Chapter 13
Visualizing Bibliometric Networks

Nees Jan van Eck and Ludo Waltman

Abstract This chapter provides an introduction to the topic of visualizing bibliometric networks. First, the most commonly studied types of bibliometric networks (i.e., citation, co-citation, bibliographic coupling, keyword co-occurrence, and coauthorship networks) are discussed, and three popular visualization approaches (i.e., distance-based, graph-based, and timeline-based approaches) are distinguished. Next, an overview is given of a number of software tools that can be used for visualizing bibliometric networks. In the second part of the chapter, the focus is specifically on two software tools: VOSviewer and CitNetExplorer. The techniques used by these tools to construct, analyze, and visualize bibliometric networks are discussed. In addition, tutorials are offered that demonstrate in a step-by-step manner how both tools can be used. Finally, the chapter concludes with a discussion of the limitations and the proper use of bibliometric network visualizations and with a summary of some ongoing and future developments.

13.1 Introduction

The idea of visualizing bibliometric networks, often referred to as "science mapping," has received serious attention since the early days of bibliometric research. Visualization has turned out to be a powerful approach to analyze a large variety of bibliometric networks, ranging from networks of citation relations between publications or journals to networks of coauthorship relations between researchers or networks of co-occurrence relations between keywords. Over time, researchers have started to analyze larger and larger networks, leading to the need for more advanced visualization techniques and tools. At the same time, professional users of bibliometrics, for instance research institutions, funding agencies, and publishers, have become more and more interested in bibliometric network visualizations. To make bibliometric network visualizations available to a wider public, both inside

N.J. van Eck (✉) • L. Waltman
Centre for Science and Technology Studies, Leiden University, Leiden, The Netherlands
e-mail: ecknjpvan@cwts.leidenuniv.nl; waltmanlr@cwts.leidenuniv.nl

and outside the bibliometric research community, researchers have developed a number of software tools, most of which are freely available. In this chapter, we focus mainly on two of these tools: VOSviewer and CitNetExplorer. VOSviewer is a tool that we have developed over the past few years and that offers in a relatively easy way the basic functionality needed for visualizing bibliometric networks. CitNetExplorer is a more specialized tool that we developed recently for visualizing and analyzing citation networks of publications.

This chapter offers an introduction to the topic of visualizing bibliometric networks. It first provides an overview of the literature and of the main software tools that are available. It then focuses specifically on the VOSviewer and CitNetExplorer tools. The most important techniques used by these tools are discussed, and tutorials are offered that provide step-by-step instructions on how the tools can be used. The chapter concludes with a discussion of the limitations and the proper use of bibliometric network visualizations and with a summary of some ongoing and future developments.

13.2 Literature Review

We first provide a brief overview of the literature on visualizing bibliometric networks. We start by discussing the types of bibliometric networks that have received most attention in the literature. We then discuss a number of commonly used visualization approaches. We refer to Börner, Chen, and Boyack (2003) and Börner (2010) for more extensive overviews of the literature.

13.2.1 Types of Bibliometric Networks

A bibliometric network consists of nodes and edges. The nodes can be, for instance, publications, journals, researchers, or keywords. The edges indicate relations between pairs of nodes. The most commonly studied types of relations are citation relations, keyword co-occurrence relations, and coauthorship relations. In the case of citation relations, a further distinction can be made between direct citation relations, co-citation relations, and bibliographic coupling relations. Bibliometric networks are usually weighted networks. Hence, edges indicate not only whether there is a relation between two nodes or not but also the strength of the relation. Below, we discuss the different types of relations in bibliometric networks in more detail.

Two publications are co-cited if there is a third publication that cites both publications (Marshakova, 1973; Small, 1973). The larger the number of publications by which two publications are co-cited, the stronger the co-citation relation between the two publications. Small and colleagues proposed to use co-citations to analyze and visualize relations between publications (Griffith, Small, Stonehill, &

Dey, 1974; Small & Griffith, 1974). Later on, the use of co-citations to study relations between researchers and between journals was introduced by, respectively, White and Griffith (1981) and McCain (1991). A well-known co-citation-based analysis is the study by White and McCain (1998) of researchers in the field of information science.

Bibliographic coupling is the opposite of co-citation. Two publications are bibliographically coupled if there is a third publication that is cited by both publications (Kessler, 1963). In other words, bibliographic coupling is about the overlap in the reference lists of publications. The larger the number of references two publications have in common, the stronger the bibliographic coupling relation between the publications. Although bibliographic coupling was introduced earlier than co-citation, it initially received less attention in the literature on visualizing bibliometric networks. In more recent years, however, the popularity of bibliographic coupling increased considerably (e.g., Boyack & Klavans, 2010; Jarneving, 2007; Small, 1997; Zhao & Strotmann, 2008).

Compared with co-citation and bibliographic coupling, direct citations, sometimes referred to as cross citations, offer a more direct indication of the relatedness of publications. Nevertheless, in the literature on visualizing bibliometric networks, it is relatively uncommon to work with direct citations. This is probably because the use of direct citations often leads to very sparse networks (i.e., networks with only a very small number of edges). In spite of this issue, there seems to be an increasing interest in direct citations in the more recent literature (e.g., Boyack & Klavans, 2010; Persson, 2010; Small, 1997; Waltman & Van Eck, 2012). Direct citations also play an essential role in Eugene Garfield's work on algorithmic historiography (Garfield, Pudovkin, & Istomin, 2003). We will come back to Garfield's work in Sect. 13.3.

In addition to the above-discussed citation-based bibliometric networks, networks of co-occurrences of keywords have also been studied extensively. Keywords can be extracted from the title and abstract of a publication, or they can be taken from the author-supplied keyword list of a publication. In some cases, especially in the older literature, keywords are restricted to individual words, but in other cases they also include terms consisting of multiple words. The number of co-occurrences of two keywords is the number of publications in which both keywords occur together in the title, abstract, or keyword list. For examples of early work on keyword co-occurrence networks, we refer to Callon, Courtial, Turner, and Bauin (1983), Callon, Law, and Rip (1986), and Peters and Van Raan (1993).

Finally, we briefly mention bibliometric networks based on coauthorship. In these networks, researchers, research institutions, or countries are linked to each other based on the number of publications they have authored jointly. Coauthorship networks have been studied extensively, but relatively little attention has been paid to the visualization of these networks. At the level of countries, the visualization of coauthorship networks is discussed by Luukkonen, Tijssen, Persson, and Sivertsen (1993).

13.2.2   Visualization Approaches

Many different approaches have been proposed for visualizing bibliometric networks. Our focus in this chapter is on three popular approaches. We refer to these approaches as the distance-based approach, the graph-based approach, and the timeline-based approach. We emphasize that these three approaches are definitely not the only approaches available. Alternative approaches for instance include circular visualizations (e.g., Börner et al., 2012) and self-organizing maps (e.g., Skupin, Biberstine, & Börner, 2013).

In the distance-based approach, the nodes in a bibliometric network are positioned in such a way that the distance between two nodes approximately indicates the relatedness of the nodes. In general, the smaller the distance between two nodes, the higher their relatedness. Nodes are usually positioned in a two-dimensional space. Edges between nodes are normally not shown. The most commonly used technique for determining the locations of the nodes in a distance-based visualization is multidimensional scaling (e.g., Borg & Groenen, 2005). The use of this technique for visualizing bibliometric networks has a long tradition, going back to early work on the visualization of co-citation networks by Griffith et al. (1974) and White and Griffith (1981). More recently, some alternatives to multidimensional scaling were introduced in the literature. One alternative is the VOS technique (Van Eck, Waltman, Dekker, & Van den Berg, 2010), which is used in the VOSviewer software that we will discuss in detail in the second part of this chapter. Another alternative is the VxOrd technique, also known as DrL or OpenOrd (http://www.sandia.gov/~smartin/software.html). The VxOrd technique has been used to create distance-based visualizations of very large bibliometric networks (e.g., Boyack, Klavans, & Börner, 2005; Klavans & Boyack, 2006). An example of a distance-based visualization is presented in Fig. 13.1. This example has been taken from a well-known study by White and McCain (1998). The visualization displays a co-citation network of researchers in the field of information science.

