



Universiteit
Leiden
The Netherlands

VOSviewer: putting research into context

Jong, R.M. de; Bus, D.

Citation

Jong, R. M. de, & Bus, D. (2023). VOSviewer: putting research into context. *Research Software Community Leiden*. doi:10.21428/a1847950.acdc99d6

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](https://creativecommons.org/licenses/by/4.0/)

Downloaded from: <https://hdl.handle.net/1887/3589566>

Note: To cite this publication please use the final published version (if applicable).

Research Software Community Leiden

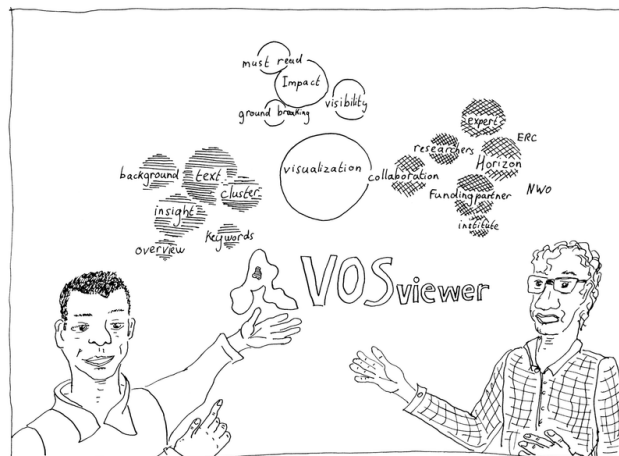
VOSviewer: putting research into context

Rutger De Jong Dennis Bus

Published on: Mar 09, 2023

URL: <https://researchsoftware.pubpub.org/pub/j3sr4bo9>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](#)



VOSviewer is an invention of Leiden scholars
Ludo Waltman and Nees-Jan van Eck.

Item	Description
Name	VOSviewer
Cost	Free download
Likely uses	Discovery; Storytelling
Benefits	Creating overviews of research topics; insight to the working of other tools
What intelligence	Clustering algorithm; Natural Language Processing; Layout algorithm
URL	https://www.vosviewer.com
Accompanying tutorials	https://gorlaeus-library.github.io/VOSviewer

We decided to kick off our series on searching with artificial intelligence with an old faithful companion: [VOSviewer](https://www.vosviewer.com). The application was actually created to make visualizations of bibliometric networks by Leiden's Centre for Science and Technology ([CWTS](https://www.cwts.nl)). This may seem like a strange choice: VOSviewer does not do any searching per se, you have to extract your search results and import them yourself. However, some techniques used in VOSviewer are also implemented in other solutions, and the information the tool provides can be very helpful to get a better understanding of your field of research, especially if you are starting out on a topic that is new for you as a researcher.

This blog post will be the most detailed of the series, as many basic information is explained here. It is also supplemented by a large tutorial exploring all different aspects of VOSviewer: <https://gorlaeus-library.github.io/VOSviewer>

Visit the web version of this article to view interactive content.

Visualization of the most common phrases in news articles about Leiden University in a number of Dutch newspapers.

How it works

At the basis of the VOSviewer is the visualization of similarities ([N. J. V. Eck and Waltman 2007](#)) ([N. J. van Eck and Waltman 2010](#)). The objects – articles, journals, organizations, authors or terms – are located relative to each other in such a way that the distance between two objects is an approximate measure of their similarity within the set.

For example if we would look at a data set of articles about food, a topic such as food processing will form a cluster of its own, while a topic such as nutritional value might form another cluster at a distance from the processing. However, there will also be some instances where the food processing method has a direct influence on the nutritional value: these articles will be found in between the two clusters or at the edge of either cluster, depending on the main topic of the article.

