VOSON Dashboard Userguide

For version 0.131 (10Sep18)

Xiaolan Cai[1], Robert Ackland[1], and Bryan Gertzel[1]

[1]VOSON Lab, Australian National University

Draft
25 September 2018

Contents

List of Figures

1 Introduction

1.1 VOSON Dashboard

VOSON Dashboard is an R/Shiny application for collecting and analysing networks and associated text data. VOSON Dashboard builds on a number of R packages, in particular igraph (for network analysis) and VOSON SML. VOSON SML is an R package that provides a suite of tools for collecting and constructing networks from social media data. It provides easy-to-use functions for collecting data across popular platforms (Facebook, Twitter, and YouTube) and generating different types of networks for analysis.

1.1.1 Credits

Bryan Gertzel is the lead developer of VOSON Dashboard, with Robert Ackland also contributing to the software development. The user documentation has been written by Xiaolan Cai and Robert Ackland.

1.1.2 Citing VOSON Dashboard

1.2 About this guide

VOSON Dashboard continues to be developed and improved largely based on feedback from users. This guide will be updated as we improve VOSON Dashboard.

2 Starting with VOSON Dashboard

2.1 Downloading the application

VOSON Dashboard has not yet been publicly released: participants in workshops or courses will be given instructions on how to download a zip file.

2.2 Preliminaries

It is recommended to use VOSON Dashboard with RStudio. Several R packages are required to be installed prior to VOSON Dashboard, including but not limited to: vosonSML, igraph, Shiny, Shinydashboard, shinyjs, htmlwidgets, sourcetools, crosstalk, SnowballC, and wordcloud.

2.3 Steps for running VOSON Dashboard in RStudio

To run VOSON Dashboard from RStudio (Figure 1):

1. Click "File > Open File" to open the file named "ui.R";

2. Set your working directory by clicking "Session > Set Working Directory", usually "To Source File Location";

3. Then click "Run App" on the right top of the top left window in RStudio: after a few seconds the user interface will popup.
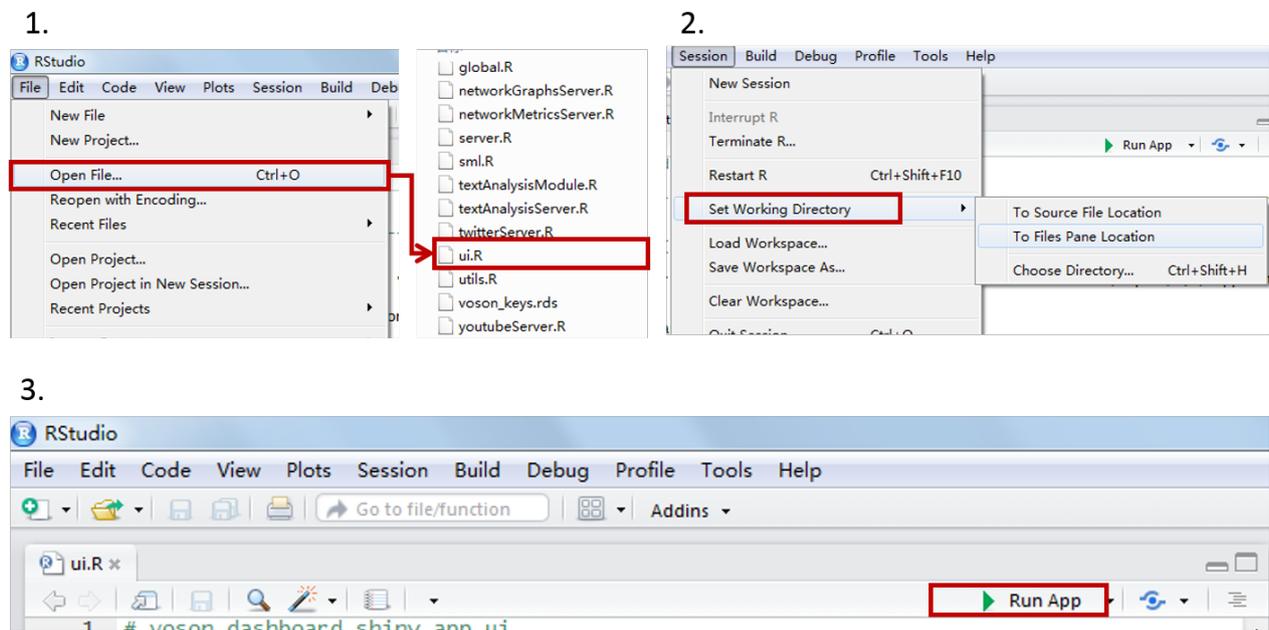

