VOSON Dashboard Userguide

For version 0.5.7

Francisca Borquez Bryan Gertzel Xiaolan Cai Robert Ackland

VOSON Lab, The Australian National University

28 August 2020

Contents

1 Introduction to VOSON Dashboard

`VOSON Dashboard` is an interactive `R/Shiny` application for collecting and analysing networks and associated text data. The app has a dashboard layout with sections for visualising and manipulating network graphs, performing text analysis, displaying network metrics and the collection of network data. The social media network and text data collection is performed using the `vosonSML` "Social Media Lab" R package, which provides easy-to-use functions for collecting data across popular platforms (Twitter, YouTube and Reddit)

Figure 1: image

and generating different types of networks for analysis. `VOSON Dashboard` builds on a number of R packages, in particular igraph for network analysis.

1.1 Credits

Bryan Gertzel is the lead developer and maintainer of `VOSON Dashboard`, with Robert Ackland also contributing to the software development. The user documentation has been written by Francisca Borquez, Xiaolan Cai and Robert Ackland. Bryan Gertzel is the lead developer and maintainer of `vosonSML`. The lead developer of the first version of `vosonSML` (which was then called SocialMediaLab) was Timothy Graham, with contributions by Robert Ackland and Chung-hong Chan.

1.2 Citing VOSON Dashboard

1.3 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 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`, `Shiny`, `Shinydashboard`, `DT`, `shinyjs`; `igraph` and `visNetwork` (graph visualisation), `tm`, `SnowballC`, `wordcloud` and `syuzhet` (text analysis). For an up-to-date list of required packages, please check the GitHub page.

In addition to that, it is recommended to have up-to-date versions of the packages to ensure the app runs smoothly.

2.2 Installing VOSON Dashboard

`VOSON Dashboard` is an R package and must be installed before the app can be run. Version 0.5.7 is presently available via GitHub and CRAN.

To install the latest Github release, please run the following command:

```
install.packages("https://github.com/vosonlab/VOSONDash/releases/download/v0.5.7/VOSONDash-0.5.7.tar.gz
      repo = NULL, type = "source")
```

To install the CRAN release, please run the following command:

```
install.packages("VOSONDash")
```

2.3 Running VOSON Dashboard in RStudio

Once `VOSON Dashboard` is installed, the Shiny web application can be run from the RStudio console using the `runVOSONDash()` function:

```
library(VOSONDash)
runVOSONDash()
```

`VOSON Dashboard` will open in the default web browser.

When run, `VOSON Dashboard` will first check that all of the required R packages are installed. If any packages are missing, the app will print a message indicating the missing packages and a command that can be used to install them.

2.4 User interface

The navigator is on the left side (black), and the workspace is on the right side. You can access the different workspaces by clicking on menu items in the navigator (Figure 2).