In the graph-based approach, nodes are positioned in a two-dimensional space, just like in the distance-based approach. The difference between the two approaches is that in the graph-based approach edges are displayed to indicate the relatedness of nodes. The distance between two nodes need not directly reflect their relatedness. The graph-based approach is most suitable for visualizing relatively small networks. Visualizing larger networks using the graph-based approach often does not give good results because of the large number of edges that need to be displayed. The most commonly used technique for creating graph-based visualizations of bibliometric networks is the graph drawing algorithm of Kamada and Kawai (1989). An alternative technique is the algorithm of Fruchterman and Reingold (1991). Graph drawing algorithms are sometimes used in combination with the pathfinder network technique for graph pruning (Schvaneveldt, Dearholt, & Durso, 1988). For examples of graph-based visualizations of bibliometric networks, we refer to Chen (1999), De Moya-Anegón et al. (2007), Leydesdorff and Rafols (2009), and White (2003). Figure 13.2 shows an example of a graph-based
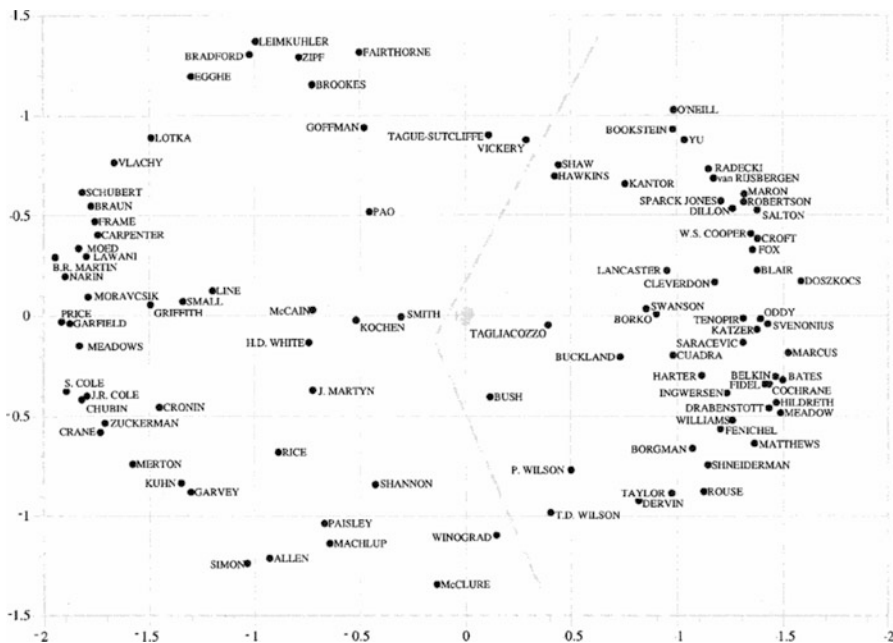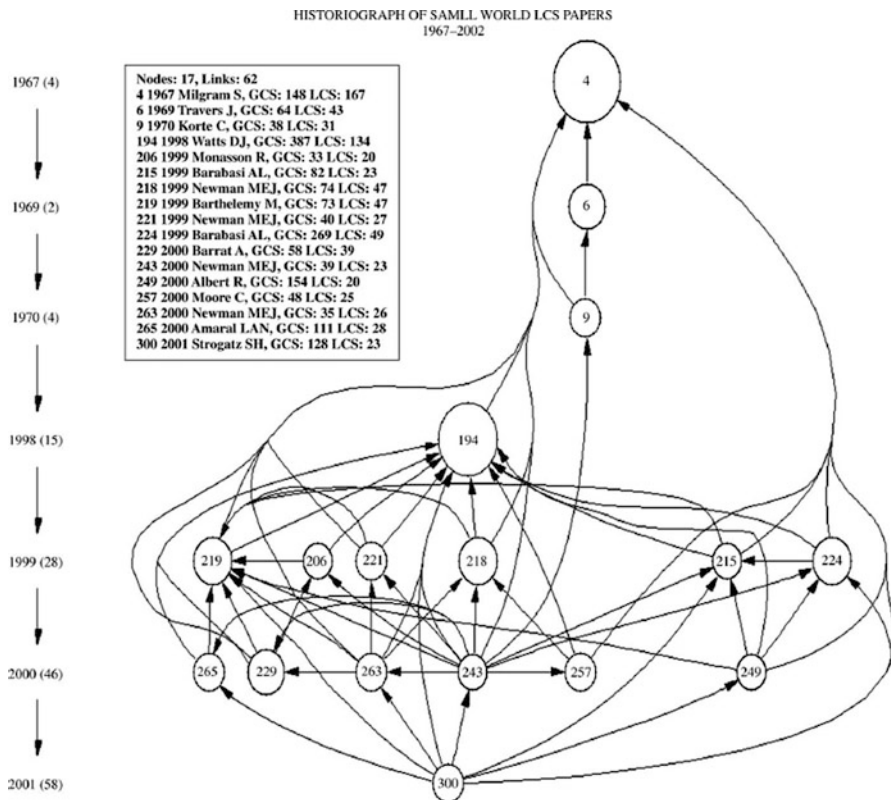
Fig. 13.1 Example of a distance-based visualization. Co-citation network of researchers in the field of information science. *Source*: White & McCain (1998, Fig. 4). Copyright 1998 by John Wiley & Sons, Inc. Reprinted with permission

visualization. This visualization displays the same researcher co-citation network as the visualization presented in Fig. 13.1, but instead of a distance-based approach a graph-based approach is used. The example shown in Fig. 13.2 has been taken from White (2003).

A third approach for visualizing bibliometric networks is the timeline-based approach. Unlike the distance-based and graph-based approaches, the timeline-based approach assumes that each node in a bibliometric network can be linked to a specific point in time. The timeline-based approach is especially suitable for visualizing networks of publications, since a publication can be easily linked to a specific point in time based on its publication date. In a timeline-based visualization, there are two dimensions, one of which is used to represent time. The other dimension can be used to represent the relatedness of nodes. The location of a node in the time-dimension is determined by the specific point in time to which the node is linked. The location of a node in the other dimension can be determined based on the relatedness of the node to other nodes. Timeline-based visualizations have for instance been used by Chen (2006), Garfield et al. (2003), and Morris, Yen, Wu, and Asnake (2003). They are also used in the CitNetExplorer software that will be discussed later on in this chapter. An example of a timeline-based visualization is provided in Fig. 13.3. The example, taken from Garfield (2004), displays a citation network of publications on the so-called small world phenomenon.

Fig. 13.2 Example of a graph-based visualization. Co-citation network of researchers in the field of information science. *Source*: White (2003, Fig. 1). Copyright 2003 by Wiley Periodicals, Inc. Reprinted with permission

13.3 Software Tools

There are many software tools that can be used for visualizing bibliometric networks. Some of these tools are general statistical or network analysis tools. Other tools have been developed specifically for visualizing bibliometric networks. In this section, we do not aim to give a comprehensive overview of all available tools. Instead, we have selected a number of tools that seem to be among the most important or most popular ones. We briefly discuss the main features of each of these tools. We refer to Cobo, López-Herrera, Herrera-Viedma, and Herrera (2011) for another recent overview of software tools for visualizing bibliometric networks. This overview includes a number of tools that are not considered in this section.

Below, we first discuss two general network analysis tools (Pajek and Gephi). We then discuss three tools (CiteSpace, Sci2, and VOSviewer) that have been developed specifically for analyzing and visualizing bibliometric networks. Finally, we consider two tools (HistCite and CitNetExplorer) that focus entirely on the analysis and visualization of one specific type of bibliometric network, namely citation networks of publications. The tools that we discuss are all freely available, at least for specific types of use. Some tools are open source, but most are not. Table 13.1 provides for each tool the URL of its website.

HISTORIOGRAPH OF SAMLL WORLD LCS PAPERS
1967–2002

Nodes: 17, Links: 62
4 1967 Milgram S, GCS: 148 LCS: 167
6 1969 Travers J, GCS: 64 LCS: 43
9 1970 Korte C, GCS: 38 LCS: 31
194 1998 Watts DJ, GCS: 387 LCS: 134
206 1999 Monasson R, GCS: 33 LCS: 20
215 1999 Barabasi AL, GCS: 82 LCS: 23
218 1999 Newman MEJ, GCS: 74 LCS: 47
219 1999 Barthelemy M, GCS: 73 LCS: 47
221 1999 Newman MEJ, GCS: 40 LCS: 27
224 1999 Barabasi AL, GCS: 269 LCS: 49
229 2000 Barrat A, GCS: 58 LCS: 39
243 2000 Newman MEJ, GCS: 39 LCS: 23
249 2000 Albert R, GCS: 154 LCS: 20
257 2000 Moore C, GCS: 48 LCS: 25
263 2000 Newman MEJ, GCS: 35 LCS: 26
265 2000 Amaral LAN, GCS: 111 LCS: 28
300 2001 Strogatz SH, GCS: 128 LCS: 23

Fig. 13.3 Example of a timeline-based visualization. Citation network of publications on the small world phenomenon. *Source*: Garfield (2004, Fig. 12). Copyright 2004 by SAGE Publications. Reprinted with permission

Table 13.1 URLs of the websites of a number of software tools that can be used for visualizing bibliometric networks

CitNetExplorer	http://www.citnetexplorer.nl
CiteSpace	http://cluster.cis.drexel.edu/~cchen/citespace/
Gephi	http://www.gephi.org
HistCite	http://www.histcite.com
Pajek	http://pajek.imfm.si
Sci2	https://sci2.cns.iu.edu
VOSviewer	http://www.vosviewer.com

13.3.1   General Network Analysis Tools

Two general network analysis tools that have been used for visualizing bibliometric networks are Pajek (De Nooy, Mrvar, & Batagelj, 2005) and Gephi (Bastian, Heymann, & Jacomy, 2009). Since these are general network analysis tools, they have no specific functionality for processing bibliographic data, for instance for extracting bibliometric networks from output files of the Web of Science bibliographic database. Other software tools are required to take care of this. In the case of Pajek, a tool called WoS2Pajek is available for this purpose. In the case of Gephi, Sci2 (discussed below) can be used.

Pajek is a popular software tool for social network analysis. Many different techniques for social network analysis are available in Pajek. From a bibliometric perspective, important techniques offered by Pajek include clustering and main path analysis (Hummon & Doreian, 1989). The visualization capabilities of Pajek are somewhat limited in comparison with some of the tools discussed below. Pajek provides graph-based visualizations, for instance based on the techniques of Kamada and Kawai (1989) and Fruchterman and Reingold (1991). An interesting feature of Pajek is the support it offers for three-dimensional visualizations. Recently, a link to VOSviewer (discussed below) was included in Pajek, allowing networks analyzed in Pajek to be visualized in VOSviewer.

Compared with Pajek, Gephi is focused less on network analysis and more on network visualization. Gephi offers extensive visualization capabilities, making it possible to customize visualizations in great detail. Gephi also supports various network analysis techniques, but not as many as Pajek. A comparison between the use of Gephi and the use of VOSviewer (discussed below) for visualizing bibliometric networks is made by Leydesdorff and Rafols (2012).

13.3.2   Tools for Analyzing and Visualizing Bibliometric Networks

We now discuss three software tools that have been developed specifically for analyzing and visualizing bibliometric networks: CiteSpace (Chen, 2004, 2006), Sci2, and our own VOSviewer (Van Eck & Waltman, 2010). All three tools can take Web of Science output files as input.

Unlike other software tools, CiteSpace has a strong focus on dynamic visualizations that show how bibliometric networks evolve over time. CiteSpace offers both graph-based and timeline-based visualizations. It also provides users with a large range of options. Because of the many options that are available, the user interface of CiteSpace may appear a bit overwhelming to new users, and it may take some effort to learn how to use the tool.

Sci2 is a general software tool for analyzing bibliometric networks. Many different analysis techniques are available in Sci2. Some of these techniques are

implemented in external programs that have been linked to Sci2. To visualize a bibliometric network, Sci2 also relies on external software tools. By default, Sci2 uses a tool called GUESS (http://graphexploration.cond.org). Using this tool, graph-based visualizations of bibliometric networks are provided. Two other net-work visualization tools that can be used together with Sci2 are Gephi (discussed above) and Cytoscape (http://www.cytoscape.org).

Our aim with VOSviewer is to offer an easy-to-use software tool that is completely focused on the visualization of bibliometric networks. Unlike the above-discussed tools, VOSviewer provides distance-based visualizations of bibliometric networks. By default, VOSviewer therefore displays only the nodes in a bibliometric network and does not display the edges between the nodes. In the visualizations provided by VOSviewer, the distance between two nodes approxi-mately indicates the relatedness of the nodes. By providing distance-based visual-izations rather than graph-based ones, VOSviewer is especially suitable for visualizing larger networks. Because of its strong focus on visualization, VOSviewer offers less functionality for analyzing bibliometric networks than other tools. However, VOSviewer does have some special text mining features. VOSviewer will be discussed in more detail in the second part of this chapter.

13.3.3   Tools for Analyzing and Visualizing Publication Citation Networks

Finally, we consider two software tools that focus exclusively on the analysis and visualization of networks of direct citation relations between publications. These tools are Eugene Garfield's HistCite (Garfield et al., 2003) and our own recently developed CitNetExplorer. Both tools offer timeline-based visualizations of publi-cation citation networks.

