SciMAT: A New Science Mapping Analysis Software Tool

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: {mjcobo, lopez-herrera, viedma, herrera}@decsai.ugr.es

This article presents a new open-source software tool, SciMAT, which performs science mapping analysis within a longitudinal framework. It provides different modules that help the analyst to carry out all the steps of the science mapping workflow. In addition, SciMAT presents three key features that are remarkable in respect to other science mapping software tools: (a) a powerful preprocessing module to clean the raw bibliographical data, (b) the use of bibliometric measures to study the impact of each studied element, and (c) a wizard to configure the 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 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, 1999). 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, 2008; Noyons, Moed, & Luwel, 1999).

It is common to find scientific papers and reports that contain a science mapping analysis to show and uncover the hidden key elements (e.g., documents, authors, institutions, topics, etc.) in a specific interest area (Bailón-Moreno, Jurado-Alameda, & Ruíz-Baños, 2006; López-Herrera, Cobo, Herrera-Viedma, & Herrera, 2010; López-Herrera et al., 2009; Porter & Youtie, 2009; van Eck & Waltman, 2007). Some of these works were undertaken for academic purposes and others with competitive animus, such as those related to patent analysis in R&D business departments (Porter & Cunningham, 2004).

Currently, some studies use nonspecific science mapping software (e.g., Pajek, Gephi, or UCINET), and others use specific (and sometimes ad hoc) science mapping software tools (e.g., CoPalRed, Science of Science Tool, or VOSviewer). A list of software tools widely used in the literature can be found in Börner et al. (2010) and Cobo, López-Herrera, Herrera-Viedma, and Herrera (2011b).

In Cobo et al. (2011b), we presented an analysis of the features, advantages, and drawbacks of the different science mapping software tools available. As a result, we concluded that there was no single science mapping software tool powerful and flexible enough to incorporate all the key elements (data retrieval, preprocessing, network extraction, normalization, mapping, analysis, visualization, and interpretation) in any science mapping workflow (Cobo et al., 2011b). Therefore, researchers usually have to use more than one (and sometimes several) software tools to perform a deep science mapping analysis. For example, it is common practice to use an ad hoc software tool to clean the data in the preprocessing stage, then to apply another tool to build the science maps, and sometimes it is necessary to use a third-party software tool to visualize, navigate, and interact with the results.

Bibliometric measures and indicators can be employed to carry out a performance analysis of the generated maps (Cobo, López-Herrera, Herrera-Viedma, & Herrera, 2011a). This kind of analysis allows us to quantify and measure the performance, quality, and impact of the generated maps and their components, as shown in Cobo, López-Herrera, Herrera, and Herrera-Viedma (2012).

In this article, we present a new open-source[1] science mapping software tool called *SciMAT*[2] (Science Mapping Analysis software Tool) which incorporates methods, algorithms, and measures for all the steps in the general science mapping workflow, from preprocessing to the visualization of the results. SciMAT allows the user to carry out studies based on several bibliometric networks (co-word, cocitation,

[1]With license GPLv3 (for more information, see http://www.gnu.org/licenses/gpl-3.0.html).

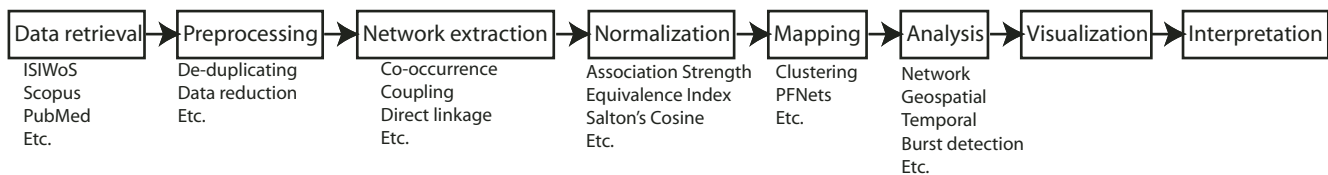[2]Its associated Web site can be seen at http://sci2s.ugr.es/scimat

FIG. 1. Workflow of science mapping.

author cocitation, journal cocitation, coauthor, bibliographic coupling, journal bibliographic coupling, and author bibliographic coupling). Different normalization and similarity measures can be used over the data (association strength, Equivalence Index, Inclusion Index, Jaccard Index, and Salton's cosine). Several clustering algorithms can be chosen to cut up the data. In the visualization module, three representations (strategic diagrams, cluster networks, and evolution areas) are jointly used, which allows the user to better understand the results. Furthermore, SciMAT presents three key features that other science mapping software tools either do not have or have only in limited form:

- *Powerful preprocessing module:* SciMAT implements a wide range of preprocessing tools such as detecting duplicate and misspelled items, time slicing, data reduction, and network preprocessing.
- *Use of bibliometric measures:* SciMAT is based on the science mapping approach presented in Cobo et al. (2011a), which allows us to carry out science mapping studies under a longitudinal framework (Garfield, 1994; Price & Gürsey, 1975) and to build science maps enriched with bibliometric measures. Therefore, SciMAT also implements a wide range of bibliometric measures (mainly based on citations) such as the h-index (Alonso, Cabrerizo, Herrera-Viedma, & Herrera, 2009; Hirsch, 2005), g-index (Egghe, 2006), hg-index (Alonso, Cabrerizo, Herrera-Viedma, & Herrera, 2010), or q^2-index (Cabrerizo, Alonso, Herrera-Viedma, & Herrera, 2010). These bibliometric measures give information about the interest in and impact on the specialized research community of each detected cluster or evolution area.
- *A wizard to configure the analysis:* SciMAT incorporates a wizard that allows the user to configure the different steps of science mapping analysis. In this way, the user can select the measures, algorithms, and analysis techniques used to carry out the science mapping analysis.

This article is organized as follows. The general workflow of a science mapping analysis is described, and some representative software tools developed for this kind of analysis are briefly shown. Next, we describe SciMAT, together with its main characteristics, functionalities, architecture, and technologies used. We then present some possible scenarios where SciMAT could be employed and conclude this section with a real-use case, showing the flow of steps needed to carry out the analysis with SciMAT, and some conclusions that an analyst could draw from interpretation of the results. The results of a formal

validation of SciMAT are shown, and some concluding remarks are made.

Foundations of Science Mapping

In this section, different aspects of the science mapping analysis are described. First, the general workflow in a science mapping analysis is shown, describing each step. Second, a brief review of science mapping software tools is carried out.

The Workflow of Science Mapping

The general workflow in a science mapping analysis has different steps (Börner et al., 2003; Cobo et al., 2011b) (see Figure 1): data retrieval, data preprocessing, network extraction, network normalization, mapping, analysis, and visualization. At the end of this process, the analyst has to interpret and obtain conclusions from the results.

Nowadays, there are several online bibliographic (and also bibliometric) databases where the data can be retrieved, with the ISI Web of Science[3] (ISIWoS), Scopus,[4] Google Scholar,[5] and the National Library of Medicine's MEDLINE[6] being the most important. Moreover, a science mapping analysis can be made using patents (e.g., by downloading the data from the U.S. Patent and Trademark Office[7] or from the European Patent Office[8]) or funding data (e.g., from the National Science Foundation[9]).

Usually, the data retrieved from the bibliographic sources contain errors, so a preprocessing process must be applied first. In fact, the preprocessing step is one of the most important to obtain good results in science mapping analysis. Different preprocessing processes can be applied to the raw data, such as detecting duplicate and misspelled items, time slicing, data reduction, and network reduction (for more information, see Cobo et al., 2011b). The data reduction is carried out to select the most representative data for the analysis, so it is performed after the de-duplicating process.

---

[3]http://www.webofknowledge.com
[4]http://www.scopus.com
[5]http://scholar.google.com
[6]http://www.ncbi.nlm.nih.gov/pubmed
[7]http://www.uspto.gov/
[8]http://www.epo.org/
[9]http://www.nsf.gov/

Once the data has been preprocessed, a network is built using a unit of analysis, with journals, documents, cited references (full reference, author's reference, or source reference can be used), authors (author's affiliation also can be used), and descriptive terms or words (Börner et al., 2003) being the most common.

Several relations among the units of analysis can be established, such as co-occurrence, coupling, or direct linkage. A co-occurrence relation is established between two units (authors, terms, or references) when they appear together in a set of documents; that is, when they co-occur throughout the corpus. A coupling relation is established between two documents when they have a set of units (authors, terms, or references) in common. Furthermore, the coupling can be established using a higher level unit of aggregation, such as authors or journals. That is, a coupling between two authors or journals can be established by counting the units shared by their documents (using the author's or journal's oeuvres). Finally, a direct linkage establishes a relation between documents and references, particularly a citation relation.

These relations can be represented as a graph or network, where the units are the nodes, and the relations between them represent an edge between two nodes. In the co-occurrence relation, the nodes can be authors, terms, or references whereas in the coupling relation, the nodes are documents, and in the aggregated coupling relation, the nodes can be authors or journals. (Other units can be selected as aggregation data.)

In addition, different aspects of a research field can be analyzed depending on the units of analysis used and the kind of relation selected (Cobo et al., 2011b). For example, using the authors, a coauthor or coauthorship analysis can be performed to study the social structure of a scientific field (Gänzel, 2001; Peters & van Raan, 1991). Using terms or words, a co-word (Callon, Courtial, Turner, & Bauin, 1983) analysis can be performed to show the conceptual structure and the main concepts dealt with by a field. Cocitation (Small, 1973) and bibliographic coupling (Kessler, 1963) are used to analyze the intellectual structure of a scientific research field. A description of these techniques and networks can be found in Cobo et al. (2011b).

When the network of relationships between the selected units of analysis has been built, a normalization process is needed (van Eck & Waltman, 2009). Different measures have been used in the literature to normalize a bibliometric network: Salton's cosine (Salton & McGill, 1983), Jaccard Index (Peters & van Raan, 1993), Equivalence Index (Callon, Courtial, & Laville, 1991), and association strength (Coulter, Monarch, & Konda, 1998; van Eck & Waltman, 2007).

Once the normalization process is finished, we can apply different techniques to build the science map. Dimensionality reduction techniques such as principal component analysis or multidimensional scaling, clustering algorithms (Chen et al., 2010; Chen & Redner, 2010; Coulter et al., 1998; Kandylas, Upham, & Ungar, 2010; Rosvall &

Bergstrom, 2010; Small and Sweeney, 1985), and Pathfinder networks (PFNETs) (Quirin, Cordón, Santamaría, Vargas-Quesada, & Moya-Anegón, 2008; Schvaneveldt, Durso, & Dearholt, 1989) have been widely used (Börner et al., 2003).

Analysis methods for science mapping allow us to discover useful knowledge from data, networks, and maps (Cobo et al., 2011b). There are different analysis methods, such as *network analysis* (Carrington, Scott, & Wasserman, 2005; Cook & Holder, 2006; Skillicorn, 2007; Wasserman & Faust, 1994), *temporal* or *longitudinal analysis* (Garfield, 1994; Price & Gürsey, 1975), *geospatial analysis* (Batty, 2003; Leydesdorff & Persson, 2010; Small & Garfield, 1985), *performance analysis* (Cobo et al., 2011a), and so on. Each kind of analysis allows us to discover different views and knowledge. In addition, these analyses can be applied over the maps or directly over the networks. For example, the network analysis can measure the centrality of a given node on the whole network, or the centrality of a cluster (if a cluster algorithm was applied to build the map) on the map. The results of the analysis methods can even be used to build the map. In this sense, the geospatial analysis can help to lay out the elements over a geographical map. Similarly, the network analysis can be used to lay out the map elements according to certain network measures.

As described in Cobo et al. (2011a), the performance analysis uses bibliometric measures and indicators (based on citations), such as the h-index (Alonso et al., 2009; Hirsch, 2005), g-index (Egghe, 2006), hg-index (Alonso et al., 2010), or q^2-index (Cabrerizo et al., 2010) to quantify the importance, impact, and quality of the different elements of the maps (e.g., clusters), and also of the network. For this reason, a set of documents has to be added to each element of the whole network and map.

In a bibliometric network, each node (unit of analysis) could have an associated set of documents. With this set of documents, a performance analysis could be carried out. For example, we could calculate the amount of documents associated with a node, the citations achieved by those documents, the h-index, and so on.