Figure 1: Steps to load VOSON Dashboard in RStudio

2.4 User interface

The navigator is on the left side, and the working space is on the right side. You can change working space by clicking menu items in the navigator (Figure 2).
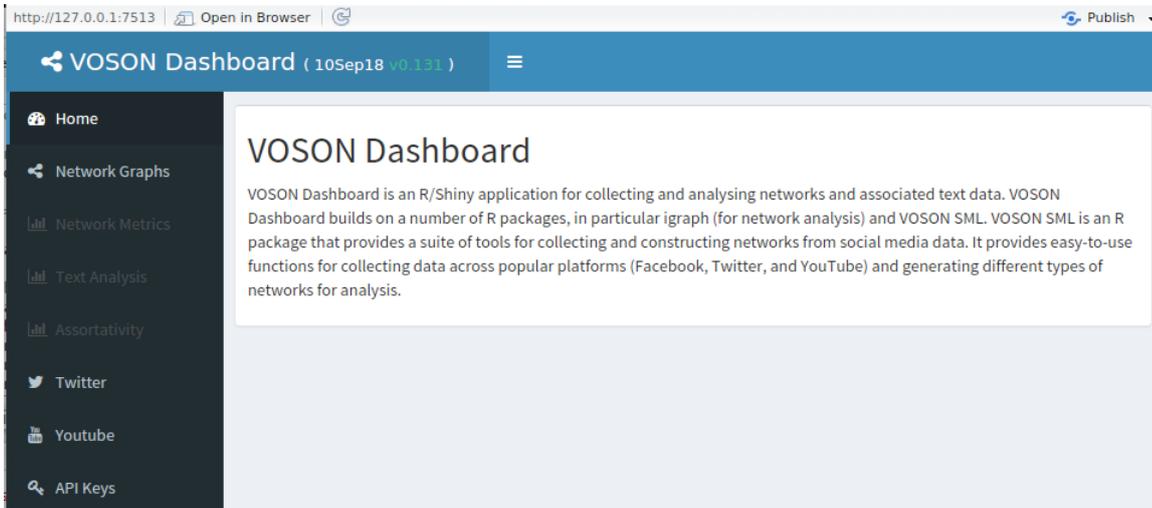
Figure 2: The interface on startup

The following menu items are listed in the navigator.

2.4.1 Home

A brief introduction to VOSON Dashboard.

2.4.2 Network Graphs

Where you can visualise network graphs and modify networks (e.g. create subnetworks).

2.4.3 Network Metrics

Node- and network-level metrics, including visualizations of node-level metrics.

2.4.4 Text Analysis

Analysis of text data for nodes/edges in the network (when it exists).

2.4.5 Assortativity

The construction of homogeneity and homophily indexes (only for networks where nodes are categorised into groups).

2.4.6 Twitter

For collecting from Twitter.

2.4.7 YouTube

For collecting data from YouTube.

2.4.8 API Keys

For storing, loading and using API keys for Twitter and YouTube.

