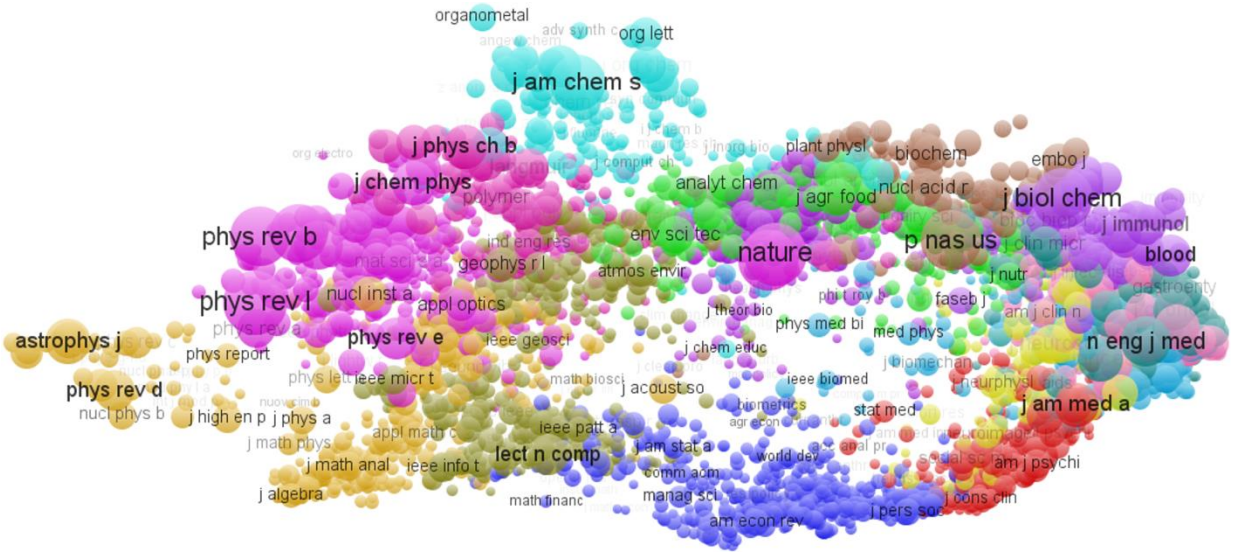
VOSviewer Manual

Nees Jan van Eck and Ludo Waltman

28 September 2016



Manual for VOSviewer version 1.6.5

Universiteit Leiden

CWTS
Meaningful metrics

Table of contents

1 Introduction 2

2 User interface 3

2.1 Main panel 4

2.1.1 Network visualization 4

2.1.2 Overlay visualization 5

2.1.3 Density visualization 6

2.1.4 Zooming and scrolling 8

2.2 Options panel 9

2.3 Information panel 12

2.4 Overview panel 12

2.5 Action panel 12

2.5.1 File tab 13

2.5.2 Items tab 15

2.5.3 Analysis tab 16

3 File types 20

3.1 Map file 20

3.2 Network file 22

3.3 Thesaurus file 22

3.4 Cluster colors file, overlay colors file, and density colors file 23

4 Advanced topics 25

4.1 Using command line parameters 25

4.2 Making a map available online 29

4.3 Increasing the availability of memory 30

References 31

1 Introduction

VOSviewer is a computer program for creating maps based on network data and for visualizing and exploring these maps. The main features of VOSviewer can be summarized as follows:

- *Creating maps based on network data*. Maps can be created based directly on a network, but it is also possible to create maps of scientific publications, scientific journals, researchers, research organizations, countries, or keywords based on co-authorship, co-occurrence, citation, bibliographic coupling, or co-citation networks extracted from Web of Science, Scopus, PubMed, or RIS files. Term maps can be created directly based on a text corpus. Maps are created using the VOS layout technique and the VOS clustering technique.
- *Visualizing and exploring maps*. Three visualizations are provided, the network visualization, the overlay visualization, and the density visualization. Zooming and scrolling functionality allows maps to be explored in full detail, which is essential when working with large maps containing hundreds or even thousands of items.

Although VOSviewer is intended primarily for analyzing bibliometric networks, the program can in fact be used to create, visualize, and explore maps based on any type of network data.

VOSviewer is written in the Java programming language, which means that it runs on most hardware and operating system platforms. VOSviewer can be obtained from www.vosviewer.com. The program can be used freely for any purpose.

This manual is concerned with version 1.6.5 of VOSviewer. The manual is organized as follows. We first introduce the user interface of VOSviewer in Chapter 2, we then explain the file types used by VOSviewer in Chapter 3, and finally we discuss some advanced topics in Chapter 4. For additional information about VOSviewer, we refer to a paper that we have written (Van Eck & Waltman, 2010). In this paper, a general introduction to VOSviewer is provided. Also, the technical implementation of specific parts of the program is discussed in considerable detail. Similar information, including a step-by-step tutorial, can also be found in a more recent book chapter (Van Eck & Waltman, 2014).

2 User interface

The main window of VOSviewer is shown in Figure 1. As can be seen in the figure, the main window consists of the following five panels:

- *Main panel*. In this panel, a visualization of the currently active map is presented. Zoom and scroll functionality can be used to determine the area in the currently active map that is shown in the main panel.
- *Options panel*. This panel can be used to change the way in which the currently active map is visualized in the main panel.
- *Information panel*. In this panel, the description of an item in the currently active map can be shown.
- *Overview panel*. In this panel, an overview of the currently active map is presented. A rectangular frame is displayed in the overview panel to indicate the area in the currently active map that is shown in the main panel.
- *Action panel*. This panel can be used to undertake all kinds of actions, such as creating a new map, opening or saving an existing map, making a screenshot, and updating the layout or the clustering of a map.

VOSviewer provides three visualizations, which are referred to as the network visualization, the overlay visualization, and the density visualization. As shown in Figure 1, the **Network Visualization**, **Overlay Visualization**, and **Density Visualization** tabs can be used to switch between the three visualizations.

In the next sections, we discuss the five panels of VOSviewer in more detail.

To get some hands-on experience with VOSviewer, we encourage the reader to use the map file `journal_map.txt` and the network file `journal_network_sparse.txt`, which are both distributed together with VOSviewer. The files contain a map of a co-citation network of 232 journals in the fields of economics, management, and operations research (for more details, see Van Eck & Waltman, 2010). To open the map, click the **Open** button on the **File** tab in the action panel, select the map file `journal_map.txt` and the network file `journal_network_sparse.txt`, and click the **OK** button.

Figure 1. The main window of VOSviewer. The numbers indicate (1) the main panel, (2) the options panel, (3) the information panel, (4) the overview panel, and (5) the action panel.

2.1 Main panel

As can be seen in Figure 1, the main panel is used to present a visualization of the currently active map. Zoom and scroll functionality can be used to determine the area in the currently active map that is shown in the main panel. The way in which the currently active map is visualized depends on whether the network visualization, the overlay visualization, or the density visualization is selected.