As mentioned earlier, the whole network is usually split into subnetworks or clusters. To obtain performance indicators of each subnetwork, we need a list/set of documents associated to the whole subnetwork. So, we need to aggregate the set of documents of all nodes in the subnetwork into a single one. To do that, an aggregation function, called *document mapper* in this article, has to be defined.

Different document mapper functions can be defined:

- *Core document mapper:* returns the documents that are present in at least two nodes (Cobo et al., 2011a).
- *Secondary document mapper:* returns the documents that are present in only one node (Cobo et al., 2011a).
- *k-core document mapper:* is a generalization of the core document mapper. It returns the documents that are present in at least *k* nodes.
- *Union (∪) document mapper:* returns the algebraic union of the set of documents associated with the subset of nodes.
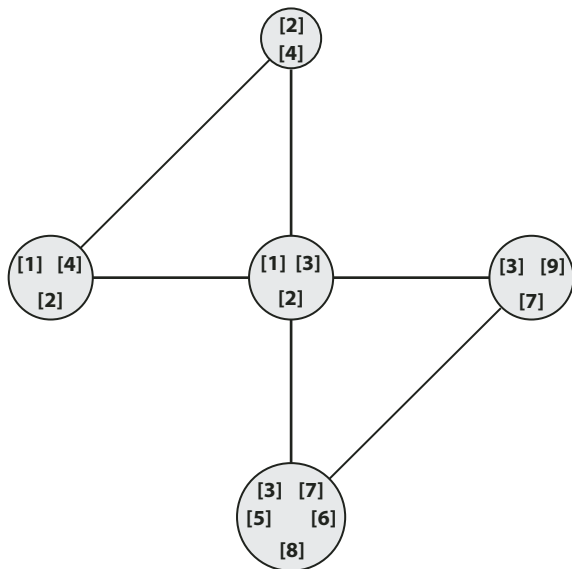
FIG. 2.   A subnetwork with the documents associated with each node.

TABLE 1.   Results returned by different document mappers.

Document mapper function	Documents returned
Core	[1], [2], [3], [4], [7]
Secondary	[5], [6], [8], [9]
k-core ($k = 3$)	[2], [3]
Union	[1], [2], [3], [4], [5], [6], [7], [8], [9]
Intersection	ϕ

- *Intersection (∩) document mapper:* returns the algebraic intersection of the set of documents associated with the subset of nodes.

As a visual example, suppose the subnetwork shown in Figure 2. Each sphere represents a node, and its associated documents are placed inside. If we want to assign a set of document to the subnetwork, a document mapper function must be applied. The result of the five aforementioned document mappers are shown in Table 1.

Following the science mapping workflow, the visualization techniques are used to represent a science map and the result of the different analyses. The visualization technique employed is very important to allow a good understanding and better interpretation of the output. The network results from the mapping step can be represented with different visualization tools such as, for example, *heliocentric maps* (Moya-Anegón et al., 2005), *geometrical models* (Skupin, 2009), *thematic networks* (Bailón-Moreno et al., 2006; Cobo et al., 2011a), or maps where the proximity between items represents their similarity (Davidson, Hendrickson, Johnson, Meyers, & Wylie, 1998; Fabrikant, Montello, & Mark, 2010; Polanco, François, & Lamirel, 2001; van Eck &

Waltman, 2010). The clusters detected in a network can be categorized using a *strategic diagram* (Callon et al., 1991; Cobo et al., 2011a). To show the evolution of detected clusters in successive time periods (temporal or longitudinal analysis), different visualization 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., 2011a). Furthermore, visualization can be improved using the results of a performance analysis, which allows us to add a third dimension to the visualized elements. For example, the strategic diagram can show spheres whose volume is proportional to the citations achieved by each cluster.

Note that although the visualization and mapping steps are different, they are also interdependent. The visualization technique used will vary depending on the method selected to build the map. For example, the strategic diagram only visualizes maps built with a clustering algorithm.

Finally, when the science mapping analysis is finished, the analysts have to interpret the results and maps using their experience and knowledge. In the interpretation step, the analyst aims to discover and extract useful knowledge that could be used to make decisions.

Tools for Science Mapping Analysis

Science mapping analysis can be carried out with different software tools. Some general software tools not specifically designed for science mapping analysis can be employed for this task (Börner et al., 2010), such as Pajek (Batagelj & Mrvar, 1998), Gephi (Bastian, Heymann, & Jacomy, 2009), UCINET (Borgatti, Everett, & Freeman, 2002), or Cytoscape (Shannon et al., 2003). However, there are a variety of software tools specifically developed to perform a science mapping analysis.

In Cobo et al. (2011b), we describe and compare nine representative science mapping software tools: Bibexcel (Persson, Danell, & Wiborg Schneider, 2009), CiteSpace II (Chen, 2004, 2006), CoPalRed (Bailón-Moreno et al., 2006; Bailón-Moreno, Jurado-Alameda, Ruíz-Baños, & Courtial, 2005), IN-SPIRE (Wise, 1999), Loet Leydesdorff's software, Network Workbench Tool (Börner et al., 2010; Herr, Huang, Penumarthy, & Börner, 2007), Science of Science Tool (Sci²Team, 2009), VantagePoint (Porter & Cunningham, 2004), and VOSviewer (van Eck & Waltman, 2010).

These tools have different characteristics and implement different methods and algorithms. Consequently, we can make the following points:

- The science mapping software tools available have different characteristics, and no single tool implements all the steps necessary to carry out a science mapping analysis. For example, VOSviewer is mainly focused on the mapping and

visualization steps, and Loet Leydesdorff's software is not able to perform the preprocessing step or to visualize the generated maps.

- Considering that the de-duplicating step could be the most important preprocessing process, CoPalRed and VantagePoint have a good module to carry out this task. Network Workbench Tool and Science of Science Tool have this module, too, but it requires the use of an external software tool. Other software tools such as BibExcel or VOSviewer do not include this kind of preprocessing.

- Another important characteristic is the bibliometric networks available. Although the majority of software tools can construct a huge variety of bibliometric networks, none can construct all kinds of bibliometric networks. Indeed, some tools (e.g., CoPalRed) focus on only one kind of network, and other software tools (e.g., VOSviewer) are not able to build any.

- Each science mapping software tool employs different visualization techniques. For example, VOSviewer uses proximity maps, CoPalRed uses strategic diagrams to categorize the detected clusters, and VantagePoint has different maps such as factor maps.

- Furthermore, these software tools do not enable the user to define different configurations to develop a study. Most allow only one kind of analysis, one kind of similarity measure, one kind of clustering algorithm, and so on.

- As noted in Cobo et al. (2011a), it would be desirable for a science mapping software tool to enrich the output with quality and impact measures (e.g., bibliometric measures) that could help the analyst analyze the results and detect their importance and impact in the research area considered. For example, if a co-word analysis has been carried out to map a specific research area, using bibliometric measures (published documents or citations achieved), the analyst could assess the output topics that have attracted the most attention in the research community and, consequently, which are the most important for the studied research area. To our knowledge, CiteSpace is the only software tool that includes in its analysis the citations count, but in any case, this measure is only used in a basic way.

We therefore think it would be desirable to develop a science mapping software tool that satisfies the following requirements: (a) it should incorporate modules to carry out all the steps of the science mapping workflow, (b) it should present a powerful de-duplicating module, (c) it should be able to build a large variety of bibliometric networks, (d) it should be designed with good visualization techniques, and (e) it should enrich the output with bibliometric measures. We have taken into account all these requirements in the development of SciMAT.

SciMAT

SciMAT is a new, open-source science mapping software tool that implements the aforementioned software requirements. It can be freely downloaded, modified, and redistributed according to the terms of the GPLv3 license. The executable file, user guide, and source code can be downloaded through its Web site (http://sci2s.ugr.es/scimat).

SciMAT is based on the science mapping analysis approach presented in Cobo et al. (2011a), which allows us to carry out science mapping studies under a longitudinal framework (Garfield, 1994; Price & Gürsey, 1975). Although this approach was originally developed to carry out a conceptual science mapping analysis, we have extended it in SciMAT to perform any kind of science mapping analysis (including intellectual and social). This science mapping analysis approach establishes the following steps (Cobo et al., 2011a):

1. To detect the substructures contained (mainly, clusters of authors, words, or references) in the research field by means of a bibliometric analysis (bibliographic coupling, journal bibliographic coupling, author bibliographic coupling, coauthor, cocitation, author cocitation, journal cocitation, or co-word analysis) for each studied period.
2. To lay out in a low dimensional space the results of the first step (clusters).
3. To analyze the evolution of the detected clusters through the different periods studied to detect the main general evolution areas of the research field, their origins, and their interrelationships.
4. To carry out a performance analysis of the different periods, clusters, and evolution areas, by means of bibliometric measures.

The main characteristics of SciMAT are:

- SciMAT incorporates all modules necessary to perform all the steps of the science mapping workflow, which can be configured ad hoc. It helps the analyst to carry out the different steps of the science mapping workflow, from data loading and preprocessing of the raw data to the visualization and interpretation of the results.
- SciMAT incorporates methods to build the majority of the bibliometric networks, different similarity measures to normalize them and build the maps using clustering algorithms, and different visualization techniques useful for interpreting the output.
- SciMAT implements a wide range of preprocessing tools such as detecting duplicate and misspelled items, time slicing, and data and network reduction.
- SciMAT allows the analyst to perform a science mapping analysis in a longitudinal framework to analyze and track the conceptual, intellectual, or social evolution of a research field through consecutive time periods.
- SciMAT builds science maps enriched with bibliometric measures based on citations such as the sum, maximum, minimum, and average citations. Furthermore, SciMAT uses advanced bibliometric indexes such as the h-index (Alonso et al., 2009; Hirsch, 2005), g-index (Egghe, 2006), hg-index (Alonso et al., 2010), and q^2-index (Cabrerizo et al., 2010).

In the following subsections, we describe the SciMAT software tool. First, the structure of its knowledge base is analyzed in detail. Second, the architecture and the different algorithms and methods provided by SciMAT to perform a science mapping analysis are described. The technologies used in the development of the tool then are summarized.

Knowledge Base: Entities

SciMAT generates a knowledge base from a set of scientific documents where the relations of the different entities related to each document (authors, keywords, journal, references, etc.) are stored. This structure helps the analyst to edit and preprocess the knowledge base to improve the quality of the data and, consequently, obtain better results in the science mapping analysis.

The knowledge base is composed of 16 entities: *Affiliation*, *Author*, *Author Group*, *Author-Reference*, *Author-Reference Group*, *Document*, *Journal*, *Publish Date*, *Period*, *Reference*, *Reference Group*, *Reference-Source*, *Reference-Source Group*, *Subject-Category*, *Word*, and *Word Group*.

The principal entity is the *Document*, which represents a scientific document (usually articles, letters, reviews, or proceedings papers). It contains information such as the title, abstract, doi, citations, and so on. The Document has a variety of information associated with it, such as the authors, affiliations, keywords, cited references, the journal (or conference), and the publication year. Each one is considered an entity in the knowledge base.

The *Author* is the entity that represents the person who has been involved in the development of a Document. An Author can be associated with a set of Documents, and similarly a Document can have a set of Authors. Furthermore, an Author has an associated position in his or her Documents.

The *Affiliation* represents the author's affiliations. Given that the authors may work in different places (universities, institutes, etc.) during their research, an Author has a set of associated Affiliations.

Usually, the scientific documents have a set of keywords associated with them, commonly provided by the authors (author's words). Moreover, depending on the bibliometric database used to retrieve the data, the documents may contain descriptive words provided by the database (source's words). For example, ISIWoS adds a set of keywords called ISI Keywords PLUS to each document. In addition, sometimes the analyst needs to add more words to those documents which contain few descriptive terms (words). These words can be selected from the title, abstract, or body of the document, or they can be added manually. In our context, this set of words will be called *added words*. In this sense, the entity *Word* represents a descriptive term of a document. A set of Words can appear in different Documents, and each Document can have a set of Words. Each Word can have different roles in the Documents in which it appears. In this way, a Document can have words provided by the authors (author's words role), provided by the database (source's words role), or added in the preprocessing step (added words role).

The entity *Reference* represents the intellectual base of a scientific document. Similarl to the entity Word, a Document has a set of References associated with it, and each Reference can be presented in different Documents. The References can often be divided into small pieces of information.
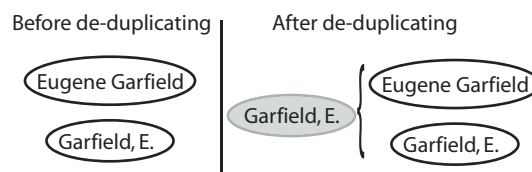

FIG. 3.   An author group after a de-duplicating process.