HistCite takes Web of Science output files as input. Based on the bibliographic data in these files, various statistics on publications, researchers, journals, etc. are presented. The bibliographic data can also be edited, for instance to correct errors and to add missing data elements. After the bibliographic data has been processed, the next step is to visualize the publication citation network. A visualization of a publication citation network is referred to as a historiograph in HistCite, and the algorithmic construction of such a visualization is called algorithmic historiography (Garfield et al., 2003). By default, HistCite includes the 30 most frequently cited publications in the visualization of a publication citation network. An example of a visualization produced by HistCite is shown in Fig. 13.3.

Compared with HistCite, CitNetExplorer offers more extensive analysis and visualization capabilities. Most importantly, CitNetExplorer provides functionality for drilling down into a publication citation network. This functionality makes it possible to explore a publication citation network in an interactive fashion. Initially, for instance, we may start with a large network that includes several millions of

publications. We may then gradually drill down into this network until for instance we have reached a small subnetwork that includes no more than, say, 100 publications, all dealing with a specific topic that we are interested in. Like HistCite, CitNetExplorer is able to process Web of Science output files. CitNetExplorer will be discussed in more detail in the next two sections.

13.4   Techniques

In this section, we discuss the main techniques used in the two software tools on which we focus in the rest of this chapter: VOSviewer and CitNetExplorer. We first consider VOSviewer and then CitNetExplorer. The discussion relates to version 1.5.5 of VOSviewer and version 1.0.0 of CitNetExplorer. Readers who are interested mainly in the practical application of the two tools and not so much in the underlying techniques may skip this section and proceed to the next one.

13.4.1   VOSviewer

VOSviewer takes a distance-based approach to visualizing bibliometric networks. Any type of bibliometric network can be visualized. Directed networks, for instance networks based on direct citation relations, are treated as undirected. Networks may consist of several thousand nodes. Because of computational limitations and memory constraints, networks with more than 10,000 nodes tend to be difficult to handle in VOSviewer.

13.4.1.1   Normalization, Mapping, and Clustering

In a bibliometric network, there are often large differences between nodes in the number of edges they have to other nodes. Popular nodes, for instance representing highly cited publications or highly prolific researchers, may have several orders of magnitude more connections than their less popular counterparts. In the analysis of bibliometric networks, one usually performs a normalization for these differences between nodes. VOSviewer by default applies the association strength normalization. This normalization is discussed in detail by Van Eck and Waltman (2009).

After a normalized network has been constructed, the next step is to position the nodes in the network in a two-dimensional space in such a way that strongly related nodes are located close to each other while weakly related nodes are located far away from each other. For this purpose, VOSviewer uses the VOS mapping technique, where VOS stands for "visualization of similarities." A detailed discussion of the VOS mapping technique is provided by Van Eck et al. (2010).

VOSviewer by default also assigns the nodes in a network to clusters. A cluster is a set of closely related nodes. Each node in a network is assigned to exactly one cluster. The number of clusters is determined by a resolution parameter. The higher the value of this parameter, the larger the number of clusters. In the visualization of a bibliometric network, VOSviewer uses colors to indicate the cluster to which a node has been assigned. The clustering technique used by VOSviewer is discussed by Waltman, Van Eck, and Noyons (2010). The technique requires an algorithm for solving an optimization problem. For this purpose, VOSviewer uses the smart local moving algorithm introduced by Waltman and Van Eck (2013).

We refer to the Appendix for a technical summary of the normalization, mapping, and clustering techniques used by VOSviewer.

13.4.1.2   Displaying a Bibliometric Network

After the nodes in a bibliometric network have been positioned in a two-dimensional space and have been assigned to clusters, the network can be displayed. VOSviewer uses various techniques to optimize the way in which networks are displayed. In order to ensure that labels of nodes do not overlap each other, labels are displayed only for a selection of all nodes. This selection is determined in such a way that as many labels as possible are displayed while labels of more important nodes (i.e., nodes that have more edges) are given priority over labels of less important nodes. Like computer software for exploring geographical maps (e.g., Google Maps), VOSviewer offers zooming and panning (scrolling) functionality. This is especially useful for exploring larger networks consisting of hundreds or thousands of nodes. When zooming in, the selection of nodes for which labels are displayed is updated and labels that previously were not shown may become visible.

VOSviewer also supports overlay visualizations. In an overlay visualization, the color of a node indicates a certain property of the node. For instance, nodes may represent journals and the color of a node may indicate the number of times a journal has been cited. We refer to Van Eck, Waltman, Van Raan, Klautz, and Peul (2013) for an example of the use of overlay visualizations in VOSviewer. Another visualization supported by VOSviewer is the density visualization. In this visualization, colors indicate how nodes are distributed in the two-dimensional space underlying the visualization. The density visualization allows one to immediately identify dense areas in which many nodes are located close to each other. We refer to Van Eck and Waltman (2010) for a discussion of the technical details of the density visualization.

13.4.1.3   Fractional Counting Methodology

As will be demonstrated in Sect. 13.5, VOSviewer is able to process output files of the Web of Science bibliographic database. Based on these files, VOSviewer can

construct co-citation and bibliographic coupling networks of publications, journals, and researchers. A special feature of VOSviewer is the possibility to construct co-citation and bibliographic coupling networks using a fractional counting methodology. To explain this methodology, we take bibliographic coupling networks of publications as an example, but the methodology works in a similar way for other types of networks.

Suppose we have 100 publications that all cite the same publication. Using an ordinary full counting methodology, each of the 100 publications has a bibliographic coupling relation with a weight of 1 with each of the 99 other publications. This yields a total bibliographic coupling weight of $99 \times 1 = 99$ for each of the 100 publications. Using the fractional counting methodology, on the other hand, each of the 100 publications has a bibliographic coupling relation with a weight of 1/99 with each of the 99 other publications, yielding a total bibliographic coupling weight of $99 \times (1/99) = 1$ for each publication. In other words, using the fractional counting methodology, giving a citation to a publication always results in a total bibliographic coupling weight of 1, irrespective of the number of other publications that also cite the same publication (although there must of course be at least one other citing publication). Hence, in the case of the fractional counting methodology, highly cited publications play a less important role in the construction of a bibliographic coupling network. In the same way, publications with a long reference list (e.g., review articles) play a less important role in the construction of a co-citation network. In general, we recommend the use of VOSviewer's fractional counting methodology instead of the ordinary full counting methodology.

13.4.1.4   Text Mining Techniques

Finally, we briefly mention VOSviewer's text mining functionality for constructing co-occurrence networks of terms extracted from English-language textual data, for instance from titles and abstracts of publications (Van Eck & Waltman, 2011). VOSviewer relies on the Apache OpenNLP toolkit (http://opennlp.apache.org) to perform part-of-speech tagging (i.e., to identify verbs, nouns, adjectives, and so on). It then uses a linguistic filter to identify noun phrases. The filter selects all word sequences that consist exclusively of nouns and adjectives and that end with a noun. Plural noun phrases are converted into singular ones. Some noun phrases (e.g., "conclusion," "interesting result," and "new method") are very general, and one usually does not want these noun phrases to be included in one's co-occurrence network. VOSviewer therefore calculates for each noun phrase a relevance score. Essentially, noun phrases have a low relevance score if their co-occurrences with other noun phrases follow a more or less random pattern, while they have a high relevance score if they co-occur mainly with a limited set of other noun phrases. Noun phrases with a low relevance score tend to be quite general, while noun phrases with a high relevance score typically have a more specific meaning. VOSviewer allows one to leave out noun phrases with a low relevance score. In this way, one gets rid of many general noun phrases. The remaining noun phrases

usually represent relevant terms in the domain of interest. VOSviewer can be used to visualize the co-occurrence network of these terms.

13.4.2 CitNetExplorer

CitNetExplorer visualizes networks of direct citation relations between publications. A timeline-based approach is taken. CitNetExplorer supports very large networks. Networks may include millions of publications and citation relations. The maximum size of the networks that can be handled by CitNetExplorer depends mainly on the amount of computer memory that is available.

13.4.2.1 Constructing a Publication Citation Network

CitNetExplorer is able to construct a publication citation network based on Web of Science output files. Web of Science output files contain bibliographic data on publications, but they do not directly indicate the citation relations that exist between these publications. For each publication in a Web of Science output file, the list of cited references is given. To find out which publications cite which other publications, the cited references in a Web of Science file need to be matched with the publications in the file. For some cited references, it will not be possible to match them with a publication. This usually means that these cited references point to publications that are not included in the Web of Science file. For other cited references, matching will be possible. Cited references that can be matched with a publication indicate the existence of a citation relation between two publications in the Web of Science file.

Citation matching can be performed in different ways. CitNetExplorer first attempts to match based on DOI. However, DOI data often is not available. In that case, matching is done based on first author name (last name and first initial only), publication year, volume number, and page number. A perfect match is required for each of these data elements. Data on the title of the cited journal usually is available as well, but this data is not used by CitNetExplorer. This is because in many cases the title of a journal is not written consistently in the same way.

CitNetExplorer assumes publication citation networks to be acyclic. This for instance means that it is not allowed to have both a citation from publication A to publication B and a citation from publication B to publication A. Likewise, it is not allowed to have a citation from publication A to publication B, a citation from publication B to publication C, and a citation from publication C to publication A. In other words, when moving through a publication citation network by following citation relations from one publication to another, one should never get back again at a publication that has already been visited. In practice, publication citation networks are not always perfectly acyclic. There may for instance be publications in

the same issue of a journal that mutually cite each other. CitNetExplorer solves this problem by checking whether a publication citation network is acyclic and by removing citation relations that cause the network to have cycles. CitNetExplorer also removes all citation relations that point forward in time, for instance a publication from 2013 citing a publication from 2014.

13.4.2.2 Displaying a Publication Citation Network

If a publication citation network consists of more than 50 or 100 publications, displaying all publications and all citation relations is typically of little or no use. There will usually be lots of citation relations, and when displaying these citation relations, many of them will inevitably cross each other, leading to a visualization that is hard to interpret. For this reason, in the case of a larger network, CitNetExplorer displays only a selection of all publications. By default, this selection consists of the 40 most frequently cited publications in the network, but the selection of publications to be displayed may be changed by the user. To keep things simple, in the discussion below, we assume that we are dealing with a small network and that all publications in the network are displayed. Larger networks are visualized in the same way as described below, with the exception that only a selection of all publications are included in the visualization.