3 Network analysis with VOSON Dashboard

3.1 Workign with networks

The first step is to load a graphml file by clicking "Choose graphml file > Browse" in "Network Graphs" (Figure 3).
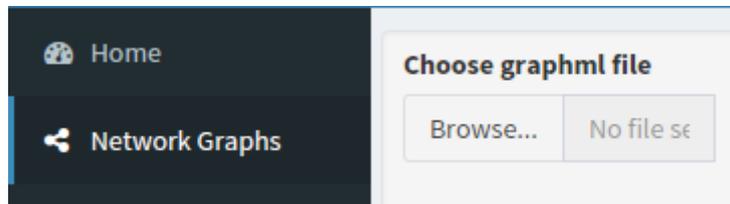

Figure 3: Load graphml data

3.1.1 Network visualization

The loaded network will be visualized with the default plot in the canvas on the right side. Three plot types are available: Plot, D3 force, D3 simple, but at present only the standard "plot" canvas provides full functionality (Figure 4)
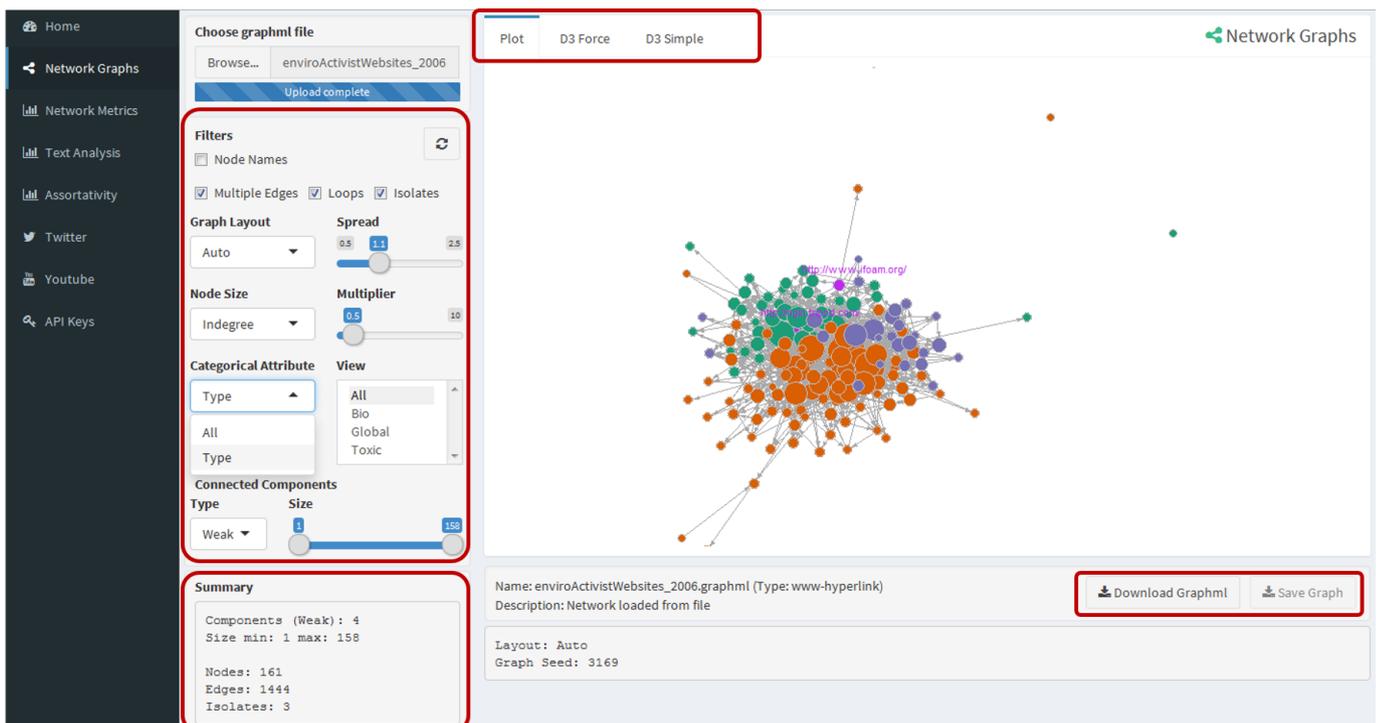
Figure 4: The Network Graphs window

A brief summary of network metrics is shown.

"Filters" can be used to control the visualization of a network and to manipulate the network (Figure **??**):

- Node names, multiple edges, loops and isolates can be shown or hidden.
- The layout of the visualization can be changed by choosing different layouts from "Graph Layout". The "Spread" bar controls the distances between nodes.
- The size of nodes can be made to reflect node-level metrics (e.g. indegree centrality or betweenness

centrality) via the "Node Size" checkbox. The "Multiplier" bar allows for control over the scaling of node size.