2.1.1 Network visualization

When the network visualization is selected, items are indicated by their label and, by default, also by a circle. For each item, the size of the item's label and the size of the item's circle depend on the weight of the item.[1] The color of an item is determined by the cluster to which the item belongs. We note that for some items

[1] The weight of an item can be determined by the *weight* (or *normalized weight*) column in a map file (see Section 3.1). When a network is available, two weights are provided automatically. One weight is the number of links of an item. The other weight is the total strength of the links of an item.

the label may not be visible. This is done in order to avoid overlapping labels. Also, by default, no lines between items are displayed. However, this can be changed by increasing the number of lines in the **Lines** text box in the options panel.

An example of the network visualization is shown in Figure 2. The distance between two journals in the visualization approximately indicates the relatedness of the journals in the co-citation network. In general, the closer two journals are located to each other, the stronger their relatedness in terms of co-citation links. Lines indicate the strongest co-citation links between journals.

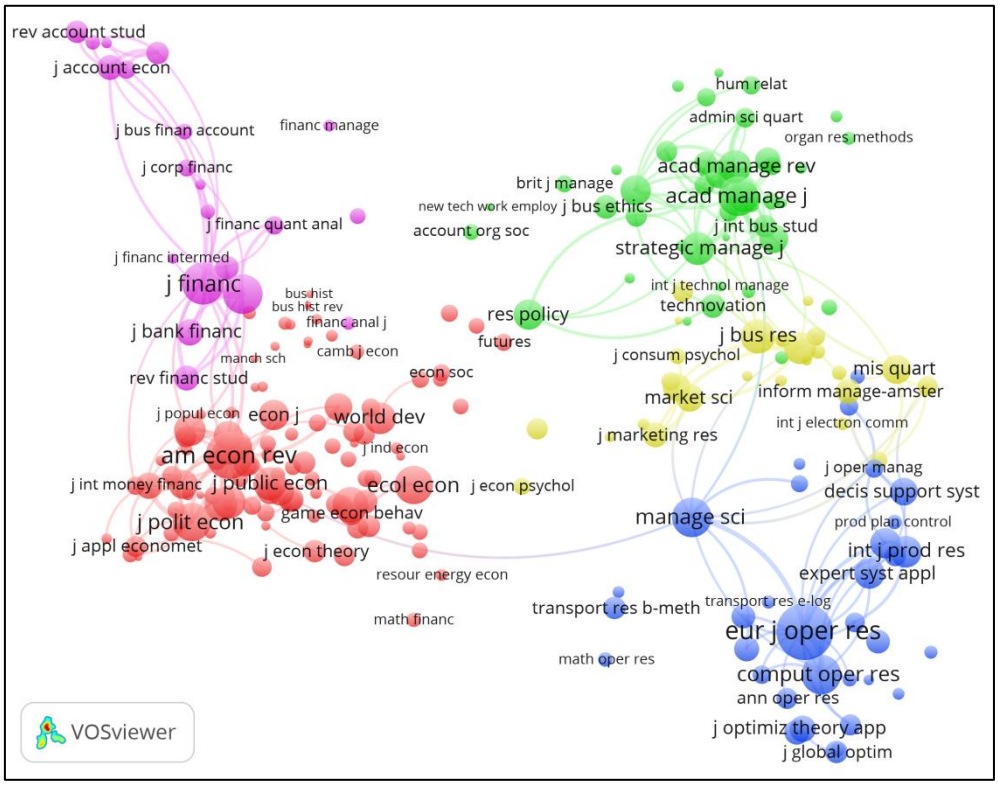


Figure 2. The network visualization.

2.1.2 Overlay visualization

The overlay visualization is identical to the network visualization except that items are colored differently. There are two ways in which items can be colored in the overlay visualization. If scores have been given to items (using the *score* column in a map file; see Section 3.1), the color of an item is determined by the item's score, where by default colors range from blue (low score) to green (average score) to red (high score). On the other hand, if user-defined colors have been specified for items (using the *red*, *green*, and *blue* columns in a map file; see Section 3.1), the

color of an item is determined by the item's user-defined color. If neither scores nor user-defined colors are available, the overlay visualization is not available.

An example of the overlay visualization is shown in Figure 3. A color bar is shown in the lower right corner of the visualization. This color bar is shown only if colors are determined by the scores of items. The color bar indicates the correspondence between colors and scores. In the overlay visualization shown in Figure 3, colors indicate the impact factors of journals.

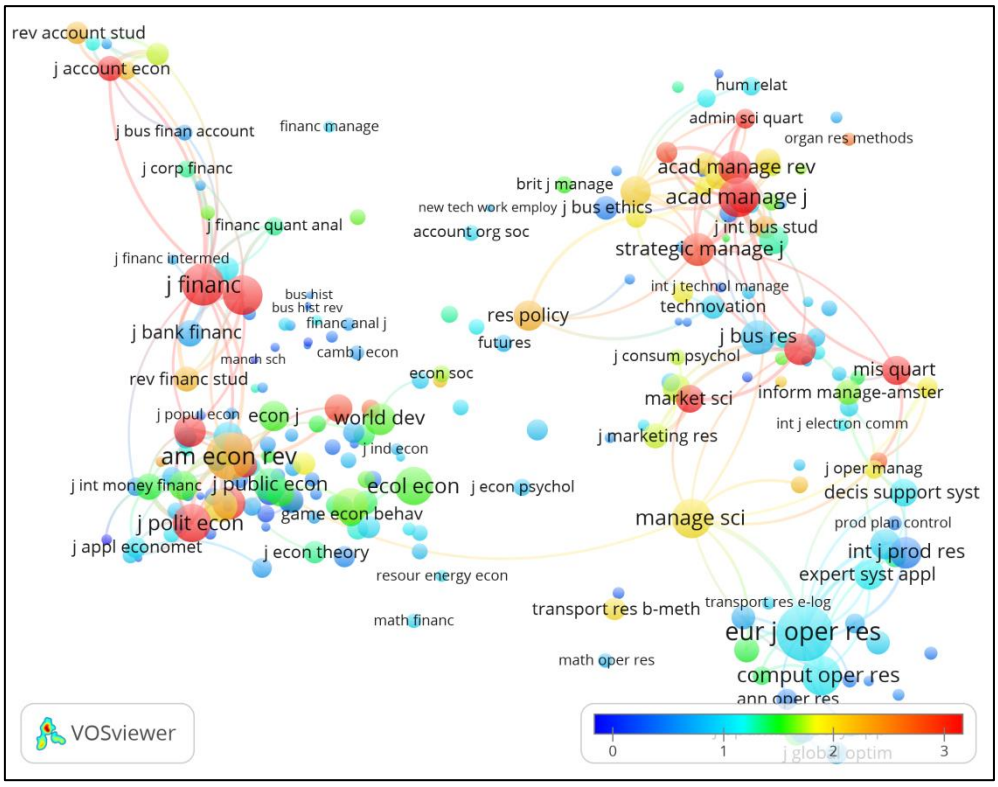


Figure 3. The overlay visualization.

2.1.3 Density visualization

There are two variants of the density visualization. We first discuss the item density visualization, followed by the cluster density visualization. The **Item density** and **Cluster density** radio buttons in the options panel can be used to switch between the two variants of the density visualization. We refer to Van Eck and Waltman (2010) for a detailed discussion of the technical implementation of the density visualization.