In the timeline-based visualization of CitNetExplorer, the vertical dimension is used to represent time, with more recent years being located below older years. Publications are positioned in the vertical dimension based on the year in which they appeared. The vertical dimension is organized into layers, each of equal height. A year is represented by at least one layer, but some years may be represented by multiple layers. Multiple layers are used if there are citation relations between publications from the same year. The horizontal dimension in the timeline-based visualization of CitNetExplorer is used to provide an indication of the relatedness of publications. In general, publications that are strongly related to each other, based on citation relations, are positioned close to each other in the horizontal dimension.

Positioning the publications in a publication citation network in the horizontal and vertical dimensions of a timeline-based visualization is a hierarchical graph drawing problem. Following the literature on hierarchical graph drawing (e.g., Healy & Nikolov, 2013), CitNetExplorer first assigns each publication to a layer in the vertical dimension. This is done based on the year in which a publication appeared. In addition, publications are assigned to layers in such a way that citations always flow in an upward direction in the visualization. In other words, for any citation relation, the layer to which the citing publication is assigned must be located below the layer of the cited publication. After each publication has been assigned to a layer, CitNetExplorer positions the publications in the horizontal dimension.

Publications are assigned to layers in the vertical dimension in a year-by-year fashion. If the publications from a given year do not cite each other, they are all assigned to the same layer, unless the number of publications exceeds the maximum

number of publications per layer. By default, the maximum number of publications per layer is 10. If the publications from a given year do cite each other, they are assigned to multiple layers in such a way that the layer of a citing publication is always located below the layer of the corresponding cited publication. This ensures that citations always flow in an upward direction. Assigning publications to multiple layers is done using a simple heuristic algorithm that aims to minimize the number of layers that are needed.

Publications are positioned in the horizontal dimension based on their citation relations with other publications. In addition to direct citation relations, various types of indirect citation relations (e.g., co-citation and bibliographic coupling relations) are taken into account as well. In general, the higher the relatedness of two publications, the closer the publications are located to each other in the horizontal dimension. However, publications that have been assigned to the same layer in the vertical dimension must have at least a certain minimum distance from each other in the horizontal dimension. This is done in order to minimize the problem of overlapping publications in the visualization. The technique used by CitNetExplorer to position publications in the horizontal dimension is fairly similar to the VOS mapping technique discussed in Sect. 13.4.1 and in the Appendix.

To optimize the way in which a publication citation network is displayed, CitNetExplorer uses similar techniques as VOSviewer. CitNetExplorer labels publications by the last name of the first author. To prevent labels from overlapping each other, labels may sometimes be displayed only for a selection of all publications. Like VOSviewer, CitNetExplorer offers zooming and panning functionality.

13.4.2.3 Analysis Techniques

CitNetExplorer offers various techniques for analyzing publication citation networks. At the moment, techniques are available for extracting connected components, for clustering publications, for identifying core publications, and for finding shortest and longest paths between publications. More techniques are expected to become available in future versions of CitNetExplorer.

Clustering of publications is accomplished following the methodology proposed by Waltman and Van Eck (2012). This methodology optimizes a variant of the modularity function of Newman and Girvan (2004) and Newman (2004). The level of detail of the clustering is determined by a resolution parameter. Like VOSviewer, CitNetExplorer uses the optimization algorithm introduced by Waltman and Van Eck (2013).

The identification of core publications is based on the idea of k-cores introduced by Seidman (1983). A core publication is a publication that has at least a certain minimum number of citation relations with other core publications. Both incoming and outgoing citation relations are counted. Hence, for the purpose of identifying core publications, a publication citation network is treated as undirected. The identification of core publications makes it possible to get rid of unimportant publications in the periphery of a publication citation network.

13.5 Tutorials

We now offer two tutorials, one for VOSviewer and one for CitNetExplorer. The aim of the tutorials is to provide a basic introduction to both software tools. For each tool, step-by-step instructions are given on how the tool can be used. The bibliometric networks that are analyzed and visualized in the two tutorials deal with the field of scientometrics and closely related fields. Analyses for other fields of science can be performed in a similar way. To run VOSviewer and CitNetExplorer, one needs to have a computer system that offers Java support. Java version 6 or higher needs to be installed.

13.5.1 Data Collection

Data was collected from the Web of Science bibliographic database produced by Thomson Reuters.[1] Bibliographic data was downloaded for all 25,242 publications in the 13 journals listed in Table 13.2. To select these journals, we started with *Scientometrics* and *Journal of Informetrics*, which we regard as the two core journals in the field of scientometrics. We then used the 2012 edition of Thomson Reuters' Journal Citation Reports to identify closely related journals. We took all journals listed among the five most closely related journals to either *Scientometrics* or *Journal of Informetrics*, excluding journals that seem to be mainly nationally oriented. For each of the selected journals, we determined whether it has any predecessors. These predecessors were included in the selection as well, provided that they are indexed in the Web of Science database. In this way, the list of 13 journals shown in Table 13.2 was obtained.

Bibliographic data for the 25,242 publications in the journals listed in Table 13.2 was downloaded from the Web of Science database. The database supports various file formats. We used the tab-delimited format. For each publication, the full record including cited references was obtained. In the Web of Science database, bibliographic data can be downloaded for at most 500 publications at a time. Downloading therefore took place in batches. We ended up with a large number of Web of Science output files, each containing bibliographic data for at most 500 publications.

As can be seen in Table 13.2, the publications included in the data collection cover the period 1945–2013. Bibliographic data was downloaded separately for publications from the period 1945–1999 and for publications from the period 2000–2013. In the VOSviewer tutorial, only publications from the latter period will be considered. In the CitNetExplorer tutorial, publications from both periods will be included in the analysis.

---

[1] The data collection took place on November 7, 2013.

Table 13.2 Journals included in the data collection

Journal	Time period	No. of pub.
American Documentation	1956–1969	796
ASLIB Proceedings	1956–2013	2,697
Information Processing &Management	1975–2013	3,036
Information Scientist	1968–1978	254
Information Storage and Retrieval	1963–1974	372
Journal of Documentation	1945–2013	3,778
Journal of Information Science	1979–2013	1,855
Journal of Informetrics	2007–2013	399
Journal of the American Society for Information Science	1970–2000	2,995
Journal of the American Society for Information Science and Technology	2001–2013	2,486
Research Evaluation	2000–2013	383
Research Policy	1974–2013	2,596
Scientometrics	1978–2013	3,595

13.5.2 VOSviewer

In this tutorial, we demonstrate the visualization of three bibliometric networks: A bibliographic coupling network of researchers, a co-citation network of journals, and a co-occurrence network of terms.

13.5.2.1 Bibliographic Coupling Network of Researchers

To construct and visualize a bibliographic coupling network of researchers, we take the following steps:

1. Launch VOSviewer. If VOSviewer has not yet been downloaded, then first download it from http://www.vosviewer.com.
2. Press the **Create** button on the **Action** tab to open the **Create Map** dialog box.
3. Select the **Create a map based on a network** option button and press the **Next** button.
4. Go to the **Web of Science** tab to open the **Select Web of Science File** dialog box. In this dialog box, the Web of Science output files that we want to work with can be selected. The CTRL key can be used to select multiple files. We select the files that we obtained in Sect. 13.5.1. We only include files containing bibliographic data for publications from the period 2000–2013. After the files have been selected, first press the **OK** button to close the **Select Web of Science File** dialog box and then press the **Next** button.
5. Select the **Bibliographic coupling of authors** and **Fractional counting** option buttons and press the **Next** button. By selecting the **Fractional counting** option

button, we indicate that we want to use the fractional counting methodology discussed in Sect. 13.4.1.

6. VOSviewer asks for the minimum number of publications a researcher must have in order to be included in the bibliographic coupling network. We choose the default value of five publications. In our data set, there turn out to be 792 researchers with at least five publications. To go to the next step, press the **Next** button.

7. Some researchers may have no or almost no bibliographic coupling relations with other researchers. It is usually best to exclude these researchers from a bibliographic coupling network. VOSviewer asks for the number of researchers to be included in the bibliographic coupling network. We choose to include 500 researchers. This means that the $792 - 500 = 292$ researchers with the smallest number of bibliographic coupling relations will be excluded. To go to the next step, press the **Next** button.

8. VOSviewer lists the 500 researchers included in the bibliographic coupling network and offers the possibility to remove individual researchers from the network. We choose not to remove anyone. To finish the construction of the bibliographic coupling network and to close the **Create Map** dialog box, press the **Finish** button.

VOSviewer now applies the normalization, mapping, and clustering techniques discussed in Sect. 13.4.1. Since we are working with a relatively small network, this will take at most a few seconds. VOSviewer then provides us with the visualization shown in Fig. 13.4.
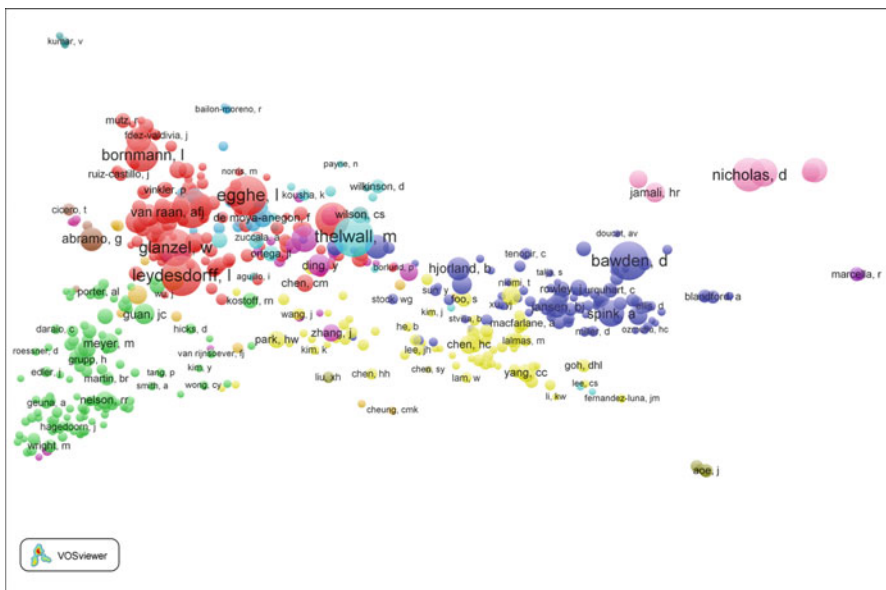
Fig. 13.4 VOSviewer visualization of a researcher bibliographic coupling network

In the visualization presented in Fig. 13.4, each circle represents a researcher. Large circles represent researchers that have many publications. Small circles represent researchers with only a few publications. In general, the closer two researchers are located to each other in the visualization, the more strongly they are related to each other based on bibliographic coupling. In other words, researchers that are located close to each other tend to cite the same publications, while researchers that are located far away from each other usually do not cite the same publications. Colors indicate clusters of researchers that are relatively strongly related to each other. There are 15 clusters, most of which are fairly small. It may be helpful to reduce the number of clusters a bit. This can be done as follows:

9. Go to the **Map** tab in the left part of the VOSviewer window.
10. In the **Clustering resolution** text box, decrease the value of the resolution parameter (see Sect. 13.4.1) from 1.00 to 0.50.
11. Select the **Clustering only** option button and press the **Run** button.