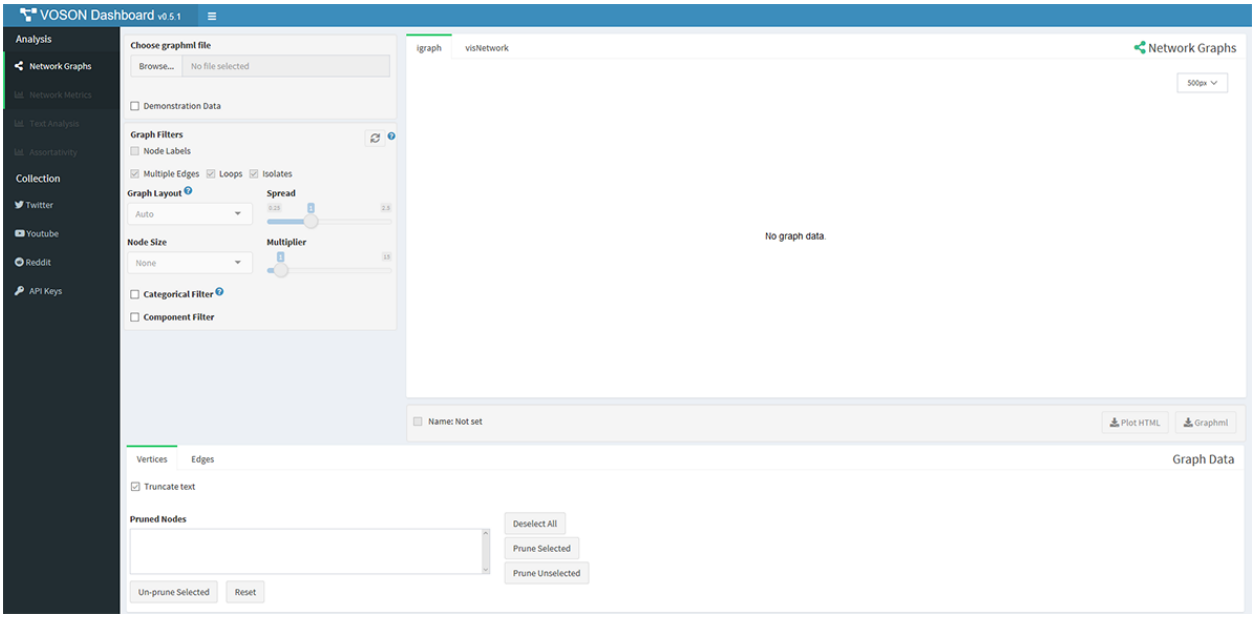


Figure 2: The interface on startup

The following items are listed in the navigator. Further details of these items are discussed in the following sections.

2.4.1 Network Graphs

In this workspace you can load files, visualise network graphs and modify networks (e.g. create subnetworks).

2.4.2 Network Metrics

This option becomes available when a Graphml file is loaded. It displays node- and network-level metrics, including visualisations of node-level metrics.

2.4.3 Text Analysis

This option becomes available when a Graphml file is loaded, and where text data are stored as either node or edge attributes (see below). It displays basic text analysis (e.g. frequency counts, wordclouds) and sentiment analysis.

2.4.4 Assortativity

The construction of homogeneity and homophily indexes (only for those networks where there is a categorical node attribute).

2.4.5 Twitter

For collecting tweets and creating Twitter networks ("Actor", "Activity", "Two-mode" and "Semantic" networks).

2.4.6 YouTube

For collecting comments from YouTube videos and creating YouTube networks ("Actor" and "Activity" networks).

2.4.7 Reddit

For collecting comments on Reddit posts and creating Reddit networks ("Actor" and "Activity" networks).

2.4.8 API Keys

For storing, loading and using API keys for Twitter and YouTube. VOSON Dashboard collects social media data via vosonSML, which in turn uses rtweet, for Twitter data collection. YouTube accesses the Google Data API directly. The Reddit collection is based on an approach found in the RedditExtractoR package, and uses an open public API that does not require to enter credentials.

3 Network analysis with VOSON Dashboard

3.1 Working with networks

To work with networks in `VOSON Dashboard`, you can (1) open a Graphml file, for example one of the demo files provided, or (2) collect data and create a network, which will be discussed in section 5.

On the 'Network Graph' workspace, load a Graphml file through clicking on "Choose Graphml file > Browse" (Figure 3). Alternatively, you can access demo files by clicking on the "Package datasets" box. Select a demo dataset and click on the "Load Graphml" button.

3.1.1 Network visualisation

After the Graphml file is loaded, a network visualisation will be presented on the "Network Graph" area (white canvas, on the right side). Two fully functional plot types are available within this version: `igraph` and `visNetwork`. General network data (number of nodes, edges and isolates) are reported in the Network Graph area (bottom LHS) (Figure 4).

The Network Graph area can be re-sized by changing the pixel options provided in the top RHS.

Additional commands and information will be displayed at the left side of the visualisation.

