

# STI 2018 Leiden

*23rd International Conference on Science and Technology Indicators  
"Science, Technology and Innovation Indicators in Transition"*

## **STI 2018 Conference Proceedings**

*Proceedings of the 23rd International Conference on Science and Technology Indicators*

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

### **Chair of the Conference**

Paul Wouters

### **Scientific Editors**

Rodrigo Costas  
Thomas Franssen  
Alfredo Yegros-Yegros

### **Layout**

Andrea Reyes Elizondo  
Suze van der Luijt-Jansen

The articles of this collection can be accessed at <https://hdl.handle.net/1887/64521>

ISBN: 978-90-9031204-0

© of the text: the authors

© 2018 Centre for Science and Technology Studies (CWTS), Leiden University, The Netherlands



This ARTICLE is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License

## Science maps and strategic thinking

Flemming Skov<sup>\*</sup>, Tobias Wang<sup>\*\*</sup> and Jens Peter Andersen<sup>\*\*\*</sup>

*\*fs@bios.au.dk*

Department of Bioscience, Aarhus University, Grenaaavej 14, DK-8410 Rønne, (Denmark)

*\*\*tobias.wang@bios.au.dk*

Department of Bioscience, Aarhus University, C.F. Møllers Allé 3, DK-8000 Aarhus C, (Denmark)

*\*\*\*jpa@ps.au.dk*

The Danish Centre for Studies in Research and Research Policy (Department of Political Science), Aarhus University, Bartholins Allé 7, DK-8000 Aarhus, (Denmark)

### Abstract

In this paper, we discuss how to use topical, scientific reference maps to overview and navigate in dynamic research landscapes and how to use such maps as input to strategic thinking. An automated procedure is developed that creates a reference map from a set of selected scholarly papers covering a specified subject area. The method is based on co-occurrences of keywords and distributes keywords on a reference grid where related keywords appear close to another and vice versa. Once the reference map is created other entities such as individual scientific papers, researchers, and research groups may be projected to the map based on the keywords they use. An article may be represented as a point calculated as the mean x and y of all the keywords it contains. Similarly, individual researchers are plotted as the mean of all articles they have written; a research group as the mean of all researchers it includes and so forth. The centre of a distribution is supplemented with point cloud- or density maps in order to map the actual form and size of the distribution in the reference map. The maps can also show how a scientific field has evolved over the years, and potentially the current direction of research. A research institute, a research group or an individual researcher can map their abilities and competencies to the reference map and compare that to potential collaborators and competitors to visualise strengths and weaknesses. These maps may be used as an input or inspiration to strategic discussions where strategic positions or moves may be discussed, evaluated, and presented as strategic narratives: (1) stay and fortify; (2) invade and conquer; (3) merge and fuse subjects; (4) re-apply methodology from other areas; (5) explore and discover new land outside the reference map; and (6) invent new research areas.

### Introduction

The production of knowledge in the form of scholarly papers is growing at an exponential rate and on average doubling every 24 years (Bornmann & Mutz, 2015). Scientific research is dynamic and aims, goals, directions and methods are constantly changing. New areas appear, and existing established scientific domains may merge or split. The rapidly increasing scientific production and the constant reshaping of the scientific landscape makes it difficult to maintain an overview of a given knowledge domain. Researchers, research managers, funding agencies and knowledge users in general need tools to follow shifts in research focus. Most research

institutions use hierarchical diagrams or simply text to describe their research in strategic documents, in brochures, or on their web sites. This approach, however, is conservative and static and often represents what used to be, rather than the actual research done.

We see science mapping as one way to provide the necessary overview. The purpose of science maps is to visualise, understand and communicate the structure and content of scientific knowledge domains (e.g. Börner et al., 2012; Boyack & Klavans, 2014; Klavans & Boyack, 2009, 2011; Rosvall & Bergstrom, 2007). Science mapping is inspired by the traditional cartographic map that has been used for centuries to project selected parts of the globe on a two-dimensional surface. Science maps are (like all maps) abstract in the sense that they do not represent a one to one relationship between reality and the map, but are scaled up versions of the reality using symbols and various projections. In this paper, we describe a general method for the development of knowledge domain visualization based on the use of keywords in scientific papers and present a framework for the visual analyses of these maps and a discussion of strategic options based hereon. The aim of our paper is not to present novel mapping techniques, we build on well-established methods (see e.g. Hinze, 1994; Rip & Courtial, 1984), and some previous visualisations are related to what we present here (Su & Lee, 2010). The purpose rather is to illustrate a dynamic set of visualizations, and their combination with strategic approaches, providing novel insight into the application of bibliometric co-word maps for research planning and leadership.

