

# Science Mapping Software Tools: Review, Analysis, and Cooperative Study Among Tools

M.J. Cobo, A.G. López-Herrera, E. Herrera-Viedma, and F. Herrera

Department of Computer Science and Artificial Intelligence, CITIC-UGR (Research Center on Information and Communications Technology), University of Granada, E-18071 Granada, Spain. E-mail: {mjcobero, lopez-herrera, viedma, herrera}@decsai.ugr.es

**Science mapping aims to build bibliometric maps that describe how specific disciplines, scientific domains, or research fields are conceptually, intellectually, and socially structured. Different techniques and software tools have been proposed to carry out science mapping analysis. The aim of this article is to review, analyze, and compare some of these software tools, taking into account aspects such as the bibliometric techniques available and the different kinds of analysis.**

## Introduction

Science mapping, or bibliometric mapping, is an important research topic in the field of bibliometrics (Morris & Van Der Veer Martens, 2008; van Eck & Waltman, 2010). It attempts to find representations of intellectual connections within the dynamically changing system of scientific knowledge (Small, 1997). In other words, science mapping aims at displaying the structural and dynamic aspects of scientific research (Börner, Chen, & Boyack, 2003; Morris & Van Der Veer Martens; Noyons, Moed, & Luwel, 1999a).

The general workflow in a science mapping analysis has different steps: data retrieval, preprocessing, network extraction, normalization, mapping, analysis and visualization. At the end of this process, the analyst has to interpret and obtain some conclusions from the results.

There are different bibliometric sources where the data can be retrieved, such as the ISI Web of Science (WoS) or Scopus. Moreover, a science mapping analysis can be performed using patent or funding data.

The preprocessing step is maybe one of the most important ones. The goodness of the result will depend on the quality of the data. Several preprocessing methods can be applied, for example, to detect duplicate and misspelled elements.

Different approaches have been developed to extract networks using the selected units of analysis (authors, documents, journals, and terms). *Co-word* analysis (Callon, Courtial, Turner, & Bauin, 1983) uses the most important words or keywords of the documents to study the conceptual structure of a research field. *Co-author* analyzes the authors and their affiliations to study the social structure and collaboration networks (Gänzel, 2001; Peters & van Raan, 1991). Finally, the cited references are used to analyze the intellectual base used by the research field or to analyze the documents that cite the same references. In this sense, *bibliographic coupling* (Kessler, 1963) analyzes the citing documents, whereas *co-citation* analysis (Small, 1973) studies the cited documents. Other approaches such as author bibliographic coupling (Zhao & Strotmann, 2008), author co-citation (White & Griffith, 1981), journal bibliographic coupling (Gao & Guan, 2009; Small & Koenig, 1977), and journal co-citation (McCain, 1991) are examples of macro analysis using aggregated data.

Once the network has been built, a normalization process is commonly performed over the relation (edges) between its nodes by using similarity measures. A review of similarity measures used in science mapping was carried out in (van Eck & Waltman, 2009).

With the normalized data different techniques can be used to build the map (mapping process; Börner et al., 2003). Dimensionality reduction techniques such as principal component analysis or multidimensional scaling (MDS), clustering algorithms and Pathfinder networks (PFNETs) are widely used.

Analysis methods for science mapping allow us to extract useful knowledge from data. *Network analysis* (Carrington, Scott, & Wasserman, 2005; Cook & Holder, 2006; Skillicorn, 2007; Wasserman & Faust, 1994) allows us to perform a statistical analysis over the generated maps to show different measures of the whole network or measures of the relationship or overlapping (the Jaccard's Index can be used to do that) of the different detected clusters (if a clustering algorithm has been applied). *Temporal analysis* (Garfield, 1994;

---

Received October 26, 2010; revised February 10, 2010; accepted February 10, 2010

© 2011 ASIS&T • Published online 2 May 2011 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/asi.21525

Price & Gursay, 1975) aims to show the conceptual, intellectual, or social evolution of the research field, discovering patterns, trends, seasonality, and outliers. *Burst detection* (Kleinberg, 2003), a particular temporal analysis, aims to find features that have high intensity over finite durations of time periods. Finally, *geospatial analysis* (Batty, 2003; Leydesdorff & Persson, 2010; Small & Garfield, 1985) aims to discover where something happens and what its impact on neighbouring areas is.

Additionally, visualization techniques are used to represent a science map and the result of the different analyses, for example, the networks can be represented using *heliocentric maps* (Moya-Anegón et al., 2005), *geometrical models* (Skupin, 2009), *thematic networks* (Bailón-Moreno, Jurado-Alameda, & Ruíz-Baños, 2006; Cobo, López-Herrera, Herrera-Viedma, & Herrera, 2011), or maps where the proximity between items represents their similarity (Davidson, Wylie, & Boyack, 1998; Polanco, François, & Lamirel, 2001; van Eck & Waltman, 2010). To show the evolution in different time periods, *Cluster string* (Small, 2006; Small & Upham, 2009; Upham & Small, 2010), and *thematic areas* (Cobo et al., 2011) can be used.

Although the science mapping analysis can be performed using generic social network analysis tools such as Pajek (Batagelj & Mrvar, 1998) and UCINET (Borgatti, Everett, & Freeman, 2002), or bioinformatics software such as Cytoscape (Shannon et al., 2003), there are other software tools specifically developed for this purpose. Some of these software tools are specifically conceived for scientific science mapping and others may be used in nonscientific domains. Some of these software tools have been implemented only for visualizing science maps and others allow us to visualize and also build the maps. A list of generic software tools used in scientometrics research is shown in Börner et al. (2010).

The aim of this article is to present a deep comparative study of nine representative science mapping software tools by showing their advantages, drawbacks and most important differences. We analyze the following software tools: Bibexcel (Persson, Danell, & Wiborg Schneider, 2009), CiteSpace II (Chen, 2004, 2006), CoPalRed (Bailón-Moreno, Jurado-Alameda, Ruíz-Baños, & Courtial, 2005; Bailón-Moreno et al., 2006), IN-SPIRE (Wise, 1999), Leydesdorff's Software, Network Workbench Tool (Börner et al., 2010; Herr, Huang, Penumarthy, & Börner, 2007), Science of Science (Sci<sup>2</sup>) Tool (Sci<sup>2</sup> Team, 2009), VantagePoint (Porter & Cunningham, 2004), and VOSViewer (van Eck & Waltman, 2010). Each one provides us with its own view of the data due to the fact that they implement different analysis techniques and algorithms. We should point out that they present complementary characteristics, and, therefore, it could be desirable to take their synergies to perform a complete science mapping analysis. We complete our analysis by showing the performance of all software tools with an example, and analyze some positive synergies among them.

This article is organized as follows. In the Science Mapping section, some concepts on science mapping are presented. The Software Tools Designed to Perform a Science

Mapping Analysis: A Survey section describes the software tools to be analyzed. In the Comparative Study section, a comparison is made among the described software tools. In the Analysis of Generated Maps: A Cooperative Study Among Tools section, we show the performance of the software tools with a set of data and analyze their possible positive synergies. In the Lessons Learned section, we note some lessons learned. Finally, some concluding remarks are made.

## Science Mapping

Science mapping or bibliometric mapping is a spatial representation of how disciplines, fields, specialties, and individual documents or authors are related to one another (Small, 1999). It is focused on monitoring a scientific field and delimiting research areas to determine its cognitive structure and its evolution (Noyons, Moed, & van Raan, 1999b).

In this section, different important aspects of a science mapping analysis are described, such as: (a) the data sources, (b) the units of analysis, (c) the data preprocessing, (d) the similarity measures that can be used to normalize the relations between the units of analysis, (e) the mapping steps, (f) the types of methods of analysis that can be employed, (g) some visualization techniques, and finally, (h) interpretation of results.

### Data Sources

Nowadays, there are several online bibliographic (and also bibliometric) databases where scientific works and documents and their citations are stored. These sources of bibliographic information allow us to search and retrieve information about the majority of scientific fields. Undoubtedly, the most important bibliographic databases are ISI WoS (<http://www.webofknowledge.com>), Scopus (<http://www.scopus.com>), Google Scholar (<http://scholar.google.com>), and NLM's MEDLINE (<http://www.ncbi.nlm.nih.gov/pubmed>).

ISI WoS, Scopus, and Google Scholar do not cover the scientific fields and journals in the same way, as different studies show. There are different studies (Bar-Ilan, 2010; Falagas, Pitsouni, Malietzis, & Pappas, 2008; Mikki, 2010) that relate this fact. Moreover, downloading large datasets from Google Scholar is difficult and a dump of the entire dataset is not available.

There are other bibliographic sources that can be used, such as: arXiv (<http://arxiv.org>), CiteSeerX (<http://citeseerx.ist.psu.edu/>), Digital Bibliography & Library Project (DBPL; <http://dblp.uni-trier.de/>), SAO/NASA Astrophysics Data System (ADS; <http://adswww.harvard.edu/>), Science Direct (<http://www.sciencedirect.com/>)

Patent data and funding data are also frequently used. Patent data can be retrieved from specific data sources such as the United States Patent and Trademark Office (USPTO; <http://www.uspto.gov/>) or the Derwent Innovations Index provided by ISI WoS. Funding data can be downloaded from the National Science Foundation (<http://www.nsf.gov/>)

TABLE 1. Bibliometric techniques taxonomy.

Bibliometric technique		Unit of analysis used	Kind of relation
Bibliographic coupling	Author	Author's oeuvres	Common references among author's oeuvres
	Document	Document	Common references among documents
	Journal	Journal's oeuvres	Common references among journal's oeuvres
Co-author	Author	Author's name	Authors' co-occurrence
	Country	Country from affiliation	Countries' co-occurrence
	Institution	Institution from affiliation	Institutions' co-occurrence
Co-citation	Author	Author's reference	Co-cited author
	Document	Reference	Co-cited documents
	Journal	Journal's reference	Co-cited journal
Co-word		Keyword, or term extracted from title, abstract or document's body	Terms' co-occurrence

### Units of Analysis in Bibliometric Techniques

The most common units of analysis in science mapping are journals, documents, cited references, authors (the author's affiliation can also be used), and descriptive terms or words (Börner et al., 2003). The words can be selected from the title, abstract, body of the document, or some combinations of them. Furthermore, we can select the original keywords of the documents (author's keywords) or the indexing ones provided by the bibliographic data sources (e.g., ISI Keywords Plus) as words to analyze.

Several relations among the units of analysis can be established. Usually, the units of analysis are used as a co-occurrence data by the science mapping process, that is, the similarity between the units of analysis is usually measured counting the times that two units appear together in the documents. Furthermore, direct linkage can be used to obtain the relations among units.

The relation among units can be represented as a graph or network, where the units are the nodes and the relations among them represent an edge between two nodes, that is, by using relationships among units of analysis, different bibliometric networks can be built.

In Table 1, a taxonomy of the most common bibliometric techniques according to the units of analysis used and the established relationships among them is presented.

Different aspects of a research field can be analyzed depending on the selected units of analysis, for example, by using the authors (co-author or co-authorship analysis) the social structure of a scientific field can be analyzed (Gänzel, 2001; Peters & van Raan, 1991). Likewise, by using the author's affiliations—co-institution, co-university, or co-country—the international dimension of the research field is studied. On the other hand, co-word (Callon et al., 1983) analysis is used to show the conceptual structure and the main concepts treated by a field. Co-citation (Small, 1973) and bibliographic coupling (Kessler, 1963) are used to analyze the intellectual structure of a scientific research field. The difference between bibliographic coupling and co-citation is that bibliographic coupling is a fixed and permanent relationship because it depends on the references contained in coupled documents, whereas co-citation will vary over time (Jarneving, 2005).

Bibliographic coupling and co-citation can be extended using journals and authors. Particularly, author bibliographic coupling (Zhao & Strotmann, 2008) aims at discovering co-author relationships between authors that cite the same references, whereas journal bibliographic coupling (Gao & Guan, 2009; Small & Koenig, 1977) aims at discovering the journals that cite the same references. On the other hand, author co-citation (White & Griffith, 1981) aims to discover the authors that are frequently cited together, whereas journal co-citation analysis (McCain, 1991) discovers the journals that are co-cited frequently. Furthermore, journal bibliographic coupling and journal co-citation can be extended to a category journal level. This supra-level of journal co-citation has been used to study the marrow of science (Moya-Anegón et al., 2007) using the ISI categories.

Finally, a relation between units can be established using direct linkages, for example, a document-document, author-author, or journal-journal citation network. Furthermore, a relation can be established using different units, for example, an author-paper (consumed/produced) network.

### Data Preprocessing

The data retrieved from the bibliographic sources normally contains errors, for example, misspelling in the author's name, in the journal title, or in the references list. Sometimes, additional information has to be added to the original data, for example, if the author's address is incomplete or wrong. For this reason, a science mapping analysis cannot be applied directly to the data retrieved from the bibliographic sources, that is to say, a preprocessing process over the retrieved data is necessary. In fact, the preprocessing step is perhaps one of the most important for improving the quality of the units of analysis (mainly authors and words) and thus to obtain better results in the science mapping analysis.