In the item density visualization, items are indicated by their label in a similar way as in the network visualization and the overlay visualization. Each point in a map

has a color that depends on the density of items at that point. By default, this color is somewhere in between red and blue. The larger the number of items in the neighborhood of a point and the higher the weights of the neighboring items, the closer the color of the point is to red. Conversely, the smaller the number of items in the neighborhood of a point and the lower the weights of the neighboring items, the closer the color of the point is to blue. An example of the item density visualization is shown in Figure 4.

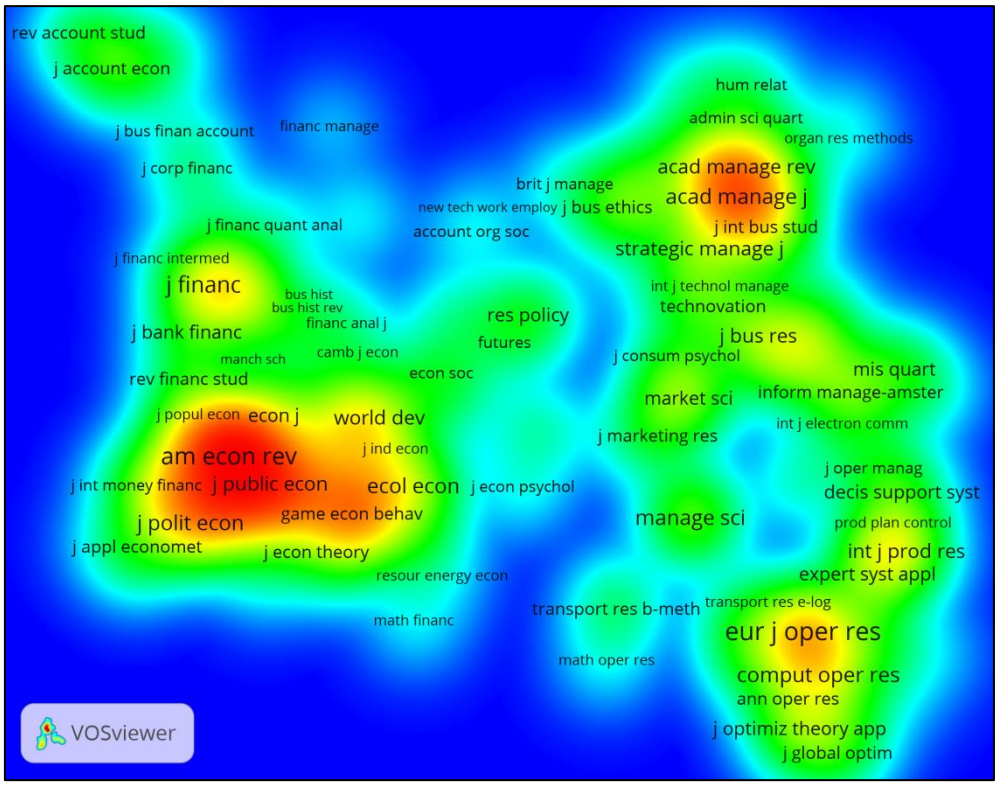


Figure 4. The item density visualization.

The cluster density visualization is available only if items have been assigned to clusters. The cluster density visualization is similar to the item density visualization except that the density of items is displayed separately for each cluster of items. In the cluster density visualization, the color of a point in a map is close to the color of a certain cluster if there are a large number of items belonging to that cluster in the neighborhood of the point. Like in the item density visualization, items with high weights count more heavily than items with low weights. An example of the cluster density visualization is shown in Figure 5.

Figure 5. The cluster density visualization.

2.1.4 Zooming and scrolling

To facilitate the detailed examination of a map, VOSviewer offers zoom and scroll functionality. In the main panel, zooming and scrolling can be done in the following three ways:

- *Using the mouse*. To zoom in, move the mouse upwards while keeping the right mouse button pressed. Conversely, to zoom out, move the mouse downwards while keeping the right mouse button pressed. As an alternative, the mouse wheel can be used to zoom in and out. To scroll through a map, move the mouse while keeping the left mouse button pressed.
- *Using the navigation buttons in the upper right corner of the main panel (see Figure 1)*. Use the plus and minus buttons to zoom in and out. Use the arrow buttons to scroll through a map.
- *Using the keyboard*. Use the plus and minus keys to zoom in and out. Use the arrow keys to scroll through a map.

2.2 Options panel

The options panel can be used to change the way in which the currently active map is visualized in the main panel. The network visualization, the overlay visualization, and the density visualization each provide somewhat different options. Some of the options are not always available. The options panel shows only the options that are relevant given the currently available map data.

When the network visualization or the overlay visualization is selected, the following options may be available:

- **Visualization**.
  - **Scale**. This slider determines the size of the labels of items in the main panel. The **Scale** slider also determines the thickness of the lines used to display links between items.
  - **Weights**. When items have multiple weights, the **Weights** drop down list determines the weights that control the size of the labels of items in the main panel.
  - **Scores**. This drop down list is available only if the overlay visualization is selected. When items have multiple scores, the **Scores** drop down list determines the scores that are used to color items in the main panel.
- **Labels**.
  - **Size variation**. The higher the weight of an item, the larger the item's label in the main panel. The **Size variation** slider determines the strength of this effect.
  - **Circles** and **Frames**. These radio buttons determine how items are indicated in the main panel. If the **Circles** radio button is selected, items are indicated both by their label and by a circle. If the **Frames** radio button is selected, items are indicated by their label displayed within a rectangular frame.
  - **Max. length**. This text box determines the maximum length of a label displayed in the main panel. If the length of a label exceeds the maximum length, only the first part of the label is displayed.
  - **Font**. This drop down list determines the font that is used to display the labels of items in the main panel.
- **Lines**.

- **Size variation**. The stronger the link between two items, the thicker the line that is used to display the link in the main panel. The **Size variation** slider determines the strength of this effect.
- **Lines**. This text box determines the number of links to be displayed in the main panel using lines. If the number of links in the network exceeds the number of links to be displayed, only the strongest links are displayed.
- **Colored lines**. This check box determines whether links are displayed using gray lines or colored lines.
- **Curved lines**. This check box determines whether links are displayed using straight lines or curved lines.

- **Colors**.
  - **Cluster Colors**. This button is available only if the network visualization is selected. The button offers four options:
    - **Edit**. This option is the default choice. Choose this option to edit the current cluster colors in the **Edit Cluster Colors** dialog box.
    - **Import**. Choose this option to import cluster colors from a cluster colors file (see Section 3.4).
    - **Export**. Choose this option to export the current cluster colors to a cluster colors file (see Section 3.4).
    - **Restore original**. Choose this option to restore the original cluster colors.
  - **Overlay Colors**. This button is available only if the overlay visualization is selected. The button offers four options:
    - **Set range**. This option is the default choice. Choose this option to bring up the **Set Overlay Colors Range** dialog box. This dialog box can be used to change the minimum and maximum score that determine how scores are mapped to colors. By default, scores less than or equal to the minimum score are mapped to blue, scores equal to the average of the minimum and the maximum score are mapped to green, and scores greater than or equal to the maximum score are mapped to red. The **Set Overlay Colors Range** dialog box can also be used to determine how scores are normalized.
    - **Import**. Choose this option to import overlay colors from an overlay colors file (see Section 3.4).

