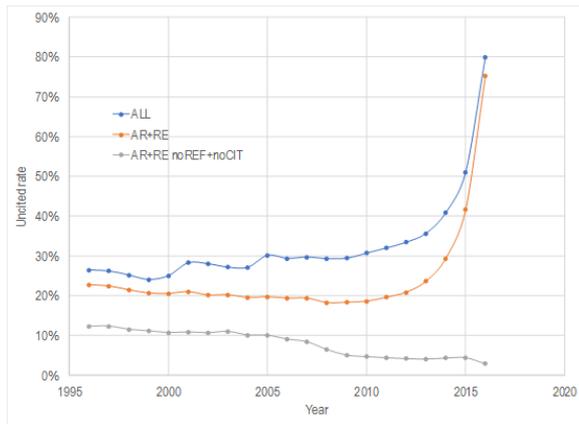
# Topic Prominence in Science: FAQ 6/9

## How many publications don't have any citations?

Uncited rates depend on age. Here is a chart based on Scopus data. AR+RE is articles and reviews. Uncited rates for these are historically around 20%. If all indexed documents are included the rate is around 30%.

noREF+noCIT refers to **uncited and no references**. In recent years this has been < 5%

NB. only the noREF+noCIT are excluded when clustering Topics.



# Topic Prominence in Science: FAQ 7/9

## What are Representative publications?

Representative publications are very strongly linked within the topic, and are intended to give us a feel for the central research question of a topic. They typically have many within-topic links and a high fraction of their links within the topic, and are also relatively highly cited for their age

## How are Representative Papers calculated?

$$\text{score} = \text{nsame} * (\text{nsame}/\text{nlinks}) * \ln((\text{nc9615}+1)/\text{greatest}(1,\text{year-pubyr}))$$

There are three pieces that are multiplied together to get the core articles score. The top core papers are those with the top scores.

1.  $\text{nsame}$  = the number of links (references + citations) to papers in the same topic – this favors review papers and highly cited papers that are strongly linked within the topic.
2.  $\text{nsame}/\text{nlinks}$  = fraction of links within the topic – this varies from 0 to 1, and favors papers where most/all of their links are within the topic
3.  $\ln(\text{nc9615}+1)/\text{greatest}(1,\text{yr-pubyr})$  = this is the log-transformed citation count divided by age – this favors highly cited papers, but we use the log transform so that citation counts don't overwhelm the other two pieces.

Although scores are calculated for all papers within a topic, we limit the ranking of core papers to those that are in the SciVal time window because the intent is to understand “what the topic is about right now”, not what it was 15-20 years ago (which is what the most highly cited papers would indicate).

# Topic Prominence in Science: FAQ 8/9

## **What are Related Topics?**

Related Topics are those that are most closely related to the current topic in terms of the text (titles and abstracts) used by the authors in the topic. Related topics may be from the same field or from other fields. There are many cases where researchers from different fields are doing similar work and use similar words, but do not cite each other. Related Topics uncovers these relationships.

## **How are Related Topics calculated?**

Related Topics are computed using deep learning embeddings based on the titles and abstracts of all Scopus documents. These are long vectors that express the semantics of a document. Note these vectors are dense in contrast to the standard TF/IDF token vectors that contain many zeroes. The embeddings for a topic are created by aggregating those of the documents in that Topic. Then for each Topic the top N closest vectors are computed.

# Topic Prominence in Science: FAQ 9/9

**I would like to know more about the underlying research behind Topic Prominence in Science, where should I start?**

You can read more about the underlying methodology and background research which went into the development of Topic Prominence in Science

## **Topics and prominence.**

Klavans, R. and K.W. Boyack, [Research portfolio analysis and topic prominence](#). Journal of Informetrics, 11 (4): pp 1158-1174 2017

## **Accuracy of competing methods**

Klavans, R. and K.W. Boyack, [Which type of citation analysis generates the most accurate taxonomy of scientific and technical knowledge?](#) Journal of the Association for Information Science and Technology, 2017. 68(4): p. 984-998.

## **Emerging topics.**

Small, H., K.W. Boyack, and R. Klavans, [Identifying emerging topics in science and technology](#). Research Policy, 2014. 43: p. 1450-1467.

## **How Topics are created**

Ludo Waltman and Nees Jan van Eck [A New Methodology for Constructing a Publication-Level Classification System of Science](#). Journal of the American Society for Information Science and Technology 63(12): 2378-2392, 2012