## Methods

The general method consists of three steps: (1) defining the boundaries of the research domain; (2) creating a reference map based on co-occurrence of keywords; and (3) creating profiles for researchers, research groups or research topics by overlay and visualisation of research dynamics

### *Defining the boundaries of the research domain*

Reference maps are constructed from basic bibliographic information (keywords, title, year, Web of Science ID, etc). Within any given research domain, the number of published papers easily amounts to tens of thousands each year. The method described in this paper works best with a medium number of articles (10-25,000). The number of papers is not a computational problem as such, but we want the reference map to cover the research area at a reasonable scale, e.g., where it is neither too detailed, nor too broad. In the first case, patterns would be lost in detail and complexity whereas the latter would be much too general to be useful.

### *Building the reference map*

The procedure we present here can be thought of as a generic approach, in which some elements can be swapped for other approaches. The specific approach here, however; has been automated in Python and includes the following major steps: (1) raw txt-files (export from Web of Science or similar sites) are imported and keywords cleaned and standardised using the Porter Stemmer (Porter, 1980); (2) keywords (authors own plus keywords plus) that appear in the same paper are linked and a node- and a link file are created; (3) these files are imported to Gephi<sup>1</sup> and a graph network constructed using the force-atlas mapping algorithm (Jacomy, Venturini, Heymann, & Bastian, 2014), which distributes keywords in a two-dimensional reference landscape; (4) finally a file with each keyword and it's x- and y-coordinate in the reference landscape is stored for further use. Existing software for visualising these types of maps, e.g.

---

<sup>1</sup> <https://gephi.org>

VosViewer (Van Eck & Waltman, 2018), can produce overlay maps somewhat similar to those provided here, however, the Python setup allows systematic production of multiple specialized maps visualizing different dimensions, based on the same reference map.

### *Profiling and dynamics*

The reference file can be used to map the position of keywords directly, but any set of scholarly articles that broadly belong to the science domain mapped may be projected to the reference map. The centre position of an article, e.g., is calculated as the mean coordinates of all keyword it contains and will thus be found in the centre of the ‘cloud’ of keywords. Similarly, the centre location of a researcher is defined in the centre of his or her articles. The position of a group of researchers is calculated in a similar way. The method may also be used to map the distribution of a given subject area. The centre position is a very simple representation and more depth and information is gained when an author is shown by the cloud that represents all articles written by her or him.