        - **Export**. Choose this option to export the current overlay colors to an overlay colors file (see Section 3.4).
        - **Restore original**. Choose this option to restore the original overlay colors.
    - **Black background**. This check box determines whether the main panel has a white or a black background color.

When the density visualization is selected, the following options may be available:

- **Visualization**.
    - **Scale**. This slider determines the size of the labels of items in the main panel.
    - **Weights**. When items have multiple weights, the **Weights** drop down list determines the weights that are used in the calculation of the density of items.
- **Labels**. These options are identical to the options provided when the network visualization or the overlay visualization is selected (except that the **Circles** and **Frames** radio buttons are not available).
- **Density**.
    - **Kernel width**. This slider determines the value of the kernel width parameter. We refer to Van Eck and Waltman (2010) for more information about this parameter.
    - **Item density** or **Cluster density**. These radio buttons determine whether the item density visualization or the cluster density visualization is selected.
- **Colors**.
    - **Density Colors**. This button is available only if the item density visualization is selected. The button offers three options:
        - **Import**. This option is the default choice. Choose this option to import density colors from a density colors file (see Section 3.4).
        - **Export**. Choose this option to export the current density colors to a density colors file (see Section 3.4).
        - **Restore original**. Choose this option to restore the original density colors.

- **Cluster Colors**. This button is available only if the cluster density visualization is selected. The button offers four options:
  - **Edit**. This option is the default choice. Choose this option to edit the current cluster colors in the **Edit Cluster Colors** dialog box.
  - **Import**. Choose this option to import cluster colors from a cluster colors file (see Section 3.4).
  - **Export**. Choose this option to export the current cluster colors to a cluster colors file (see Section 3.4).
  - **Restore original**. Choose this option to restore the original cluster colors.
- **White background**. This check box is available only if the cluster density visualization is selected. The check box determines whether the main panel has a black or a white background color.

2.3 Information panel

The information panel is available only if item descriptions are available. Item descriptions may for example have been made available using the *description* column in a map file (see Section 3.1). In the information panel, the description of an item in the currently active map can be shown. For example, when the mouse pointer is moved over the label of an item in the main panel, the description of the item will be shown in the information panel.

2.4 Overview panel

In the overview panel, an overview of the currently active map is presented. Each item in the map is indicated by a small colored dot. A rectangular frame is displayed in the overview panel to indicate the area in the currently active map that is shown in the main panel. By left-clicking in the overview panel, the area in the currently active map that is shown in the main panel can be changed.

2.5 Action panel

The action panel can be used to undertake all kinds of actions. The panel consists of three tabs: The **File** tab, the **Items** tab, and the **Analysis** tab. These tabs are discussed in Subsections 2.5.1, 2.5.2, and 2.5.3.

2.5.1 File tab

The **File** tab can be used to perform a number of basic actions. The following buttons are available on the **File** tab:

- **Map**.
  - **Create**. Use this button to create a new map. The button brings up the **Create Map** wizard. There are three ways in which a new map can be created using this wizard:
    - **Create a map based on network data**. This option requires a network. The network indicates which pairs of items are linked to each other, and for each pair of linked items it indicates the strength of their link. To create a map based on network data, a network file and optionally also a map file are needed. For a detailed discussion of these files, we refer to Sections 3.1 and 3.2. Instead of map files and network files, it is also possible to work with GML files and Pajek files.[2]
    - **Create a map based on bibliographic data**. This option requires bibliographic data. The data can be read from Web of Science, Scopus, PubMed, or RIS files. Using this option, it is possible to create maps of scientific publications, scientific journals, researchers, research organizations, countries, or keywords based on co-authorship links (i.e., multiple items co-authoring the same publication), co-occurrence links (i.e., multiple items co-occurring in the same publication), citation links (i.e., one item citing another item), bibliographic coupling links (i.e., multiple items citing the same publication), or co-citation links (i.e., multiple items being cited by the same publication).[3]
    - **Create a map based on text data**.[4] This option requires a text corpus. The corpus can be read from a corpus file. A corpus file is a text file that contains on each line the text of a document. This text is assumed to be

[2] GML (graph modeling language) files are supported by various computer programs for network analysis and visualization, for example by Gephi, a popular computer program for network visualization available at https://gephi.org. Pajek is a well-known computer program for social network analysis (De Nooy, Mrvar, & Batagelj, 2011). It is available at http://pajek.imfm.si/doku.php. VOSviewer supports Pajek matrix, network, partition, and vector files.

[3] When a map is created based on bibliographic data, either the full counting method or the fractional counting method needs to be used. We refer to Perianes-Rodriguez, Waltman, and Van Eck (2016) for more information about these two counting methods.

[4] We refer to Van Eck and Waltman (2011) for more information about this option.

in English. Using natural language processing techniques, VOSviewer extracts terms from the corpus file, where a term is defined as a sequence of nouns and adjectives (ending with a noun). Based on the extracted terms, VOSviewer creates a term map. This is a map in which terms are located in such a way that the distance between two terms provides an indication of the number of co-occurrences of the terms. In general, the smaller the distance between two terms, the larger the number of co-occurrences of the terms. Two terms are said to co-occur if they both occur on the same line in the corpus file.

In addition to a corpus file, a scores file may also be provided. A scores file is a text file that contains on each line the score of a document. Based on the scores in a scores file, VOSviewer calculates a score for each term in the term map. The score of a term equals the average score of the documents in which the term occurs. In the overlay visualization, colors are used to indicate the scores of terms.

It is also possible to create a term map based on Web of Science, Scopus, PubMed, or RIS files instead of a corpus file. In that case, terms are extracted from the titles and abstracts of scientific publications.

- **Open**. Use this button to open an existing map. The button brings up the **Open Map** dialog box. To open a map, a map file and optionally also a network file are needed. For a detailed discussion of these files, we refer to Sections 3.1 and 3.2. Instead of map files and network files, it is also possible to work with GML files and Pajek files.
- **Save**. Use this button to save the currently active map. The button brings up the **Save Map** dialog box. A map can be saved in a map file and a network file. For a detailed discussion of these files, we refer to Sections 3.1 and 3.2. Instead of map files and network files, it is also possible to work with GML files and Pajek files.
- **Print**. Use this button to print a screenshot of the main panel.
- **Screenshot**. This button offers three options:
  - **Save to file**. This option is the default choice. Choose this option to save a screenshot of the main panel. The screenshot resembles the main panel as closely as possible. However, the navigation buttons in the upper right corner of the main panel are not shown in the screenshot. If the **Optimize labeling** check box in the **Screenshot Options** dialog box (see below) is checked, the number of labels visible

in the screenshot is optimized. This means that some labels not visible in the main panel may be visible in the screenshot. Screenshots can be saved in a number of graphic file formats. For most applications, we recommend the PNG format. Some formats, such as EPS, PDF, and SVG, use vector graphics when saving a screenshot. This has the advantage that the screenshot can be resized without loss of quality.