Different preprocessing processes can be applied to prepare the data to get a good performance in the science mapping analysis:

- Detecting duplicate and misspelling items. Sometimes, there are items in the data that represent the same object or concept but with different spelling, for example, an author's name can be written in different ways (e.g., Garfield, E.; Eugene

Garfield), and yet each way represents the same author. In other cases, a concept can be represented with different words (lexical forms) or acronyms, and yet represent the same concept. To detect duplicate items and misspelling enables these errors to be fixed.

- The time slice process is useful to divide the data into different time subperiods, or time slices, to analyze the evolution of the research field under study. This process is only necessary if the science mapping analysis is made in the context of a longitudinal study (Garfield, 1994; Price & Gürsey, 1975).
- Data reduction aims to select the most important data. Normally, we have a large amount of data. With such a quantity of data, it could be difficult to get good and clear results in the science mapping analysis. For this reason, it is ordinarily carried out using a portion of the data. This portion could be, for example, the most cited articles, the most productive authors, and the journals with the best performance metrics.
- Networks preprocessing can be used to select the most important nodes of the network of relationships between the units of analysis (bibliometric networks) according to different measures, removing the isolated nodes, removing the less important links between nodes, etc.

### Normalization Process

When the network of relationships between the selected units of analysis has been built, a transformation is first applied to the data to derive similarities from the data or, more specifically, to normalize the data (van Eck & Waltman, 2009).

Different similarity measures have been used in the literature, the most popular being *Salton's Cosine* (Salton & McGill, 1983), *Jaccard's Index* (Peters & van Raan, 1993), *Equivalence Index* (Callon, Courtial, & Laville, 1991), and *Association Strength* (Coulter, Monarch, & Konda, 1998; van Eck & Waltman, 2007), which is also known as *Proximity Index* (Peters & van Raan, 1993; Rip & Courtial, 1984) or *Probabilistic Affinity Index* (Zitt, Bassecoulard, & Okubo, 2000).

Usually, a normalization of the document's terms is needed; for example, if a co-citation analysis is performed and various clusters are detected, then a label should be set to each one. This label should be selected using the most important document's terms of the cluster. The text normalization sets a weight to each term according to its importance in the corpus. Different text normalization measures can be applied (Baeza-Yates & Ribeiro-Neto, 1999; Chen, Ibekwe-SanJuan, & Hou, 2010; Salton & McGill, 1983): tf-idf, latent semantic analysis, log-likelihood ratio tests, log entropy, mutual information, etc.

### The Mapping Step

The mapping step is the most important one. The process itself is responsible for building the map by applying a mapping algorithm to the whole network formed using the relationships among the selected units of analysis.

Different techniques have been proposed to build the map (Börner et al., 2003). Dimensionality reduction techniques

such as principal component analysis or MDS are used to transform the network into a low-dimension space (often two-dimension). Clustering algorithms are used to perform community detection, splitting the global network into different subnetworks. Recently, some authors have proposed new and different clustering algorithms to carry out this task: *Stremer* (Kandylas, Upham, & Ungar, 2010), *spectral clustering* (Chen et al., 2010), *modularity maximization* (Chen & Redner, 2010), and a *bootstrap resampling* with a significance clustering (Rosvall & Bergstrom, 2010), among others. Finally, Pathfinder networks (PFNETs) are used to identify the backbone of the network (Quirin, Cordón, Santamaría, Vargas-Quesada, & Moya-Anegón, 2008; Schvaneveldt, Durso, & Dearholt, 1989). Furthermore, general graph mining techniques (Cook & Holder, 2006; Skillicorn, 2007) or social network analysis (Carrington et al., 2005; Wasserman & Faust, 1994) can be used in the mapping step.

The information obtained and the kind of map built will depend of the applied technique.

### Analysis Methods

Once the map has been built, different analyses can be applied to extract useful knowledge.

*Network analysis* (Carrington et al., 2005; Cook & Holder, 2006; Skillicorn, 2007; Wasserman & Faust, 1994) allows us to perform a statistical analysis over the generated map in the later step, for example, different measures on the network, such as the total number of nodes and isolated nodes, average degree, the number of weakly connected components, or the graph density can be measured. If a community detection algorithm was applied to build the map, then Callon's centrality and density (Callon et al., 1991; Cobo et al., 2011) or other values that measure the relationships among the detected clusters can be used. Moreover, the overlapping between the clusters can be measured using, for example, the Jaccard's Index. Furthermore, if documents are assigned to each cluster, a performed analysis can be carried out to obtain quantitative or qualitative measures of each cluster (Cobo et al., 2011).

Another important analysis that can be performed in a science mapping process is the *temporal analysis*, which aims to identify the nature of phenomena represented by a sequence of observations such as patterns, trends, seasonality, and outliers. In other words, it aims to analyze the evolution of the research field across different periods of time. This task can be performed using a longitudinal framework (Garfield, 1994; Price & Gürsey, 1975).

*Burst detection* is a kind of temporal analysis. It aims to find features that have a high intensity over finite durations of time periods. In Kleinberg (2003), an algorithm to deal with this problem was described.

Finally, *geospatial analysis* (Batty, 2003; Leydesdorff & Persson, 2010; Small & Garfield, 1985) aims to answer the question of where something happens and with what impact on neighbouring areas. Geospatial analysis requires spatial

TABLE 2. General information.

Software tool	Last version	Year	Developed by
Bibexcel	2010-09-22	2010	University of Umeå (Sweden)
CiteSpace	2.2.R9	2010	Drexel University (USA)
CoPalRed	1.0 beta	2005	University of Granada (Spain)
IN-SPIRE	5	2010	Pacific Northwest National Laboratory
Leydesdorff's Software	N/A	N/A	University of Amsterdam (The Netherlands)
Network Workbench Tool	1.0.0	2009	Indiana University (USA)
Science of Science (Sci <sup>2</sup> ) Tool	0.0.3 alpha	2010	Indiana University (USA)
VantagePoint	7	2010	Search Technology, Inc.
VOSViewer	1.2.1	2010	Leiden University (The Netherlands)

attribute values or geolocations for the units of analysis; this data is usually extracted from affiliation data.

### Visualization Techniques

As we have shown in the previous subsection, the output of each analysis method is different. The visualization technique employed is very important to a good understanding and better interpretation of the output.

The networks and subnetworks detected in the mapping step can be represented using *heliocentric maps* (Moya-Anegón et al., 2005), *geometrical models* (Skupin, 2009), and *thematic networks* (Bailón-Moreno et al., 2006; Cobo et al., 2011). Another approach consists of representing the networks in a map, where the distance between two items reflects the strength of the relation between both (Davidson et al., 1998; Fabrikant, Montello, & Mark, 2010; Polanco et al., 2001; van Eck & Waltman, 2010). A smaller distance generally indicates a stronger relation (van Eck & Waltman, 2010).

If a community detection is applied, then the different clusters detected (subnetworks) can be categorized using a strategic diagram. A *strategic diagram* (Callon et al., 1991; Cobo et al., 2011) is a two-dimensional space built by plotting themes according to different measures extracted using a post network analysis.

To show the evolution of detected clusters in successive time periods (temporal analysis), different techniques have been used: *Cluster string* (Small, 2006; Small & Upham, 2009; Upham & Small, 2010), *rolling clustering* (Kandylas et al., 2010), *alluvial diagrams* (Rosvall & Bergstrom, 2010), *ThemeRiver visualization* (Havre, Hetzler, Whitney, & Nowell, 2002), and *thematic areas* (Cobo et al., 2011). Other authors propose laying out the graph of a given time period, taking into account the previous and subsequent ones (Leydesdorff & Schank, 2008), or to pack synthesized temporal changes into a single graph (Chen, 2004; Chen et al., 2010).

Geospatial results are usually visualized over a world or a thematic map. As an example, if a co-author analysis is applied and then a community detection is performed, the detected clusters of authors can be represented as a network in which each node is laid out over the author's country.

### Interpretation

When the science mapping analysis has finished, the analyst has to interpret the results and maps using their experience and knowledge.

In the interpretation step, the analyst looks to discover and extract useful knowledge that could be used to make decisions on which policies to implement.

### Software Tools Designed to Perform a Science Mapping Analysis: A Survey

Although the science mapping analysis can be performed using generic software for social network analysis (Börner et al., 2010), there are other software tools specifically developed for science mapping analysis.

In this section, we present nine representative software tools specifically developed to analyze scientific domains by means of science mapping. These software tools are as follows:

- Bibexcel (Persson et al., 2009)
- CiteSpace II (Chen, 2004, 2006)
- CoPalRed (Bailón-Moreno et al., 2005, 2006)
- IN-SPIRE (Wise, 1999)
- Leydesdorff's Software
- Network Workbench Tool (Börner et al., 2010; Herr et al., 2007)
- Sci<sup>2</sup> Tool (Sci<sup>2</sup> Team, 2009)
- VantagePoint (Porter & Cunningham, 2004)
- VOSViewer (van Eck & Waltman, 2010)

In Table 2 some details of these software tools are described.

#### Bibexcel

Bibexcel (<http://www.umu.se/inforsk/Bibexcel>; Persson et al., 2009) is a bibliometric tool developed at the University of Umeå (Sweden). This tool was specifically developed to manage the bibliometric data and build maps, which can be read by software such as Excel, SPSS, UCINET (Borgatti et al., 2002), and Pajek (Batagelj & Mrvar, 1998). Bibexcel is freely accessible for academic nonprofit use.

Bibexcel can read data retrieved from different bibliographic sources, such as ISI Web of Science (WoS), Scopus, and the Procite export format.

Bibexcel allows different preprocessing over the textual data to be performed, for example, an English word stemmer can be applied and duplicate documents can be deleted. Moreover, Bibexcel enables the deletion of low frequency items and keeps only the strongest links.

Different bibliometric networks can be extracted. The principal ones are as follows: co-citation, bibliographic coupling, co-author, and co-word. Furthermore, different co-occurrence matrices can be extracted using any document's field, or some combination of fields. The matrices can be normalized using three different measures: Salton's Cosine, Jaccard's Index, and the Vladutz and Cook measures.

To the normalized data, the user can apply a clustering algorithm or prepare a matrix for an MDS analysis (using external software). Bibexcel does not present an adequate visualization tool for the output, but it presents different export options that make data visualization possible using external software like Pajek, UCINET or SPSS. The bibliometric networks can also be exported.

### *CiteSpace II*

CiteSpace (<http://cluster.cis.drexel.edu/~cchen/citespace>; Chen, 2004, 2006) was developed at Drexel University (USA) and it can be freely downloaded. It is a software tool developed to detect, analyze, and visualize patterns and trends in scientific literature. Its primary goal is to facilitate the analysis of emerging trends in a knowledge domain.

CiteSpace can read different formats of bibliographic sources, such as WoS, PubMed, arXiv, and SAO/NASA Astrophysics Data System (ADS). Furthermore, CiteSpace is able to read grants data such as NSF Awards and patent data from Derwent Innovations Index.

Different kinds of bibliometric networks can be constructed: co-author, co-author institutions, co-author countries, co-grants,<sup>1</sup> subject categories co-occurrence, co-word, documents co-citation, author co-citation, journal co-citation, and documents bibliographic coupling. The networks, or graphs, can be built for different time periods to analyze the evolution of the studied domain. Moreover, the analyst can filter the items with which the networks are built to select the most important of them (e.g., select the 100 most cited items from each time slice). The matrix (network) is normalized using Salton's Cosine, Dice, or the Jaccard's Index.

Once the networks have been built, CiteSpace allows us to visualize them and perform several analyses on them. CiteSpace allows the analyst to perform a spectral clustering and a citation burst detection. In addition, CiteSpace has three visualization modes (Chen, 2006): cluster view, time line, and time zone.

If clusters are detected, CiteSpace can assign labels to each one using the most important terms extracted from the keywords, title, or abstract. The terms are measured using the

tf-idf, log-likelihood ratio tests, or mutual information (Chen et al., 2010).

### *CoPalRed*

CoPalRed (<http://ec3.ugr.es/copalred/>; Bailón-Moreno et al., 2005, 2006) is a commercial software developed by the research group EC<sup>3</sup> at the University of Granada (Spain). It is specifically designed to perform co-word analysis using the keywords of scientific documents. It is described as a Knowledge System, which collects the information contained in databases and transforms it into new knowledge.

This software tool reads files in comma-separated values format (csv), generated through the reference manager software Procite.

One of the strengths of CoPalRed is that it contains a preprocessing module that allows users to normalize the keywords in a simple way. With this module, the user can unify items (lexical items) that represent the same concept. Once the keywords are unified, CoPalRed builds a co-occurrence matrix and normalizes it using the equivalence index (Callon et al., 1991).

CoPalRed performs three kinds of analysis: structural analysis, strategic analysis, and dynamic analysis.