Instead of 15 clusters, we now have only six, of which three are very small. The three larger clusters consist of researchers in scientometrics (upper left area in the visualization), researchers in information science and information retrieval (right area), and researchers in technology and innovation studies (lower left area).

13.5.2.2   Co-citation Network of Journals

We now demonstrate the construction and visualization of a co-citation network of journals. The first four steps that we take are the same as described above for analyzing a bibliographic coupling network of researchers. After these four steps have been taken, we proceed as follows:

1. Select the **Co-citation of sources** and **Fractional counting** option buttons and press the **Next** button.
2. VOSviewer asks for the minimum number of citations a journal must have received in order to be included in the co-citation network. The default value is 20 citations, but we choose to require at least 50 citations. This means that a journal can be included in the co-citation network only if in our Web of Science output files there are at least 50 cited references that point to the journal. There turn out to be 619 journals that satisfy this requirement. To go to the next step, press the **Next** button.
3. VOSviewer asks for the number of journals to be included in the co-citation network. The journals with the smallest number of co-citation relations will be excluded. Since the 619 journals selected in the previous step have all been cited quite significantly (i.e., at least 50 times), we do not think there is a need to exclude any journals. We therefore simply choose to include all 619 journals. To go to the next step, press the **Next** button.
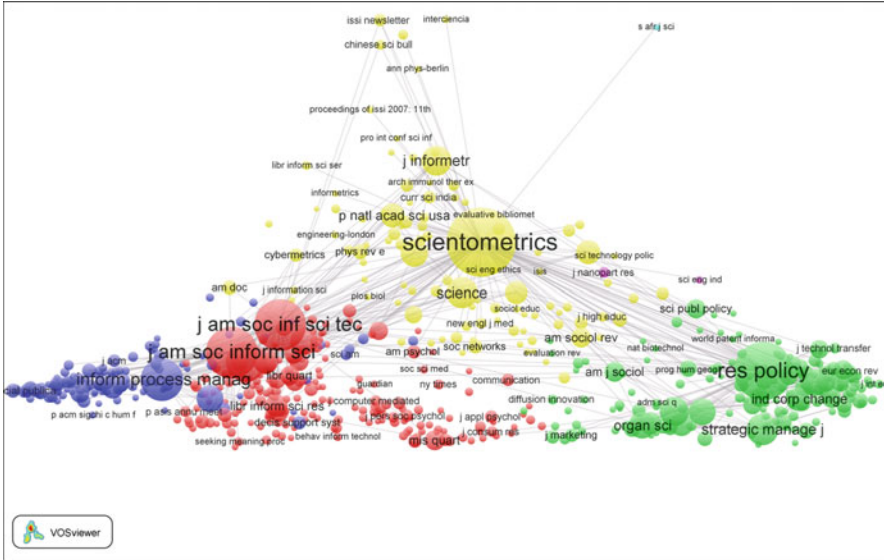
Fig. 13.5 VOSviewer visualization of a journal co-citation network

4. VOSviewer lists the 619 journals included in the co-citation network and offers
   the possibility to remove individual journals from the network. We do not
   remove any journal. To finish the construction of the co-citation network and
   to close the **Create Map** dialog box, press the **Finish** button.

The visualization that we obtain of our journal co-citation network is shown in
Fig. 13.5.[2]

Each circle in the visualization presented in Fig. 13.5 represents a journal. The
size of a circle reflects the number of citations a journal has received. Journals that
are located close to each other in the visualization tend to be more strongly related,
based on co-citations, than journals that are located far away from each other. Three
broad groups of journals can be distinguished: Journals that publish scientometrics
research in the upper area of the visualization, journals that publish information
science and information retrieval research in the lower left area, and journals that
publish technology and innovation studies research in the lower right area. These
three groups of journals can also be easily recognized in the density visualization
that is available on the **Density View** tab.

---

[2] To improve the visualization, the **Size variation** parameter in the **Options** dialog box has been
set to a value of 0.40. In addition, the **No. of lines** parameter has been set to a value of 500. This has
the effect that 500 lines, representing the 500 strongest co-citation relations between journals, are
displayed in the visualization.

VOSviewer has identified six clusters of journals.[3] These clusters are indicated using colors in the visualization shown in Fig. 13.5. Two clusters are very small and can be ignored. Of the four larger clusters, the yellow one consists mainly of journals publishing scientometrics research, while the green one covers journals focusing on technology and innovation studies research. The red and the blue cluster both consist of journals publishing information science and information retrieval research. By zooming in on the lower left area in the visualization, it can be seen that the journals in the blue cluster are more computer science oriented, focusing mainly on technical information retrieval research, while the journals in the red cluster are more social science oriented, focusing on general information science and library science research.

13.5.2.3   Co-occurrence Network of Terms

Finally, we consider the construction and visualization of a co-occurrence network of terms extracted from the titles and abstracts of publications. We take the following steps:

1. Launch VOSviewer.
2. Press the **Create** button on the **Action** tab to open the **Create Map** dialog box.
3. Select the **Create a map based on a text corpus** option button and press the **Next** button.
4. Go to the **Web of Science** tab, select the Web of Science output files to be used, and press the **Next** button. We use the same files as we did above in the analysis of researcher bibliographic coupling and journal co-citation networks.
5. Select the **Title and abstract fields** option button and press the **Next** button. VOSviewer will now extract noun phrases from the titles and abstracts of the publications in our Web of Science output files (see Sect. 13.4.1). This may take some time.
6. Select the **Binary counting** option button and press the **Next** button. The use of a binary counting methodology means that in the construction of a co-occurrence network the number of times a noun phrase occurs in the title and abstract of a publication plays no role. A noun phrase that occurs only once in the title and abstract of a publication is treated in the same way as a noun phrase that occurs for instance ten times.
7. VOSviewer asks for the minimum number of occurrences a noun phrase must have in order to be included in the co-occurrence network. We choose the default value of ten occurrences. There turn out to be 3,158 noun phrases that occur in the title or abstract of at least ten publications. To go to the next step, press the **Next** button.

---

[3] The resolution parameter of VOSviewer's clustering technique is set to its default value of 1.00, not to the value of 0.50 that was used in the case of the author bibliographic coupling network.

8. VOSviewer asks for the number of noun phrases to be included in the co-occurrence network. By default, VOSviewer suggests to include 60 % of the noun phrases selected in the previous step. We follow this suggestion and choose to include 1,894 noun phrases in the co-occurrence network. To determine the noun phrases to be excluded, VOSviewer will calculate for each noun phrase a relevance score (see Sect. 13.4.1). The noun phrases with the lowest relevance scores will be excluded. To go to the next step, press the **Next** button. VOSviewer will now perform the calculation of the relevance scores. This may take some time.

9. VOSviewer lists the 1,894 noun phrases included in the co-occurrence network and offers the possibility to remove individual noun phrases from the network. We order the noun phrases alphabetically and indicate that we want to remove all noun phrases starting with "Elsevier." These noun phrases result from copyright statements in the abstracts of publications in Elsevier journals. After removing six noun phrases that start with "Elsevier," we end up with a co-occurrence network of 1,888 noun phrases. We refer to these noun phrases as terms. To finish the construction of the co-occurrence network and to close the **Create Map** dialog box, press the **Finish** button.

Figure 13.6 shows the visualization of our term co-occurrence network.[4]

In the visualization presented in Fig. 13.6, each circle represents a term. The size of a circle indicates the number of publications that have the corresponding term in their title or abstract. Terms that co-occur a lot tend to be located close to each other in the visualization. VOSviewer has grouped the terms into six clusters, of which four are of significant size. The red cluster, located in the upper left area in the visualization, consists of scientometric terms. The green cluster, located in the lower left area, covers terms related to technology and innovation studies. In the right area in the visualization, the blue and yellow clusters consist of terms related to information science and information retrieval. Similar to what we have seen in the case of the journal co-citation network, one cluster (the blue one) is more computer science-oriented while the other (the yellow one) is more focused on the social sciences.

Comparing the visualizations shown in Figs. 13.4, 13.5, and 13.6, it can be concluded that we have obtained a quite consistent picture of the structure of the field of scientometrics and closely related fields. The three visualizations all suggest a similar division into subfields. The differences between the visualizations are fairly small and relate mainly to the positioning of the subfields relative to each other and to the level of detail that is provided.

---

[4] To improve the visualization, the **Size variation** parameter in the **Options** dialog box has been set to a value of 0.40. A few terms in the upper part of the visualization are not visible in Fig. 13.6.
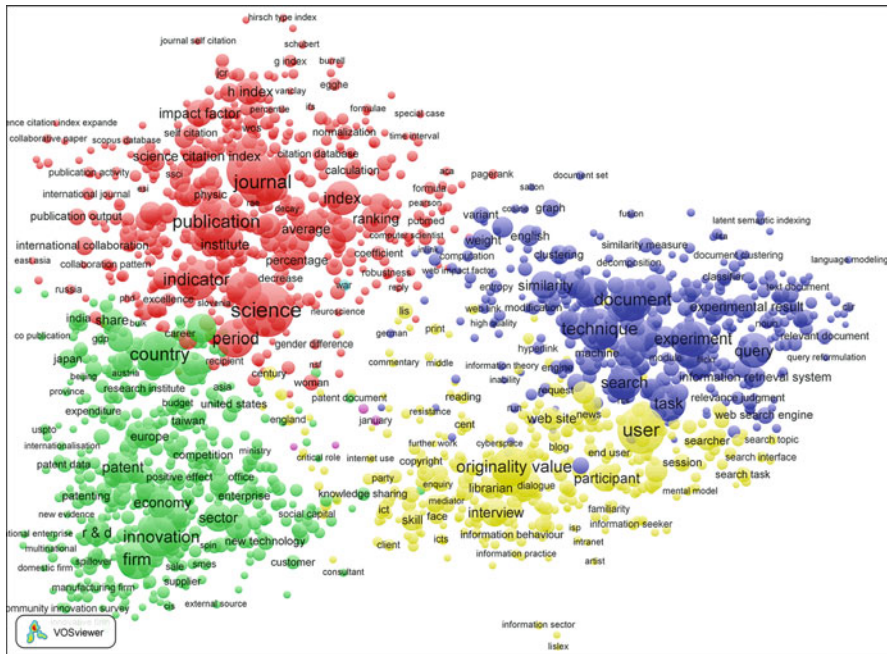
Fig. 13.6 VOSviewer visualization of a term co-occurrence network

13.5.3 CitNetExplorer

In this CitNetExplorer tutorial, we analyze and visualize the citation network of publications in the field of scientometrics.