- **Copy to clipboard**. Choose this option to copy a screenshot of the main panel to the clipboard. The screenshot can for example be pasted into a Word document or a PowerPoint presentation.
- **Screenshot options**. Choose this option to bring up the **Screenshot Options** dialog box. This dialog box can be used to change some screenshot-related settings. The **Scaling** drop down list determines the resolution (i.e., the number of pixels) of a screenshot. The resolution is calculated relative to the resolution of the main panel. Using a scaling of 100%, screenshots have the same resolution as the main panel. Using the default scaling of 200%, screenshots have a resolution that is twice as high (i.e., twice as many pixels horizontally and vertically) as the resolution of the main panel. The **Scaling** drop down list has no effect on screenshots that are saved in a file format that uses vector graphics. The **Optimize labeling** check box determines whether the number of labels visible in a screenshot is optimized. Optimizing the number of labels visible in a screenshot means that some labels not visible in the main panel itself may be visible in a screenshot of the main panel. The **Include border** check box determines whether a border is included around a screenshot.

- **Info**.
  - **Manual**. Use this button to open the VOSviewer manual. This requires an internet connection.
  - **About VOSviewer**. This button brings up the **About VOSviewer** dialog box. This dialog box provides some general information about VOSviewer, such as the version number, a copyright notice, a license text, the address of the VOSviewer web site, and a list of software libraries used by VOSviewer.

2.5.2 Items tab

The **Items** tab provides a list of items in the currently active map. By default, a list of all items in the map is provided. However, a filter can be used to restrict the list

to a subset of the items in the map. To do so, enter a filter string in the **Filter** text box. This yields a list of all items whose label contains the filter string.

The **Group items by cluster** check box determines how items are listed. If the check box is unchecked, items are simply listed alphabetically. If the check box is checked, items are first grouped by cluster and then listed alphabetically within each cluster.

By double-clicking on an item on the **Items** tab, the item is shown in the main panel.

2.5.3 Analysis tab

The **Analysis** tab can be used to update the layout and clustering of the currently active map. This is done using the VOS layout technique and the VOS clustering technique.[5] The **Analysis** tab can also be used to change the parameters of these techniques. The following parameters and controls are available on the **Analysis** tab:

- **Normalization**. Use the normalization method drop down list to determine how the strength of the links between items is normalized. The following options are available:
    - **No normalization**. If this option is selected, no normalization is performed. We generally do not recommend this option.
    - **Association strength**. If this option is selected, the association strength method is used for normalizing the strength of the links between items. Apart from a multiplicative constant, this method is identical to Eq. (6) in Van Eck and Waltman (2009). This option is selected by default.
    - **Fractionalization**. If this option is selected, the fractionalization method is used for normalizing the strength of the links between items. Apart from a multiplicative constant, this method is identical to Eq. (13) in Van Eck and Waltman (2009).
    - **LinLog/modularity**. If this option is selected, normalization is performed in the same way as in the LinLog layout technique and the modularity clustering technique. For more information about these techniques, we refer to Newman (2004) and Noack (2007, 2009).

[5] Together, these two techniques provide a unified framework for layout and clustering. We refer to Van Eck, Waltman, Dekker, and Van den Berg (2010), Waltman, Van Eck, and Noyons (2010), Waltman and Van Eck (2013), and Van Eck and Waltman (2014) for more information about these techniques.

- **Layout**.
  - **Attraction** and **Repulsion**. These parameters influence the way in which items are located in a map by the VOS layout technique. The **Attraction** parameter must have an integer value between -9 and +10. The **Repulsion** parameter must have an integer value between -10 and +9. The value of the **Repulsion** parameter must be lower than the value of the **Attraction** parameter. For most purposes, our recommendation is to set the **Attraction** and **Repulsion** parameters to values of, respectively, 2 and 1. Values of 2 and 0 or values of 1 and 0 also sometimes yield good results.
  - **Use default values**. This check box determines whether default values of the attraction and repulsion parameters will be used when creating a new map using the **Create Map** wizard (see Subsection 2.5.1). If the check box is checked, default parameters values will be used. These default parameters values depend on the type of map that is created.[6] If the check box is not checked, the parameter values specified in the **Attraction** and **Repulsion** text boxes will be used.
  - **Advanced Parameters**. This button brings up the **Advanced Layout Parameters** dialog box. This dialog box can be used to change a number of more advanced parameters of the VOS layout technique. The following parameters are available:
    - **Random starts**. This parameter determines the number of times the optimization algorithm of the VOS layout technique is run. Each time the optimization algorithm is run, a different layout may be obtained. The best layout obtained in all runs of the optimization algorithm will be used as the final layout. The larger the value of the **Random starts** parameter, the higher the quality of the final layout that will be obtained.
    - **Max. iterations**. This parameter determines the maximum number of iterations performed by the optimization algorithm of the VOS layout technique. The larger the value of the parameter, the higher the quality of the layout that will be obtained. In general, the default value of the parameter works well and does not need to be changed.

[6] The repulsion parameter has a default value of -1 when creating a co-authorship map of researchers and a default value of 0 when creating a co-occurrence map of keywords or a citation map of publications. In all other cases, the repulsion parameter has a default value of 1. The attraction parameter always has a default value of 2.

        - **Initial step size**, **Step size reduction**, and **Step size convergence**. These are technical parameters of the optimization algorithm of the VOS layout technique. The parameters must have values between 0.000001 and 1. The smaller the value of the **Step size convergence** parameter, the higher the quality of the layout that will be obtained. In general, the default values of the parameters work well and do not need to be changed.
        - **Random seed**. This parameter determines the seed of the random number generator used by the optimization algorithm of the VOS layout technique. The seed must be a non-negative integer.
    - **Update Layout**. This button is available only if a network is available. Use this button to update the layout of the currently active map using the VOS layout technique.
- **Clustering**.
    - **Resolution**. This parameter determines the level of detail of the clustering produced by the VOS clustering technique. The parameter must have a non-negative value. The larger the value of the parameter, the larger the number of clusters that will be obtained. Our recommendation is to try out different values for the parameter and to use the value that yields the level of detail considered most satisfactory for the particular application at hand.
    - **Min. cluster size**. This parameter determines the minimum size of clusters produced by the VOS clustering technique. In principle, each cluster produced by the VOS clustering technique must include at least the minimum number of items specified by this parameter. The **Min. cluster size** parameter can be useful to simplify the clustering results obtained from the VOS clustering technique by getting rid of small and uninteresting clusters.
    - **Merge small clusters**. This check box determines how the VOS clustering technique handles small clusters. Small clusters are clusters that include fewer items than the **Min. cluster size** parameter. If the check box is checked, small clusters are merged into larger clusters. If the check box is not checked, small clusters are discarded and their items have no cluster assignment.
    - **Advanced Parameters**. This button brings up the **Advanced Clustering Parameters** dialog box. This dialog box can be used to change a number of

more advanced parameters of the VOS clustering technique. The following parameters are available:

- **Random starts**. This parameter determines the number of times the optimization algorithm of the VOS clustering technique is run. Each time the optimization algorithm is run, a different clustering may be obtained. The best clustering obtained in all runs of the optimization algorithm will be used as the final clustering. The larger the value of the **Random starts** parameter, the higher the quality of the final clustering that will be obtained.
- **Iterations**. This parameter determines the number of iterations performed by the optimization algorithm of the VOS clustering technique. The larger the value of the parameter, the higher the quality of the clustering that will be obtained. In general, the default value of the parameter works well and does not need to be changed.
- **Random seed**. This parameter determines the seed of the random number generator used by the optimization algorithm of the VOS clustering technique. The seed must be a non-negative integer.