- The "Categorical Attribute" drop-box allows the visualization of subgroups (sets of nodes with common values of a categorical node attribute), when they exist. Colors are automatically assigned to nodes in different categories, and a second drop-box allows for subnetworks to be created based on particular values of the node categorical attribute.

- The "Connected Components" tool allows for the network to be filtered so only components matching particular criteria are included. For example, it is possible to filter the network so only nodes contained in components of size greater than 5 are included in the network. It is possible to filter the network using both "strong" and "weak" connected types. The "size" bar generate sub-networks containing components within certain size range by changing the minimum and maximum component size.

3.1.2 Viewing the network data

The vertex and edge data are shown in tables below the canvas Figure 5) and Figure 6. Nodes and edges can be sorted (e.g. sorting on indegree centrality) and it is possible to search for nodes/edges matching particular criteria (e.g. nodes with particular names).

Figure 5: The vertices table

3.1.3 Manipulating the network

There are two ways to manipulate networks: using "Filter" parameters (discussed above) and via "Pruning" to exclude particular nodes from the network.

3.1.4 Saving networks

It is possible to save a network to a graphml file by clicking the "Download Graphml" button.

Figure 6: The edges table

3.2 Network metrics

The main network-level metrics are displayed, as well as bar charts component distribution, degree distribution(only for undirected network), indegree distribution and outdegree distribution (only for directed network).

3.3 Assortativity

For network with subgroups, the computation of homogeneity and homophily indexes are shown.

4 Text analysis with VOSON Dashboard

For a network with text fields (either node or edge attributes), it is possible to conduct basic text analysis.

At present, VOSON Dashboard provides two types of text analysis: word frequency bar charts and word clouds. For network with subgroups, a comparison word cloud will also be generated.

4.1 Text filters

The "Filters" controls allow you to filter and manipulate the text before analysis (Figure 8):

- "Remove standard stopwords": this removes (English) stopwords such as "and", "the", and "but".

- "User-define stopwords": here you can create your own list of stopwords (use commas or spaces between stopwords.

- "Apply word stemming": this attempts to reduce words to their stems.

- When you are working with a Twitter network, two other options are available: "Remove Twitter hashtags" and "Remove Twitter Usernames".