"Graph Filters" can be used to control the visualisation and to manipulate the network.

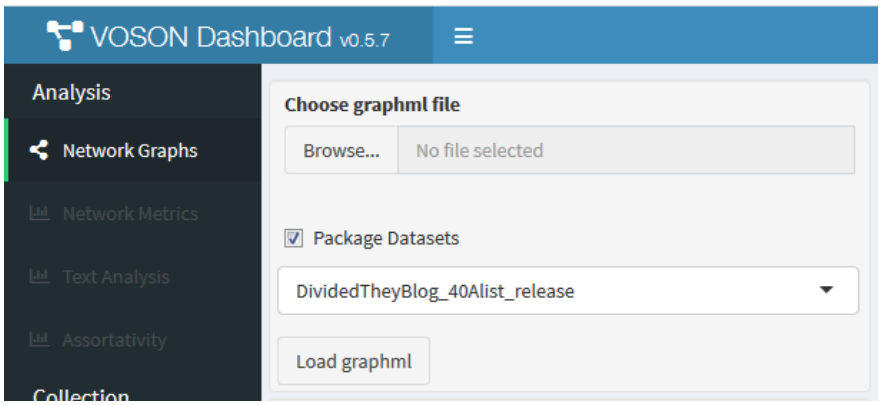


Figure 3: Load graphml data

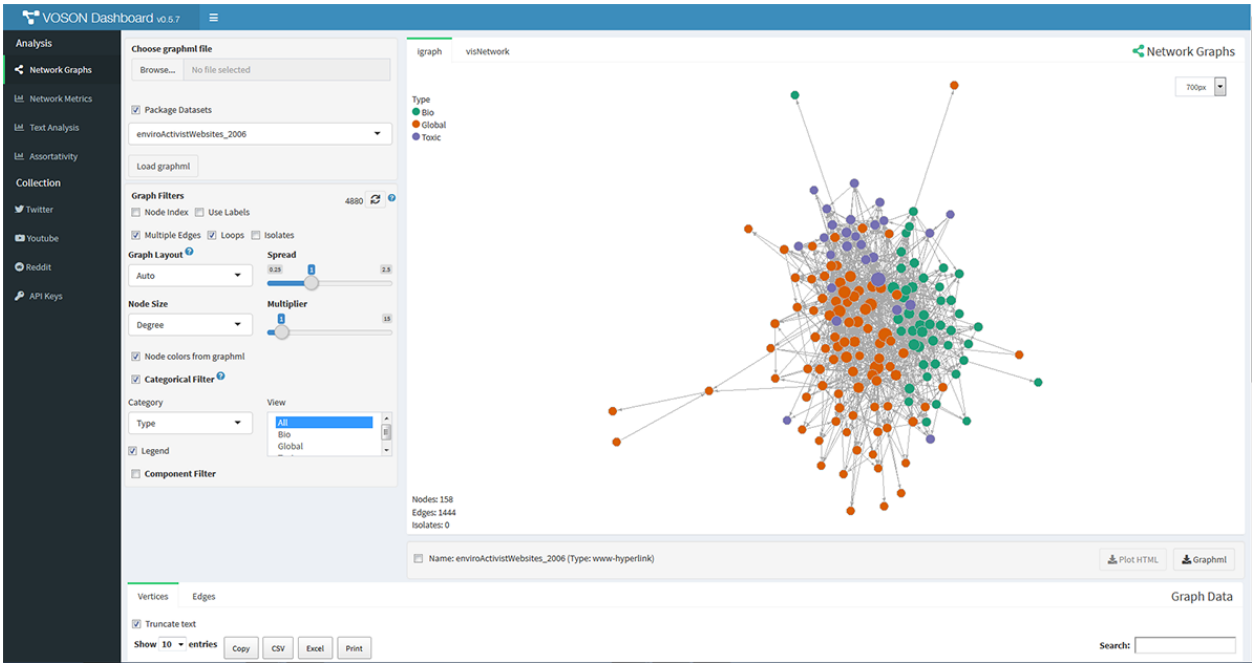


Figure 4: The Network Graphs window

- "Node index", "Node labels", "Multiple edges", "Loops" and "Isolates" may be shown or hidden. Note that at present, the "Multiple edges" function displays all types of edges in the network, as well as, all the attributes in the edges table. By deselecting this option, the network will be simplified to reflect an unweighted edge. The attributes in the edges table will be simplified too. Simplifiying multiple edges is a default `igraph` action, as the user would have to manually input the edge attributes that would like to keep in the network. In a future versions, we plan to introduce edge filters, to create e.g. networks of retweets only, and to allow the user to select which edge attributes to keep.
- The "Reseed" button changes the seed for the random number generator, and the plot is re-drawn. Plot seed value is displayed next to the re-seed button.
- The layout of the visualisation may be changed by choosing one of the "Graph Layout" algorithms.
- The number of iterations "niter" input field may be adjusted for 'FR' and GraphOpt plots.
- Input fields for GraphOpt plots: "charge", "mass", "spring.length" and "constant".
- The "Spread" select box controls the distances between nodes.
- "Node size" can be adjusted to reflect node-level metrics (e.g. indegree centrality or betweenness centrality) via the "Node Size" select box.