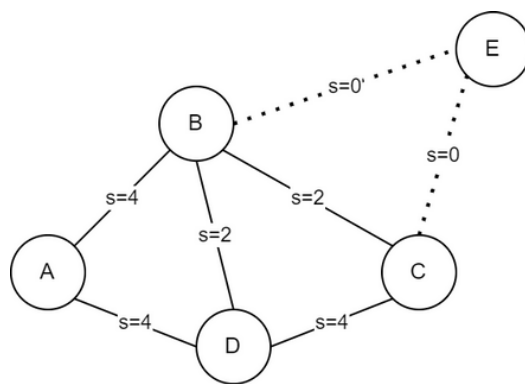
So how exactly does VOS know if something is similar? Within the program we can choose from several options to base the similarity upon, depending on the data set and type of analysis we want to do:

- Co-authorship, to build collaboration networks and find partners for research projects
- Co-occurrence (based on the documents terms occurring together), to discover the best search keywords and define popularity, age and impact of topics
- Citation analysis (direct citation), to discover seminal papers that everyone refers to
- Bibliographic coupling (based on the number of shared references), to find articles with a common knowledge base
- Co-citations (the number of times articles are cited together), to find complementary articles

The information VOSviewer uses is all contained within the set of articles you start from. So let's say we want to find the most popular articles within a topic. We will start by downloading all of these articles including their citation information. We see documents as similar if they refer to or are cited by the same other documents (similarity because they link each other, the direction is not important). Thus we can calculate for all pairs of documents in our set how similar document i is to document j . VOS uses the following formula for this:

$$similarity_{ij} = \frac{2n_{documents} citations_{shared_{ij}}}{citations_{total_i} citations_{total_j}} = \frac{relation_between_items}{normalisation_factor}$$

So how do we go from a high dimensional similarity- matrix - score to a visual representation? The really difficult task of VOS is to map all of these documents on a two dimensional chart, so that the distance between the documents is related to the inverse of their similarity (they are displayed closer to each other if they are more similar), while at the same time the documents should not overlap in the visualization. As the sets get bigger you may imagine how the complexity builds and the exact position becomes more of an estimation, but still provides us with a clear picture of overlap within linked articles. The distance between fully unlinked articles or clusters of articles is less clear: in fact they should be as far from the other documents they are not (indirectly) linked to as possible.



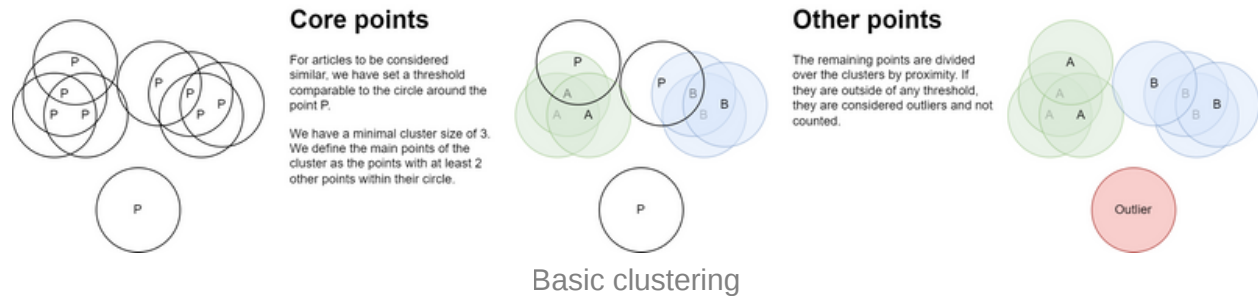
Distribution

The more links they share, the higher the similarity. The higher the similarity, the closer to each other they are in space.

TIP: Layout is created by optimizing the possibilities. The computer algorithm in use tries all combinations until it finds a semi-optimal presentation. As the algorithm starts at a random position, the results may differ if you run the layout multiple times. To prevent this, set a specific number (not 0) for random seed. This option can be found in the analysis tab, under the ‘Advanced parameters’ of the Layout section.

After the mapping to x,y-coordinates on a map, VOSviewer goes one step further and also indicates to us some clusters of highly related terms visualized with colours. These clusters make the maps so easy to interpret (and sell in presentations to funders).