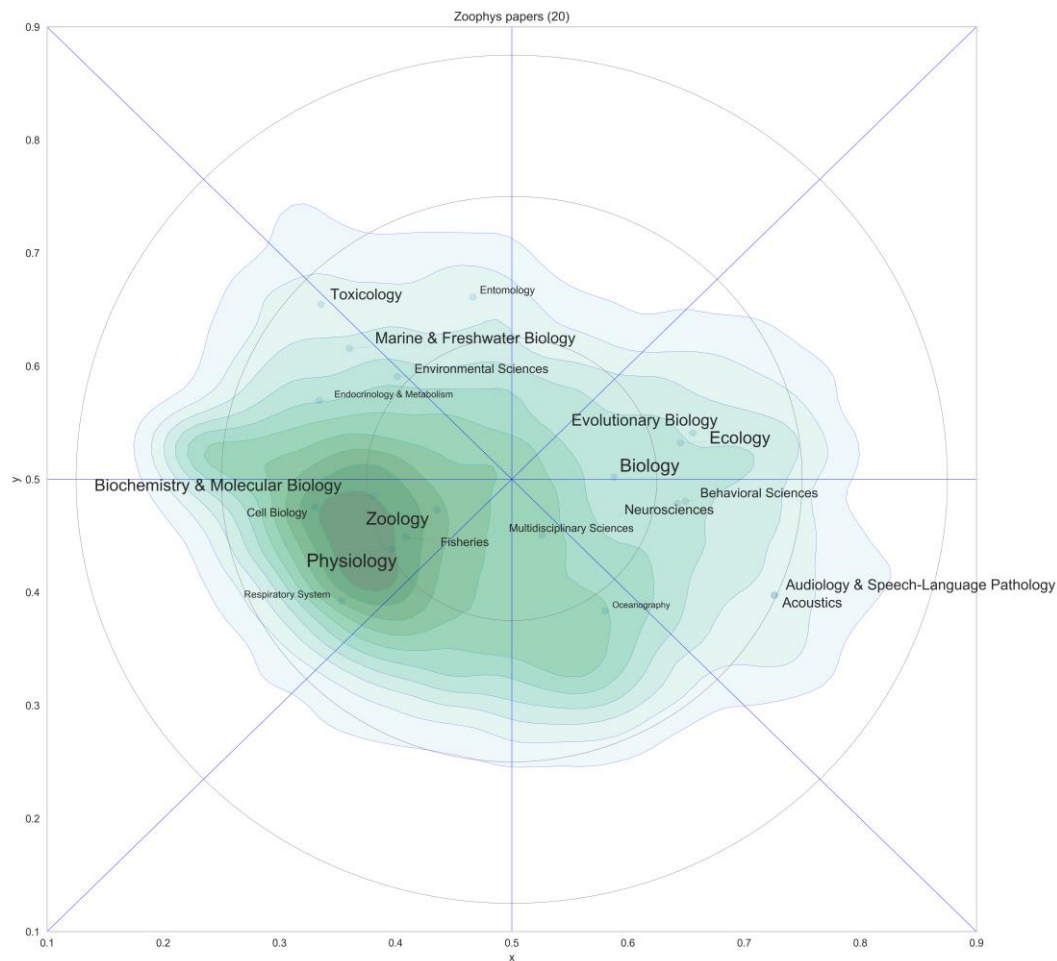
### **An example - Zoophysiology**

In the following we present an example map for the research area ‘zoophysiology’, to demonstrate how the reference map can be used to provide an overview of the domain and visualise how it has evolved over time (here since 1995). A few basic guidelines may be useful to mention here: in general, the centre of such maps contains commonly used keywords shared by many papers while rarely used keywords are closer to the periphery. Distance and direction are meaningful in this type of science map and keywords (or other objects) that appear close to each other in the map may be considered more related than entities far apart (The so called first principle of geography: “everything is related to everything else, but near things are more related than distant things” (Tobler, 1970)).

### *Reference landscape with place names*

The figure below shows the basic reference map for our example. The map is an abstract representation of a knowledge domain of zoophysiology and viewers should be able to recognize zones and regions in the map before they can navigate it and use it for any practical purpose. Fig. 1 shows one way of doing this. Note the linear guides that can be used for convenient references (e.g., Northwest or Southeast). The circles represent distance to the centre of the map and makes it easier to compare maps. Colours represent research activity in the map (darker colours correspond to many articles). The centre position of the Web of Science subject categories is used to provide a first impression on how different research subjects are distributed in the landscape (the centre is calculated as the mean x- and y-position for all articles belonging to a certain subject category).

Figure 1: A reference maps for zoophysiology with major zones and areas named. In this case based on where in the landscape we find the centre of the Web of Science subject categories.