- **Update Clustering**. This button is available only if a network is available. Use this button to update the clustering of the currently active map using the VOS clustering technique.

- **Rotate/flip**.
  - **Rotate**. Use this button to rotate the currently active map. The **Degrees to rotate** parameter determines the number of degrees by which the map is rotated.
  - **Flip Horizontally**. Use this button to flip the currently active map in horizontal direction.
  - **Flip Vertically**. Use this button to flip the currently active map in vertical direction.

3 File types

The two primary file types used by VOSviewer are the map file and the network file. Map files and network files are simple text files that can easily be viewed and edited using a text editor (e.g., Notepad) or a spreadsheet program (e.g., Excel). Map files and network files consist of multiple columns. Hence, each line in a map file or a network file contains multiple information elements. Different information elements on the same line are separated from each other by a comma, a semicolon, or a tab. If an information element (e.g., the label of an item) itself contains a comma or a semicolon, the information element needs to be enclosed within double quotes.

We discuss map files and network files in more detail in Sections 3.1 and 3.2. Four additional file types, the thesaurus file, the cluster colors file, the overlay colors file, and the density colors file, are discussed in Sections 3.3 and 3.4. We note that all example files mentioned in this chapter are distributed together with VOSviewer.

3.1 Map file

A map file is a text file that contains information about the items in a map. Each line in a map file corresponds with an item. The only exception is the first line. This is a header line that indicates what is contained in each of the columns of a map file. Below, we list the columns that can be used in a map file. For each column, we provide the column header and we specify what the column contains.

id	The ID of an item. Items need to have an ID only if a map file is used in combination with a network file.
label	The label of an item.
sublabel	The sublabel of an item. In the main panel, the sublabel of an item is displayed below the item's ordinary label. Sublabels are displayed in a smaller font.
description	The description of an item. The description of an item is used to provide information about the item in the information panel. HTML formatting can be used in this column.
url	The URL of an item. This column can be used to associate a web page with an item. Clicking on the label of an item in the main panel will cause the web page associated with the item to be opened in a web browser.
x	The horizontal coordinate of an item.
y	The vertical coordinate of an item.
cluster	The number of the cluster to which an item belongs. Only integers between 1 and 1000 are allowed in this column.

weight The weight of an item. Only non-negative numbers are allowed in this column. The higher the weight of an item, the more prominent the item is presented in the main panel.

normalized weight The normalized weight of an item. Only non-negative numbers are allowed in this column. The higher the normalized weight of an item, the more prominent the item is presented in the main panel. The default presentation of an item is obtained by setting the item's normalized weight to 1.

score The score of an item. In the overlay visualization in the main panel, items can be colored based on their score, with by default colors ranging from blue to green to red.

red The red component of the user-defined color of an item. Only integers between 0 and 255 are allowed in this column. In the overlay visualization in the main panel, items can be colored based on their user-defined color.

green The green component of the user-defined color of an item. Only integers between 0 and 255 are allowed in this column. In the overlay visualization in the main panel, items can be colored based on their user-defined color.

blue The blue component of the user-defined color of an item. Only integers between 0 and 255 are allowed in this column. In the overlay visualization in the main panel, items can be colored based on their user-defined color.

In a map file, one always uses only a subset of the above columns. The order in which the columns are used is not important.

There are a number of restrictions on the columns that are used in a map file:

- There must be an *id* column or a *label* column. (If there is no *label* column, the ID of an item is used as the item's label.)
- If there is a *sublabel* column, there must be a *label* column as well.
- The *x* and *y* columns must be used together.
- The *weight* column and the *normalized weight* column cannot be used together.
- The *score* column and the *red*, *green*, and *blue* columns cannot be used together.
- The *red*, *green*, and *blue* columns must be used together.

Furthermore, it is possible to have multiple *weight* and *score* columns. Each *weight* or *score* column then must have its own label. For example, one could have a *weight<Links>* column, a *weight<Documents>* column, and a *weight<Citations>* column.

For an example of a map file, see the file `journal_map.txt`.

3.2 Network file

A network file is a text file that contains the adjacency matrix of a network. The adjacency matrix of a network is a square matrix that indicates for each pair of items in the network the strength of the link between the items. The strength of a link is given by a non-negative number. If there is no link between two items, the strength of the link between the items equals zero. VOSviewer requires networks to be undirected, and it therefore requires adjacency matrices to be symmetrical. If a network file contains an asymmetrical adjacency matrix, VOSviewer will make the matrix symmetrical by averaging corresponding elements on both sides of the main diagonal.

A network file has either a full format or a sparse format:

- *Full format*. The entire adjacency matrix, including elements that are equal to zero, is stored in the network file. The file consists of n lines and $n + 1$ columns, where n denotes the number of items in the network. The element in the ith row and the jth column of the adjacency matrix is stored on the ith line and in the $(j + 1)$th column of the network file. The first column of the network file contains IDs of items. This column indicates for each row and column of the adjacency matrix the ID of the corresponding item. For an example of a network file in full format, see the file `journal_network_full.txt`.
- *Sparse format*. Only the non-zero elements of the adjacency matrix are stored in the network file. The file consists of two or three columns. The first two columns contain IDs of items. The third column contains the non-zero elements of the adjacency matrix. This column indicates the strength of the link between the items referred to in the first two columns. If there is no third column, the strength of the link between the items referred to in the first two columns always equals one. In the case of a symmetrical adjacency matrix, it is sufficient to store only half of the matrix (either the upper triangular part or the lower triangular part) in the network file. For an example of a network file in sparse format, see the file `journal_network_sparse.txt`.

A network file is usually used in combination with a map file. For each item ID in the network file, there must then be a corresponding item ID in the map file.

3.3 Thesaurus file

A thesaurus file is a text file that can be used in the following situations:

- When creating a map based on bibliographic data (see Subsection 2.5.1), a thesaurus file can be used to merge different variants of a source title, an

author name, an organization name, a country name, or a cited reference. This may for example be useful when the name of a researcher is written in different ways in different publications (e.g., with first initial only and with all initials). A thesaurus file can then be used to indicate that different names in fact refer to the same researcher.

- When creating a map based on text data (see Subsection 2.5.1), a thesaurus file can be used to merge synonyms into a single term. This may be useful not only for merging different terms referring to the same concept, but also for merging different spellings of the same term (e.g., *behavior* and *behaviour*). It may also be useful for merging an abbreviation of a term with the term itself.

Each line in a thesaurus file contains a label and indicates an alternative label that replaces the original label. The only exception is the first line. This is a header line that indicates what is contained in each of the columns of a thesaurus file. A thesaurus file must have two columns: A *label* column and a *replace by* column. The *label* column contains a label. Depending on the situation (see above), a label may represent a source title, an author name, an organization name, a country name, a cited reference, or a term. The *replace by* column contains an alternative label that replaces the original label. The *replace by* column may also be empty. In that case, the original label is not replaced by an alternative one, but instead the original label is simply ignored. When creating a map based on a corpus file, this allows a thesaurus file to be used as a kind of stop word list. A thesaurus file may for example indicate that certain uninteresting terms (e.g., *method*, *result*, and *conclusion* in the case of a corpus of abstracts of scientific publications) should be ignored. For examples of thesaurus files, see the files `thesaurus_authors.txt` and `thesaurus_terms.txt`.