- Structural analysis. It shows the knowledge in the form of thematic networks in which words and their relationships are drawn.
- Strategic analysis. It places each thematic network in a relative position within the global thematic network using two criteria: centrality (or intensity of its external relations) and density (according to their internal cohesion density).
- Dynamic analysis. CoPalRed analyzes the transformations of the thematic networks over time. It identifies approaches, bifurcations, appearances, and disappearances of themes.

CoPalRed visualizes the results using strategic diagrams, themes, and thematic networks (Bailón-Moreno et al., 2005, 2006; López-Herrera et al., 2009, 2010). Each theme has assigned a label that is the name of the most central node (keyword) of its associated thematic network. Furthermore, each theme can be represented in the strategic diagram as a sphere, where its volume is proportional to the number of documents belonging to it. In the same way, each node (keyword) of the thematic network can be represented as a sphere where its volume is proportional to the keyword's frequency.

### *IN-SPIRE*

IN-SPIRE (<http://in-spire.pnl.gov>; Wise, 1999) is a commercial visual document analysis software tool that gives the analyst the ability to uncover relationships, trends, and themes hidden within data to obtain new knowledge and new insights. IN-SPIRE uses landscape metaphor to help the user to easily discover the relationship among documents and the sets of documents that are very similar. This tool uses statistical word patterns to characterize documents based on their context (Hetzler & Turner, 2005). IN-SPIRE derived from the

<sup>1</sup>Using the funding field of the document and analyzing the sponsors' names of the grants that co-occur in the funding data.

SPIRE project funded by the Department of Energy and the U.S. intelligence agencies. It has been developed at Pacific Northwest National Laboratory (United States).

IN-SPIRE can read unformatted documents (ASCII text) or formatted documents such as HTML and XML. Furthermore, it can read data from Microsoft Excel documents and csv formatted files. The software allows the user to select the fields that will be used to measure the similarity among documents and other meta-fields such as the title of the documents and the associated date.

Unlike the other analyzed software tools, IN-SPIRE does not extract bibliometric networks from the selected field. It uses a field or a set of fields to calculate the similarity among documents using its own text engine (Wise, 1999). In short, it uses the vector space model (Salton & McGill, 1983), and thus each document is represented as a vector of terms. So, if keywords are selected as the field, then the similarity measure will show whether two documents have similar keywords. Although IN-SPIRE is able to build a map using any field, its text engine works better if words are selected as the field. The text engine needs a large amount of data to correctly detect the similarities among documents.

When the similarities among documents have been calculated, IN-SPIRE performs a cluster algorithm called “*Fast Divisive Clustering*” (Wise, 1999). At the end of the clustering process, several themes (sets of documents) are generated. Each theme has as its name the more frequently appearing terms (using tf-idf) of its documents.

IN-SPIRE provides two different visualization techniques, which are the flagship of the software: *Galaxies* and *ThemeScape*<sup>TM</sup>. The *Galaxies* visualization employs the metaphor of documents as stars in the night sky. On the other hand, *ThemeScape* is constructed directly from the distribution of documents in the *Galaxies* visualization, representing themes as sedimentary layers that together create the appearance of a natural landscape. In the *ThemeScape* visualization, the height of its peaks corresponds to topic strength at those locations; the extent of its peaks corresponds to the area and brightness of the themes in the *Galaxies* visualization. In both visualizations, the proximity of two items (documents) reveals the similarity between them. Related documents are grouped together and common themes are highlighted.

IN-SPIRE provides a set of tools that help the analyst to discover knowledge within the corpus of studied documents:

- Time slicer allows us to see how particular themes grow or shrink over time and can show how the mix of themes in the galaxy changes over time.
- The Groups tool defines a collection of documents within the studied corpus.
- The Facets allows us to discover relationships between calculated themes as well as groups defined by the user.
- Robust query capabilities that support boolean, word proximity, phrase, or example-based searches.
- The Correlation tool allows us to discover correlation between the groups.

## Leydesdorff's Software

Leydesdorff's software (<http://www.leydesdorff.net>) is a set of command-line programs that enable a science mapping with different analysis functions to be performed. It was developed at the University of Amsterdam (The Netherlands). The set of programs is freely accessible to the academic community.

The different programs allow the performance of several bibliometric analyses: co-word, co-author, author bibliographic coupling, journal bibliographic coupling, and author co-citation. The results can be visualized using external software such as Pajek (Batagelj & Mrvar, 1998), UCINET (Borgatti et al., 2002), Network Workbench Tool (see Subsection 3.6), or the Sci<sup>2</sup> Tool (see Subsection 3.7). Moreover, international and institutional collaboration, and collaboration at the level of cities can be analyzed. The visualization of these collaboration networks can be done using Google Maps and external software. The different matrices are normalized using the Salton's Cosine measure.

There are programs for organizing the data downloaded from different sources (WoS, Scopus, Google Scholar, and Google) into a database. This database will be the input file of the different analysis programs.

The set of programs does not allow the data to be preprocessed; so, for example, to perform a longitudinal analysis, external software is needed to split the data into different time periods.

## Network Workbench Tool

The Network Workbench (NWB) Tool (<http://nwb.slis.indiana.edu>) is a general network analysis, modelling, and visualization toolkit for physics, biomedical, and social science researchers (Börner et al., 2010; Herr et al., 2007). It was developed by the *Cyberinfrastructure for Network Science Center* at Indiana University (USA) and is freely accessible.

The NWB Tool provides specific algorithms to deal with publications data to construct and analyze bibliometric networks and maps. The tool is able to read different bibliometric data formats such as ISI WoS, Scopus, Bibtex, and EndNote Export Format. It can also read funding data from the National Science Foundation (NFS) and other scholarly data in csv format.

The NWB Tool allows the data to be preprocessed, different kinds of networks to be built, a graph analysis over the built networks to be performed, and finally their visualization. Moreover, the tool is able to carry out a temporal analysis.

- Data preprocessing is performed removing duplicate records, dividing the data by different time periods, and detecting and unifying duplicate items with different spelling (i.e., items that represent the same author in a co-author analysis or terms that represent the same concept in a co-word analysis).
- The NWB Tool allows different kinds of networks to be built: documents co-citation, co-author, co-word, and documents bibliographic coupling. Furthermore, the tool can build networks by direct linkage; for example, it can build an

author-document network (consumed/produced) or a direct citation network.

- Several algorithms can be used to perform the mapping step and a graph analysis on the generated networks. Furthermore, the tool is able to carry out a burst detection to identify increases in the usage frequency of items.
- The visualization of the generated graphs is carried out through different external plugins (e.g., GUESS, Jung). Furthermore, several graph layouts can be applied such as the DrL algorithm, which is the opensource successor of VxOrd (Davidson, Hendrickson, Johnson, Meyers, & Wylie, 2001) that was used in the VxInsight program (Boyack, Wylie, & Davidson, 2002; Davidson et al., 1998).

### *Sci<sup>2</sup> Tool*

The Sci<sup>2</sup> Tool (<http://sci.slis.indiana.edu>) is a modular toolset specifically designed to perform the study of science. It supports temporal, geospatial, topical, and network analysis and the visualization of datasets at the micro (individual), meso (local), and macro (global) levels (Sci<sup>2</sup> Team, 2009). It was developed by *Cyberinfrastructure for Network Science Center* at Indiana University (USA) and it is freely accessible.

The Sci<sup>2</sup> Tool is similar to the NWB Tool (see Network Workbench Tool section), but it is specifically focused on science studies and has specific algorithms to deal with this topic. The most important strength of the tool could be that it provides several methods to deal with bibliometric data, to prepare it for later analysis.

Similarly to the NWB Tool, the Sci<sup>2</sup> Tool is able to read different bibliographic data formats: ISI WoS, Scopus, Bibtext, and EndNote Export Format. It can also read funding data from the National Science Foundation (NFS) and other scholarly data in csv format.

The Sci<sup>2</sup> Tool allows the data to be prepared and preprocessed, extracting different types of networks, performing a temporal, geospatial, topical, and network analysis, and finally visualizing the results through different plugins and layout algorithms. Sci<sup>2</sup> Tool includes the DrL layout algorithm.

The data preparation cleans the bibliographic data and creates different networks and tables that can be used in preprocessing, analysis, and visualization. Principally, the networks that can be extracted are as follows: co-author, co-PI (Principal investigator), co-word, document co-citation, journal co-citation, author co-citation, author bibliographic coupling, document bibliographic coupling, and journal bibliographic coupling. Moreover, the tool can build different direct linkage networks such as author-citation, document-citation, source-citation paper, and, finally, author-document (consumed/produced) network.

The tool contains several algorithms to perform the mapping step and next applies different analyses. The mapping step can be performed using community detection and backbone identification. Temporal analysis is performed slicing the data into different time periods and through a burst detection. Geospatial analysis is performed through geocoding and geospatial thematic maps. Topic analysis is performed

through a burst detection over the words and a co-word analysis. Network analysis allows a statistical analysis and the application of different algorithms over the networks.

### *VantagePoint*

VantagePoint (<http://www.thevantagepoint.com/>; Porter & Cunningham, 2004) is a powerful commercial text-mining software tool for discovering knowledge in search results from patent and literature databases, or generally from structured texts. It allows the user to analyze large volumes of structured text to discover patterns and relationships and quickly address *who*, *what*, *when*, and *where*. It was developed by Search Technology Inc. (USA). VantagePoint has been used to perform several science mapping analyses (Morel, Serruya, Penna, & Guimarães, 2009; Porter & Youtie, 2009a,b; Shapira, Youtie, & Porter, 2010).

VantagePoint has 180 import filters that allow the user to import data from almost any literature or patent database. Furthermore, it has import filters to load data from Microsoft Excel and Access, XML file format<sup>2</sup> or used-defined filters.

Once the dataset is loaded, VantagePoint shows the different fields included in the dataset; for example, if the dataset contains bibliographical information, then the fields could be the title, authors, affiliations, abstract, and references of the documents (records).

The VantagePoint graphic interface has three parts: the workspace, the title view, and the detail windows. The workspace displays all of the lists, matrices, and map views generated by the user. The title view displays the titles of the records in the dataset for a selected set of items. Finally, the detail window shows the co-occurrence of items in one field (any field can be selected) with items or nodes selected using lists or charts.

This software tool allows us to build different lists from any field. The lists show all of the field's items from the dataset. In the list view, for each item, the number of records where the item appears and the number of instances (number of times that the items appear in the dataset, taking into account the duplicate items in the records) is shown. The items of a list can be assigned to several groups. Groups are useful for defining a portion of the dataset to reduce the data used in the later analysis, for example, a group containing the top 30 authors can be built. The items can be associated with more than one group.

One strength of VantagePoint is its preprocessing and data cleaning tools. A list can be cleaned or reduced using the Cleanup function, which attempts to identify the items that may be equivalents, performing fuzzy near matches on specific fields. Moreover, a list can be cleaned, applying a thesaurus. Although VantagePoint has several predefined thesauruses that can be easily used, the user can define his/her own thesaurus or edit an existing thesaurus using

<sup>2</sup>There is a wizard that allows an import filter to be created from XML data.



the Thesaurus Editor. Any change performed over a list will generate a new list, so we always keep the original data.

VantagePoint allows several kinds of matrices to be built that show the records in the dataset contained in two given lists:

- Co-occurrence matrix: it shows the number of records in which the element *i* (from the first list) and the element *j* (from the second list) appear together.
- Auto-correlation matrix: it shows the correlations among items in a list.
- Cross-correlation matrix: it shows the correlations among items in a list based on the values of another list.
- Factor matrix: it is the result of a principal component analysis. The factor matrix shows the items in rows and the factors in columns.

VantagePoint also allows different matrices to be built that can be used as the input in the mapping process: co-author (using the author's name, affiliation or country), co-citation (using the reference, reference's author or source), and co-word (using any set of terms). Furthermore, if the selected lists to build the matrix are different, heterogeneous matrices can be built; for example, the user can build a matrix of author per year to analyze the author's productivity. The matrices can be exported into a text file, or the user can directly copy a selection of the matrix and paste it in Microsoft Excel.

The correlation matrices can be normalized using Pearson's *r*, Salton's Cosine or the Max Proportional measures. Furthermore, the co-occurrence matrix can be normalized using the tf-idf similarity measure.

VantagePoint includes three kinds of maps that correspond to the three last matrices: cross-correlation map, auto-correlation map, and factor map. These maps are a graphical representation of the corresponding matrices. In the cross-correlation maps, the similarity between items is measured using the cosine. In the factor map, and auto-correlation the similarity measure used is Pearson's *r*.

Finally, VantagePoint also includes the capability to execute Visual Basic scripts to make repetitive (and/or complex) actions that a user may require.

### VOSViewer

VOSViewer (<http://www.vosviewer.com>; van Eck & Waltman, 2010) is a software tool specifically designed for constructing and visualizing bibliometric maps, paying special attention to the graphical representation of such maps. It is appropriate to represent big maps since zoom functionality, special labelling algorithms, and density metaphors are used. The software tool was developed by the *Centre for Science and Technology Studies* at Leiden University (The Netherlands) and it is freely available to the bibliometric research community.