Depending on the database used to retrieve the data, these pieces may be different, but some information appears more often, such as author, journal, and year. For this reason, there are two entities related to the Reference: the *Author-Reference* and the *Source-Reference*.

Other entities associated with a Document are *Journal* and *Publish Date*. Logically, a Document can have only one Journal (or conference) and one Publish Date associated with it whereas both entities can have one set of Documents associated. Moreover, the Journal and Publish Date entities have an associated *Subject Category* which represents a global category, often given by the bibliometric database, that classifies the journal into the main knowledge categories. The Journal can be associated with many Subject Categories, and this relation can change over the years. That is, it is possible for a journal to have a different category associated to it each year.

The entity *Period* represents a set of (not necessarily disjointed) years. Usually, a set of Periods is defined to perform a longitudinal science mapping analysis (Garfield, 1994).

Note that five of the aforementioned entities can be used as a unit of analysis in the science mapping analysis carried out by SciMAT: Author, Word, Reference, Author-Reference, and Source-Reference. These entities should be carefully preprocessed, paying special attention to the misspelling and de-duplicating process. Usually, the de-duplicating process joins the similar items, so only one of them remains. For example, suppose that two items, Garfield, E. and Eugene Garfield, are stored in the knowledge base. Both items represent the same author and therefore should be joined (joining its association with the other entities). But, when two items are joined, only one of them is kept in the knowledge base (obviously, this item contains the association of the second item), and it is impossible to know the initial items joined. For this reason, our knowledge base provides the concept of *group* for each unit of analysis. A group is a set of items that represents the same entity. Thus the knowledge base contains five kinds of groups: *Author Group*, *Word Group*, *Reference Group*, *Author-Reference Group*, and *Source-Reference Group*. A group can be marked as *stop group*, in which case it will not be taken into account in the science mapping analysis.

An example of groups and how they help in the de-duplicating process is shown in Figure 3. On the left, we can see the items before being processed, and on the right we can see the group items. The shadow ellipse represents the
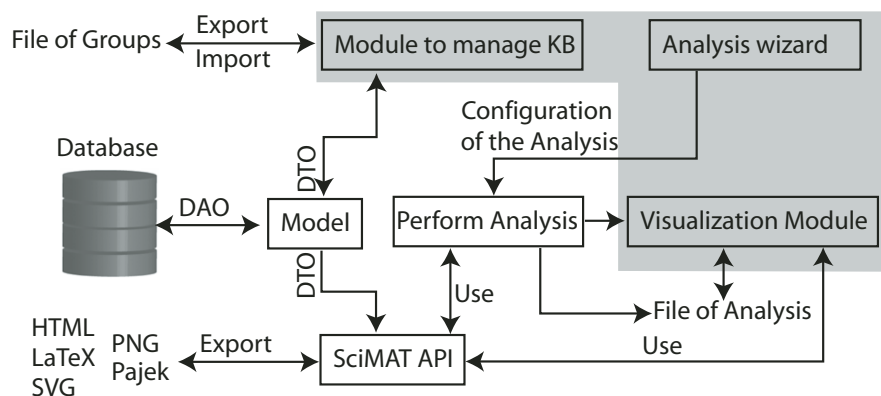
FIG. 4.   Architecture of SciMAT.

group (in this case, an Author Group), and the remaining ellipses represent the entities (Authors) associated with this group.

Architecture, Modules, Functionalities, and Algorithms

In this subsection, we describe the architecture of SciMAT, showing the tool's inner workings and how its modules interact. We also describe the different functionalities and algorithms available in SciMAT.

Internally, SciMAT is composed of several independent modules that interact with each other to carry out a science mapping analysis. Some modules are involved in the management of the graphical user interface (GUI) as well as the interaction with the user. Other modules are not visible to the user, these being the core of SciMAT. Figure 4 illustrates the architecture of SciMAT. The shaded boxes represent the modules responsible for the management of the GUI.

The main core of SciMAT is made up of:

• *The model:* The knowledge base described earlier is stored in a database. In this sense, the *model* is responsible for managing the database, and communication with it.
• *The SciMAT Application Programming Interface (SciMAT API):* The *SciMAT API* contains the necessary methods and algorithms to carry out the different steps of the science mapping analysis, from the loading of the raw data to the visualization of the results.

The *model* is the bridge between the GUI and the knowledge base stored in the database. To perform its function, the model uses two pattern designs: *Data Access Object* (DAO) and *Data Transfer Object* (DTO). The DAOs are responsible for communication with the database, performing the editing and selecting operations. The DTOs represent the different entities (discussed earlier) of the knowledge base, and they are used to transmit the information from the database to the different methods that make use of it.

Another important element of the core of SciMAT is its API, which contains all the necessary methods to carry out a

science mapping analysis with several configurations. Thanks to the object-oriented programming techniques used in the development of SciMAT, the API can be easily extended and improved to add new methods, algorithms, and measures. Specifically, the API provides methods to import data from different formats to the knowledge base, various methods to filter the data which will be used in the analysis (data and network reduction), several techniques to build the bibliometric network, the most common measures to normalize the network, different clustering algorithms to construct the map, several analysis techniques, and various kinds of visualizations. That is, the SciMAT API contains the necessary methods, techniques, and measures to develop a specific science mapping analysis.

The SciMAT API also can be used by an advanced user to develop ad hoc tools or specific scripts that allow him or her to configure and carry out a science mapping analysis. In this way, the advanced user can develop his or her own algorithms or a new loader to read his or her own data and import it into the knowledge base of SciMAT.

Taking into account the GUI, there are three important modules: (a) a module dedicated to the management of the knowledge base and its entities, (b) a module (wizard) responsible for configuring the science mapping analysis, and (c) a module to visualize the generated results and maps. These modules allow the analyst to carry out the different steps of the science mapping workflow.

Module to manage the knowledge base.   This module contains the necessary methods and algorithms responsible for the management of the knowledge base. As mentioned earlier, it communicates with the database (knowledge base) through the *model*, which responds to the request returning a DTO object. Moreover, the different actions performed by the user in the knowledge base are carried out using the model through its DAO object.

Regarding its functionalities, the module to manage the knowledge base is responsible for building the knowledge base, importing the raw data from different bibliographical
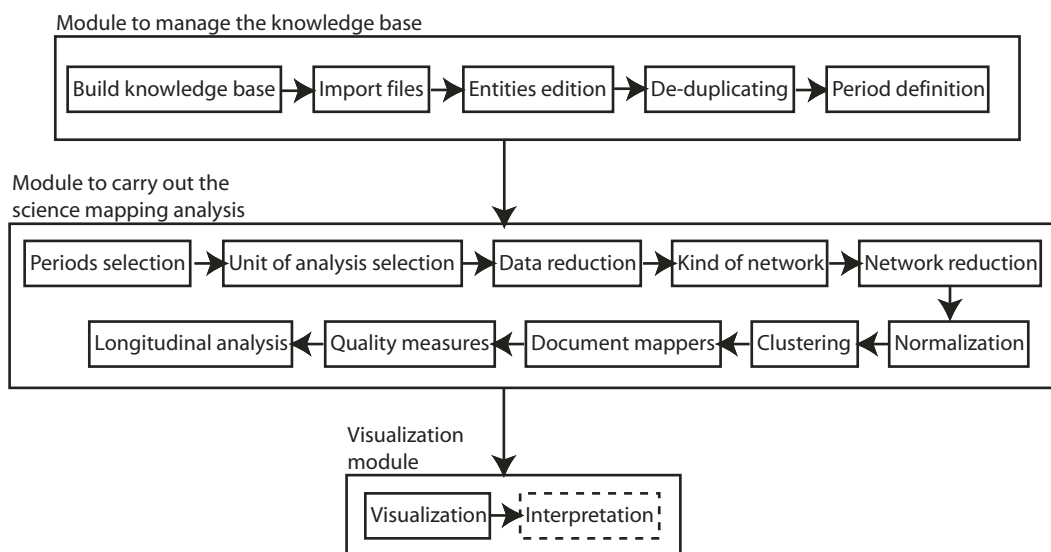
FIG. 5.   SciMAT workflow.

sources, and cleaning and fixing the possible errors in the entities. It can be considered as a first stage in the preprocessing step.

This module incorporates loaders to read bibliographical information exported from bibliometric sources, such as ISIWoS and Scopus (RIS format). Moreover, it is possible to import data from a specific CSV format. SciMAT uses these loaders to import the bibliographical data to a new or an existing knowledge base.

Each entity can be edited, and its attributes and associations can be modified. Furthermore, by means of Groups, the de-duplicating step can be performed. The user can join the items that represent the same entity, under the same group. In addition, this module incorporates methods to help the analyst in the de-duplicating process, such as finding similar items by plural or by Levenshtein distance, or importing the groups and their associated items from a file (in XML format). Finally, the time-slicing step is performed using the Period entity.

Wizard to configure the science mapping analysis. This wizard is one of the most important modules of the GUI. It is responsible for generating a particular configuration for carrying out the science mapping analysis so that the user can select the methods and algorithms that will be used to perform each step. Once the user has specified the desired configuration, it is sent to the module responsible for carrying out the analysis, which, using the SciMAT API, will perform the science mapping analysis. Finally, the results are stored in a file and are then sent to the visualization module.

Although the wizard has been implemented according to the steps of the science mapping workflow (see Figure 1), some steps are performed in a different order. For example, the de-duplicating and time-slicing preprocessing has to be

done earlier, using the knowledge base manager. To summarize, the SciMAT workflow is implemented as in Figure 5.

As shown, the workflow is divided into four main stages: (a) to build the data set, (b) to create and normalize the network, (c) to apply a cluster algorithm to get the map, and (d) to perform a set of analyses. These stages and their respective steps are described below:

1. *Build the data set:* At this stage, the user can configure the periods of time used in the analysis, the aspects that he or she wants to analyze, and the portion of the data that has to be used.
   (a) *Select the periods:* A set of periods, previously defined in the knowledge base, must be selected. The science mapping analysis is done for each selected period, and finally, they are used in the longitudinal or temporal analysis to study the structural evolution of the field.
   (b) *Select the unit of analysis:* Depending on the selected items, the conceptual (using terms or words), social (using authors), and intellectual (using references) aspects can be analyzed. As a unit of analysis, the user can select any of the five groups existing in the knowledge base: Author Group, Author-Reference Group, Source-Reference Group, Reference Group, or Word Group. Only one of them can be selected at a time. If the Word Group has been selected, the roles of the words (author's word, source's word, or added words) with which the user wants to perform the analysis have to be chosen.
   (c) *Data reduction:* SciMAT allows the user to filter the data using a minimum frequency as a threshold. For each period, a threshold also must be fixed. Therefore, only the units of analysis with a frequency greater than or equal (in a given period) to the selected threshold are considered.

TABLE 2. Bibliometric networks summary.

Relations Units	Co-occurrence	Coupling	Aggregated coupling by author	Aggregated coupling by journal
Author	coauthor	*	*	*
Reference	cocitation	bibliographic coupling	author bibliographic coupling	journal bibliographic coupling
Author-Reference	author cocitation	*	*	*
Source-Reference	journal cocitation	*	*	*
Word	co-word	*	*	*

*Other advanced and noncommon bibliometric networks.

2. *Create and normalize the network:* At this stage, the network is built using co-occurrence or coupling relations or, indeed, aggregating coupling. Then, the network is filtered to keep only the most representative items. Finally, a normalization process is performed using a similarity measure.

   (a) *Select the way in which the network will be built:* The network can be built using different methods such as co-occurrence, coupling, and aggregated coupling. Depending on the unit of analysis selected and the kind of relation chosen, different bibliometric networks can be built. For example, if the network is built using words as the unit of analysis and co-occurrence as the relation, a co-word bibliometric network will be built. Likewise, if references are selected as the unit of analysis and coupling as the relation, a bibliographic coupling network will be built. In Table 2, a summary of the kinds of networks available depending on the configuration of unit and relation is shown. As can be seen, SciMAT is able to build the most common bibliometric networks used in the literature, as well as other advanced and less common bibliometric networks (marked as an asterisk in Table 2). For example, selecting words as the unit of analysis and coupling as the relation, a conceptual coupling network could be built; moreover, it could be aggregated by authors or journals. Although this kind of bibliometric network is not frequently required, SciMAT enables its construction.