A common way in AI to find clusters is by density-based clustering. In other words, where there are many points, there is most likely a clustering of related points. There is no set number of clusters in this technique, but you have to set thresholds (which you can adjust in the program) for the document similarity so the algorithm knows when to neglect a document for a specific cluster. It works as follows:



If the clustering seems off, you can play with the thresholds to provide a better view. Overall clustering works best if the set is not completely uniform. In case you see just one big circle, it could well be that your set consists of very uniform topics or highly unrelated topics.

In VOSviewer you will notice that some items are visualized in between items of other colors. The principles mentioned above do still apply, but the real clustering uses the similarity network in the background. While the visualization is 2-dimensional, this similarity network has many dimensions.

Who can benefit from VOSviewer?

We see some clear use cases for the VOSviewer software in several roles within the university. Click the links to see an example of how one can use the technique described. The techniques can be found in our VOSviewer tutorial on Github.io.

Principal Investigator

Most likely you are already an expert in the topic at hand. However, sometimes you need to visualize your knowledge in order to get funding for example, where the funder wants to see collaboration partners or the impact of your output within the field. Or you have a research question where you need to know what has not yet been researched in depth, what topics are hot or what search terms (also known as article search engine optimization) make your work as visible as possible.

- [Collaboration maps](#) (author similarity)
- [Article overview of the field](#) (based on citations)
- [Topic overview](#) (based on search terms)

PhD

As a PhD you usually start out fresh on a new line of research when you enter university. You want to catch up quickly to the knowledge base of the professors and need field overviews. You also want to know what specific topics and keywords are of interest at the moment so you can focus your search and research interests.

- [Article overview of the field](#) (based on citations)
- [Topic overview](#) (based on search terms)

- [Journal maps](#) (where should I publish)

Policy makers

You need to be able to support research policy. Possibly you want to tell stories about how your researchers collaborate (internally and externally), how your organization is mentioned in news outlets and whether the publication output matches the goals set for the research vision.

- [Collaboration maps](#) (author similarity)
- [Topic overview](#) (based on search terms)
- [Overlays](#) (for additional information on all analyses)

If you are interested in all possibilities of the VOSviewer software, check out the other chapters for examples and information on how to use a diversity of databases to get the best results. Another good starting point to discover all of VOSviewer's possibilities is the [AIDA booklet](#).

Limitations

Based upon the examples, the technical basis and our experience as librarians, we have identified the following limitations you have to take into account while working with VOSviewer:

- Large dataset needed for most evaluations (at least 100-200 items) otherwise the results are not statistically relevant
- The parameters and dataset you provide determine what the visualization will look like. It is important to know that the software can be used mainly for qualitative indications and not so much for quantitative measurement. It is also necessary to be very specific while working with VOSviewer:
 - How did I obtain my data (search query, database, date, number of results)
 - What parameters did I use to create my maps (counting method, excluded keywords, number of occurrences/citations, etc.)
 - Did I change visualization parameters?
 - Write down what you see in the different clusters and what this tells you
 - Why did I make these choices?
- It may be difficult to assess if you are not an expert in the field yet. Always ask an expert (supervisor for a Ph.D. for example) to have a look at the map and help with the analysis. In my experience they immediately see patterns and are usually quite surprised by the results.
- The software is not automatically updated; check the VOSviewer website regularly as new functionality is rolled out frequently.
- Non-standard databases/materials without DOI need additional work to be used.
- Author analysis: the authors are defined by name, not by Orcid. This means it is unusable in fields dominated by Chinese authors.

- Term extraction is based on English language/grammar. The algorithm used for natural language processing is not suitable for other languages. In the examples we will explain how to convert text to English.

Be aware that using VOSviewer is an iterative process. Most likely your first tries will give you some information to improve your search or visualization.

Disclaimer

In this series we look at new technology to help with literature research. We are not fulltime researchers or programmers and are open to suggestions to improve the contents of these blogs. If you want to share your story/research, or want to help out by providing us your research topic as example, that would be highly appreciated. Please contact [Rutger de Jong](#) or [Dennis Bus](#) for more information.