We start by opening the network:

1. Launch CitNetExplorer. If CitNetExplorer has not yet been downloaded, then first download it from http://www.citnetexplorer.nl.
2. The **Open Citation Network** dialog box will appear. On the **Web of Science** tab, we select the Web of Science output files that we want to work with. These are the files that we obtained in Sect. 13.5.1. All files are included, covering the full period 1945–2013. By default, the **Include non-matching cited references** check box is checked. We leave it this way. If the **Include non-matching cited references** check box is not checked, only publications for which bibliographic data is available in our Web of Science output files will be included in the citation network. If the check box is checked, all publications with at least a certain minimum number of citations will be included, even if no bibliographic data (other than the data that can be extracted from cited references) is available. For the minimum number of citations, we choose the default value of ten.
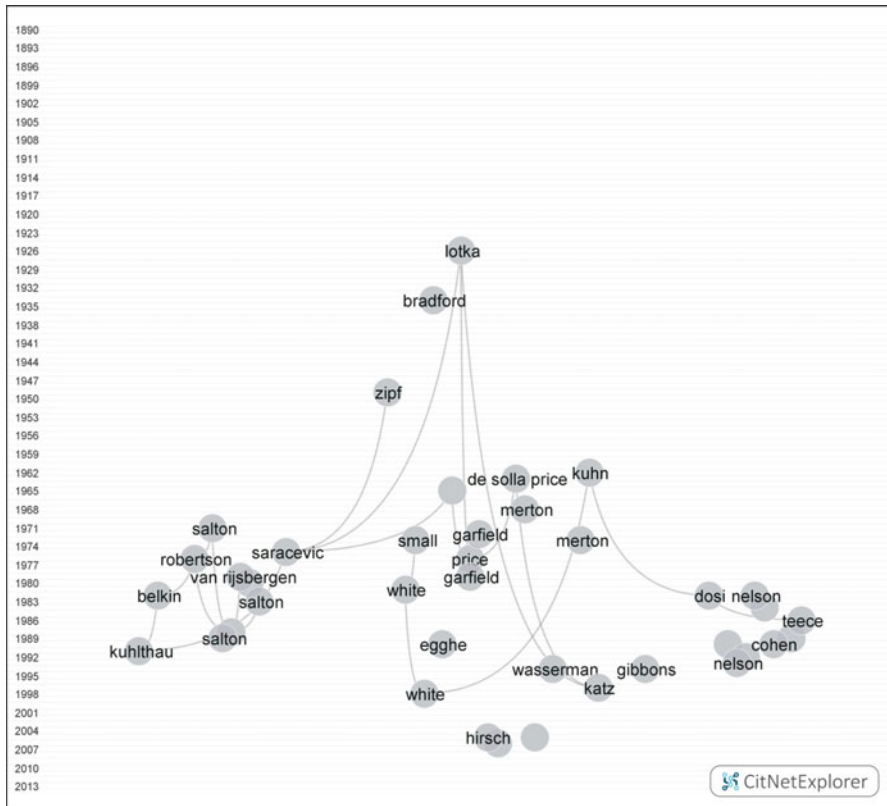
Fig. 13.7 CitNetExplorer visualization of our publication citation network

3. Press the **OK** button to close the **Open Citation Network** dialog box. The citation network will now be opened. CitNetExplorer will perform citation matching and will make sure that the network is acyclic (see Sect. 13.4.2). This may take some time.

The visualization of the citation network is shown in Fig. 13.7. Recall from Sect. 13.4.2 that only the 40 most frequently cited publications in the network are displayed. When the mouse is moved over a publication, some bibliographic information (e.g., authors, title, and journal) is displayed in the left part of the CitNetExplorer window. The visualization presented in Fig. 13.7 has a clear structure. Publications on information science and information retrieval can be found in the left part of the visualization, publications on scientometrics in the middle part, and publications on technology and innovation studies in the right part. The curved lines indicate citation relations between publications.

Suppose we are interested to learn more about the literature on visualizing bibliometric networks. Among the 40 publications displayed in Fig. 13.7, we recognize three publications on this topic: Small (1973), White and Griffith

(1981), and White and McCain (1998). All three publications were discussed in the literature review presented in Sect. 13.2. We use these publications to drill down into the citation network:

4. Click on Small (1973). This publication has now been marked. To indicate this, the publication is displayed using a square instead of a circle.
5. Click on White and McCain (1998). This publication has now been marked as well. Notice that Small (1973), White and McCain (1998), and a few other publications are displayed using a red border. This border indicates that these publications are selected. By default, a publication is selected if it has been marked or if it is on a citation path from one marked publication to another. For instance, White and Griffith (1981) is selected because it is on a citation path from White and McCain (1998) to Small (1973). White and Griffith (1981) is cited by White and McCain (1998) and is citing Small (1973). Publications that are not among the 40 displayed in the visualization may also be selected, but this is not directly visible. The number of selected publications is reported in the upper left part of the CitNetExplorer window. In our case, there turn out to be 261 selected publications. Each of these publications is on a citation path from White and McCain (1998) to Small (1973).
6. Press the **Drill down** button in the upper part of the CitNetExplorer window. By pressing this button, we drill down from the full citation network to a citation network that includes only the 261 selected publications. This network is now referred to as the current network.

After drilling down, we obtain the visualization shown in Fig. 13.8. Of the 261 publications, the 40 most frequently cited are displayed in the visualization.

Many of the publications displayed in Fig. 13.8 do not directly deal with the topic of visualizing bibliometric networks. We therefore drill down deeper into the citation network. We use a clustering technique (see Sect. 13.4.2) for this purpose:

7. Press the **Analysis** button in the upper part of the CitNetExplorer window. In the **Analysis** menu, choose the **Clustering** option. The **Clustering** dialog box will appear.
8. We use the default parameter values presented in the **Clustering** dialog box. Press the **OK** button to close the dialog box. The 261 publications in the current network will now be clustered.
9. A message box is displayed indicating that five clusters of publications have been identified. Press the **OK** button to close the message box. Publications now have colors in the visualization. The color of a publication indicates the group to which the publication belongs. There are five groups, each corresponding with one of the five clusters that have been identified.
10. Publications on the topic of visualizing bibliometric networks turn out to be concentrated in group 2. These publications have a green color. In the **Selection parameters** frame in the left part of the CitNetExplorer window, choose the **Based on groups** option in the **Selection** drop-down box. In the **Groups** list
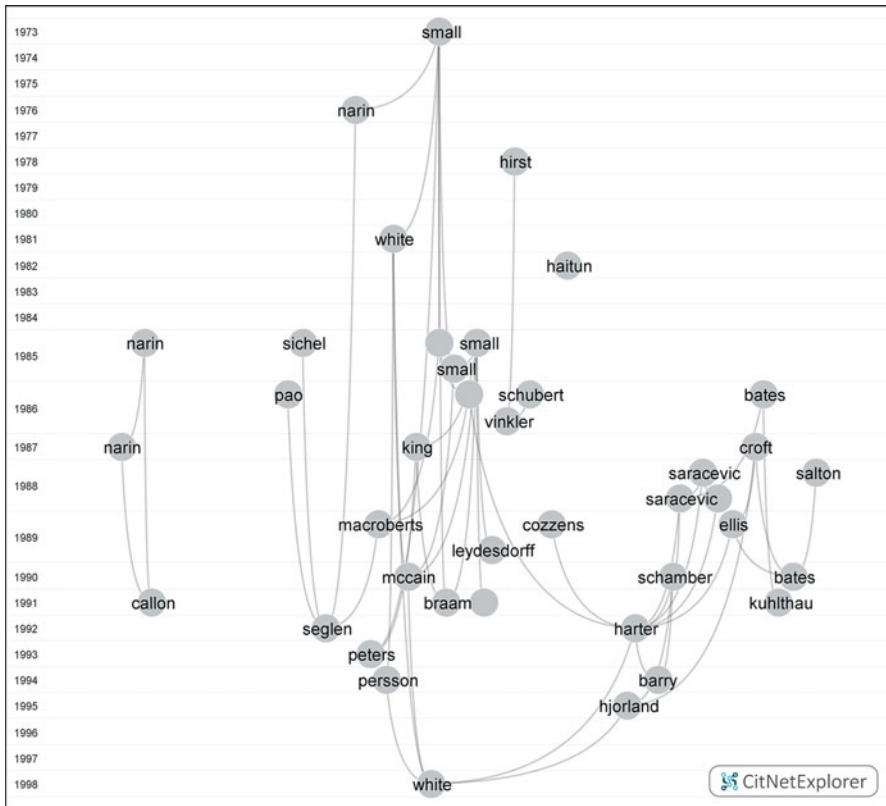
Fig. 13.8 CitNetExplorer visualization of our publication citation network after drilling down the first time

box, check **Group 2**. The 66 publications belonging to group 2 are now selected. These publications are displayed using a red border in the visualization.

11. Press the **Drill down** button.

The visualization obtained after drilling down is shown in Fig. 13.9.[5] The current network now includes 66 publications. The 40 most frequently cited ones are displayed in the visualization.