   (a) *Network reduction:* Each edge of the network will have a value, which can be the co-occurrence or coupling between the corresponding nodes. SciMAT allows the network to be filtered using a minimum threshold edge value. For each selected period, a threshold value must be set; that is, only the edges with a value greater than or equal to the selected threshold in a given period will be taken into account.

   (b) *Select the similarity measures to normalize the network:* SciMAT allows the user to choose the similarity measures commonly used in the literature to normalize networks: association strength (Coulter et al., 1998; van Eck & Waltman, 2007), Equivalence Index (Callon et al., 1991), Inclusion Index, Jaccard Index (Peters & van Raan, 1993), and Salton's cosine (Salton & McGill, 1983).

3. *Apply a clustering algorithm to get the map and its associated clusters or subnetworks:* At this stage, the clustering algorithm used to build the map has to be selected. Different clustering methods are available in SciMAT, such as the *Simple Centers Algorithm* (Cobo et al., 2011a; Coulter et al., 1998), *Single-linkage* (Small & Sweeney, 1985), and variants such as *Complete-linkage*, *Average-linkage*, and *Sum-linkage*.

4. Apply a set of analyses: The final step of the wizard consists of selecting the analyses to be performed on the generated map.

   (a) *Network analysis:* By default, SciMAT adds Callon's density and centrality (Callon et al., 1991; Cobo et al., 2011a) as network measures to each detected cluster in each selected period. *Callon's centrality* measures the degree of interaction of a network with other networks, and it can be understood as the external cohesion of the network. It can be defined as $c = 10 * \Sigma e_{uv}$, with u an item belonging to the cluster and v an item belonging to other clusters. *Callon's density* measures the internal strength of the network, and it can be understood as the internal cohesion of the network. Σ can be defined as $d = 100 \dfrac{\sum e_{ij}}{n}$, with i and j items belonging to the cluster and n the number of items in the theme. These measures are useful to categorize the detected clusters of a given period in a strategic diagram (Cobo et al., 2011a).

   (b) *Performance analysis:* SciMAT is able to assess the output according to several performance and quality measures. To do that, it incorporates into each cluster a set of documents using a document mapper function and then calculates the performance based on quantitative and qualitative measures (using citation-based measures, number of documents, etc.). SciMAT is able to add several document sets, even in the same analysis, calculated with different document mappers.

     SciMAT incorporates five different document mappers for co-occurrence networks: (a) core mapper (Cobo et al., 2011a), (b) secondary mapper (Cobo et al., 2011a), (c) *k*-core mapper, (d) union mapper, and (e) intersection mapper.

     For coupling networks, SciMAT has two kinds of document mappers depending on the kind of coupling used. That is, if a basic coupling has been selected (each item of the cluster will be a document), the basic coupling document mapper is the only one available, which adds the items of the cluster as documents. If an aggregated coupling is selected, the

aggregated coupling document mapper can be selected, which adds the documents associated with its items to each cluster (author's or journal's oeuvres). Note that each node of the cluster also has a set of associated documents. These documents correspond to the set of documents associated with the item (node) in the corresponding dataset.

Once the sets of documents have been associated to each cluster, a set of performance bibliometric measures can be added to each set. SciMAT adds by default the number of documents as the performance measure. Moreover, the citations of a set of documents are used to assess the quality and impact of the clusters. In this sense, basic measures such as the sum, minimum, maximum, and average citations or complex measures such as the h-index (Alonso et al., 2009; Hirsch, 2005), g-index (Egghe, 2006), hg-index (Alonso et al., 2010), or q^2-index (Cabrerizo et al., 2010) can be used, even simultaneously.

(c) *Temporal analysis or longitudinal analysis:* This allows the user to discover the conceptual, social, or intellectual evolution of the field. SciMAT is able to build an evolution map to detect the evolution areas (Cobo et al., 2011a) and an overlapping items graph (Price & Gürsey, 1975; Small, 1977) across the periods analyzed. Furthermore, SciMAT allows the user to choose different measures to calculate the weight of the "evolution nexus" (Cobo et al., 2011a) between the items of two consecutive periods, such as association strength (Coulter et al., 1998; van Eck & Waltman, 2007), Equivalence Index (Callon et al., 1991), Inclusion Index, Jaccard's Index (Peters & van Raan, 1993), and Salton's cosine (Salton & McGill, 1983).

Visualization module. This module is responsible for showing the results obtained by the system (using the SciMAT API) and helping the user to analyze and interpret them. It allows the user to navigate and interact with the results, focusing on those aspects that he or she wants to analyze in detail.

Different visualization techniques are available, such as the strategic diagram, cluster network, evolution map, and overlapping map. The strategic diagram (Figure 6a) shows the detected clusters of each period in a two-dimensional space, and categorizes them according to their Callon's density and centrality measures. Each cluster in the strategic diagram can be enriched by the bibliometric measures selected in the wizard. The associated network for each cluster is shown as well (for a graph of the relationship between its items, see Figure 6b).

The results of the temporal or longitudinal analysis are shown using an evolution map and an overlapping-items graph (see Figure 7). As an example, in Figure 7a, we can observe two different evolution areas delimited by differently shaded shadows. One is composed of Cluster A[1] and Cluster A[2], and the other is composed of clusters Cluster B[1], Cluster B[2], and Cluster C[2]. Cluster D[1] is discontinued, and Cluster D[2] is considered to be a new cluster. The solid lines (Lines 1 and 2) mean that the linked cluster shares the
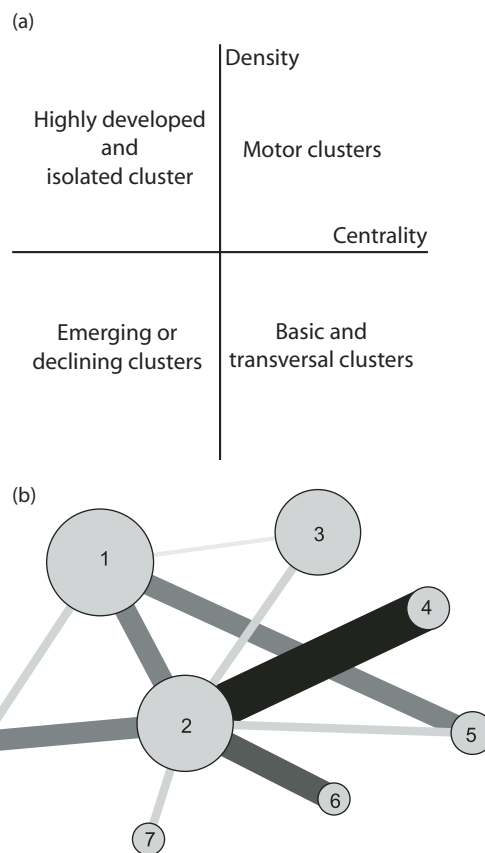
FIG. 6. The strategic diagram and cluster network. (a) The strategic diagram. (b) An example of a cluster network.

main item (usually the most significant one). A dotted line (Line 3) means that the themes share elements that are not the main item. The thickness of the edges is proportional to the Inclusion Index, and the volume of the spheres is proportional to the number of published documents associated with each cluster. Following this example, in Figure 7b, the overlapping-items graph across the two consecutive periods is shown. The circles represent the periods and their number of associated items (unit of analysis). The horizontal arrow represents the number of items shared by both periods, the Stability Index between them is shown in parentheses. The upper incoming arrow represents the number of new items in Period 2, and the upper outgoing arrow represents the items that are presented in Period 1, but not in Period 2.

The visualization module can build a report in HTML or LATEX format using the API. The images (strategic diagrams, overlapping-items map, etc.) are exported in PNG and SVG formats so the user can easily edit them. Furthermore, the cluster networks and evolution maps are exported in Pajek format.

Technologies

SciMAT has been programmed in Java. This allows the tool to run on any platform such as Windows, MacOS, Linux, and so on.

(a)

Period 1 | Period 2

Cluster A¹ — 1 — Cluster A²

Cluster B¹ — 2 — Cluster B²

— 3 — Cluster C²

Cluster D¹   Cluster D²

(b)

50 → 5 → 70

45 (0.6)

.25

Period 1   Period 2

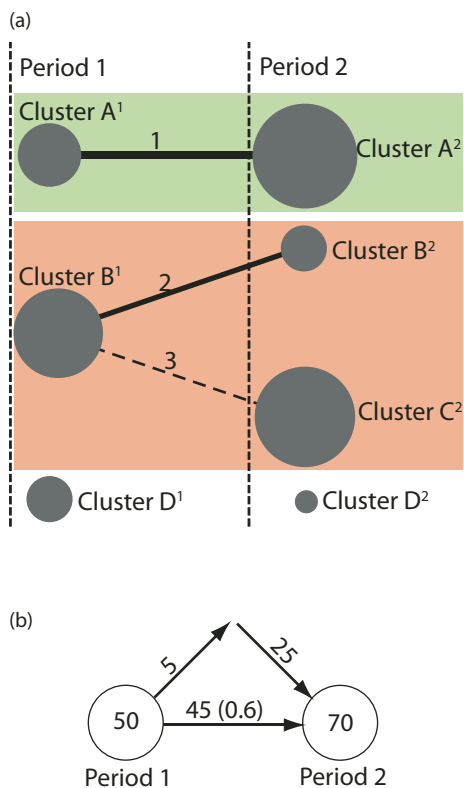FIG. 7. Examples of evolution. (a) Evolution areas. (b) Stability between periods. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Taking into account the programming technique used, SciMAT has been developed under the object-oriented methodology. Furthermore, different design patterns have been used, such as the pattern *Observable*, *Observer*, *Command*, *Edit*, *Singleton*, *Factory*, *Data Access Object*, *Data Transfer Object*, and so on. With the use of these patterns, and the abstract techniques employed, SciMAT can be extended easily to incorporate new methods and algorithms.

Furthermore, SQLite (Kreibich, 2010) has been used as database engine to store the knowledge base. SQLite is a public-domain software package that provides a *relational database management system*. It has some good properties—serverless, zero configuration, cross-platform, self-contained, small runtime footprint, transactional, highly reliable (Kreibich, 2010)—and has minimum system requirements. Thanks to these capabilities, the SciMAT knowledge base can be opened with any database browser that reads SQLite files.

Finally, SciMAT uses different technologies such as SVG, HTML, and XML to export the result obtained in the performed analysis.

Scenarios and Potential-Use Case Using SciMAT

As mentioned earlier, science mapping analysis is a useful technique to discover the social, intellectual, and conceptual aspects of research fields, specialties, or individual

documents or authors. Furthermore, the combined use of bibliometric indicators helps to quantify their impact and quality.

There are several academic scenarios in which SciMAT could be applied:

- Suppose that a journal editor needs to analyze the topics or themes covered by his or her journal and their evolution over the years. The editor also may be interested in knowing which themes have achieved a greater number of citations and which ones have been most productive (in terms of the number of documents). By means of a science mapping analysis performed using SciMAT, the editor could discover the topics covered by the journal since its inception, and how they have evolved to the current topics, so he or she could know if the scope of the journal is attracting citations, and which are the "hot" topics. Furthermore, analyzing the social aspects of the journal, the editor could find out the common co-authorships or author collaboration networks, and the authors who usually publish highly cited articles in the journal. For example, this analysis could help to determine the authors most suitable for guesting/editing a future special issue. Finally, analyzing the intellectual aspects, the editor could uncover the journals and authors most frequently cited by the journal; that is, its intellectual base. All this information could help to make high-level decisions.
- A researcher may want to analyze his or her own research field to study the impact of the different themes covered. He or she could identify a research field using the journals contained in a specific ISI Subject Category. Using SciMAT, the researcher could build a strategic diagram with which to categorize the different themes covered by the field, and an evolution map with which to study the evolution of the themes over the years. Furthermore, using the bibliometric indicators available in SciMAT, the researcher could identify those themes that were most productive and those themes with greater impact. Once this conceptual analysis has been performed, the researcher could study other structural aspects. For example, analyzing the intellectual base (cited references, authors, and journals) by means of a cocitation analysis, the researcher could uncover the key references in the field, and perhaps discover some new ones.
- If a university is interested in identifying the most important or internationally relevant research undertaken by its faculty staff, SciMAT could help to address these issues. For example, an analyst could build a map showing the evolution of the topics covered by the university or research center over the years, and learn which topics have always been most productive and cited, which topics are no longer of interest to the international community (i.e., they are not attracting citations), and which are currently the most cited and productive. This analysis could help to determine where the university should invest more resources (money, staff, facilities). Although this analysis could be carried out with all publications of the university, it also could be done by focusing on some specific departments, schools, or areas. Finally, by analyzing the intellectual base of the university (i.e., the references cited by its researchers), the university directors could identify the most and least used journals, and consequently decide to which journals they should continue to subscribe, to which journals they should no longer subscribe, and to which

new journals they should subscribe. Note that these analyses could be performed on other research institutions, on a specific set of researchers (e.g., a research group), or indeed on an R&D company.