4.2 Visualizing word frequencies

Below the "Filters" is a summary of text analysis results, including text attributes, stopwords, and word counts for all and each category.
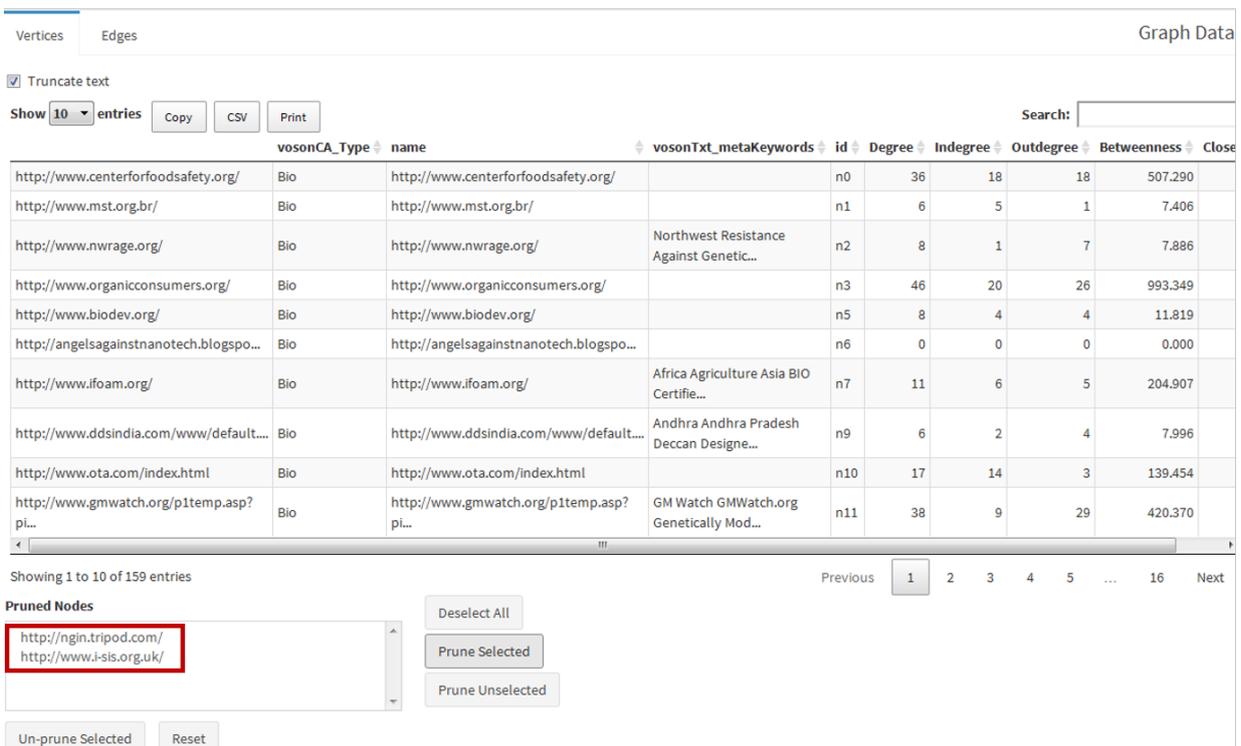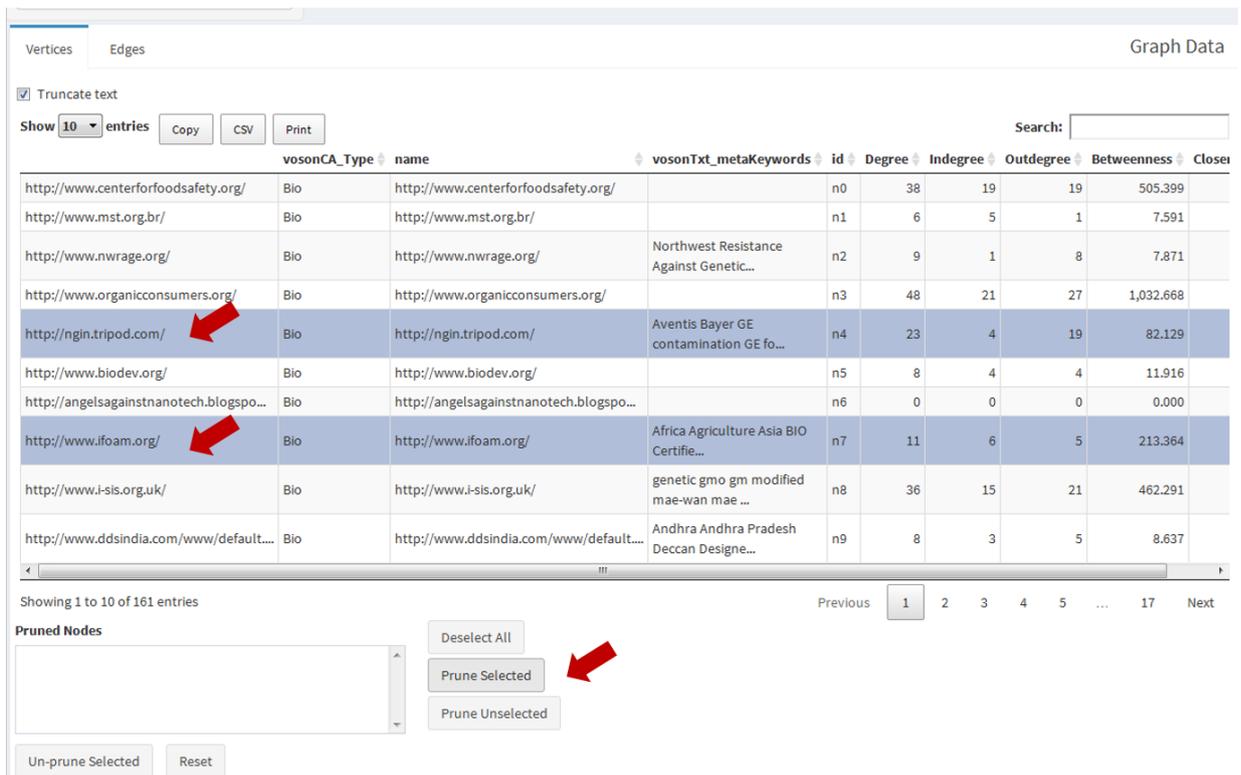
Figure 7: Manipulating network database