- The "Multiplier" slider allows for control over the node size scaling.
- "Node colour from Graphml" inherits node colors from loaded Graphml files in plots.
- The "Categorical Filter" checkbox allows for the visualisation of subgroups (sets of nodes with common values of a categorical node attribute), when they exist (i.e. when the Graphml file already contains categorical attributes). Note that the name of the categorical node attribute needs to be prepended with "vosonCA_", for categorical node attributes to be available in `VOSON Dashboard`. Colours 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. *Note that at present, it is only possible to use categorical attributes for **nodes** and not **edges**, in terms of creating and visualising subnetworks in graphs and generating wordclouds and comparison clouds.*
- The "Component Filter" 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" slider generates sub-networks containing components within certain range by changing the minimum and maximum component size.

3.1.2 Viewing the network data

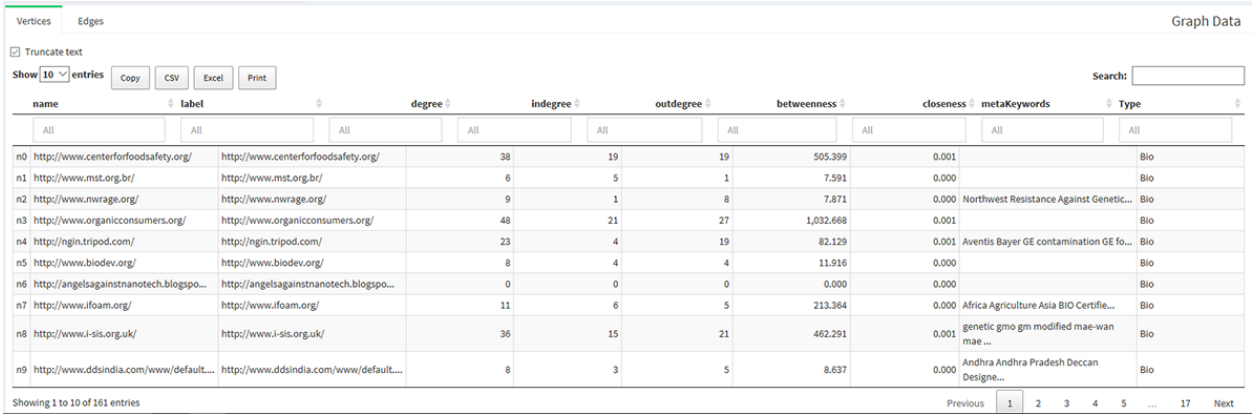
The vertex (nodes) and edge data are shown in tables below the graph, in the "Graph Data" area in (Figure 5) and (Figure 6). Nodes and edges can be sorted by any field provided in the table's top row (e.g. sorting on indegree centrality), and it is possible to search for nodes/edges matching particular criteria (e.g. nodes with particular names).