### *Profiling researchers, research groups and subjects in the landscape*

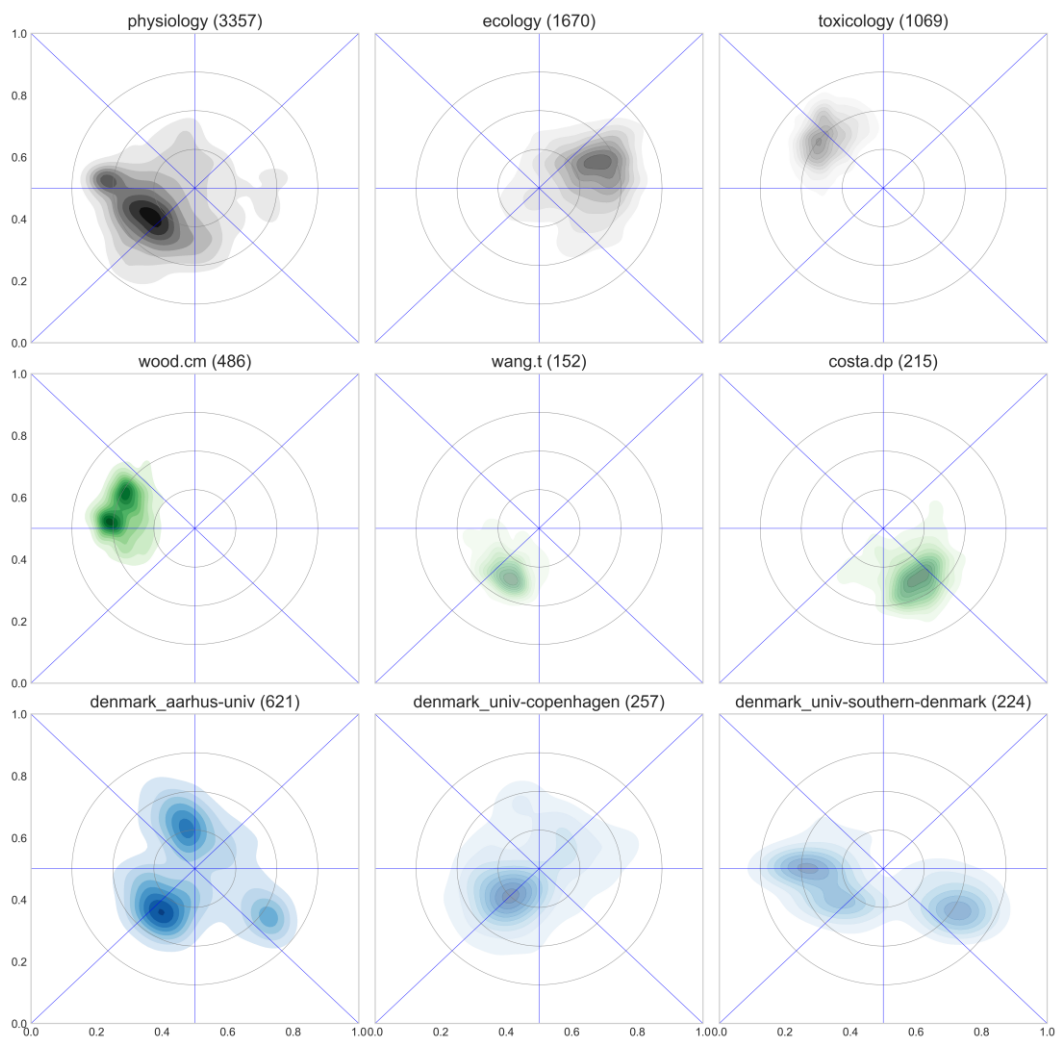
In Fig. 1 the centre of each subject category is used to position the text label to show the general characteristics of the landscapes. This is sufficient in some cases, but the centre point is a simplification that does not reveal the actual distribution of the entity mapped. The label size in the map indicates the number of articles belonging to each subject category, but it is not possible to see the shape or the extent of the actual distribution. Figure 2 demonstrates how a reference map can be used to create profiles for different entities. The upper row of profiles shows the actual distribution of articles in the reference map space for three of the subject categories. Each profile is characterized by the general position, the extent and shape of the cloud. The colour shade represent the density of the subject categories to show where there is most activity. The three profiles are very different and easily recognizable. The two subject categories 'physiology' and 'ecology' are both broad, but have centres of distribution in the lower left and upper right, respectively, and are thus very different (but still with some overlap in the centre). Subject category 'toxicology' has a distinct distribution, and fills a much narrower space in the landscape.

The middle row in Fig. 2 shows examples of individual researcher profiles. These profiles may be characterised as discussed above and illustrate how the width and breadth of the scientific field covered differs between researchers. Some researchers concentrate their research on a relatively narrow domain whereas others cover larger parts of the reference maps. Perhaps as a

result of collaboration or because their research interests have shifted over time. In cases where there are sufficient data it is possible to map the actual trajectory of a researcher over time. This type of information could be used e.g. when recruiting new staff, and provide supplementary information on the type of researchers wanted.

The bottom row in Fig. 2 shows profiles for research groups. In this example, we have used the zoophysiology research groups in three Danish universities as an example. The profiles show how much of the domain each university covers, where there is overlap and where there are unique competencies.