The canvas on the right side contains visualizations of text analysis. There are two sliders, "Display Top" is used to control how many words will appear in the visualisations (only the top x words, based on frequency counts, will be shown), and "Minimum Frequency" controls how many times a word needs to have been used in order for it to feature in the visualizations.

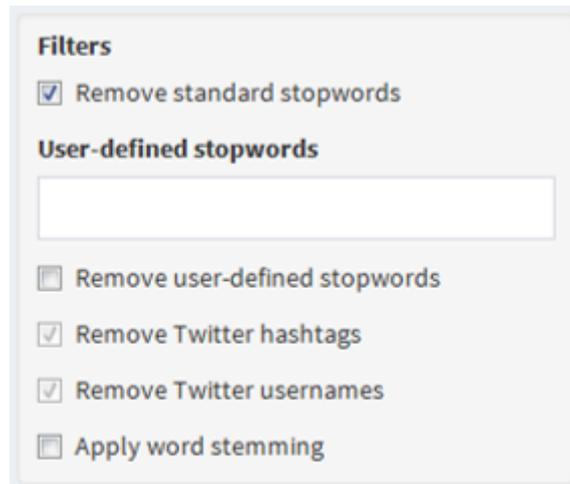There are three types of text analysis visualization:

Figure 8: Text filters

- "Word Frequency": contains frequency bar chart(s), and if applicable a bar chart for each category will be displayed (Figure 9).

- "Word Cloud": the size of a word to represents its frequency. For network with subgroups, word clouds for each category will also be displayed (Figure 10).

- "Comparison Cloud": when the network contains subgroups, this will show the words that most identified with or associated with each group. (Figure 11).