Although VOSViewer can be used to construct and visualize bibliometric maps of any kind of co-occurrence data, the software tool does not allow any co-occurrence matrix from the bibliometric data to be extracted and built. To do this, an external process is needed. Furthermore, the software tool

has no preprocessing modules to prepare the data for later analysis.

To lay out the elements on the maps, the VOS mapping technique (van Eck, Waltman, Dekker, & van den Berg, 2010) is used by VOSViewer. This technique builds a similarity matrix from a co-occurrence matrix using a similarity measure known as the association strength (van Eck & Waltman, 2007, 2009). The VOS mapping technique builds a two-dimensional map in which the elements are located in such way that the distance between any pair of items reflects their similarity as accurately as possible. The idea of the VOS mapping technique is to minimize a weighted sum of squared Euclidean distances between all pairs of items through an optimization process.

Although VOSViewer implements the VOS mapping technique, the program can also be used to view any two-dimensional map constructed with other techniques. VOSViewer allows us to perform a community detection using the VOS clustering technique, which is related to the technique of modularity-based clustering (Waltman et al., 2010). Once the map is built, VOSViewer allows its examination through four views:

- Label view. In this view each element is represented by a label and also by a circle. The more important an item, the larger its label and its circle. Thanks to an intelligent algorithm, which shows only the most important labels (most frequent) depending on the level of zoom, the software tool avoids the label overlapping. The circles that have the same color belong to the same cluster. This color is the same as the corresponding cluster's color in the cluster view.
- Density view. In this view, each item is represented by a label in a similar way as in the label view. Each point in the map has a color that depends on the density of items at that point, which depends both on the number of neighbouring items and on the weights of these items. VOSViewer calculates the density of each point according to the equation defined by (van Eck & Waltman, 2010), which uses a Gaussian kernel function. The density is translated using a color scheme (for more information see van Eck & Waltman; 2010)).
- Cluster density view. This view is available only if items have been previously assigned to a cluster. The cluster density view is similar to the ordinary density view except that the density of items is displayed separately for each cluster of items.
- Scatter view. This is a simple view in which items are indicated by a small circle and in which no labels are displayed.

### Comparative Study

As mentioned above, in this article, we also present a comparative study of the nine software tools described above. In such a way, we are able to highlight the main differences and positive synergies existing among the different software tools. To do so, we analyze the nine software tools taking into account five points of view: (a) the preprocessing methods, (b) the bibliometric networks available, (c) the normalization measures used, (d) the type of analysis, and finally, (e) other secondary aspects.

TABLE 3. Preprocessing methods.

Software tool	De-duplication	Time slicing	Data reduction	Network reduction
Bibexcel			x	x
CiteSpace		x	x	x
CoPalRed	x	x	x	
IN-SPIRE			x	
Leydesdorff's Software				
Network Workbench Tool	x	x	x	x
Science of Science Tool	x	x	x	x
VantagePoint	x	x	x	
VOSViewer				

TABLE 4. Bibliometric networks.

Software tool	Bibliographic coupling			Co-author			Co-citation			Co-word (CWA)	Direct Linkage (DL)	Others
	Author (ABCA)	Document (DBCA)	Journal (JBCA)	Author (ACAA)	Country (CCAA)	Institution (ICAA)	Author (ACA)	Document (DCA)	Journal (JCA)			
Bibexcel		x		x	x	x	x	x	x	x		x
CiteSpace		x		x	x	x	x	x	x	x		x
CoPalRed										x		
IN-SPIRE										x		
Leydesdorff's Software	x		x	x	x	x	x			x		
Network Workbench Tool		x		x				x		x	x	
Science of Science Tool	x	x	x	x			x	x	x	x	x	x
VantagePoint				x	x	x	x	x	x	x		x
VOSViewer												

### Preprocessing Methods

Special modules to perform a preprocessing of the data are important characteristics of a science mapping software tool. In Table 3, the principal preprocessing modules available in each software tool are shown.

The module to detect duplicate items is important, for example, in co-word analysis or co-author analysis. With this module, a user could decide to join two or more items that represent the same concept or the same author. This module does not only merge two items but also selects or sums up the attribute value, for example, the times cited of the original records.

A time slicing option is needed if the user wants to analyze the evolution of the domains under study. A module for reducing the data is useful if the user wants to filter the data to analyze the most important information.

Finally, network reduction is useful to filter the nodes or links of a network (similar to the reducing data module), or to apply a pruning algorithm to the networks.

Only NWB Tool and Sci<sup>2</sup> Tool have the four preprocessing modules. By contrast, Leydesdorff's Software and VOSViewer do not have any of these modules, which is a strong drawback.

IN-SPIRE performs the time slicing directly over the data. It does not need to preprocess the data to split the dataset into different slices.

### Bibliometric Relation Between Units of Analysis

An important consideration in the use of some science mapping software tools is whether they are able to establish different relationships between the units of analysis, that is, if they are able to extract different bibliometric networks.

In Table 4, the different bibliometric networks available are shown for each software tool. The column "others" means that the software tool is able to build other un-common or heterogeneous networks or matrices.

Although there are no software tools able to build all of the different varieties of bibliometric networks, Bibexcel, CiteSpace, Leydesdorff's Software, Sci<sup>2</sup> Tool, and Vantage-Point are the software tools able to build the majority of them. By contrast, VOSViewer is not able to build any of them; it is focused only on visualizing bibliometric maps. CoPalRed is focused only on one kind of bibliometric network. Finally, although IN-SPIRE can construct the maps using any field of the dataset, its way of representing the documents, by using the vector space model, makes it difficult to generate the maps

TABLE 5. Normalization measures.

Software tool	Measure
Bibexcel	Salton's Cosine, Jaccard's Index, or the Vladutz and Cook measures
CiteSpace	Salton's Cosine, Dice or Jaccard Strength
CoPalRed	Equivalence Index
IN-SPIRE	Conditional Probability
Leydesdorff's Software	Salton's Cosine
Network Workbench Tool	User defined
Science of Science Tool	User defined
VantagePoint	Pearson's r, Salton's Cosine or the Max Proportional
VOSViewer	Association Strength

TABLE 6. Methods of analysis.

Software tool	Burst detection	Geospatial	Network	Temporal
Bibexcel			x	
CiteSpace	x	x	x	x
CoPalRed			x	x
IN-SPIRE	x		x	x
Leydesdorff's Software				
Network Workbench Tool	x		x	x
Science of Science Tool	x	x	x	x
VantagePoint	x	x	x	x
VOSViewer			x	

using other fields such as the authors. It works better using words.

Some software tools allow the extraction of un-common networks, for example, the co-grant networks available in CiteSpace, co-PI networks available in Sci<sup>2</sup> Tool, or the particular matrices that are extracted by Bibexcel and VantagePoint using a set of specific documents' fields. Furthermore, some software such as Bibexcel and VantagePoint allow us to extract heterogeneous networks using different fields in the rows and columns, for example, a matrix showing the authors per years can be extracted.

Finally, NWB Tool and Sci<sup>2</sup> Tool can extract bibliometric networks using direct linkage.

### Normalization Measures

Once the bibliometric networks have been built, a normalization process can be carried out using different similarity measures. In Table 5, the measures used for each software tool are shown.

Three of the analyzed software tools use Salton's Cosine as a similarity measure. Other software tools like NWB Tool and Sci<sup>2</sup> Tool allow the users to define their own measures.

### Methods of Analysis

Different methods of analysis can be applied. In Table 6, the different methods of analysis available for each software tool are shown.

Only CiteSpace, Sci<sup>2</sup> Tool, and VantagePoint use the four kinds of analysis. Leydesdorff's Software does not carry out any of them.

CiteSpace and Sci<sup>2</sup> Tool have geocoding capabilities. CiteSpace uses Google and Yahoo!'s geocoder over the institutional data available. On the other hand, Sci<sup>2</sup> Tool uses Yahoo!'s geocoding service and an internal geocoder over any field that contains geographical data, such as institutional address and conference location.

### Other Aspects

In this subsection, we compare the software tools according to other aspects such as documentation/help, free or commercial availability, whether the source code is available, the possibility of installing the software in different platforms, and the extendability of the software.

NWB Tool and Sci<sup>2</sup> Tool have a great user guide where the tools are explained in detail. Furthermore, the user guide explains important aspects of science mapping; these are the only tools that explain this issue. VantagePoint has a good user guide and online help, and its website provides a large amount of video-tutorials. IN-SPIRE has a great website with video tutorials and online help. VOSViewer has a good manual. CiteSpace has a big wiki where important issues are described. Leydesdorff's Software has a good description and user guide for each of its command-line programs on its website.

Only three of the nine described software tools are commercial: CoPalRed, IN-SPIRE, and VantagePoint. The remaining software tools are freely available.

Taking into account the availability of the source code, only NWB Tool and Sci<sup>2</sup> make their source code available.

CiteSpace, NWB Tool, Sci<sup>2</sup> Tool, and VOSViewer were developed using the Java programming language, so they can be used with any platform (Windows, MacOS, Linux, etc.). On the other hand, Bibexcel, CoPalRed, IN-SPIRE, Leydesdorff's Software, and VantagePoint run only under Windows.

Finally, taking into account the possibilities of extending the software tools, NWB Tool and Sci<sup>2</sup> Tool are built over Cyberinfrastructure Shell (CIShell; <http://cishell.org/>), so it can be extended using this platform. According to the description given on its website, CIShell "is an open source, community-driven platform for the integration and utilization of datasets, algorithms, tools, and computing resources." VantagePoint can be extended using VisualBasic scripts.

### Analysis of Generated Maps: A Cooperative Study Among Tools

To complete the comparative study among the software tools, in this section, we develop a cooperative study of the software tools with a set of data. This cooperative use of different software tools gives us the opportunity to discover the possible positive synergies that could generate the joint use of these software tools.

To make a better comparison between software, a common science mapping analysis over a specific unit of analysis has to be performed. As was shown in Table 4, the analyzed software tools are unable to extract the same bibliometric networks, with co-word network the only one available in each software tool. For this reason, we select the words (or keywords) as the unit of analysis to perform the science mapping analysis.

As an example, we study the conceptual structure (Cobo et al., 2011) of the research field of fuzzy set theory (FST; Zadeh, 1965, 2008) by using the publications that have appeared in the most important and prestigious journals during 2005 to 2009, according to their impact factor,<sup>3</sup> on the topic: *Fuzzy Sets and Systems* and *IEEE Transactions on Fuzzy Systems*. Cobo et al. (2011), who studied their conceptual evolution across five different periods of time, made a deep analysis of these journals recently. In this section, we use the last period of time (2005-2009) of that analysis.

The amount of documents analyzed was 1,576, and they were downloaded<sup>4</sup> from the WoS. Specifically, 1,086 documents were published by the journal *Fuzzy Sets and Systems*, and 490 by *IEEE Transactions on Fuzzy Systems*.

The author's keywords and Keywords Plus of each document were used in the analysis. After a de-duplicating step (CoPalRed was used to carry out this task), there were 5,034 keywords. CoPalRed allows us to export the documents with the preprocessed items in a csv file, so this will be the input for the remaining software tools when possible. The whole

network build from the co-occurrence of these keywords contains an amount of 25,705 links.

In what follows, the different results obtained by the software tools are shown. The comparative study has been performed using the software tools able to visualize the results. For this reason, Bibexcel and Leydesdorff's software have not been used.

First, a co-word analysis was performed using CiteSpace. Given that it does not allow us to load the data in csv format, the dataset had to be loaded without any preprocessing from an ISIWoS format file. In Figure 1, the map generated by CiteSpace is shown. The map was made using the top 200 keywords. The lines between nodes represent the cosine similarity measure. The shadowed nodes represent clusters and the clusters' names were chosen selecting the most important keywords from each cluster according to the tf-idf measure. Inside each cluster there is a sphere which represents its centroid, and its volume is proportional to the size of the cluster.

We have to say that the printed out map does not show the power of CiteSpace. To make a good interpretation of the obtained results, the analyst should interact with CiteSpace's user interface, which allows us to perform a variety of analyses, different layouts, etc. Furthermore, the analyst can zoom in and zoom out on the network to appreciate the details of a local area.

Second, in Figure 2 the result obtained by CoPalRed is shown. In Figure 2a the generated strategic diagram is shown, and in Figure 2b the thematic network of a specific theme (*FUZZY-CONTROL*) is drawn. CoPalRed generated the maps using those keywords with a frequency equal to or higher than five and a co-occurrence value equal to or higher than three. The whole network contains 229 nodes and 432 links between them after this pruning. With this pruning, we maintain the most frequent and important keywords. The strategic diagram shows the main detected themes studied by the FST field in the studied period, categorizing them in four classes according to their *Callon's density* and *Callon's centrality* measures. Each theme in the strategic diagram is associated with a sphere and a label. Labels were chosen selecting the most central node of its associated thematic networks, where each node corresponds with a keyword. The volume of spheres represents the number of documents associated with each theme (or keyword in thematic networks). This information is also associated with the labels. Finally, the size of the lines in thematic networks represents the degree of association (equivalence index) between two nodes.