3.4 Cluster colors file, overlay colors file, and density colors file

A cluster colors file is a text file that contains the colors of clusters. Each line in a cluster colors file corresponds with a cluster. The only exception is the first line. This is a header line that indicates what is contained in each of the columns of a cluster colors file. A cluster colors file must have four columns: A *cluster* column, a *red* column, a *green* column, and a *blue* column. The *cluster* column contains cluster numbers. Only integers between 1 and 1000 are allowed in this column. The *red*, *green*, and *blue* columns contain the red, green, and blue components of the colors of clusters. Only integers between 0 and 255 are allowed in these columns. For an example of a cluster colors file, see the file `cluster_colors.txt`.

An overlay colors file is similar to a cluster colors file except that instead of a *cluster* column it has a *color value* column, with values between 0 and 1. In the overlay visualization in the main panel, items can be colored by transforming their scores into color values and by matching these color values with the color values in the *color value* column of the overlay colors file. Exact matching of color values is usually not possible, and in that case the colors in the overlay colors file are interpolated. For an example of an overlay colors file, see the file `overlay_colors.txt`.

A density colors file is identical to an overlay colors file. In the item density visualization in the main panel, the color of a point in a map is determined by transforming the density of items at that point into a color value and by matching this color value with the color values in the *color value* column of the density colors file. Exact matching of color values is usually not possible, and in that case the colors in the density colors file are interpolated. For an example of a density colors file, see the file `density_colors.txt`.

4 Advanced topics

In this chapter, some advanced topics are addressed. We first consider the use of command line parameters (Section 4.1). We then discuss how a map can be made available online (Section 4.2) and how the amount of memory available to VOSviewer can be increased (Section 4.3).

4.1 Using command line parameters

VOSviewer supports a large number of command line parameters. These parameters can for example be used to automatically open a map when VOSviewer is started or to override some of the default settings of VOSviewer. The command line parameters supported by VOSviewer are listed below.

Command line parameters for opening or creating a map

gml	Use this parameter to specify a GML file. Based on this file, a map will be opened or created when VOSviewer is started.
map	Use this parameter to specify a map file (see Section 3.1). Based on this file, a map will be opened or created when VOSviewer is started.
network	Use this parameter to specify a network file (see Section 3.2). A network will be read from this file when VOSviewer is started.
pajek_network	Use this parameter to specify a Pajek network (or matrix) file. Based on this file, a map will be opened or created when VOSviewer is started.
pajek_partition	Use this parameter to specify a Pajek partition file. Cluster numbers of items will be read from this file when VOSviewer is started.
pajek_vector	Use this parameter to specify a Pajek vector file. Weights of items will be read from this file when VOSviewer is started.

Command line parameters for creating a term map based on a corpus file

corpus	Use this parameter to specify a corpus file. Based on this file, a term map will be created when VOSviewer is started.
counting_method	Use this parameter to specify the counting method to be used when a term map is created when VOSviewer is started (1 for binary counting and 2 for full counting).
min_n_occurrences	Use this parameter to specify the minimum number of occurrences a term must have to be included in a term map.
n_terms	Use this parameter to specify the number of terms to be included in a term map. VOSviewer will select the terms that seem most relevant.
scores	Use this parameter to specify a scores file. When a term map is created when VOSviewer is started, this file will be used to calculate a score for each term.

thesaurus — Use this parameter to specify a thesaurus file (see Section 3.3). When a term map is created when VOSviewer is started, this file will be used to merge synonyms into a single term.

Command line parameters for saving a map

save_gml — Use this parameter to specify a GML file. When a map is opened or created when VOSviewer is started, the map will be saved in this file.

save_map — Use this parameter to specify a map file (see Section 3.1). When a map is opened or created when VOSviewer is started, the map will be saved in this file.

save_network — Use this parameter to specify a network file (see Section 3.2). When a map is opened or created when VOSviewer is started, the network will be saved in this file.

save_pajek_network — Use this parameter to specify a Pajek network (or matrix) file. When a map is opened or created when VOSviewer is started, the map will be saved in this file.

save_pajek_partition — Use this parameter to specify a Pajek partition file. When a map is opened or created when VOSviewer is started, the cluster numbers of the items in the map will be saved in this file.

save_pajek_vector — Use this parameter to specify a Pajek vector file. When a map is opened or created when VOSviewer is started, the weights of the items in the map will be saved in this file.

Command line parameters for saving a screenshot

save_screenshot_bmp — Use this parameter to specify a BMP file. When a map is opened or created when VOSviewer is started, a screenshot of the map will be saved in this file.

save_screenshot_emf — Use this parameter to specify an EMF file. When a map is opened or created when VOSviewer is started, a screenshot of the map will be saved in this file.

save_screenshot_eps — Use this parameter to specify an EPS file. When a map is opened or created when VOSviewer is started, a screenshot of the map will be saved in this file.

save_screenshot_gif — Use this parameter to specify a GIF file. When a map is opened or created when VOSviewer is started, a screenshot of the map will be saved in this file.

save_screenshot_jpg — Use this parameter to specify a JPG file. When a map is opened or created when VOSviewer is started, a screenshot of the map will be saved in this file.

save_screenshot_pdf — Use this parameter to specify a PDF file. When a map is opened or created when VOSviewer is started, a screenshot of the map will be saved in this file.

save_screenshot_png — Use this parameter to specify a PNG file. When a map is opened or created when VOSviewer is started, a screenshot of the map will be saved in this file.

save_screenshot_svg — Use this parameter to specify an SVG file. When a map is opened or created when VOSviewer is started, a screenshot of the map will be saved in this file.

save_screenshot_swf — Use this parameter to specify an SWF file. When a map is opened or created

when VOSviewer is started, a screenshot of the map will be saved in this file.

Command line parameters related to the layout and clustering techniques

attraction	Use this parameter to specify the initial value of the **Attraction** text box on the **Analysis** tab in the action panel.
largest_component	Use this parameter to indicate that only the largest component of the network must be kept when running the layout technique when VOSviewer is started.
merge_small_clusters	Use this parameter to specify the initial value of the **Merge small clusters** check box on the **Analysis** tab in the action panel ('true' for checked and 'false' for unchecked).
min_cluster_size	Use this parameter to specify the initial value of the **Min. cluster size** text box on the **Analysis** tab in the action panel.
repulsion	Use this parameter to specify the initial value of the **Repulsion** text box on the **Analysis** tab in the action panel.
resolution	Use this parameter to specify the initial value of the **Resolution** text box on the **Analysis** tab in the action panel.
run_clustering	Use this parameter to indicate that the clustering technique must be run when a map is opened or created when VOSviewer is started (even if a clustering is already available).
run_layout	Use this parameter to indicate that the layout technique must be run when a map is opened or created when VOSviewer is started (even if a layout is already available).
skip_clustering	Use this parameter to indicate that the clustering technique must not be run when a map is opened or created when VOSviewer is started.