It is possible to save the text analysis visualizations by right-clicking "Save Image" and inputting the file name and file type.
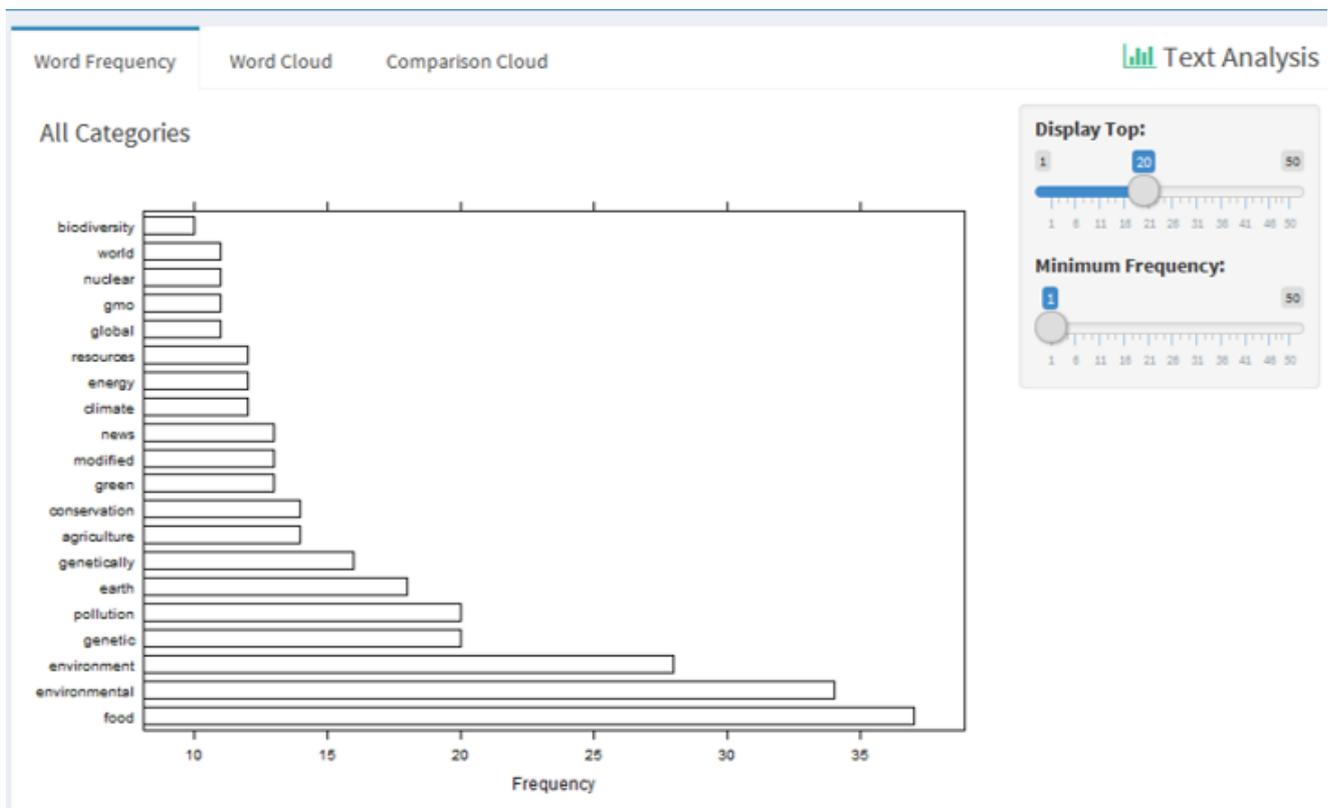

Figure 9: Word frequency bar chart

Figure 10: Word cloud

Figure 11: Comparison cloud

5 Collecting data with VOSON Dashboard

5.1 API Keys

To collect data from Twitter and YouTube, you first need to obtain API keys for Twitter and YouTube. Some instructions are provided here, but please note that the social media companies change APIs and so these instructions may be out of date.

5.1.1 Loading API keys

While it is possible to paste API keys directly in the "Twitter" or "YouTube" collection windows (Figure 12), a better approach is to use the "API Keys" window (Figure 13): here you can paste your keys and then click "Save Keys" to save your API keys to disk (to a file named "voson_keys.rds"). You will then be able to access your keys in a later session using the "Load Keys" button. *If you use a public accessible computer, please make sure to delete your "voson_keys.rds" file.*
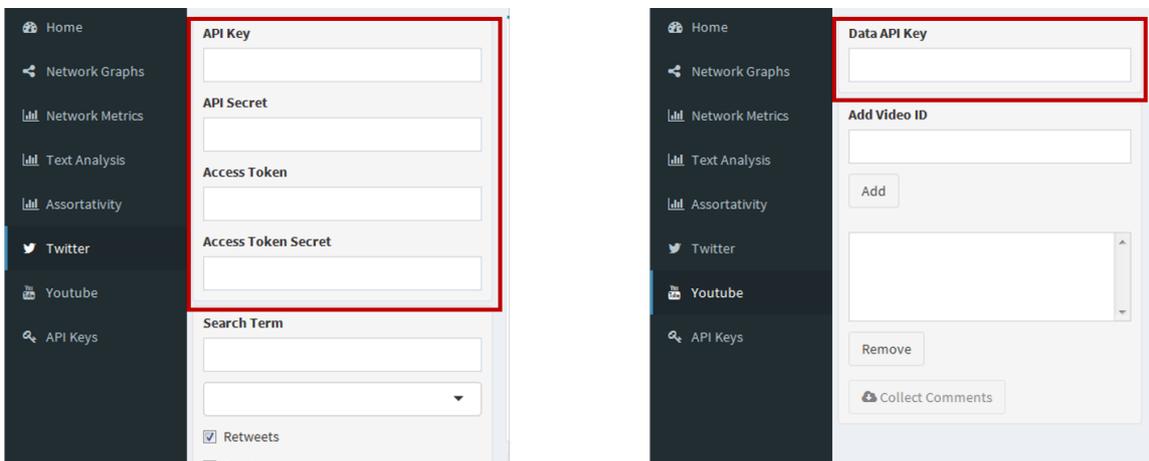