- Suppose that a government requests its scientific policy makers to identify the most fruitful areas of research, or those which seem most promising and would therefore be appropriate to focus resources. Furthermore, they may wish to determine the best researchers and their collaboration networks. As with the other scenarios, SciMAT could be useful to address these requirements. Note that this scenario also is possible in other contexts such as a city, province, or state, or indeed, a set of them.
- A business intelligence department, in a public or private company, could be interested in researching what other companies are developing. In this case, SciMAT could be useful to uncover the hidden topics (know-how) from a set of articles, technical reports, and proceedings papers published by other companies. With this information, the business intelligence department could determine which themes are enabling its direct competitors to be in a privileged position. Furthermore, the business department could determine the principal researchers in other companies, and perhaps make a job offer.

In these scenarios, SciMAT could help the analyst to obtain tentative results. However, for a better, more precise analysis, the analyst has to perform a careful preprocessing process over the units of analysis. Moreover, an adequate selection of the parameters used in the analysis must be done. In any case, the analyst should check if SciMAT's features are suitable to his or her problem.

To show how SciMAT could help in a real scenario, we next illustrate the strength of SciMAT through one of these possible scenarios of use. Specifically, we focus on a researcher who would like to know the research issues raised in a particular field or area to understand and deepen his or her knowledge of the field and identify the hot topics.

Analyzing a Field: A Practical Example

Suppose that a researcher is interested in a new research field. Moreover, he or she would like to know the themes on which the research community is working hard, those that are highly cited, and those that seem to be disappearing. With SciMAT, that researcher could discover the themes closet to his or her current research and those most appropriate for the investment of his or her effort. For example, suppose that the researcher wants to analyze the fuzzy sets theory (FST) field (Zadeh, 1965, 2008).

We could analyze FST using the papers published in the two most important and prestigious journals in the field, according to their impact factor: *Fuzzy Sets and Systems* (*FSS*) and *IEEE Transactions on Fuzzy Systems* (*IEEE-TFS*). As *FSS* was founded in 1978, we could consider the publications in both journals for the years 1978 to 2009, but slicing the data into five consecutive periods (1978–1989, 1990–1994, 1995–1999, 2000–2004, and 2005–2009) to uncover the conceptual evolution of the FST field. In the

following subsections, we show how the analysis of the field is performed using SciMAT and some interpretations that could be drawn from that analysis.

Performing the analysis with SciMAT. Initially, the researcher should retrieve the raw data from bibliographic sources. In this case, the researcher retrieved the necessary data from the ISIWoS (for years 1980–2009), Scopus (for year 1993 of the journal *IEEE-TFS*), and Science Direct[10] (for years 1978 and 1979 of the journal *FSS*), because no source covered all the years. A total of 6,823 documents are retrieved.

Once the raw bibliographic data have been downloaded from the bibliographic sources, the first step in SciMAT would be to build a knowledge base and load the retrieved data using the importation capabilities of the knowledge base management module.

The second step carried out by the researcher would be editing the knowledge base, to fix possible errors (in titles, authors, references, etc.) and improve the quality of the data. To do this, SciMAT incorporates a manager for each entity (Document, Author, Reference, Word, Journal, etc.) so that the researcher can easily edit the information associated with each entity and its relations with other entities.

Note that all the managers have the same structure: on the left side, a list of entities is shown, and on the right side, the fields of the selected entity and its relations with other entities are shown.

In Figure 8, the Document's manager is shown. In the list of documents (left side), one of the most cited articles in the knowledge base is selected. On the right side, its associated information (title, abstract, publication data, citations, etc.) and associations are shown.

As mentioned earlier, five entities can be employed as the unit of analysis using the concept of group: Author Group, Word Group, Reference Group, Author-Reference Group, and Source-Reference Group. These have special managers to perform the de-duplicating process. These managers have a common structure: The left side shows a list of defined groups, and the right side shows the entities associated with the selected group (header-table) and the entities without an associated group (foot-table).

Because of this, the researcher would be interested in carrying out a conceptual evolution analysis, with keywords used as the unit of analysis. Thus, he or she should perform a de-duplicating step over the words. To do this, he or she should define the Word Groups, joining those words that represent the same concept. This could be done using the Word Group's manual set capability; that is, the special manager to perform the de-duplicating process.

Figure 9 displays the manager to perform the manual set of the Word Groups. For example, it can be seen

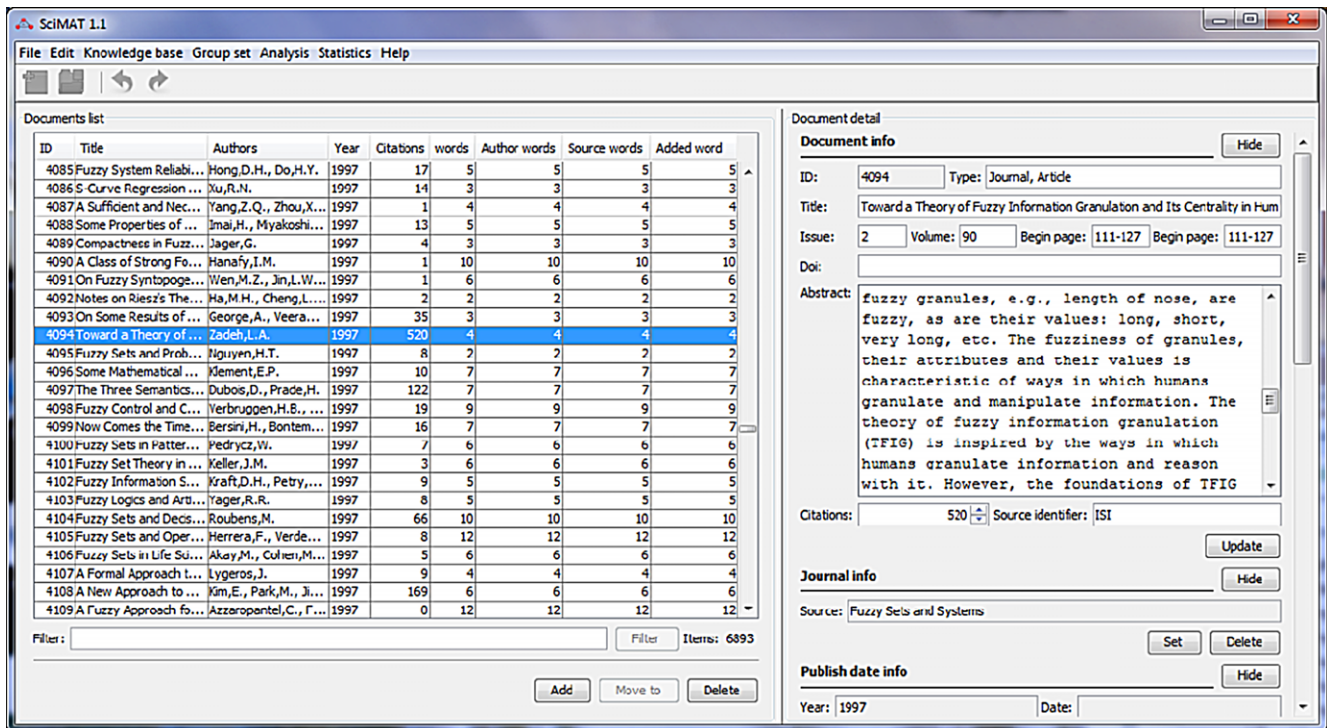---

[10]http://www.sciencedirect.com/

FIG. 8. Document's manager. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

that a particular word group with the name *GROUP-DECISION-MAKING* has been defined (left side). It also can be observed that this word group collects four different word names or variants (top right side) for the concept: *GROUP-CONSENSUS-OPINION*, *GROUP-DECISION-ANALYSIS*, *GROUP-DECISION-MAKING*, and *GROUP-DECISION-MAKING-(GDM)*. The lower right side allows the user to add more variants of the concept *GROUP-DECISION-MAKING*.

After cleaning the knowledge base, the third step should be to define the time slices in which the study is going to be performed; that is, to establish the groups of years that will be used later in the longitudinal analysis. The periods are defined using the Period manager. In particular, we consider five consecutive periods: 1978–1989, 1990–1994, 1995–1999, 2000–2004, and 2005–2009.

When the knowledge base is cleaned and the groups and periods are defined, the fourth step should be to configure all the necessary parameters that the analysis needs, using the wizard to perform the science mapping analysis. The analyst could use the groups' statistics (Figure 10) to estimate the correct parameters.

As shown earlier, the wizard allows the researcher to select the unit of analysis[11] (see Figure 11a), the similarity measure used to normalize the network, the clustering algorithm, the document mappers, the bibliometric measures (see Figure 11b), and the remaining key aspects needed to configure the science mapping analysis.

Specifically, we could fix the following configuration: Word as the unit of analysis (author, source,[12] and added keywords), co-occurrence as the way to build the network, Equivalence Index as the similarity measure to normalize the network, and the Simple Centers Algorithm as the clustering algorithm.

The bibliometric measures chosen could be the sum of the citations and the h-index, and these measures could be calculated for the documents mapped to each cluster by using the core and secondary document mappers.

At the end of all the steps in the wizard, the map would be built using the selected configuration. Then, the results would be saved to a file, and the visualization module loaded. The visualization module has two views: Longitudinal and Period.

The *Period view* (see Figure 12) shows detailed information for each period, its strategic diagram, and for each cluster, the bibliometric measures, the network, and their associated nodes. As an example, Figure 12 shows information about the period 2005 to 2009. Furthermore, it shows the cluster *H-INFINITY-CONTROL*, with its performance measures and cluster network.

Finally, in the *Longitudinal view* the overlapping map and evolution map are shown. This view helps us to detect the evolution of the clusters throughout the different periods, and study the transient and new items of each period and the items shared by two consecutive periods. In Figure 13, an example of the longitudinal view is shown.

[11]Only the unit of analysis with defined groups can be selected.

[12]ISI Keyword Plus.