Third, the csv file exported by CoPalRed was loaded in IN-SPIRE. After defining the dataset, and selecting the terms, IN-SPIRE generated two maps: the Galaxy view (Figure 3) and Theme view (Figure 4).

In the galaxy view, the shadows represent groups of documents that are considered to be similar. The names of these themes are generated using the most important keywords according to their tf-idf measure. In the Theme view, the height of each peak corresponds to topic strength at that location, and the extent of each peak corresponds to the area

<sup>3</sup>According to the 2009 Journal Citation Report (ISI Web of Science).

<sup>4</sup>The data was downloaded on January 15, 2010.

CiteSpace, v. 2.2.R10  
 December 22, 2010 4:35:22 PM CET  
 D:\MCo\Documents\Universidad\research\Análisis\Terminos\trabajos\EnDesarrollo\JASIST-SoftwareSurvey\estudioComparativo\CiteSpace-Coword\data  
 Timespan: 2005-2009 (Slice Length=5)  
 Selection Criteria: Top 200 per slice  
 Network: N=200, E=271 (Density=0.0136)  
 Modularity Q=0.733  
 Mean Silhouette=0.7918

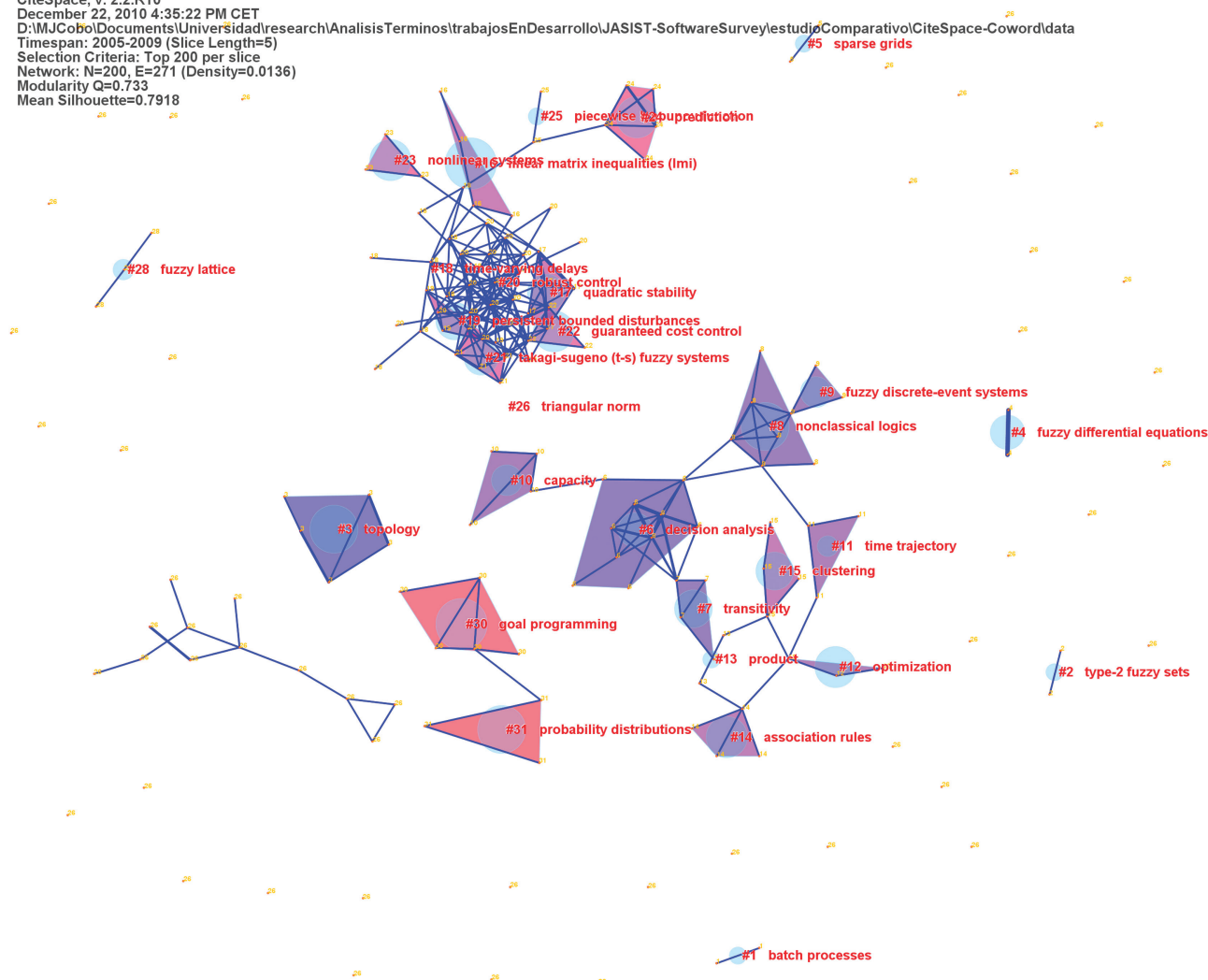


FIG. 1. Map generated by CiteSpace. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

and brightness of the corresponding theme in the Galaxy view.

As we can see in both the Galaxy and Theme views, IN-SPIRE did not detect many themes, because of the way in which it interprets the data. Unlike the other software tools analyzed, IN-SPIRE uses the vector space model to represent the documents, so it needs a large amount of terms to correctly detect the themes. In our dataset, the documents do not contain the necessary keywords, so IN-SPIRE could not determine correctly the similarity among documents. Probably if we had used the abstract or the full text of the documents, IN-SPIRE would have to obtain better results.

Now, the csv file was loaded into Sci<sup>2</sup> Tool,<sup>5</sup> and a co-occurrence network using the keywords (author's keywords and Keywords Plus) was created. We applied a weak component clustering to the whole network obtained after dropping the keywords with a frequency below five and the links with a co-occurrence value below three (the whole network is the same as the generated by CoPalRed). The biggest weak

component is shown in Figure 5. The size of the nodes is proportional to the respective keyword's frequency, and the size of the lines represents the co-occurrence (without normalization) of the linked nodes. Only the names of the top 50 keywords are shown. The color of the nodes varies in a linear way from gray to black according to their frequency, and the color of the links varies from green to black according to their co-occurrence value. The network was laid out using the GUESS plugin.

Fifth, a Factor Map was built by VantagePoint (Figure 6) using those keywords with a frequency equal to or higher than five (after this pruning the dataset contains 392 keywords). Each node represents a cluster of terms. The label of each theme was chosen selecting its most important term. The size of nodes is proportional to the number of documents, and the line between nodes represents the similarity (Pearson's *r*) between factors.

Finally, the co-occurrence matrix generated by CoPalRed was transferred to the VOSViewer format to visualize the results of a co-word analysis. In Figure 7, the cluster view is shown. We can observe how the different keywords are

<sup>5</sup>In this example, Sci<sup>2</sup>Tool and NWB Tool obtained the same results.

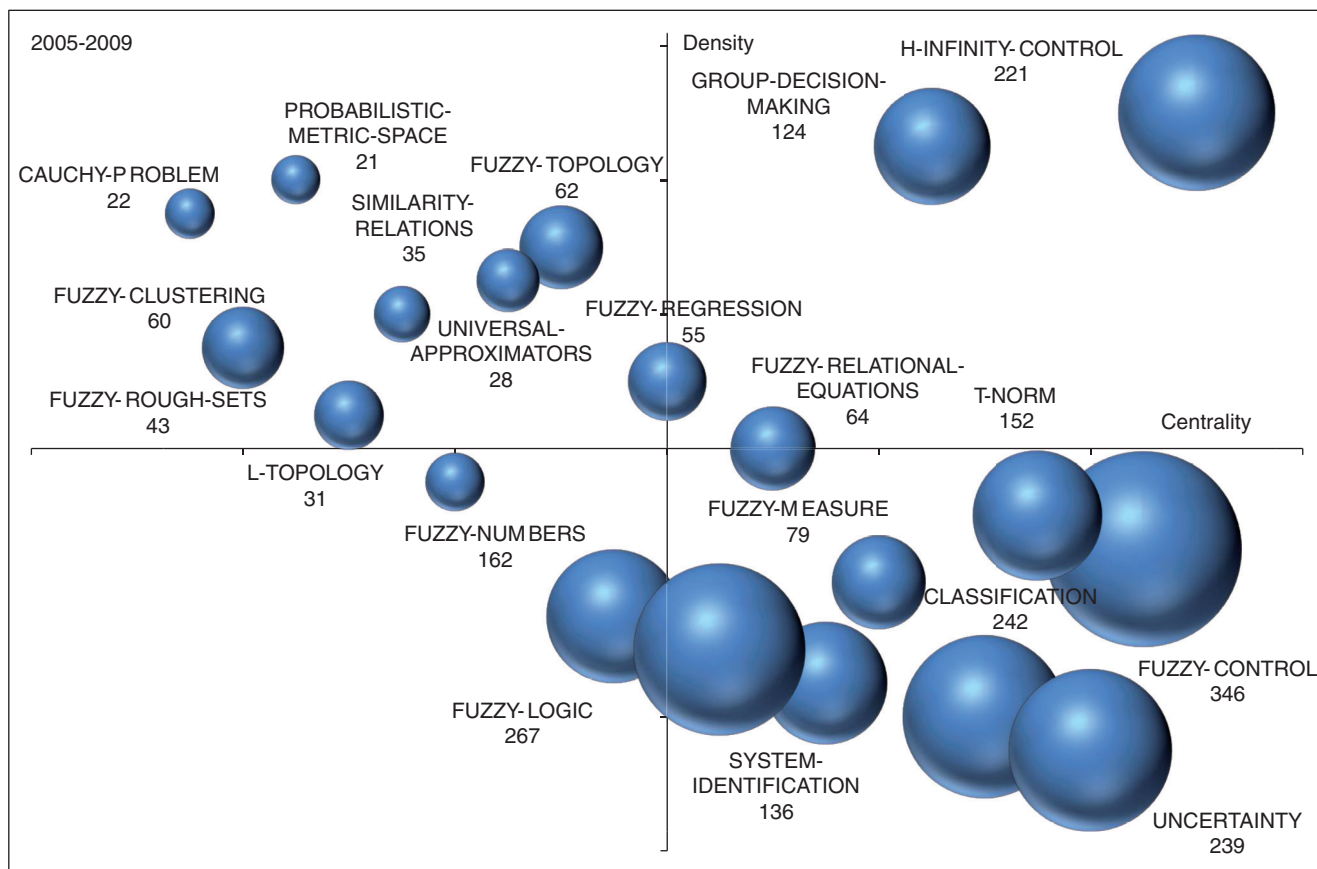


FIG. 2a. CoPalRed's results—(a) Strategic diagram. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

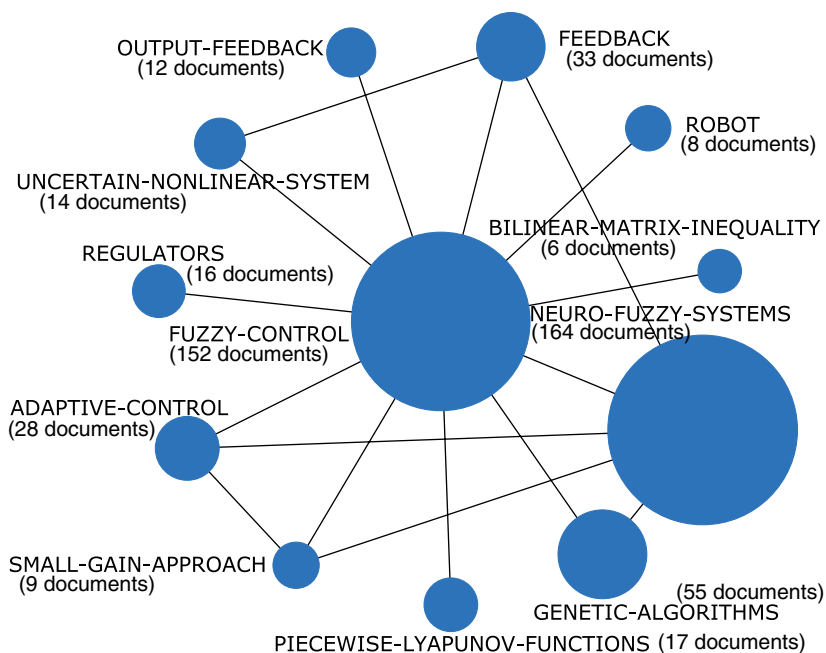


FIG. 2b. CoPalRed's results—(b) Thematic network. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