Figure 12: Input API keys directly in "Twitter" or "YouTube" menu

Figure 13: Input, save and load API keys in "API Keys" window

5.2 Collecting data from Twitter

5.2.1 Search Terms

Both hashtags and account names can be used as search criteria for Twitter.

- When using hashtags as key words, leave the drop-down menu under the key words box empty.
- When using accounts as keywords, the drop-down menu can be left empty, or choose either of "From" or "To". Leaving the menu empty will crawl tweets and re-tweets from and to the ac-

count(s). Selecting "From" will only crawl the tweets and re-tweet initiative from the account(s), and selecting "To" will only crawl tweets and re-tweets to the account(s).

If more than one search terms are defined, tweets and re-tweets contain either of the terms will be collected.

By clicking or un-clicking the boxes in front of "Re-tweets" and "Replies", the type of Twitter activities to be collected can be defined.

The "Count" bar is design to control the number of collected tweets.

The "Language" bar is design to select the language of tweets to be collected.

The "Date Range" provides a option to collect tweets during a certain time period. *Please note that the Twitter Free API only allows collection of tweets authored over the previous 7 days.*

5.2.2 Navigating the database

A table of the collected data is shown at the bottom. Basic attributes of each tweets are listed in this table, including: text, favorite and favorite count, IDs of tweets and re-tweets, time and etc.

There is a "Truncate text" option on the left top of the data table. If click the "Truncate text", a column containing truncated text will be generated.

At the top right of the table, there is a Search box: by entering text, only the matched records will be shown.

5.2.3 Saving the database

The collected data can be download by right-clicking the buttons below the "Console":

- "Download Data": Click this to save a database in csv format.
- "Download Graphml": Click this to save a database in graphml format.
- "Download Graphml (+ text)": Click this to save a graphml file that also includes the text content.

5.2.4 Working with the Twitter database

There are two ways to work with the Twitter data just collected: Click "View Graph" to visualize the network, or save the database as a graphml file and then load it from the "Network Graphs" window.

5.3 Collecting data from YouTube

5.3.1 Search term

YouTube comments can be collected by using the video ID, which is the part after "=" in the YouTube URL. For example, the ID of the video "https://www.youtube.com/watch?v=l94v4YOqxOc" is "l94v4YOqxOcs" (Figure 14). When more than one video IDs are applied to search, the comments of each video will be crawled.
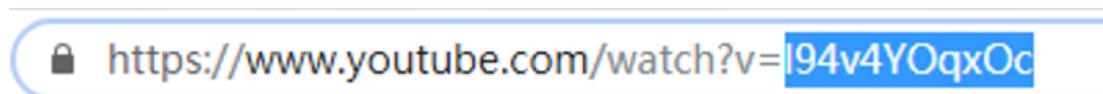

Figure 14: YouTube Video ID

5.3.2 Navigating the database

The table of collected YouTube data is listed at the bottom of the "Console". The main attributes of collected comments are listed in this table, including the user who post the comment, the text content of comment, reply to other users and etc.

By clicking the "Truncate text" on the top left of the table, a column of truncated comments will be generated.

Similar to Twitter data, there is a Search box on the top right of the table. You can enter any text and matched records will be shown.

Figure 15: Searching in YouTube data

5.3.3 Saving database

The collected YouTube database can be saved in two ways:

- By right-clicking "Download Data" and entering name and type (csv) of the file, a csv file is saved.

- By right-clicking "Download Graphml" and entering name and type (graphml) of the file, a graphml file is saved.

5.3.4 Working with collected YouTube data

As with Twitter, you can quickly view the visualization of collected YouTube data by clicking "View Graph". Alternatively, you can save the graphml file and then load into the "Network Graphs" window.