We now have an overview of an important part of the literature that appeared in the period 1973–1998 on the topic of visualizing bibliometric networks. However,

---

[5] Notice that in the visualization shown in Fig. 13.9, publications are displayed in green rather than in gray. This is because the publications included in the visualization all belong to the same cluster identified by CitNetExplorer's clustering technique.
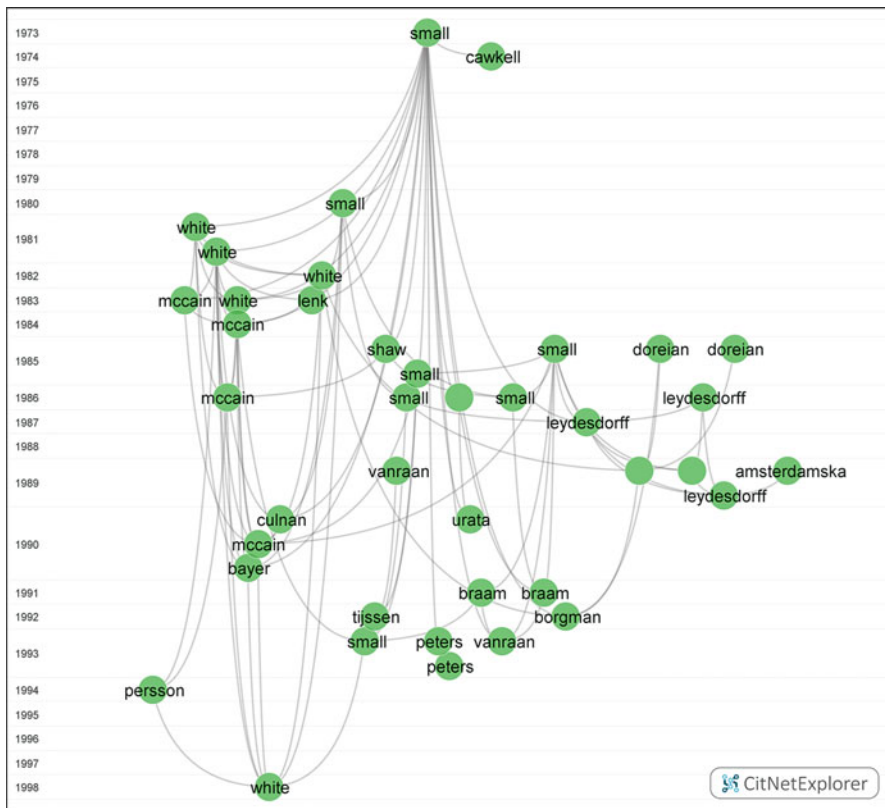
Fig. 13.9 CitNetExplorer visualization of our publication citation network after drilling down the second time

it may be that not all important publications from the period 1973–1998 are included in the current network. Moreover, publications from earlier and later periods are not included at all. To include a larger part of the literature, we expand the current network. We first look at publications that precede the 66 publications in the current network. We expand the current network with all publications cited by at least one of these 66 publications:

12. Press the **Expand** button in the upper part of the CitNetExplorer window. The **Expand Current Network** dialog box will appear.
13. In the **Publications** drop-down box, choose the **Predecessors** option. Do not change the values of the **Min. number of citation links** and **Max. distance** parameters, and do not check the **Add intermediate publications** check box. Press the **OK** button to close the **Expand Current Network** dialog box. The current network will now be expanded.
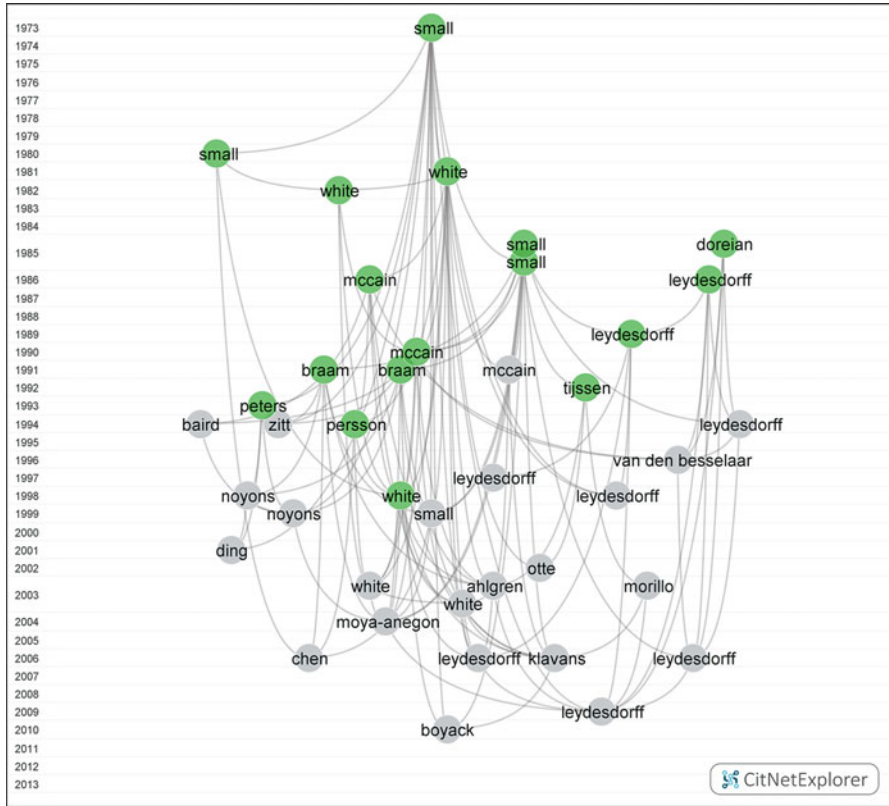
Fig. 13.10 CitNetExplorer visualization of our publication citation network after the second expansion

After expansion, the current network includes 342 publications. Among these publications, we for instance find classical works by Alfred Lotka, Eugene Garfield, and Derek de Solla Price.

Suppose we also want to see publications that are preceded by the above-mentioned 66 publications. In other words, we want to see publications by which the 66 publications are cited. The number of publications that cite at least one of our 66 publications will probably be quite large, and many of these publications may not be directly related to the topic of visualizing bibliometric networks. We therefore require publications to cite at least three of our 66 publications. We take the following steps:

14. Press the **Back** button in the upper part of the CitNetExplorer window. The expansion of the current network from 66 to 342 publications will be undone.
15. Press the **Expand** button to open the **Expand Current Network** dialog box.
16. In the **Publications** drop-down box, choose the **Successors** option. Set the **Min. number of citation links** parameter to a value of three. Do not change the

value of the **Max. distance** parameter, and do not check the **Add intermediate publications** check box. Press the **OK** button to close the **Expand Current Network** dialog box and to perform the expansion.

The visualization obtained after the expansion is shown in Fig. 13.10. The current network now includes 248 publications. Among the more recent publications, we find publications on the topic of visualizing bibliometric networks by for instance Kevin Boyack, Chaomei Chen, Richard Klavans, and Loet Leydesdorff.

The above tutorial has demonstrated the most essential elements of the user interface of CitNetExplorer. However, there are many features of the software that have not been discussed in this short tutorial. Some of these features are discussed in a more extensive tutorial available at http://www.citnetexplorer.nl/gettingstarted/.

Conclusion

In this chapter, we have provided an introduction into the topic of visualizing bibliometric networks. An overview has been given of the literature on this topic and of the main software tools that are available. Our focus has been mostly on two software tools: VOSviewer and CitNetExplorer. We have discussed the most important techniques used by these tools, and we have provided tutorials with step-by-step instructions on the use of the tools.

To conclude this chapter, we first discuss the limitations of bibliometric network visualizations and the proper use of these visualizations. We then summarize some ongoing and future developments in the visualization of bibliometric networks.

Limitations and Proper Use of Bibliometric Network Visualizations

The main idea of bibliometric network visualization is to allow large amounts of complex bibliographic data to be analyzed in a relatively easy way by visualizing core aspects of data. The strength of bibliometric network visualization is in the simplification it provides, but simplification comes at a cost. It typically implies a loss of information.

Loss of information takes place in reducing bibliographic data to a bibliometric network. For example, when textual data is reduced to a co-occurrence network of terms, information on the context in which terms co-occur is lost. Similarly, when we have constructed a citation network, we can see who is citing whom, but we can no longer see why someone may be citing someone else.

Loss of information also occurs in the visualization of a bibliometric network. In the case of a distance-based visualization, for instance, it is usually not possible to position the nodes in a two-dimensional space in such a way that for any pair of nodes the distance between the nodes reflects the relatedness of the nodes with perfect accuracy. Distances reflect relatedness only approximately, and we therefore lose information. In the case of graph-based and timeline-based visualizations, we may need to restrict

(continued)

ourselves to visualizing a limited number of nodes, for instance, the nodes with the highest degree in a network. This means that we lose information on the other nodes in the network.

Loss of information is especially problematic because it is often difficult to assess how much information is lost and to what extent this may affect the conclusions that can be drawn from a bibliometric network visualization. For instance, to what extent do the distances between the nodes in a distance-based visualization accurately reflect the relatedness of the nodes? To what extent does a visualization of a term co-occurrence network change if the selection of terms included in the visualization is changed? Even if we are aware that there may be inaccuracies in a bibliometric network visualization, it remains difficult to assess the magnitude and the consequences of these inaccuracies.

Related to this, it is often difficult to assess the sensitivity of a bibliometric network visualization due to various technical choices. Would other technical choices have resulted in a completely different visualization, or would the differences have been minor? How strongly does a visualization depend on the values of all kinds of technical parameters, and is it possible to justify the choice of particular parameter values? Is a certain structure suggested by a visualization a reflection of the underlying data, or is it merely an artifact of the techniques used to produce the visualization? Researchers who regularly work with bibliometric network visualizations develop an intuition that helps them to give approximate answers to these types of questions, but most users of bibliometric network visualizations lack such an intuition, making it difficult for them to assess the accuracy of a visualization.

Given the above difficulties, our general recommendation is to use bibliometric network visualizations as a complement rather than as a substitute to expert judgment. When expert judgment and bibliometric network visualizations are in agreement with each other and point in the same direction, they strengthen each other. When they do not agree, this may be a reason for experts to reconsider their opinion, it may also be a reason to ask for the opinion of additional experts, or it may be a reason to check whether the visualizations may be inaccurate because important information has been lost or because of methodological issues. Bibliometric network visualizations are most useful when they are interpreted in a careful manner and used in combination with expert judgment. Also, visualization should be a means to an end, not an end in itself. For instance, when dealing with only a small amount of data, there often is no added value in the use of visualization. It may be much better to simply study the data directly.

(continued)

(continued)

Ongoing and Future Developments in the Visualization of Bibliometric Networks

Finally, we discuss some ongoing and future developments in the visualization of bibliometric networks.

An important development, made possible by the enormous growth in computational resources that has taken place, is the increasing attention that is given to visualizing large bibliometric networks. Examples include the work by Boyack et al. (2005) on large journal networks, the work by Klavans and Boyack (2006) on large publication networks, and the work by Skupin et al. (2013) on large term networks. Interesting alternative data sources are being explored as well (e.g., Bollen et al., 2009).

A second and related development is the increasing use of interactive visualizations. The use of interactive visualizations is especially important given the above-mentioned trend toward visualizing increasingly large bibliometric networks. Static visualizations of large networks tend to be of limited use. It is rather difficult to show the detailed structure of a large network in a static visualization. Interactive visualizations allow large networks to be visualized and explored in a much more powerful way, for instance by allowing users to drill down from a general high-level overview to a very detailed low-level picture. The drill down functionality offered by CitNetExplorer can be seen as an example of this idea.

A third development is the increasing interest in dynamic visualizations that show how bibliometric networks have evolved over time. Traditionally, most attention has been paid to visualizations that offer a static picture showing the structure of a bibliometric network at a given point in time. Dynamic visualizations have received less attention, although some interesting work has been done, for instance in the CiteSpace tool (Chen, 2004, 2006) and in general network analysis tools such as Gephi and Visone (http://visone.info). Like visualizations of large networks, dynamic visualizations may benefit a lot from interactive visualization approaches.