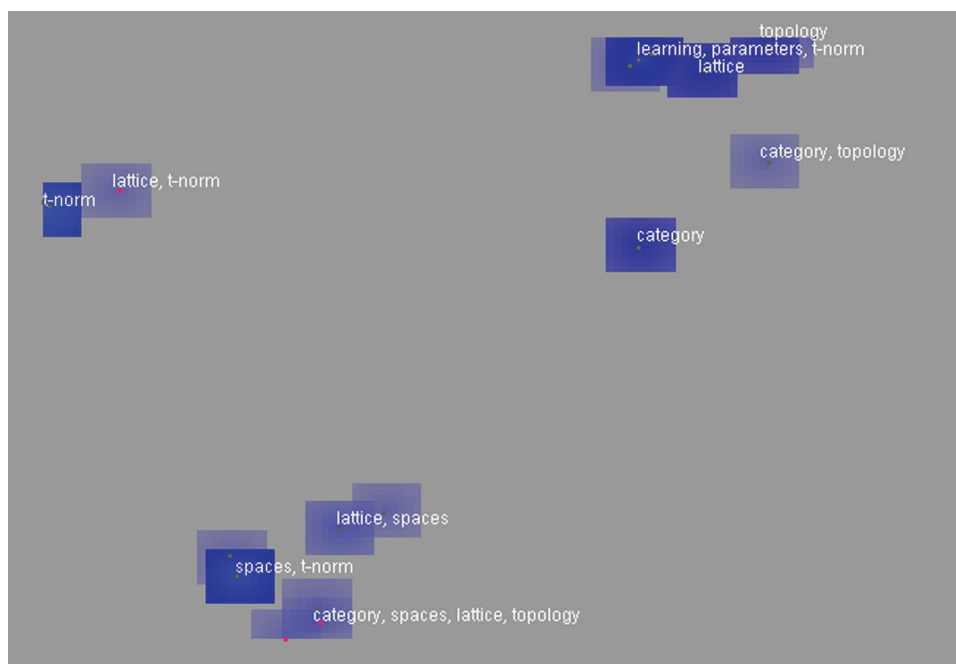


FIG. 3. IN-SPIRE's Galaxy view. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

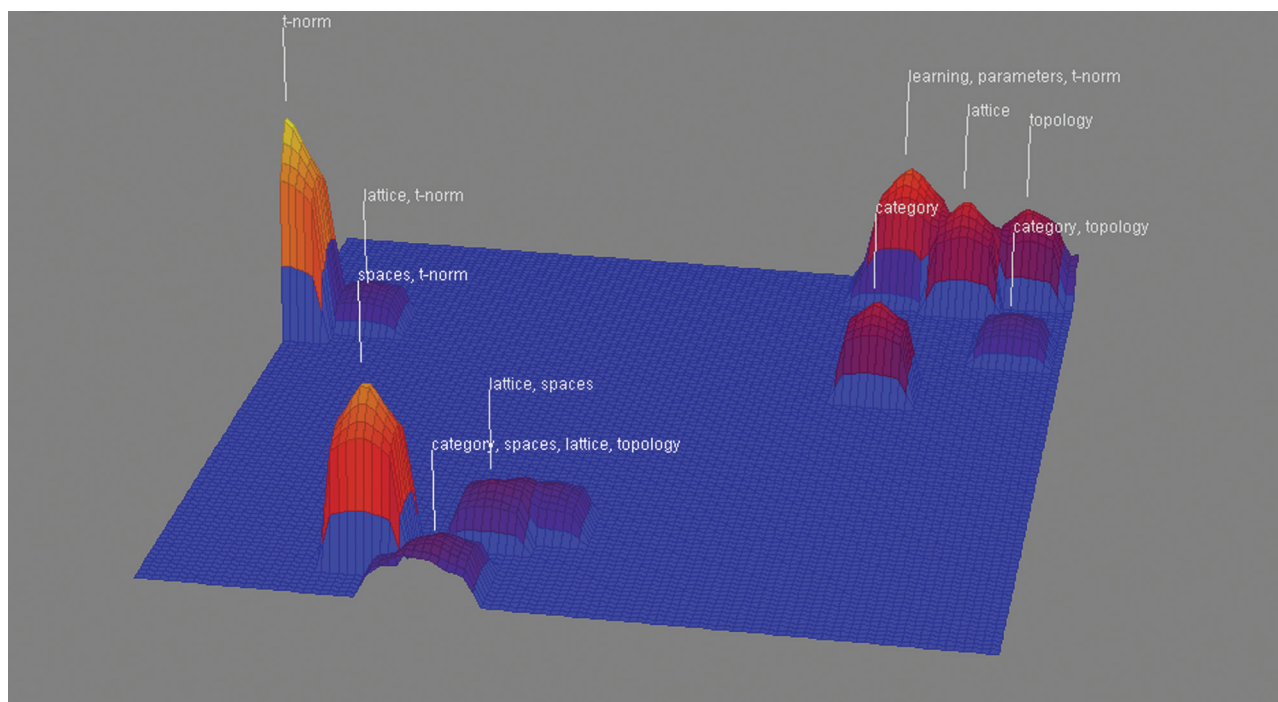


FIG. 4. IN-SPIRE's Theme view. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

laid out over a horizontal line. This means that the keywords placed on the left are very dissimilar to those placed on the right side of the maps. The size of the keywords' labels is proportional to their frequency, VOSViewer visualizes only the labels of the most important ones (most frequent) in the higher zoomed view. VOSViewer selects a random different color for each cluster. Inside each cluster, the strength of a

color at one point represents the density of this point. The density is measured using a Gaussian kernel function (van Eck & Waltman, 2010).

Similarly to CiteSpace, the printed out map of VOSViewer does not show the power of its graphic user interface. In each view, the user can zoom in on a specific area to discover the items hidden under the most important ones. As an



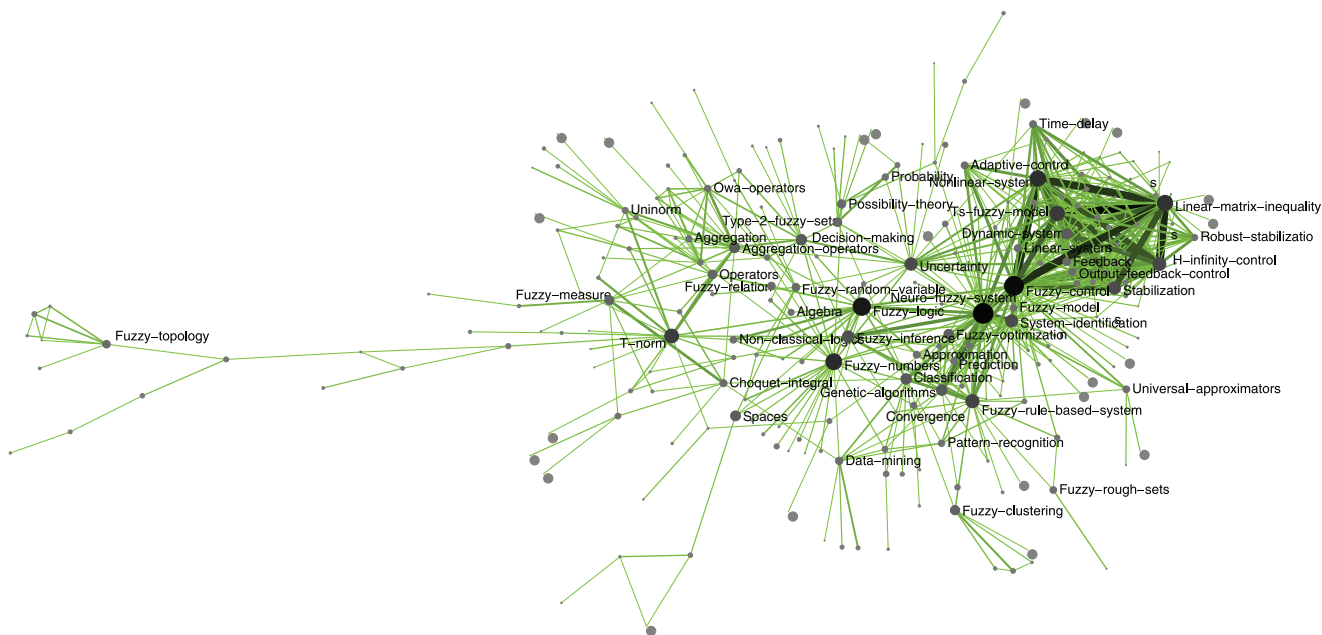


FIG. 5. Map generated by Science of Science (Sci<sup>2</sup>) Tool. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

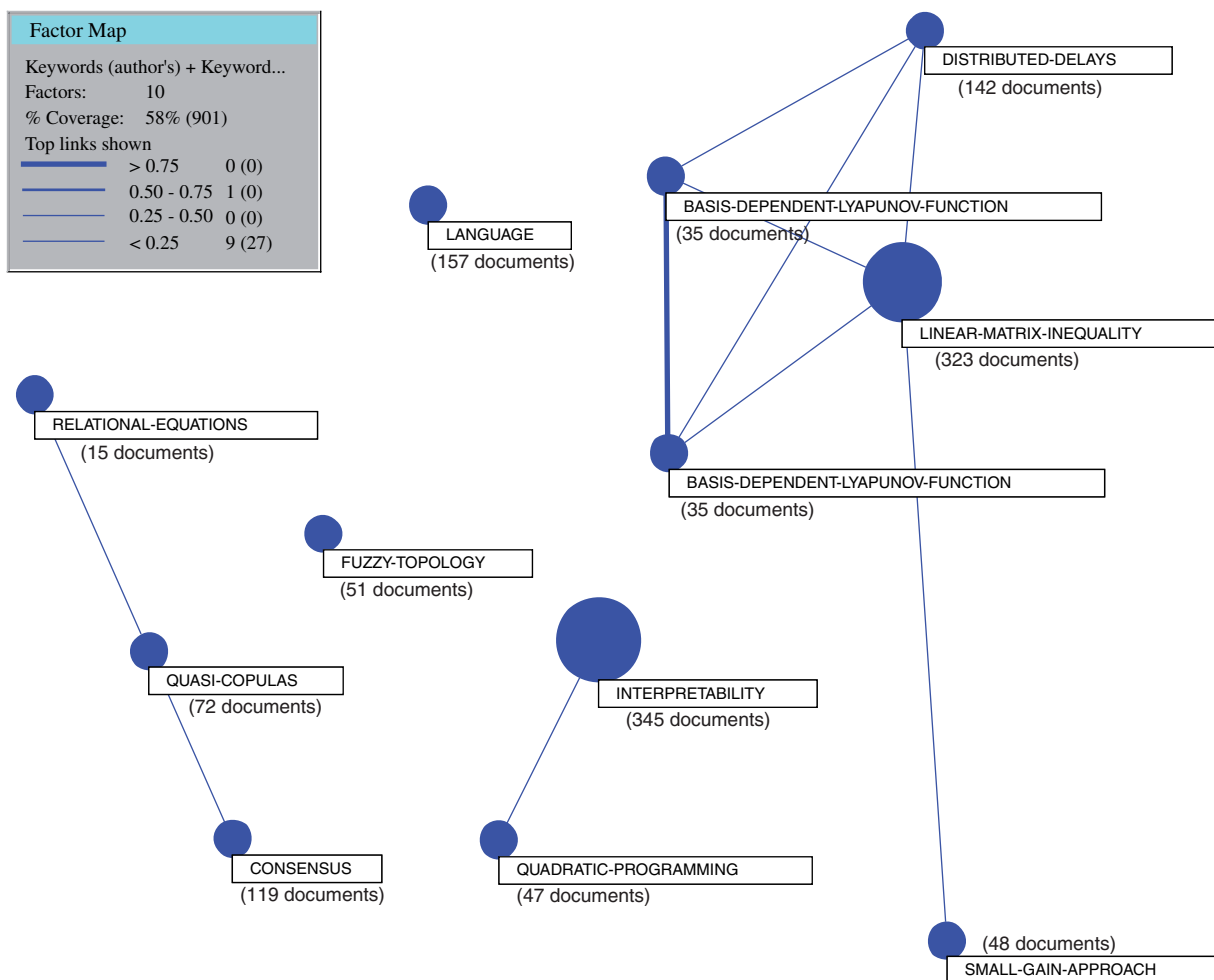


FIG. 6. Map generated by VantagePoint. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]



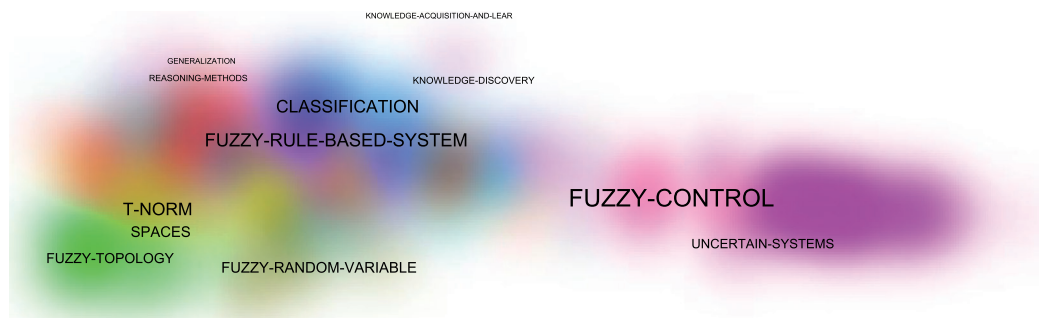


FIG. 7. VOSViewer's cluster view. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