FIG. 9 — SciMAT 1.1

Word groups

ID	Group name	Stop group	Items	Documents
8698	GRAY-LEVEL	false	1	1
11656	GRAY-LEVEL-THRESHOLDING	false	1	1
10754	GRAYNESS-AND-SPATIAL-AMBI...	false	1	1
8539	GRAY-RELATIONAL-ANALYSIS	false	1	1
2536	GRAY-SCALE	false	1	1
6478	GRAY-SCALE-IMAGES	false	1	1
3607	GREATEST-(LEAST)-INTERPRET...	false	1	1
5910	GRECO'S-REPRESENTATION-TH...	false	1	1
11331	GREEDY-ALGORITHM	false	2	4
4206	GREENS-FUNCTION	false	1	1
7521	GREEN-HOUSE-TEMPERATURE-C...	false	1	1
11664	GREEN-RELATIONS	false	1	1
7870	GREY-PREDICTOR	false	1	1
10216	GREY-RELATIONAL-SPACE	false	1	1
7927	GREY-SYSTEM	false	1	2
11103	GRILL-OPERATORS	false	1	1
2708	GRILLS	false	2	2
5123	GRINDING	false	1	2
11492	GROTHENDIECK-TOPOLOGY	false	1	1
7235	GROUND-PENETRATING-RADAR	false	1	1
3445	GROUND-VEHICLES	false	1	1
3706	GROUNDWATER-FLOW	false	1	1
152	GROUP-DECISION-MAKING	false	4	45
9702	GROUP-FUZZY-CONGRUENCE	false	1	1
4299	GROUP-METHOD-OF-DATA-HAN...	false	1	3
3296	GROUP-OBJECTS	false	1	1
10736	GROUP-OF-FUZZY-NUMBERS	false	1	1
571	GROUPOID	false	1	2

Filter: Items: 12157

Add — Edit — Move to — Delete

Words of the group

ID	Name	Documents	Role author	Role source	Role added
179	GROUP-DECISION-MAKING	39	39	39	39
1340	GROUP-DECISION-MAKING-(GDM)	3	3	3	3
7426	GROUP-CONSENSUS-OPINION	2	2	2	2
9213	GROUP-DECISION-ANALYSIS	1	1	1	1

Items: 4

Words without group

ID	Name	Documents	Role author	Role source	Role added
14	ALGORITHM	144	144	144	144
17	FUZZY-SYSTEMS	140	140	140	140
24	SYSTEMS	398	398	398	398
29	FUZZY-SYSTEM-IDENTIFICATION	2	2	2	2
33	DESIGN	311	311	311	311
38	MODELS	145	145	145	145
39	FUZZY-SYSTEM-MODELS	36	36	36	36
58	ALGORITHMS	96	96	96	96
60	MODEL	195	195	195	195
93	FUZZY-SETS	405	405	405	405
112	SETS	357	357	357	357
126	STABILITY	198	198	198	198
138	SYSTEM	87	87	87	87

Filter: Items: 93

Delete — To new group — To different group
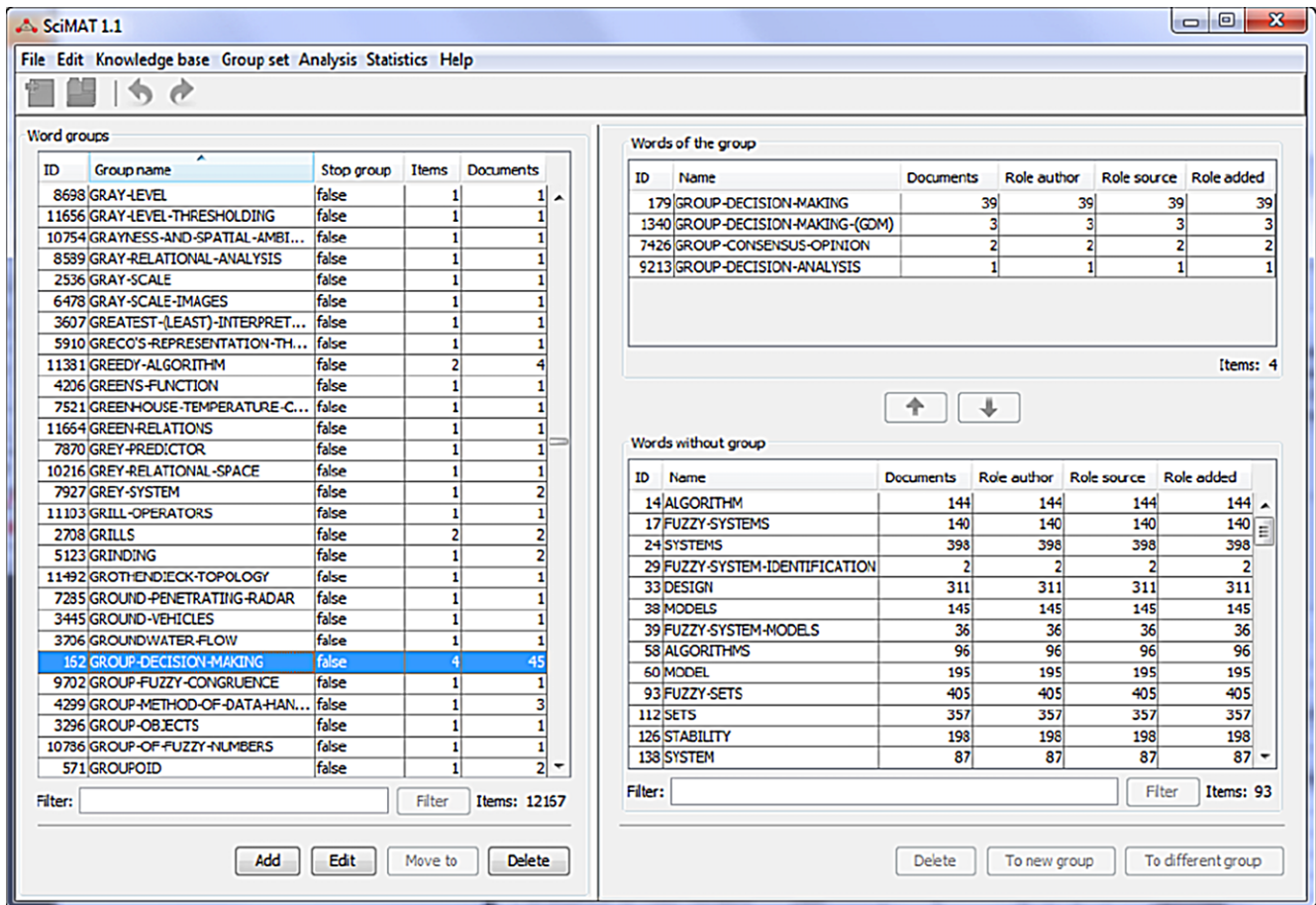
FIG. 9. Manual set Words Groups. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

FIG. 10 — SciMAT v1.1 - Word groups statistics

Statistic information

Period	Documents	Units	Max	Min	Mean	Median	Standard des...	Variance
Subperiod 1990...	1157	2787	18	0	4.77	5	2.44	5.93
Subperiod 1995...	1680	3782	24	0	4.89	5	2.42	5.84
Subperiod 2000...	1716	4528	19	0	5.81	5	2.53	6.4
Subperiod 2005...	1576	5034	31	0	7.49	7	3.16	9.98
Subperiod 1978...	764	762	8	0	2.09	2	1.13	1.29

Filter: Items: 5

Charts

- Documents
- Unique units
- Max
- Min
- Average
- Median
- Standard desviation
- Variance

Documents per period — **Frequency distribution**

Subperiod 1990-1994 — Subperiod 1995-1999
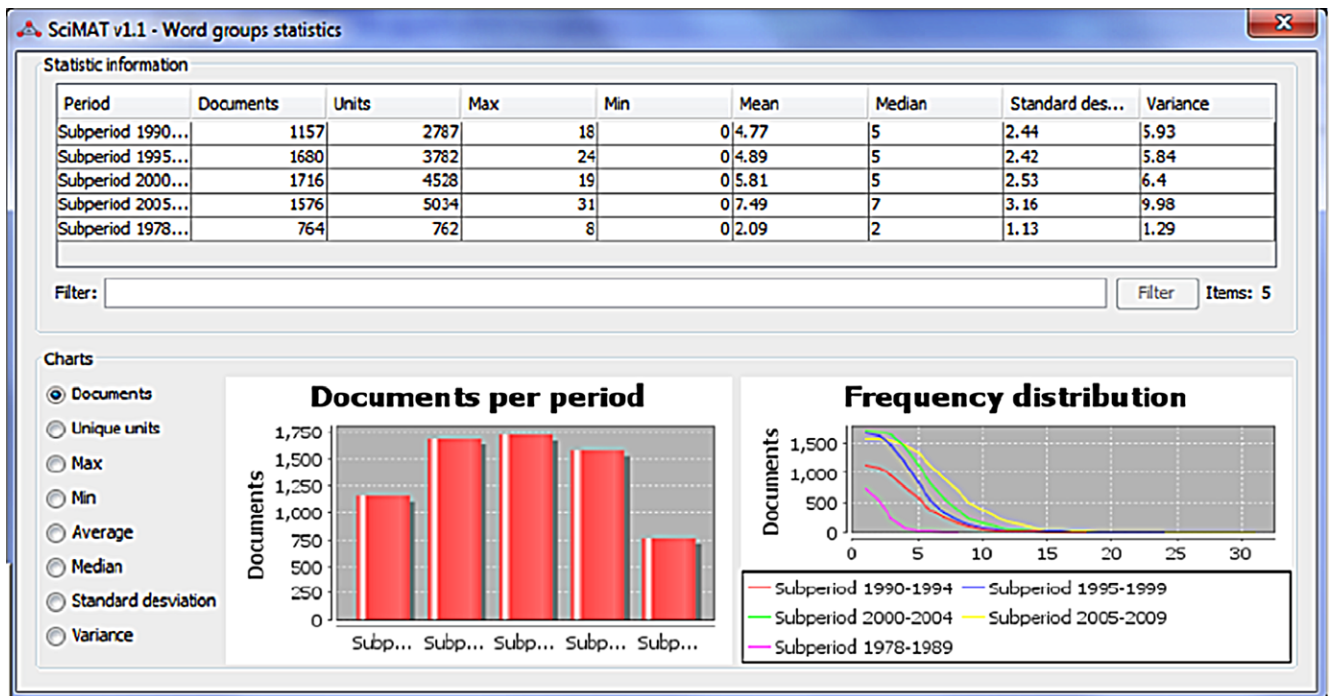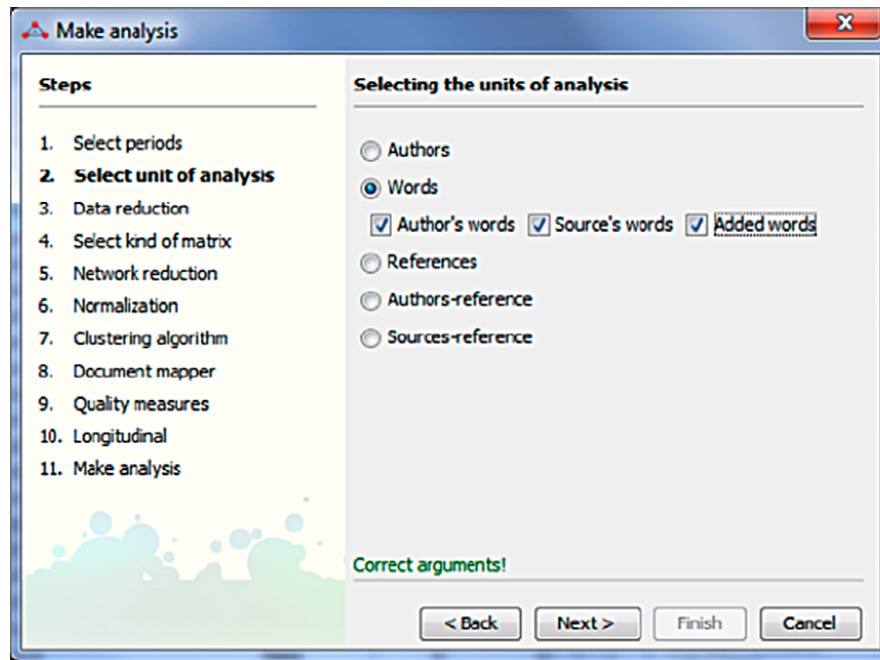Subperiod 2000-2004 — Subperiod 2005-2009
Subperiod 1978-1989

FIG. 10. Word groups' statistics. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
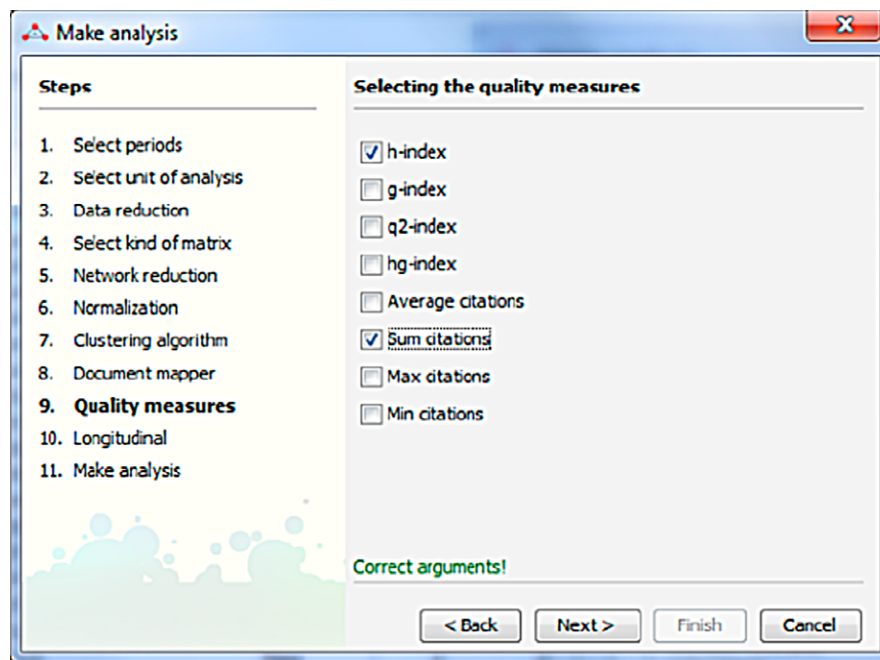
FIG. 11. The wizard to configure the analysis. (a) Choosing the unit of analysis. (b) Selecting the quality measures. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

As an example, we can observe the evolution of the cluster *FUZZY-CONTROL* and detect the themes that have been present in the majority of periods, such as *T-NORM*.