In summary, we expect a trend toward more and more interactive and dynamic visualizations that involve increasingly large bibliometric networks. Clearly, an exciting and highly challenging research agenda lies ahead of us.

Appendix: Normalization, Mapping, and Clustering Techniques Used by VOSviewer

In this appendix, we provide a more detailed description of the normalization, mapping, and clustering techniques used by VOSviewer.

Normalization

We first discuss the association strength normalization (Van Eck & Waltman, 2009) used by VOSviewer to normalize for differences between nodes in the number of edges they have to other nodes. Let a_{ij} denote the weight of the edge between nodes i and j, where $a_{ij} = 0$ if there is no edge between the two nodes. Since VOSviewer treats all networks as undirected, we always have $a_{ij} = a_{ji}$. The association strength normalization constructs a normalized network in which the weight of the edge between nodes i and j is given by

$$s_{ij} = \frac{2ma_{ij}}{k_i k_j}, \tag{1}$$

where k_i (k_j) denotes the total weight of all edges of node i (node j) and m denotes the total weight of all edges in the network. In mathematical terms,

$$k_i = \sum_j a_{ij} \quad \text{and} \quad m = \frac{1}{2}\sum_i k_i. \tag{2}$$

We sometimes refer to s_{ij} as the similarity of nodes i and j. For an extensive discussion of the rationale of the association strength normalization, we refer to Van Eck and Waltman (2009).

Mapping

We now consider the VOS mapping technique used by VOSviewer to position the nodes in the network in a two-dimensional space. The VOS mapping technique minimizes the function

$$V(\mathbf{x}_1, \ldots, \mathbf{x}_n) = \sum_{i<j} s_{ij} \left\| \mathbf{x}_i - \mathbf{x}_j \right\|^2 \tag{3}$$

subject to the constraint

$$\frac{2}{n(n-1)}\sum_{i<j}\left\|\mathbf{x}_i - \mathbf{x}_j\right\| = 1, \tag{4}$$

where n denotes the number of nodes in a network, $\mathbf{x}_i$ denotes the location of node i in a two-dimensional space, and $\|\mathbf{x}_i - \mathbf{x}_j\|$ denotes the Euclidean distances between nodes i and j. VOSviewer uses a variant of the SMACOF algorithm (e.g., Borg & Groenen, 2005) to minimize (3) subject to (4). We refer to Van Eck et al. (2010) for a more extensive discussion of the VOS mapping technique, including a comparison with multidimensional scaling.

Clustering

Finally, we discuss the clustering technique used by VOSviewer. Nodes are assigned to clusters by maximizing the function

$$V(c_1, \ldots, c_n) = \sum_{i<j} \delta\big(c_i, c_j\big)\big(s_{ij} - \gamma\big), \tag{5}$$

where c_i denotes the cluster to which node i is assigned, $\delta(c_i, c_j)$ denotes a function that equals 1 if $c_i = c_j$ and 0 otherwise, and γ denotes a resolution parameter that determines the level of detail of the clustering. The higher the value of γ, the larger the number of clusters that will be obtained. The function in (5) is a variant of the modularity function introduced by Newman and Girvan (2004) and Newman (2004) for clustering the nodes in a network. There is also an interesting mathematical relationship between on the one hand the problem of minimizing (3) subject to (4) and on the other hand the problem of maximizing (5). Because of this relationship, the mapping and clustering techniques used by VOSviewer constitute a unified approach to mapping and clustering the nodes in a network. We refer to Waltman et al. (2010) for more details. We further note that VOSviewer uses the recently introduced smart local moving algorithm (Waltman & Van Eck, 2013) to maximize (5).

References

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *International AAAI Conference on Weblogs and Social Media* [Online]

Bollen, J., Van de Sompel, H., Hagberg, A., Bettencourt, L., Chute, R., Rodriguez, M. A., et al. (2009). Clickstream data yields high-resolution maps of science. *PLoS ONE, 4*(3), e4803.

Borg, I., & Groenen, P. J. F. (2005). *Modern multidimensional scaling* (2nd ed.). New York, NY: Springer.

Börner, K. (2010). *Atlas of science: Visualizing what we know*. Cambridge, MA: MIT Press.

Börner, K., Chen, C., & Boyack, K. W. (2003). Visualizing knowledge domains. *Annual Review of Information Science and Technology, 37*(1), 179–255.

Börner, K., Klavans, R., Patek, M., Zoss, A. M., Biberstine, J. R., Light, R. P., et al. (2012). Design and update of a classification system: The UCSD map of science. *PLoS ONE, 7*(7), e39464.

Boyack, K. W., & Klavans, R. (2010). Co-citation analysis, bibliographic coupling, and direct citation. Which citation approach represents the research front most accurately? *Journal of the American Society for Information Science and Technology, 61*(12), 2389–2404.

Boyack, K. W., Klavans, R., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics, 64*(3), 351–374.

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.

Callon, M., Law, J., & Rip, A. (Eds.). (1986). *Mapping the dynamics of science and technology*. London: MacMillan.

Chen, C. (1999). Visualising semantic spaces and author co-citation networks in digital libraries. *Information Processing & Management, 35*(3), 401–420.

Chen, C. (2004). Searching for intellectual turning points: Progressive knowledge domain visualization. *Proceedings of the National Academy of Sciences, 101*(suppl. 1), 5303–5310.

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology, 57*(3), 359–377.

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

De Moya-Anegón, F., Vargas-Quesada, B., Chinchilla-Rodríguez, Z., Corera-Álvarez, E., Munoz-Fernández, F. J., & Herrero-Solana, V. (2007). Visualizing the marrow of science. *Journal of the American Society for Information Science and Technology, 58*(14), 2167–2179.

De Nooy, W., Mrvar, A., & Batagelj, V. (2005). *Exploratory social network analysis with Pajek*. Cambridge: Cambridge University Press.

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience, 21*(11), 1129–1164.

Garfield, E. (2004). Historiographic mapping of knowledge domains literature. *Journal of Information Science, 30*(2), 119–145.

Garfield, E., Pudovkin, A. I., & Istomin, V. S. (2003). Why do we need algorithmic historiography? *Journal of the American Society for Information Science and Technology, 54*(5), 400–412.

Griffith, B. C., Small, H., Stonehill, J. A., & Dey, S. (1974). The structure of scientific literatures II: Toward a macro- and microstructure for science. *Science Studies, 4*(4), 339–365.

Healy, P., & Nikolov, N. S. (2013). Hierarchical drawing algorithms. In R. Tamassia (Ed.), *Handbook of graph drawing and visualization* (pp. 409–453). Boca Raton, FL: CRC Press.

Hummon, N. P., & Doreian, P. (1989). Connectivity in a citation network: The development of DNA theory. *Social Networks, 11*(1), 39–63.

Jarneving, J. (2007). Bibliographic coupling and its application to research-front and other core documents. *Journal of Informetrics, 1*(4), 287–307.

Kamada, T., & Kawai, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters, 31*(1), 7–15.

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

Klavans, R., & Boyack, K. W. (2006). Quantitative evaluation of large maps of science. *Scientometrics, 68*(3), 475–499.

Leydesdorff, L., & Rafols, I. (2009). A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology, 60*(2), 348–362.

Leydesdorff, L., & Rafols, I. (2012). Interactive overlays: A new method for generating global journal maps from Web-of-Science data. *Journal of Informetrics, 6*(2), 318–332.

Luukkonen, T., Tijssen, R. J. W., Persson, O., & Sivertsen, G. (1993). The measurement of international scientific collaboration. *Scientometrics, 28*(1), 15–36.

Marshakova, I. (1973). System of documentation connections based on references (SCI). *Nauchno-TekhnicheskayaInformatsiya Seriya, 2*(6), 3–8.

McCain, K. W. (1991). Mapping economics through the journal literature: An experiment in journal cocitation analysis. *Journal of the American Society for Information Science, 42*(4), 290–296.

Morris, S. A., Yen, G., Wu, Z., & Asnake, B. (2003). Time line visualization of research fronts. *Journal of the American Society for Information Science and Technology, 54*(5), 413–422.

Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E, 69*(6), 066133.

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E, 69*(2), 026113.

Persson, O. (2010). Identifying research themes with weighted direct citation links. *Journal of Informetrics, 4*(3), 415–422.

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.

Schvaneveldt, R. W., Dearholt, D. W., & Durso, F. T. (1988). Graph theoretic foundations of pathfinder networks. *Computers & Mathematics with Applications, 15*(4), 337–345.

Seidman, S. B. (1983). Network structure and minimum degree. *Social Networks, 5*(3), 269–287.

Skupin, A., Biberstine, J. R., & Börner, K. (2013). Visualizing the topical structure of the medical sciences: A self-organizing map approach. *PLoS ONE, 8*(3), e58779.

Small, H. (1973). Co-citation in the scientific literature: A new measure of the relationship between two documents. *Journal of the American Society for Information Science, 24*(4), 265–269.

Small, H. (1997). Update on science mapping: Creating large document spaces. *Scientometrics, 38*(2), 275–293.

Small, H., & Griffith, B. C. (1974). The structure of scientific literatures I: Identifying and graphing specialties. *Science Studies, 4*(1), 17–40.

Van Eck, N. J., & Waltman, L. (2009). How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology, 60*(8), 1635–1651.

Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538.

Van Eck, N. J., & Waltman, L. (2011). Text mining and visualization using VOSviewer. *ISSI Newsletter, 7*(3), 50–54.

Van Eck, N. J., Waltman, L., Dekker, R., & Van den Berg, J. (2010). A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. *Journal of the American Society for Information Science and Technology, 61*(12), 2405–2416.

Van Eck, N. J., Waltman, L., Van Raan, A. F. J., Klautz, R. J. M., & Peul, W. C. (2013). Citation analysis may severely underestimate the impact of clinical research as compared to basic research. *PLoS ONE, 8*(4), e62395.

Waltman, L., & Van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology, 63*(12), 2378–2392.

Waltman, L., & Van Eck, N. J. (2013). A smart local moving algorithm for large-scale modularity-based community detection. *European Physical Journal B, 86*, 471.

Waltman, L., Van Eck, N. J., & Noyons, E. C. M. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics, 4*(4), 629–635.

White, H. D. (2003). Pathfinder networks and author cocitation analysis: A remapping of paradigmatic information scientists. *Journal of the American Society for Information Science and Technology, 54*(5), 423–434.

White, H. D., & Griffith, B. C. (1981). Author cocitation: A literature measure of intellectual structure. *Journal of the American Society for Information Science, 32*(3), 163–171.

White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972–1995. *Journal of the American Society for Information Science, 49*(4), 327–355.

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.