Graph data shown in the vertex (nodes) and edge tables can be copied, downloaded (CSV, Excel) or printed.

The Network Graph area can be re-sized by changing the pixel options provided in the top RHS.

3.1.3 Manipulating the network

There are two ways to manipulate networks: (1) via the "Filter" parameters (discussed above), and (2) via "Pruning", to manually exclude particular nodes from the network, as shown in (Figure 7).



Vertices Edges Graph Data

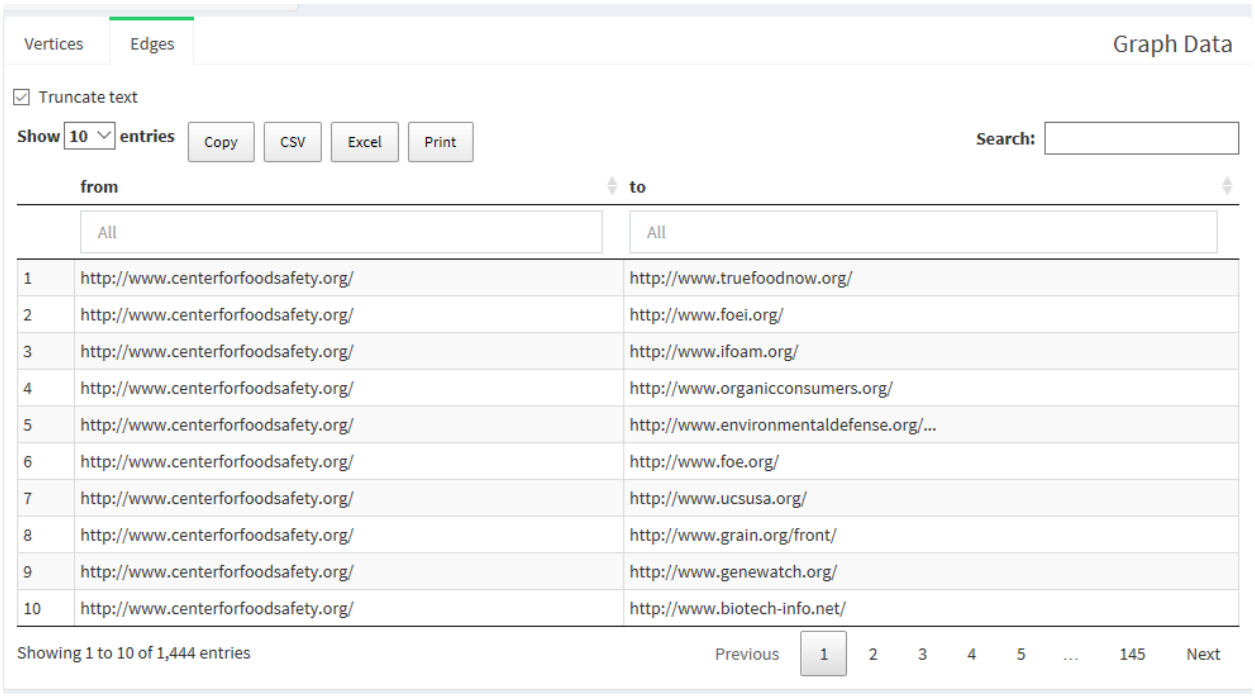
Truncate text

Show 10 entries Copy CSV Excel Print Search:

	name	label	degree	indegree	outdegree	betweenness	closeness	metaKeywords	Type
	All	All	All	All	All	All	All	All	All
n0	http://www.centerforfoodsafety.org/	http://www.centerforfoodsafety.org/	38	19	19	505.399	0.001		Bio
n1	http://www.mst.org.br/	http://www.mst.org.br/	6	5	1	7.591	0.000		Bio
n2	http://www.nwrage.org/	http://www.nwrage.org/	9	1	8	7.871	0.000	Northwest Resistance Against Genetic...	Bio
n3	http://www.organicconsumers.org/	http://www.organicconsumers.org/	48	21	27	1,032.668	0.001		Bio
n4	http://ngin.tripod.com/	http://ngin.tripod.com/	23	4	19	82.129	0.001	Aventis Bayer GE contamination GE fo...	Bio
n5	http://www.biodev.org/	http://www.biodev.org/	8	4	4	11.916	0.000		Bio
n6	http://angelsagainstnanotech.blogspo...	http://angelsagainstnanotech.blogspo...	0	0	0	0.000	0.000		Bio
n7	http://www.ifoam.org/	http://www.ifoam.org/	11	6	5	213.364	0.000	Africa Agriculture Asia BIO Certifie...	Bio
n8	http://www.i-sis.org.uk/	http://www.i-sis.org.uk/	36	15	21	462.291	0.001	genetic gmo gm modified mae-wan mae ...	Bio
n9	http://www.ddsindia.com/www/default....	http://www.ddsindia.com/www/default....	8	3	5	8.637	0.000	Andhra Andhra Pradesh Deccan Designe...	Bio

Showing 1 to 10 of 161 entries Previous 1 2 3 4 5 ... 17 Next

Figure 5: The vertices table



Vertices Edges Graph Data

Truncate text

Show 10 entries Copy CSV Excel Print Search:

	from	to
	All	All
1	http://www.centerforfoodsafety.org/	http://www.truefoodnow.org/
2	http://www.centerforfoodsafety.org/	http://www.foei.org/
3	http://www.centerforfoodsafety.org/	http://www.ifoam.org/
4	http://www.centerforfoodsafety.org/	http://www.organicconsumers.org/
5	http://www.centerforfoodsafety.org/	http://www.environmentaldefense.org/...
6	http://www.centerforfoodsafety.org/	http://www.foe.org/
7	http://www.centerforfoodsafety.org/	http://www.ucsusa.org/
8	http://www.centerforfoodsafety.org/	http://www.grain.org/front/
9	http://www.centerforfoodsafety.org/	http://www.genewatch.org/
10	http://www.centerforfoodsafety.org/	http://www.biotech-info.net/