Interpretation of the results. Once the analysis has been performed, we have to interpret the results and obtain conclusions about the analyzed research field. In this case,

Figure 12 shows the strategic diagram for the period 2005 to 2009. We can observe that there are two important motor themes (*H-INFINITY-CONTROL* and *GROUP-DECISION-MAKING*), and many basic and transversal themes which are the base of the remaining ones. Taking into account quantitative measures such as the number of documents associated with each theme (cluster), we can discover
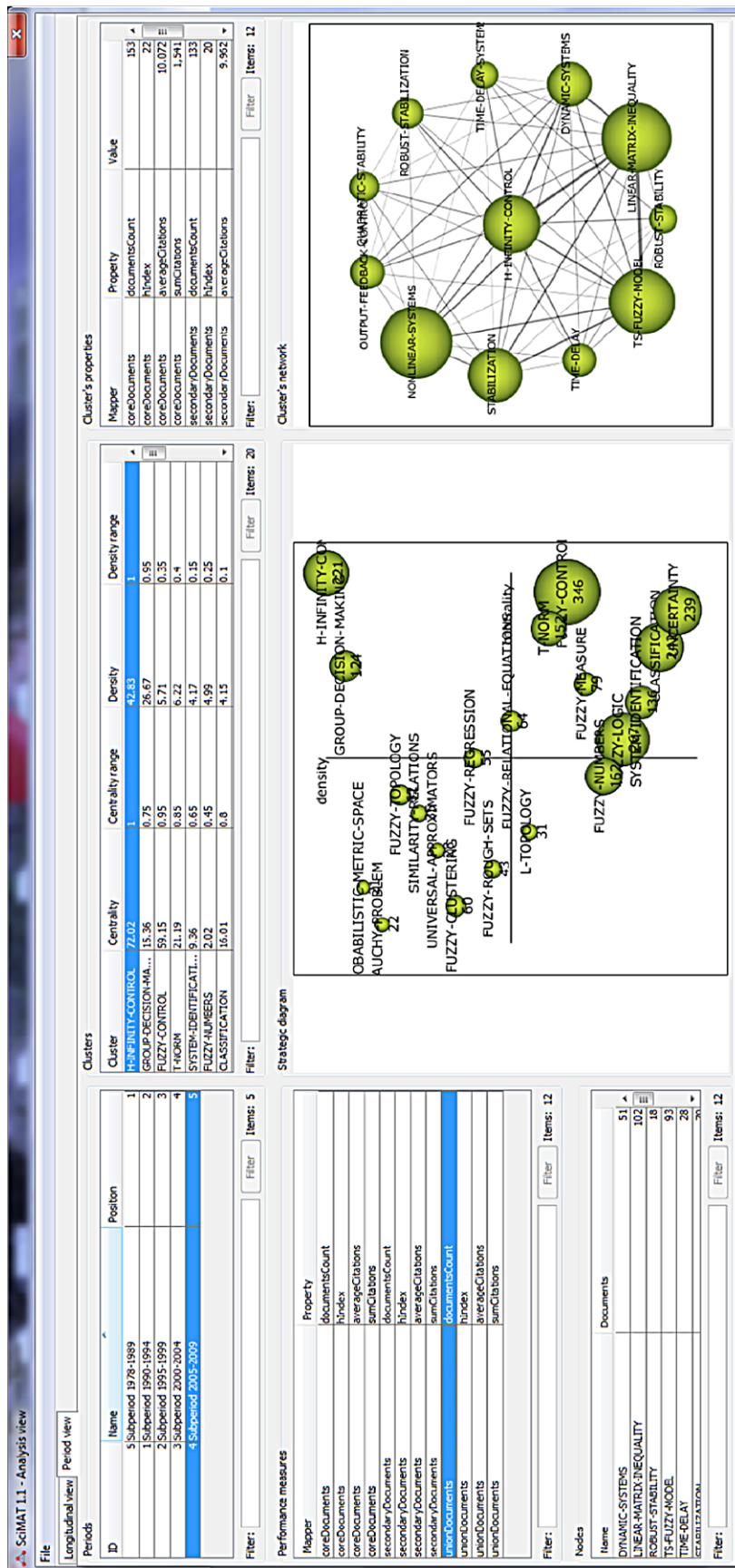
FIG. 12.   Period view. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
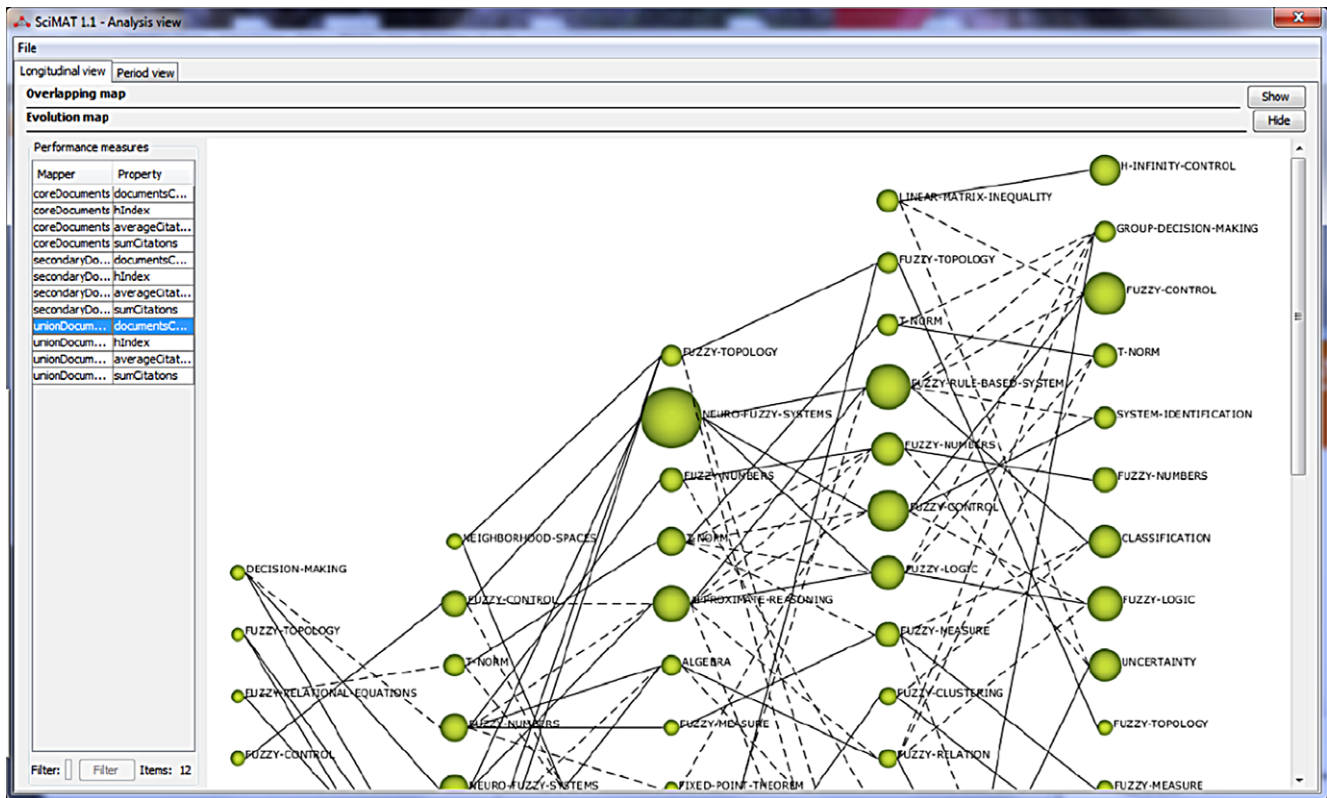
FIG. 13.   Longitudinal view. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

where the fuzzy community has been employing a great effort (e.g., *H-INFINITY-CONTROL*, *FUZZY-CONTROL*, *T-NORM*, etc.). Similarly, taking into account the qualitative measure, we could identify the themes with a greater impact; that is, the themes that have been highly cited.

In Figure 13, we can observe the conceptual evolution of the FST field. In this sense, it is easy to identify the themes that have been treated in all the periods (*FUZZY-CONTROL*, *FUZZY-TOPOLOGY*, *FUZZY-RELATION*), those that have disappeared (e.g., *FUZZY-SUBGROUP*), or those that have emerged in the last periods (e.g., *GROUP-DECISION-MAKING*).

Note that the analysis could continue by examining other structural aspects such as the main coauthorship networks in the area, to thereby raise possible collaborations with these researchers, or the main references of the FST field; that is, those most used and those that are commonly co-cited. The former could be performed selecting the authors as the unit of analysis in the wizard. The latter could be performed selecting the references as the unit of analysis. In both cases, the co-occurrence network should be selected in the wizard.

Validating SciMAT

To check the practical utility and usability of SciMAT, a user validation test has been performed. Version 1.0 of SciMAT was provided to a variety of potential end users,

including senior researchers, PhD students, heads of research groups, and technical staff of the Research and Policy Research Office of the University of Granada. Thus, 15 people have used SciMAT, and they have given us valuable suggestions and comments after using SciMAT with their own data set, including five different (science and social) research topics (computer science, information science, psychology, marketing, and chemistry).

For systematic and objective data acquisition, we used an adapted version of the Questionnaire for User Interface Satisfaction (Chin, Diehl, & Norman, 1988), which is a widely used and extended user questionnaire for evaluating software (Shneiderman, Plaisant, Cohen, & Jacobs, 2009; Vilar, 2010). The questionnaire used for testing SciMAT is shown in the Appendix. Five dimensions are considered: (a) Overall Reaction to the Software, (b) Screen, (c) Terminology and System Information, (d) Learning, and (e) System Capabilities. Several aspects are queried for each dimension, with 10-point scales as the response method. The user questionnaire is completed with two "free comment" boxes to collate both negative and positive aspects of SciMAT.

The majority of suggestions and comments offered were oriented toward improving the interface and/or interactivity of SciMAT, specifically the navigation flow and the information displayed on the interface. The users suggested adding more useful information in several modules, such as

the number of documents associated with each item (words, authors, journals, etc., and their related groups). Another important suggestion was to incorporate in SciMAT a module to visualize (with illustrative graphs) statistical information about each unit of analysis, with the main aim of using this statistical information for a better configuration of the various parameters of the analysis.

With respect to the functionality of SciMAT, the majority of the users were very satisfied with the different options, techniques, algorithms, and units of analysis present in SciMAT, although one user asked us to incorporate a more flexible module to load data.

Several users required a more complete report in both HTML and LATEX formats, including more statistical information, adding more information for each detected cluster or evolution area. As a result of these suggestions, the HTML and LATEX reports have been improved, building a detailed subsection for each cluster and showing the specific configuration of the performed analysis. Furthermore, a new advanced report (in both HTML and LATEX formats) has been added. The advanced report completes the information with the documents (showing the full reference, including citations) associated with each period and cluster.

As to the facility of use, several users reported that it is difficult to learn to use SciMAT in comparison with other science mapping software (e.g., CoPalRed or VOSviewer), although they also informed us after prolonged use that they were comfortable with the tool and the results provided.

Finally, all the suggestions given by the users' test were incorporated into Version 1.1 of SciMAT.

Concluding Remarks

We have presented here a new open-source software tool called SciMAT, to perform longitudinal science mapping, SciMAT has been developed on the basis of the science mapping approach proposed in Cobo et al. (2011a).

SciMAT presents three key features that other science mapping software tools do not have (or have in limited form): (a) a powerful preprocessing module, (b) the use of bibliometric measures, and (c) a wizard to configure the analysis. Different preprocessing processes can be applied, such as detecting duplicate and misspelled items, time slicing, data reduction, and network preprocessing. Bibliometric measures (mainly based on citations) such as the h-index (Alonso et al., 2009; Hirsch, 2005), g-index (Egghe, 2006), hg-index (Alonso et al., 2010), or q^2-index (Cabrerizo et al., 2010) are used by SciMAT to give information about the interest in and impact on the specialized research community of each detected cluster. Finally, the analysis is configured using a powerful wizard which allows the analyst to choose the algorithms, methods, and measures to be used in the analysis.