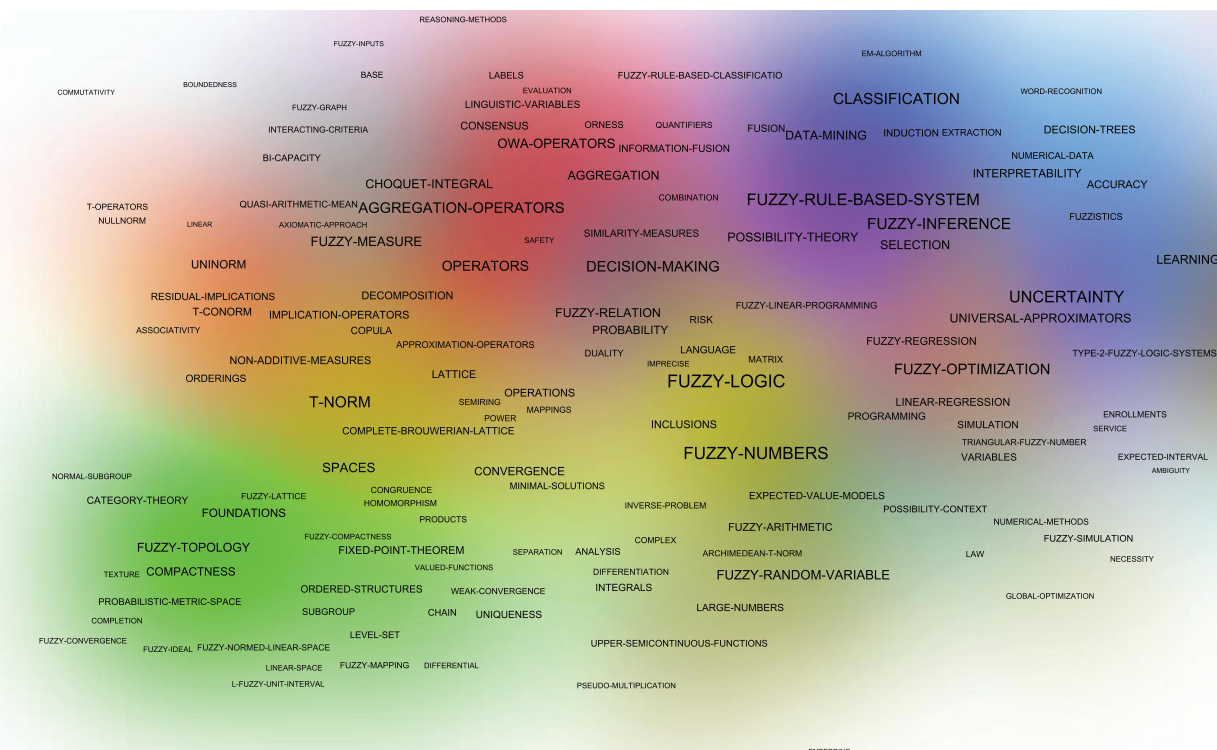


FIG. 8. VOSViewer's cluster zoom-in view. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://wileyonlinelibrary.com).]

example, in Figure 8, an enlarged, zoom-in view of the cluster visualization focused on the keywords *FUZZY-TOPOLOGY* and *T-NORM* is shown.

## Lessons Learned

As has been shown in both the Comparative Study section and the Analysis of Generated Maps: A Cooperative Study among Tools section, each software tool has different characteristics. Several software tools have powerful preprocessing techniques, others allow the generation of a high quantity of bibliometric networks, and others are focused only on one kind of bibliometric network. Finally, not all the processes of analysis are available in each software tool. For this reason, a deep science mapping analysis requires the use of different tools.

The preprocessing capabilities of VantagePoint are one of its main strengths. It incorporates a high quantity of import filters that allows us to load data from almost all the bibliographical sources. Moreover, the clean-up list method and the possibility of applying a thesaurus to carry out this task, helps the preprocessing task, especially the de-duplicating process. Vantage-Point allows us to export the results into a csv file, so other software tools can read this data to perform their own science mapping analysis over the preprocessed data.

CoPalRed has a good de-duplicating process too, but it is focused only on one kind of unit, the keywords. NWB Tool and Sci<sup>2</sup> Tool have a de-duplicating module, but this needs an external process to be performed using external software. However, both NWB Tool and Sci<sup>2</sup> Tool have a good network reduction process.

TABLE 7. Characteristics summary.

Software tool	Preprocessing	Networks	Normalization	Analysis
Bibexcel	Data and networks reduction	DBCA, ACAA, CCAA, ICAA, ACA, DCA, JCA, CWA, Others	Salton's Cosine, Jaccard's Index, or the Vladutz and Cook measures	Network
CiteSpace	Time slicing and data and networks reduction	DBCA, ACAA, CCAA, ICAA, ACA, DCA, JCA, CWA, Others	Salton's Cosine, Dice or Jaccard Strength	Burst detection, geospatial, network, temporal
CoPalRed	De-duplication, Time slicing, data reduction	CWA	Equivalence index	Network, temporal
IN-SPIRE	Data reduction	CWA	Conditional probability	Burst detection, network, temporal
Leydesdorff's Software		ABCA, JBCA, ACAA, CCAA, ICAA, ACA, CWA	Salton's Cosine	
Network Workbench Tool	De-duplication, time slicing and data and networks reduction	DBCA, ACAA, DCA, CWA, DL	User defined	Burst detection, network, temporal
Science of Science Tool	De-duplication, time slicing and data and networks reduction	ABCA, DBCA, JBCA, ACAA, ACA, DCA, JCA, CWA, DL, Others	User defined	Burst detection, geospatial, network, temporal
VantagePoint	De-duplication, time slicing and data reduction	ACAA, CCAA, ICAA, ACA, DCA, JCA, CWA, Others	Pearson's r, Salton's Cosine or the Max Proportional	Burst detection, geospatial, network, temporal
VOSViewer			Association Strength	Network

The software tools allow us to generate various kinds of bibliometric networks, but as was shown in Table 4, there is no single software tool able to extract all of them.

Taking into account the maps and visualizations generated by each software tool, as shown in the Analysis of Generated Maps: A Cooperative Study among Tools section, there are several differences between them:

- CiteSpace is able to visualize the networks using different layouts. The name of the detected clusters can be assigned using different metrics. Finally, the user graphic-interface allows us to interact with the network to carry out a good exploration of it.
- CoPalRed groups the items (keywords) under themes, and they are categorized in a strategic diagram according to their centrality and density. This categorization allows us to detect the *motor themes* of the field. For each theme, CoPalRed generates a thematic network where the relation between its keywords is shown.
- IN-SPIRE allows us to visualize two kinds of map, if sufficient data are provided. In the Theme view, the analyst can detect the most important zones of the map (where more documents are localized). The Galaxy view allows us to easily detect similar documents based on their content.
- NWB Tool and Sci<sup>2</sup> Tool generate similar visualizations. They allow us to visualize the networks using different plugins and applying different layouts and scripts to customize the view. Sci<sup>2</sup> Tool incorporates thematic maps where the information is shown over a world map.
- VantagePoint has three kinds of map that allow several views to be created. In the map view, VantagePoint shows a legend that explains the size of the lines, being the only software that produces this legend. Maybe one strength of this software tool is the user graphic-interface that allows the user to select a set of items from the map, whereupon it shows the documents associated with these items and other information in the detail window.

- VOSViewer has a powerful user graphic-interface that allows us to examine the generated maps easily. Detecting (in a visual way) the most important themes is not always easy, and in the cluster view it is difficult to say to which cluster the keywords that are between two clusters (borderline keywords) belong.

According to the methods of analysis available there are differences between the software tools; for example, the geospatial analysis is available only in CiteSpace, Sci<sup>2</sup> Tool, and VantagePoint, and only the first two have geocoding capabilities that allow us to represent the network over a world map (using Google Maps or Yahoo! Maps).

In Table 7, we show a short summary of their characteristics according to the four aspects considered in the analysis developed. As we can observe, the software tools CiteSpace, IN-SPIRE, NWB Tool, Sci<sup>2</sup> Tool, and VantagePoint could be identified as the more complete ones.

We should point out that NWB Tool and Sci<sup>2</sup> Tool have a great deal in common because they share algorithms and have several algorithms in common. NWB Tool is a network analysis, modelling, and visualization toolkit, whereas Sci<sup>2</sup> Tool is a modular toolset specifically designed to perform the study of science, but the Cyberinfrastructure for Network Science Center has developed them both and they share several algorithms and methods. Nevertheless, some capabilities such as geocoding are unique to Sci<sup>2</sup> Tool.

It is sometimes difficult to import a specific dataset into the nine described software tools. At other times, it is difficult to modify them or incorporate new measures, algorithms, and visualizations. For this reason, extension capabilities such as those provided by NWB Tool, Sci<sup>2</sup> Tool, and VantagePoint are very useful.

As mentioned above, each software tool has different characteristics and implements different techniques that are carried out with different algorithms. Consequently, each

software tool gives its particular view of the studied field. The combined use of different science mapping software tools could allow us to develop a complete science mapping analysis. Therefore, we think that the cooperation among tools could generate a positive synergy that would give us the possibility of extracting unknown knowledge that would otherwise remain undiscovered.

## Concluding Remarks

An analysis of science mapping software tools has been carried out. Specifically, we have analyzed nine representative science mapping software tools: Bibexcel, CiteSpace II, CoPalRed, IN-SPIRE, Leydesdorff's Software, Network Workbench Tool, Sci<sup>2</sup>Tool, VantagePoint, and VOSViewer.

These software tools present different characteristics; for example, some of them are focused only on visualization and others have different preprocessing modules. There is not one software tool that we could consider the best one. Consequently, we think that a complete science mapping analysis of a particular field should be made using a variety of these software tools to gather all the important knowledge and different perspectives; for example, the preprocessing step is very important and the analyst has to use a software tool to help carry out this task. Not all the software tools are able to extract all the bibliometric networks, and, so, different tools have to be used to analyze a field from different perspectives (intellectual, social, or conceptual). The software tools have different analysis methods (although some of them are common), which allow the analyst to discover different knowledge. Finally, because the visualizations are different in each one, different views of the field can be generated and these help to interpret and analyze the results. This cooperation among tools gives a positive synergy, which allows us to extract the knowledge hidden behind the data.

Considering the results obtained in the Analysis of Generated Maps: A Cooperative Study among Tools section, where co-word was the unique analysis technique used to analyze the FST field, and the positive synergies of using several software tools drawn in the Lessons Learned section, we think that a thorough analysis of any field could be carried out using the powerful of each tool. So, for example, a co-word analysis performed by CoPalRed could be complemented by IN-SPIRE using the terms extracted from abstracts and titles. Moreover, IN-SPIRE could show the conceptual changes over time using its Time tools. In addition, CiteSpace and Sci<sup>2</sup> could perform an intellectual and social analysis. CiteSpace could be used for a document co-citation analysis and Sci<sup>2</sup> for a co-author analysis. The resulting network of authors could be displayed over a world map using the geolocation capabilities of Sci<sup>2</sup>. Finally, VantagePoint could be used to build a factor map on keywords, and show the institutional affiliation related to the most interesting detected factors.

We should point out that this study does not incorporate all of the science mapping software tools used around the world. This is because researchers usually use their own ad hoc software tools and algorithms, perhaps motivated by the

lack of flexibility of available software tools. Although these could have similar characteristics to the ones presented here, they remain unpublished. Sometimes, these tools are unpublished because they were developed to perform a specific and ad hoc analysis and these developed pieces of software remained in the background.

## Acknowledgments

This work has been developed with the support of Project 90/07 (Ministry of Public Works and Transport) and Excellence Andalusian Project TIC5299. We thank the Pacific Northwest National Laboratory and Search Technology Inc. for providing us a trial version of IN-SPIRE and VantagePoint, respectively, and for their help and support.