Figure 2: Research profiles for subject categories (upper row), individual researchers (middle row), and research groups/institutions (lower row)

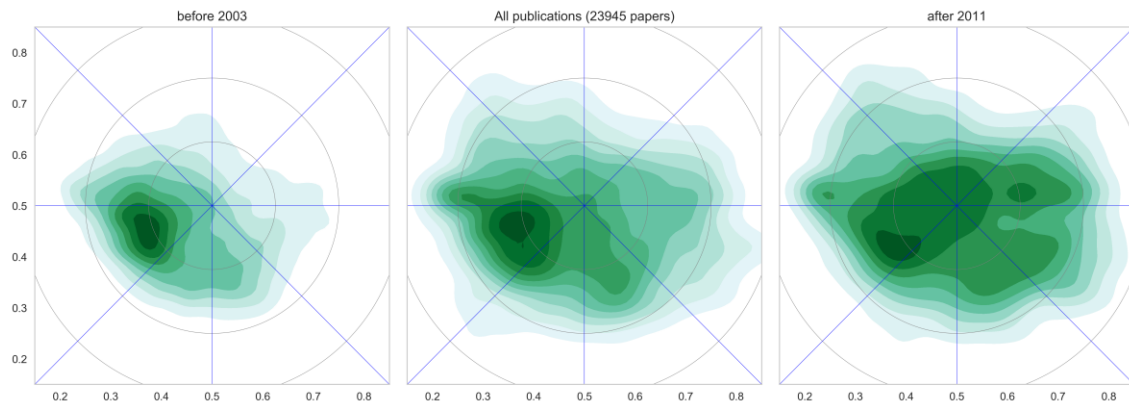


### *Dynamics - change over time*

The examples above show static maps that cover a specified time span (in this case a summary of the past approx. 20 years). It is also interesting, however, to visualise the dynamics of the landscape over time to detect if there are certain patterns or trends in their development. Fig. 3 below shows density maps for the entire research domain. The three maps show the approximate same centre of gravity in the Southwestern area, but there is, at the same time, a clear trend with an increasing proportion of articles being published towards the Northeast. This area borders the traditional disciplines of ecology and is interpreted as a result of an increasing

focus on how to combine traditional zoophysiology with ecology and conservation science. The particular branch of zoophysiology has been branded as ecophysiology or environmental physiology that and focusses on the adaptation of an organism's physiology to environmental conditions.

Figure 3: Dynamics in the research domain of zoophysiology. Each map shows the density of publications in three time slices: before 2003; 2003-2011; and after 2011. The splits are chosen to have an equal distribution of papers in each category.



#### *Which way to the front – and where is the base?*

A main purpose of research and science is to explore and scientists are drawn towards discoveries and new results, insights and techniques. Novel results and knowledge are easier to publish and such areas are often rewarded when new funding is distributed. It is therefore important to be able to recognize the research fronts as well as it is important to find the research centre, where the majority of the research is done and where it is expected to find research stability, tradition and solidity. Mapping novelty of keywords using the above mapping techniques has proven inconclusive as an approach to identifying research fronts, as novelty might indicate a front as much as it can be caused by peripheral inclusions in the field. Using the time-dimension of Figure 3, it is possible to perceive the movement as one type of front. Ideally, this can be combined with more traditional indices of research fronts (Price, 1965), however, this is a topic to unfold in further research.

#### **Science maps for strategic thinking**

Science mapping provides alternative views on the status, content and dynamics of scientific research and can be used to supplement bibliometric and other types of analyses that are used traditionally as input to strategic discussions. For the context of this paper we view 'strategic' in the widest possible sense from decisions made by individual researchers on what to work on and who to work with to strategic research management on higher administrative levels.

Any strategic decision requires some basic information regarding the landscape or environment where the action occurs. It is essential to know where you are in the landscape and where you want to go. In order to plan how to reach the goal, it is also important to know where potential opponents and collaborators are. It is also crucial to know about their strengths and weaknesses – and your own – if the strategy is going to be successful. Science mapping may help visualise

relationships, visualise position and directions, and as such provide a two-dimensional board that may provide the overview that may help facilitate strategic thinking.

### *Strategic thinking*

There is no generally accepted definition of strategic thinking, but the concept is described in Wikipedia as: “Strategic thinking includes finding and developing a strategic foresight capacity for an organization, by exploring all possible organizational futures, and challenging conventional thinking to foster decision making ...” (Wikipedia, 2018). Compared to the formal and goal oriented process of strategic planning, it can be seen as a more free and open process, where the exploration of future possibilities and brainstorming are in focus and can be used as a preparation for actual strategic planning.

The use of maps in this process have a long history and it is easy to visualise a war room, where tacticians are bent over a large map, where small figures representing enemies and allies are moved around to provide an overview of the situation and to test and develop plans for the next step. A map almost automatically forces you to think in patterns instead of linear relationships and perhaps stimulate innovation and help develop new, alternative solutions. Furthermore, the ideas may be formulated as strategic narratives which are easier to understand and present to others. A graphic, visual presentation is very strong and might help present complex scenarios and carry new ideas forward.