The main characteristics, methods, and algorithms present in SciMAT are:

- *Bibliographic sources:* SciMAT is able to import data downloaded from several bibliographic sources. Particularly, SciMAT reads the native format of ISIWoS (ISI-CE), the RIS format, which is used by Scopus and other bibliographic sources, and a specific CSV format.
- *Preprocessing:* Several preprocessing methods can be applied to improve the quality of the data.
- *Unit of analysis:* SciMAT is able to use five different units of analysis such as authors, words, references, author-reference, and source-reference.
- *Bibliographic relations:* The tool is able to set co-occurrence, coupling, and aggregated coupling (by author or journal) relations among the units of analysis.
- *Bibliographic networks:* Combining the units of analysis and the bibliographic relations among them, SciMAT can extract 20 kinds of bibliographic networks, including the common bibliographic networks used in the literature, such as coauthor (Gänzel, 2001; Peters & van Raan, 1991), bibliographic coupling (Kessler, 1963), journal bibliographic coupling (Small & Koenig, 1977), author bibliographic coupling (Zhao & Strotmann, 2008), cocitation (Small, 1973), journal cocitation (McCain, 1991), author cocitation (White & Griffith, 1981), and co-word (Callon et al., 1983).
- *Normalization of the bibliographic network:* SciMAT provides several measures to normalize the network, such as association strength (Coulter et al., 1998; van Eck & Waltman, 2007), Equivalence Index (Callon et al., 1991), Inclusion Index, Jaccard Index (Peters & van Raan, 1993), and Salton's cosine (Salton & McGill, 1983).
- *Clustering algorithms:* The tool provides different algorithms to build the maps, such as the Simple Centers Algorithm (Coulter et al., 1998; Cobo et al., 2011a), Single-linkage (Small & Sweeney, 1985) and variants, such as Complete-linkage, Average-linkage, and Sum-linkage.
- *Document Mappers:* Various sets of documents can be associated with each detected cluster. To do that, SciMAT provides different document mapper functions: core, secondary, union, intersection, and *k*-core.
- *Analysis methods:* SciMAT can perform a temporal analysis to study the evolution of the examined structural aspects (social, conceptual, or intellectual). In addition, a network analysis using different statistical network measures can be applied. Furthermore, SciMAT is able to apply several bibliometric measures to the sets of documents associated with each detected cluster. In this sense, SciMAT provides several bibliometric measures based on citations, such as the sum, minimum, maximum, and average citations, or complex measures such as the h-index (Alonso et al., 2009; Hirsch, 2005), g-index (Egghe, 2006), hg-index (Alonso et al., 2010), or q^2-index (Cabrerizo et al., 2010).
- *Visualization techniques:* The tool provides several visualization techniques such as strategic diagrams, cluster networks, evolution maps, and overlapping maps.
- *Export capabilities:* SciMAT is able to make HTML and LATEX reports with the results obtained in the analysis. Moreover, the tool can generate an advanced report with information about the documents associated with each period and cluster.

In Table 3, a comparative summary of the characteristics of SciMAT versus other science mapping software tools is shown. We can see that SciMAT is one of the most complete

TABLE 3.   Summary of SciMAT's characteristics versus other science mapping software tools.

Software tool	Preprocessing	Networks	Normalization	Analysis
BibExcel	Data and networks reduction	DBCA, ACAA, CCAA, ICAA, ACA, DCA, JCA, CWA, Others	Salton's cosine, Jaccard 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	Bust detection, network, temporal
Loet 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
SciMAT	De-duplication, time slicing, and data reduction	DBCA, ABCA, JBCA, ACAA, ACA, DCA, JCA, CWA, Others	Association strength, Equivalence Index, Inclusion Index, Jaccard Index, and Salton's cosine	Network, temporal, performance

ABCA = author bibliographic coupling; DBCA = document bibliographic coupling; JBCA = journal bibliographic coupling; ACAA = author coauthor; CCAA = country coauthor; ICAA = institution coauthor; ACA = author cocitation; DCA = document cocitation; JCA = journal cocitation; CWA = co-word; DL = direct linkage.

tools taking into account all the previously enumerated characteristics.

Although SciMAT is a complete and powerful science mapping tool, other software tools have impressive and notable characteristics (Cobo et al., 2011b). For example, regarding the preprocessing methods, VantagePoint (Porter & Cunningham, 2004) is able to import data from many kinds of formats. The Network Workbench Tool (Börner et al., 2010; Herr et al., 2007) and Science of Science Tool (Sci²Team, 2009) have good network-reduction processes. CiteSpace (Chen, 2004, 2006), Science of Science Tool, and VantagePoint have several analysis methods (geospatial, burst detection, etc.). Furthermore, SciMAT is able to calculate only two network measures (Callon's centrality and density) whereas The Network Workbench Tool and Science of Science Tool are able to add more network measures. Taking into account visualization capabilities, VOSviewer (van Eck & Waltman, 2010) has a powerful GUI that allows us to easily examine the generated maps. Similarly, Network Workbench Tool and Science of Science Tool allow us to configure the visual output with different scripts.

The main differences of SciMAT with respect to other science mapping tools are: (a) the capability to choose the methods, algorithms, and measures used to perform the analysis through the configuration wizard; (b) the use of impact measures to quantify the results; (c) the ability to perform all the steps of the science mapping workflow (Figures 1 and 5); (d) the integration of the whole process in a longitudinal framework; and (e) the methodological foundation, due to SciMAT being based on an extension of the science mapping approach presented in Cobo et al. (2011a).

Finally, SciMAT has been tested and improved according to the suggestions and comments of a wide variety of potential users, including senior researchers, PhD students, heads of research groups, and technical staff of the Research and Policy Research Office of the University of Granada.

Acknowledgments

References

Alonso, S., Cabrerizo, F., Herrera-Viedma, E., & Herrera, F. (2009). h-index: A review focused in its variants, computation and standardization for different scientific fields. Journal of Informetrics, 3(4), 273–289.

Alonso, S., Cabrerizo, F., Herrera-Viedma, E., & Herrera, F. (2010). hg-index: A new index to characterize the scientific output of researchers based on the h- and g-indices. Scientometrics, 82(2), 391–400.

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.

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. In Proceedings of the International Association for the Advancement of Artificial Intelligence (AAAI) Conference on Weblogs and Social Media (pp. 361–362). Available at http://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154

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 for Windows: Software for social network analysis. Harvard, MA: Analytic Technologies.

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., et al. (2010). Rete-netzwerk-red: Analyzing and visualizing scholarly networks using the network workbench tool. Scientometrics, 83(3), 863–876.

Cabrerizo, F.J., Alonso, S., Herrera-Viedma, E., & Herrera, F. (2010). q^2-index: Quantitative and qualitative evaluation based on the number and impact of papers in the Hirsch core. Journal of Informetrics, 4(1), 23–28.

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, No. 28. New York: Cambridge University Press.

Chen, C. (2004). Searching for intellectual turning points: Progressive knowledge domain visualization. Proceedings of the National Academy of Sciences, USA, 101(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.

Chin, J., Diehl, V., & Norman, K. (1988). Development of an instrument measuring user satisfaction of the human–computer interface. In Proceedings of the (SIGCHI) Conference on Human Factors in Computing Systems (pp. 213–218). New York: ACM Press.

Cobo, M.J., López-Herrera, A.G., Herrera, F., & Herrera-Viedma, E. (2012). A note on the ITS topic evolution in the period 2000–2009 at T-ITS. IEEE Transactions on Intelligent Transportation Systems, 13(1), 413–420.

Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E., & Herrera, F. (2011a). 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.

Cobo, M.J., López-Herrera, A.G., Herrera-Viedma, E., & Herrera, F. (2011b). Science mapping software tools: Review, analysis and cooperative study among tools. Journal of the American Society for Information Science and Technology, 62(7), 1382–1402.

Cook, D.J., & Holder, L.B. (2006). Mining graph data. Hoboken, NJ: John Wiley.

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.

Egghe, L. (2006). Theory and practise of the g-index. Scientometrics, 69(1), 131–152.

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.

Gänzel, W. (2001). National characteristics in international scientific co-authorship relations. Scientometrics, 51(1), 69–115.

Garfield, E. (1994). Scientography: Mapping the tracks of science. Current Contents: Social & Behavioral 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.

Hirsch, J. (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences, USA, 102, 16569–16572.

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), Article No. 7. doi:10.1145/1754428.1754430

Kessler, M.M. (1963). Bibliographic coupling between scientific papers. American Documentation, 14(1),10–25.

Kreibich, J. (2010). Using SQLite. Sebastopol, CA: O'Reilly Media.

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.

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. International 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 No. 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.

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 Fernndez, 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.

Noyons, E.C.M., Moed, H.F., & Luwel, M. (1999). 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. (1999). 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-organizing-maps 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.

Porter, A.L., & Youtie, J. (2009). Where does nanotechnology belong in the map of science? Nature Nanotechnology, 4, 534–536.

Price, D., & Gürsey, 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.

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). NewYork: Academic Press.

Sci²Team. (2009). Science of Science (Sci²) Tool. Indiana University and SciTech Strategies. Available at: http://sci.slis.indiana.edu

Shannon, P., Markiel, A., Ozier, O., Baliga, N., Wang, J., Ramage, D., et al. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. Genome Research, 13(11), 2498–2504.

Shneiderman, B., Plaisant, C., Cohen, M., & Jacobs, S. (2009). Designing the user interface: Strategies for effective human–computer interaction. Reading, MA: Addison-Wesley.

Skillicorn, D. (2007). Understanding complex datasets: Data mining with matrix decompositions. Data Mining and Knowledge Discovery Series: Boca Raton, FL: 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. (1977). A co-citation model of a scientific specialty: A longitudinal study of collagen research. Social Studies of Science, 7, 139–166.

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 & Management, 13(5), 277–288.

Small, H., & Sweeney, E. (1985). Clustering the Science Citation Index using co-citations. Scientometrics, 7(3), 391–409.

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.

Vilar, P. (2010). Designing the user interface: Strategies for effective human–computer interaction. Journal of the American Society for Information Science and Technology, 61(5), 1073–1074.

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.

Appendix

Questionnaire Used for Validating SciMAT

OVERALL REACTION TO THE SOFTWARE		0	1	2	3	4	5	6	7	8	9		NA
1. **SciMAT** itself	terrible											wonderful	
2. Starting up	difficult											easy	
3. Perceived utility	frustrating											satisfying	
4. Obtaining Results	inadequate power											adequate power	
5. Setting the analysis	dull											stimulating	
6. Analysis capabilities	rigid											flexible	
SCREEN		0	1	2	3	4	5	6	7	8	9		NA
7. Reading characters on the screen	hard											easy	
8. Highlighting simplifies task	not at all											very much	
9. Organization of information	confusing											very clear	
10. Sequence of screens	confusing											very clear	
TERMINOLOGY AND SYSTEM INFORMATION		0	1	2	3	4	5	6	7	8	9		NA
11. Use of terms throughout **SciMAT**	inconsistent											consistent	
12. Terminology related to task	never											always	
13. Position of messages on screen	inconsistent											consistent	
14. Prompts for input	confusing											clear	
15. Computer informs about its progress	never											always	
16. Error messages	unhelpful											helpful	
LEARNING		0	1	2	3	4	5	6	7	8	9		NA
17. Learning to operate **SciMAT**	difficult											easy	
18. Exploring new features by trial and error	difficult											easy	
19. Remembering names and use of commands	difficult											easy	
20. Performing tasks is straightforward	never											always	
21. Help messages on the screen	unhelpful											helpful	
22. Supplemental reference materials	confusing											clear	
SciMAT CAPABILITIES		0	1	2	3	4	5	6	7	8	9		NA
23. **SciMAT** speed	too slow											fast enough	
24. **SciMAT** reliability	unreliable											reliable	
25. **SciMAT** tends to be	noisy											quiet	
26. Correcting your mistakes	difficult											easy	
27. Designed for all levels of users	never											always	

Negative Comments:

Positive Comments:

FIG. A1.   Questionnaire used for validating SciMAT. It is an adapted version of the Questionnaire for User Interface Satisfaction (Chin et al., 1988).