## References

- Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval*. Boston, MA: Addison-Wesley.
- Bailón-Moreno, R., Jurado-Alameda, E., & Ruíz-Baños, R. (2006). The scientific network of surfactants: Structural analysis. *Journal of the American Society for Information Science and Technology*, 57(7), 949–960.
- Bailón-Moreno, R., Jurado-Alameda, E., Ruíz-Baños, R., & Courtial, J.P. (2005). Analysis of the scientific field of physical chemistry of surfactants with the unified scientometric model. Fit of relational and activity indicators. *Scientometrics*, 63(2), 259–276.
- Bar-Ilan, J. (2010). Citations to the “introduction to informetrics” indexed by WOS, Scopus and Google Scholar. *Scientometrics*, 82(3), 495–506.
- Batagelj, V., & Mrvar, A. (1998). Pajek-Program for large network analysis. *Connections*, 21(2), 47–57.
- Batty, M. (2003). The geography of scientific citation. *Environment and Planning A*, 35(5), 761–765.
- Borgatti, S.P., Everett, M.G., & Freeman, L.C. (2002). *Ucinet 6 for Windows: Software for social network analysis, analytic technologies*, Harvard, MA.
- Börner, K., Chen, C., & Boyack, K. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology*, 37, 179–255.
- Börner, K., Huang, W., Linnemeier, M., Duhon, R., Phillips, P., Ma, N., & Price, M. (2010). Rete-netzwerk-red: Analyzing and visualizing scholarly networks using the network workbench tool. *Scientometrics*, 83(3), 863–876.
- Boyack, K.W., Wylie, B.N., & Davidson, G.S. (2002). Domain visualization using VxInsight for science and technology management. *Journal of the American Society for Information Science and Technology*, 53(9), 764–774.
- Callon, M., Courtial, J., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research—The case of polymer chemistry. *Scientometrics*, 22(1), 155–205.
- Callon, M., Courtial, J.P., Turner, W.A., & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, 22(2), 191–235.
- Carrington, P.J., Scott, J., & Wasserman, S. (2005). *Models and methods in social network analysis. Structural Analysis in the Social Sciences*. New York: Cambridge University Press.
- Chen, C. (2004). Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Science of the United States of America (PNAS)*, 101(suppl. 1), 5303–5310.
- Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57(3), 359–377.
- Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010). The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. *Journal*

- of the American Society for Information Science and Technology, 61(7), 1386–1409.
- Chen, P., & Redner, S. (2010). Community structure of the physical review citation network. *Journal of Informetrics*, 4(3), 278–290.
- Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E., & Herrera, F. (2011). An approach for detecting, quantifying, and visualizing the evolution of a research field: A practical application to the fuzzy sets theory field. *Journal of Informetrics*, 5(1), 146–166.
- Cook, D.J., & Holder, L.B. (2006). *Mining graph data*. Hoboken, NJ: John Wiley & Sons, Inc.
- Coulter, N., Monarch, I., & Konda, S. (1998). Software engineering as seen through its research literature: A study in co-word analysis. *Journal of the American Society for Information Science*, 49(13), 1206–1223.
- Davidson, G.S., Hendrickson, B., Johnson, D.K., Meyers, C.E., & Wylie, B.N. (1998). Knowledge mining with vxinsight: Discovery through interaction. *Journal of Intelligent Information Systems*, 11(3), 259–285.
- Davidson, G.S., Wylie, B.N., & Boyack, K.W. (2001). Cluster stability and the use of noise in interpretation of clustering. *Proceedings of the IEEE Symposium on Information Visualization* (pp. 23–30). doi: 10.1109/INFVIS.2001.963275
- Fabrikant, S.I., Montello, D., & Mark, D.M. (2010). The natural landscape metaphor in information visualization: The role of commonsense geomorphology. *Journal of the American Society for Information Science and Technology*, 61(2), 253–270.
- Falagas, M.E., Pitsouni, E.I., Malietzis, G.A., & Pappas, G. (2008). Comparison of Pubmed, Scopus, Web of Science, and Google Scholar: Strengths and weaknesses. *FASEB Journal*, 22(2), 338–342.
- Gänzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics*, 51(1), 69–115.
- Gao, X., & Guan, J. (2009). Networks of scientific journals: An exploration of Chinese patent data. *Scientometrics*, 80(1), 283–302.
- Garfield, E. (1994). *Scientography: Mapping the tracks of science*. Current Contents: Social & Behavioural Sciences, 7(45), 5–10.
- Havre, S., Hetzler, E., Whitney, P., & Nowell, L. (2002). Themeriver: Visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics*, 8(1), 9–20.
- Herr, B., Huang, W., Penumarthy, S., & Börner, K. (2007). Designing highly flexible and usable cyberinfrastructures for convergence. In W.S. Bainbridge & M.C. Roco (Eds.), *Progress in convergence: Technologies for human wellbeing* (Vol. 1093, pp. 161–179). Boston: Annals of the New York Academy of Sciences.
- Hetzler, E., & Turner, A. (2005). Analysis experiences using information visualization. *IEEE Computer Graphic and Applications*, 24(5), 22–26.
- Jarneving, B. (2005). A comparison of two bibliometric methods for mapping of the research front. *Scientometrics*, 65(2), 245–263.
- Kandylas, V., Upham, S.P., & Ungar, L.H. (2010). Analyzing knowledge communities using foreground and background clusters. *ACM Transactions on Knowledge Discovery from Data*, 4(2), art. no. 7.
- Kessler, M.M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14(1), 10–25.
- Kleinberg, J. (2003). Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery*, 7(4), 373–397.
- Leydesdorff, L., & Persson, O. (2010). Mapping the geography of science: Distribution patterns and networks of relations among cities and institutes. *Journal of the American Society for Information Science and Technology*, 61(8), 1622–1634.
- Leydesdorff, L., & Schank, T. (2008). Dynamic animations of journal maps: Indicators of structural changes and interdisciplinary developments. *Journal of the American Society for Information Science and Technology*, 59(11), 1810–1818.
- López-Herrera, A.G., Cobo, M.J., Herrera-Viedma, E., & Herrera, F. (2010). A bibliometric study about the research based on hybridating the fuzzy logic field and the other computational intelligent techniques: A visual approach. *Internacional Journal of Hybrid Intelligent Systems*, 17(7), 17–32.
- López-Herrera, A.G., Cobo, M.J., Herrera-Viedma, E., Herrera, F., Bailón-Moreno, R., & Jimenez-Contreras, E. (2009). Visualization and evolution of the scientific structure of fuzzy sets research in Spain. *Information Research*, 14(4), paper 421.
- McCain, K. (1991). Mapping economics through the journal literature: An experiment in journal co-citation analysis. *Journal of the American Society for Information Science*, 42(4), 290–296.
- Mikki, S. (2010). Comparing Google Scholar and ISI Web of Science for earth sciences. *Scientometrics*, 82(2), 321–331.
- Morel, C.M., Serruya, S., Penna, G., & Guimarães, R. (2009). Co-authorship network analysis: A powerful tool for strategic planning of research, development and capacity building programs on neglected diseases. *PLoS Neglected Tropical Diseases*, 3(8), art. no. e501.
- Morris, S., & Van Der Veer Martens, B. (2008). Mapping research specialties. *Annual Review of Information Science and Technology*, 42(1), 213–295.
- Moya-Anegón, F., Vargas-Quesada, B., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., Herrero-Solana, V., & Muñoz-Fernández, F. (2005). Domain analysis and information retrieval through the construction of heliocentric maps based on ISI-JCR category cocitation. *Information Processing & Management*, 41(6), 1520–1533.
- Moya-Anegón, F., Vargas-Quesada, B., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., Muñoz-Fernández, F., & Herrero-Solana, V. (2007). Visualizing the marrow of science. *Journal of the American Society for Information Science and Technology*, 58(14), 2167–2179.
- Noyons, E.C.M., Moed, H.F., & Luwel, M. (1999a). Combining mapping and citation analysis for evaluative bibliometric purposes: A bibliometric study. *Journal of the American Society for Information Science*, 50(2), 115–131.
- Noyons, E.C.M., Moed, H.F., & van Raan, A.F.J. (1999b). Integrating research performance analysis and science mapping. *Scientometrics*, 46(3), 591–604.
- Persson, O., Danell, R., & Wiborg Schneider, J. (2009). How to use Bibexcel for various types of bibliometric analysis. In F. Åström, R. Danell, B. Larsen, & J. Wiborg Schneider (Eds.), *Celebrating scholarly communication studies: A festschrift for Olle Persson at his 60th birthday* (Vol. 5, pp. 9–24). Leuven, Belgium: International Society for Scientometrics and Informetrics.
- Peters, H.P.F., & van Raan, A.F.J. (1991). Structuring scientific activities by co-author analysis an exercise on a university faculty level. *Scientometrics*, 20(1), 235–255.
- Peters, H.P.F., & van Raan, A.F.J. (1993). Co-word-based science maps of chemical engineering. Part I: Representations by direct multidimensional scaling. *Research Policy*, 22(1), 23–45.
- Polanco, X., François, C., & Lamirel, J.C. (2001). Using artificial neural networks for mapping of science and technology: A multi-self-organizingmaps approach. *Scientometrics*, 51(1), 267–292.
- Porter, A.L., & Cunningham, S.W. (2004). *Tech mining: exploiting new technologies for competitive advantage*. Hoboken, NJ: John Wiley & Sons, Inc.
- Porter, A.L., & Youtie, J. (2009a). How interdisciplinary is nanotechnology? *Journal of Nanoparticle Research*, 11(5), 1023–1041.
- Porter, A.L., & Youtie, J. (2009b). Where does nanotechnology belong in the map of science? *Nature Nanotechnology*, 4, 534–536.
- Price, D., & Gursay, S. (1975). Studies in scientometrics I: Transience and continuance in scientific authorship. *Ci. Informatics Rio de Janeiro*, 4(1), 27–40.
- Quirin, A., Cordón, O., Santamaría, J., Vargas-Quesada, B., & Moya-Anegón, F. (2008). A new variant of the pathfinder algorithm to generate large visual science maps in cubic time. *Information Processing & Management*, 44(4), 1611–1623.
- Rip, A., & Courtial, J. (1984). Co-word maps of biotechnology: An example of cognitive scientometrics. *Scientometrics*, 6(6), 381–400.
- Rosvall, M., & Bergstrom, C.T. (2010). Mapping change in large networks. *PLoS ONE*, 5(1), e8694.
- Salton, G., & McGill, M.J. (1983). *Introduction to modern information retrieval*. New York: McGraw-Hill.
- Schvaneveldt, R.W., Durso, F.T., & Dearholt, D.W. (1989). Network structures in proximity data. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 24, pp. 249–284). New York: Academic Press.

- Sci<sup>2</sup> Team. (2009). Science of Science (Sci<sup>2</sup>) Tool. Indiana University and SciTech Strategies. Retrieved from <http://sci.slis.indiana.edu>
- Shannon, P., Markiel, A., Ozier, O., Baliga, N., Wang, J., Ramage, D., & Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11), 2498–2504.
- Shapira, P., Youtie, J., & Porter, A.L. (2010). The emergence of social science research on nanotechnology. *Scientometrics*, 85(2), 595–611.
- Skillicorn, D. (2007). Understanding complex datasets: Data mining with matrix decompositions (Chapman & Hall/Crc Data Mining and Knowledge Discovery Series). London: Chapman & Hall.
- Skupin, A. (2009). Discrete and continuous conceptualizations of science: Implications for knowledge domain visualization. *Journal of Informetrics*, 3(3), 233–245.
- Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4), 265–269.
- Small, H. (1997). Update on science mapping: Creating large document spaces. *Scientometrics*, 38(2), 275–293.
- Small, H. (1999). Visualizing science by citation mapping. *Journal of the American Society for Information Science*, 50(9), 799–813.
- Small, H. (2006). Tracking and predicting growth areas in science. *Scientometrics*, 68(3), 595–610.
- Small, H., & Garfield, E. (1985). The geography of science: Disciplinary and national mappings. *Journal of Information Science*, 11(4), 147–159.
- Small, H., & Koenig, M.E.D. (1977). Journal clustering using a bibliographic coupling method. *Information Processing and Management*, 13(5), 277–288.
- Small, H., & Upham, S.P. (2009). Citation structure of an emerging research area on the verge of application. *Scientometrics*, 79(2), 365–375.
- Upham, S.P., & Small, H. (2010). Emerging research fronts in science and technology: Patterns of new knowledge development. *Scientometrics*, 83(1), 15–38.
- van Eck, N.J., & Waltman, L. (2007). Bibliometric mapping of the computational intelligence field. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 15(5), 625–645.
- van Eck, N.J., & Waltman, L. (2009). How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology*, 60(8), 1635–1651.
- van Eck, N.J., & Waltman, L. (2010). Software survey: Vosviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538.
- van Eck, N.J., Waltman, L., Dekker, R., & van den Berg, J. (2010). A comparison of two techniques for bibliometric mapping: Multidimensional scaling and vos. *CoRR*, abs/1003.2551.
- Waltman, L., van Eck, N.J., & Noyons, E.C.M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4(4), 629–635.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.
- White, H.D., & Griffith, B.C. (1981). Author co-citation: A literature measure of intellectual structure. *Journal of the American Society for Information Science*, 32, 163–172.
- Wise, J.A. (1999). The ecological approach to text visualization. *Journal of the American Society for Information Science*, 50(13), 1224–1233.
- Zadeh, L. (1965). Fuzzy sets. *Information and Control*, 8(3), 338–353.
- Zadeh, L. (2008). Is there a need for fuzzy logic? *Information Sciences*, 178(13), 2751–2779.
- Zhao, D., & Strotmann, A. (2008). Evolution of research activities and intellectual influences in information science 1996–2005: Introducing author bibliographic-coupling analysis. *Journal of the American Society for Information Science and Technology*, 59(13), 2070–2086.
- Zitt, M., Bassecoulard, E., & Okubo, Y. (2000). Shadows of the past in international cooperation: Collaboration profiles of the top five producers of science. *Scientometrics*, 47(3), 627–657.