Visualization-related command line parameters

black_background	Use this parameter to specify the initial value of the **Black background** check box in the options panel ('true' for checked and 'false' for unchecked).
circles_frames	Use this parameter to specify how items will be presented in the network and overlay visualizations (1 for circles and 2 for frames).
cluster_colors	Use this parameter to specify a cluster colors file (see Section 3.4). Cluster colors will be imported from this file when VOSviewer is started.
colored_lines	Use this parameter to specify the initial value of the **Colored lines** check box in the options panel ('true' for checked and 'false' for unchecked).
curved_lines	Use this parameter to specify the initial value of the **Curved lines** check box in the options panel ('true' for checked and 'false' for unchecked).
density	Use this parameter to specify whether the item density visualization or the cluster density visualization will be selected in the density visualization (1 for item density visualization and 2 for cluster density visualization).
density_colors	Use this parameter to specify a density colors file (see Section 3.4). Density colors will be imported from this file when VOSviewer is started.

density_visualization	Use this parameter to select the density visualization when VOSviewer is started.
kernel_width	Use this parameter to specify the initial value of the **Kernel width** slider in the options panel.
label_size_variation	Use this parameter to specify the initial value of the **Size variation** slider for labels in the options panel.
line_size_variation	Use this parameter to specify the initial value of the **Size variation** slider for lines in the options panel.
max_label_length	Use this parameter to specify the initial value of the **Max. length** text box in the options panel.
max_score	Use this parameter to specify the initial value of the **Max. score** text box in the **Set Overlay Colors Range** dialog box.
min_score	Use this parameter to specify the initial value of the **Min. score** text box in the **Set Overlay Colors Range** dialog box.
n_lines	Use this parameter to specify the initial value of the **Lines** text box in the options panel.
network_visualization	Use this parameter to select the network visualization when VOSviewer is started.
overlay_colors	Use this parameter to specify an overlay colors file (see Section 3.4). Overlay colors will be imported from this file when VOSviewer is started.
overlay_visualization	Use this parameter to select the overlay visualization when VOSviewer is started.
scale	Use this parameter to specify the initial value of the **Scale** slider in the options panel.
scores_normalization	Use this parameter to specify how scores are normalized (1 for no normalization, 2 for normalization by dividing by the mean, 3 for normalization by subtracting the mean, and 4 for normalization by subtracting the mean and dividing by the standard deviation).
show_item	Use this parameter to zoom in on a specific item when VOSviewer is started. The item is indicated by its ID.
white_background	Use this parameter to specify the initial value of the **White background** check box in the options panel ('true' for checked and 'false' for unchecked).
zoom_level	Use this parameter to specify the initial zoom level in the main panel. The zoom level must have a value of at least 1. The higher the zoom level, the more the main panel will be zoomed in on the center of a map.

Miscellaneous command line parameters

encoding	Use this parameter to specify the character encoding that is used by VOSviewer to read and write text files. For a list of the available encodings, see http://docs.oracle.com/javase/6/docs/technotes/guides/intl/encoding.doc.html. If this parameter is not used, VOSviewer will attempt to automatically

	determine the correct encoding when reading a text file (which in some cases may result in the use of an incorrect encoding), while it will use the default encoding when writing a text file.
file_location	Use this parameter to specify the folder that is used by VOSviewer as the default file location.

To use the above command line parameters, VOSviewer needs to be run from the command line. When using the Windows executable of VOSviewer, this can for example be done as follows:

```
VOSviewer –map map.txt –density_visualization –zoom_level 2.5
```

In this way, VOSviewer will be started and the map in the map file `map.txt` will be opened. Also, the density visualization will be selected, and the main panel will be zoomed in on the center of the map. When instead of the Windows executable of VOSviewer the VOSviewer JAR file is used, VOSviewer can for example be run in the following way:

```
java –jar VOSviewer.jar –map map.txt –density_visualization
   –zoom_level 2.5
```

We note that some command line parameters cannot be used together. For example, the `map` and `pajek_network` parameters and the `map` and `corpus` parameters cannot be used together. On the other hand, some parameters can only be used in combination with another parameter. The `pajek_partition` parameter for example can only be used in combination with the `pajek_network` parameter. Similarly, the `counting_method` parameter can only be used in combination with the `corpus` parameter.

4.2 Making a map available online

Suppose one wants to make a map available online. One way to do this is simply by making a map file and a network file (or only a map file) available on the internet. To open the map, one then needs to take two steps. In the first step, the map file and the network file are downloaded from the internet. In the second step, VOSviewer is started and the downloaded map file and network file are used to open the map.

The above two-step approach is not very convenient, and therefore an alternative approach is available as well. In this alternative approach, a map can be opened in VOSviewer directly from a web page. The alternative approach works as follows. A map file and a network file (or only a map file) need to be made available on the internet. Suppose these files are located at

```
http://www.example.com/map.txt
```

and

```
http://www.example.com/network.txt
```

The map can then be made available on a web page by creating a hyperlink that points to

```
http://www.vosviewer.com/vosviewer.php?map=http://www.example.com/map.
   txt&network=http://www.example.com/network.txt
```

Using this hyperlink, the map can be opened directly in VOSviewer.

We note that all command line parameters discussed in Section 4.1 can also be used in a hyperlink. For example, to open a map, to select the density visualization, and to zoom in on the center of the map, the following hyperlink can be used:

```
http://www.vosviewer.com/vosviewer.php?map=http://www.example.com/map.
   txt&network=http://www.example.com/network.txt&density_visualizatio
   n&zoom_level=2.5
```

4.3 Increasing the availability of memory

When using VOSviewer to work with large maps (e.g., maps containing several thousands of items), the memory requirements can be quite substantial. VOSviewer will produce an out of memory error if there is not enough memory available. The availability of memory can be increased by running the VOSviewer JAR file from the command line and specifying the amount of memory that needs to be available. For example, if 2000 MB of memory needs to be available, the VOSviewer JAR file can be run as follows:

```
java –Xmx2000m -jar VOSviewer.jar
```

When working with large maps, VOSviewer may also produce a stack overflow error. To prevent this error from occurring, the stack size needs to be increased. This can be done by running the VOSviewer JAR file from the command line in the following way:

```
java -Xss1000k -jar VOSviewer.jar
```

In this case, the stack size is set to 1000 KB, but other values are possible as well.

References

De Nooy, W., Mrvar, A., & Batagelj, V. (2011). *Exploratory social network analysis with Pajek* (2nd ed.). Cambridge University Press.

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

Noack, A. (2007). Energy models for graph clustering. *Journal of Graph Algorithms and Applications*, *11*(2), 453–480.

Noack, A. (2009). Modularity clustering is force-directed layout. *Physical Review E*, *79*, 026102.

Perianes-Rodriguez, A., Waltman, L., & Van Eck, N.J. (2016). *Constructing bibliometric networks: A comparison between full and fractional counting*. arXiv:1607.02452.

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. (2014). Visualizing bibliometric networks. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact: Methods and practice* (pp. 285–320). Springer.

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.

Waltman, L., & Van Eck, N.J. (2013). A smart local moving algorithm for large-scale modularity-based community detection. *European Physical Journal B*, *86*(11), 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.