The process is in principle completely open, and that is preferred in some situations, but it could be guided by providing participants some typical options or narratives to help start exploring the strategic options:

- a. **Stay (and fortify)** (a research institute may decide that its current positions is optimal and desirable – this may be supplemented with a discussion on how to maintain current strength)
- b. **Invade** new areas (the analysis shows that a research group is not present in a new and promising part of the map – should they try to include it, e.g., by letting a member of the group specialise in it, hiring an established researcher or setting up a collaboration with another institute?)
- c. **Merge and fuse** subjects (the research area ecophysiology, mentioned earlier, is an example of how two separate research fields are combined to create something new. In the case of zoophysiology a quite successful merger that has attracted many researchers).
- d. **Apply methods** from other areas (somewhat similar to (c), but here successful methodology are borrowed from other disciplines such as statistical tools, models or new sensors originally designed and applied for something else )
- e. **Explore** what’s beyond the border (a reference map is based on a limited number of articles and is on purpose kept relatively narrow – perhaps there are new research disciplines to invade, borrow from or merge with )
- f. **Invent.** This is pure innovation and revolution: invent new territory that does not yet exist.

These narratives may be used separately or in any combination. It is worth noticing that the last four, if successful, eventually will contribute to an alteration of the reference map by merging, that will pull parts of the landscape together or by adding new areas to the map. Such dynamics are an innate property of science, but it is our hope that science maps help understand and navigate in the ever-changing world of science.



## References and Citations

- Börner, K., Klavans, R., Patek, M., Zoss, A. M., Biberstine, J. R., Light, R. P., ... Boyack, K. W. (2012). Design and update of a classification system: The ucsc map of science. *PLoS ONE*, 7(7), 1–10. <http://doi.org/10.1371/journal.pone.0039464>
- Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, 66(11), 2215–2222. <http://doi.org/10.1002/asi.23329>
- Boyack, K. W., & Klavans, R. (2014). Creation of a highly detailed, dynamic, global model and map of science. *Journal of the Association for Information Science and Technology*, 65(4), 670–685. <http://doi.org/10.1002/asi.22990>
- Hinze, S. (1994). Bibliographical cartography of an emerging interdisciplinary discipline: The case of bioelectronics. *Scientometrics*, 29(3), 353–376. <http://doi.org/10.1007/BF02033445>
- Jacomy, M., Venturini, T., Heymann, S., & Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE*, 9(6), 1–12. <http://doi.org/10.1371/journal.pone.0098679>
- Klavans, R., & Boyack, K. W. (2009). Toward a consensus map of science. *Journal of the American Society for Information Science and Technology*, 60(3), 455–476. <http://doi.org/10.1002/asi.20991>
- Klavans, R., & Boyack, K. W. (2011). Using global mapping to create more accurate document-level maps of research fields. *Journal of the American Society for Information Science and Technology*, 62(1), 1–18. <http://doi.org/10.1002/asi.21444>
- Porter, M. F. (1980). An algorithm for suffix stripping. *Program*, 14(3), 130–137.
- Price, D. J. D. S. (1965). Networks of Scientific Papers. *Science*, 149(3683), 510–515.
- Rip, A., & Courtial, J. P. (1984). Co-word maps of biotechnology: An example of cognitive scientometrics. *Scientometrics*, 6(6), 381–400. <http://doi.org/10.1007/BF02025827>
- Rosvall, M., & Bergstrom, C. T. (2007). Maps of random walks on complex networks reveal community structure. *PNAS*, 105(4). <http://doi.org/10.1073/pnas.0706851105>
- Su, H. N., & Lee, P. C. (2010). Mapping knowledge structure by keyword co-occurrence: A first look at journal papers in Technology Foresight. *Scientometrics*, 85(1), 65–79. <http://doi.org/10.1007/s11192-010-0259-8>
- Tobler, W. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(Supplement), 234–240.
- Van Eck, N. J., & Waltman, L. (2018). VosViewer.
- Wikipedia. (2018). Strategic thinking. Retrieved March 24, 2018, from [https://en.wikipedia.org/wiki/Strategic\\_thinking](https://en.wikipedia.org/wiki/Strategic_thinking)