Showing 1 to 10 of 1,444 entries Previous 1 2 3 4 5 ... 145 Next

Figure 6: The edges table

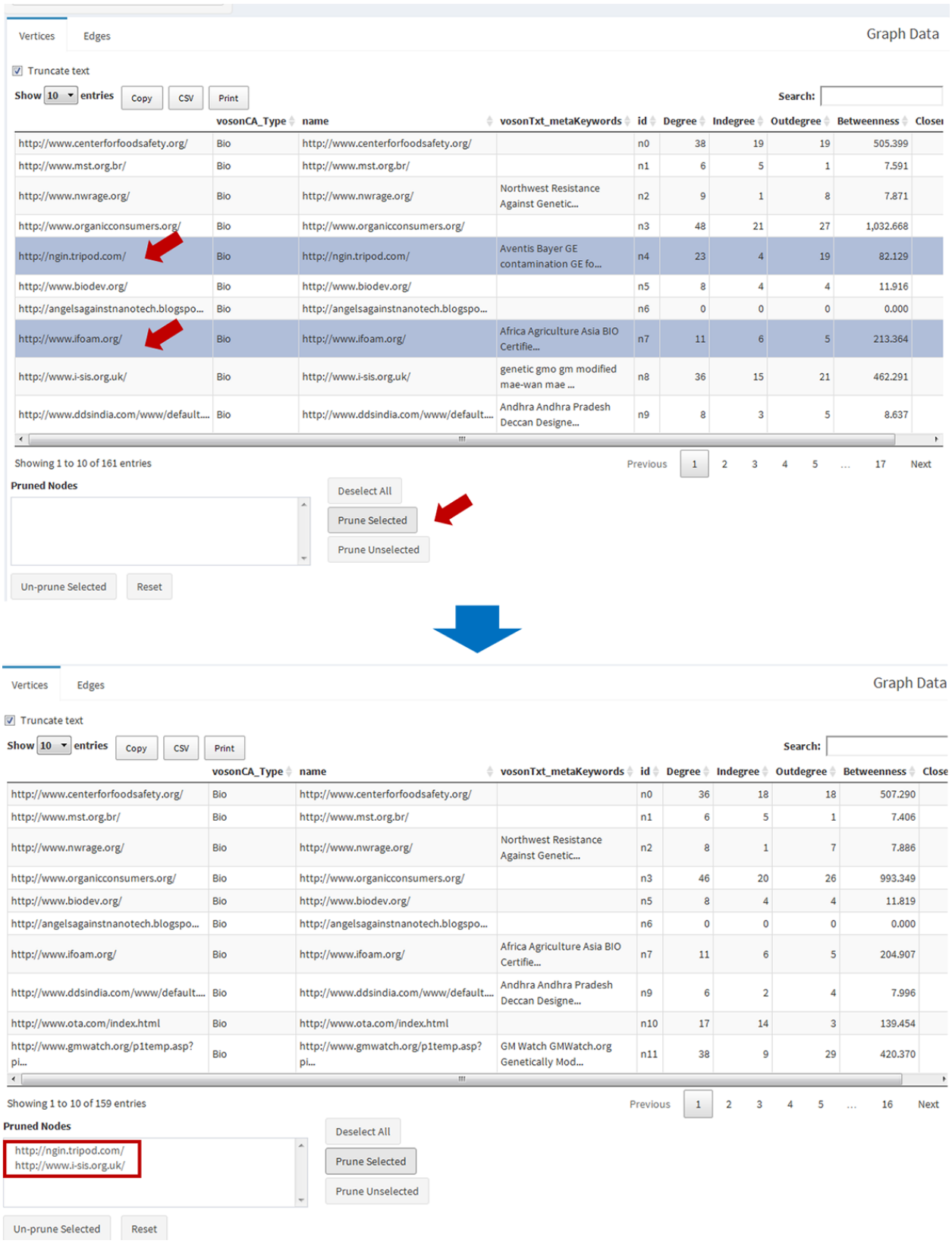


Figure 7: Pruning the network

3.1.4 Saving networks

To save a network to a Graphml file, click on the "Download Graphml" button below the visualisation canvas.

3.2 Network metrics

The main network-level metrics are displayed via the workspace "Network Metrics", as well as bar charts including component distribution, degree distribution (only for undirected networks), indegree distribution and outdegree distribution (only for directed networks).

3.3 Assortativity

For networks with subgroups, i.e. when the Graphml dataset already contains categorical node attributes, the "Assortativity" function (main menu) displays Homogeneity and Homophily indexes, including mixing matrix and population share.

3.4 Network Type

This option is available for collections conducted using `VOSON Dashboard`. The operationalisation of nodes and edges will be discussed in the context of the data sources, in section 5.

`VOSON Dashboard` enables the creation of two types of network across all data sources: "Actor networks" and "Activity networks". In "Actor networks", nodes are users and edges represent commenting activity such as, a reply comment in YouTube or a retweet in Twitter. "Actor networks" are useful to understand social behaviour and human interaction.

In "Activity networks", nodes represent comment activities, e.g. a tweet in Twitter or a Reddit comment, and edges represent responses to that comment or its diffusion (e.g. retweet, quote tweets). "Activity networks" are particularly useful to understand content diffusion and networks around issue spaces.

There are two additional supported networks available for Twitter data: "Two-mode networks", where nodes are actors (Twitter users) and hashtags, and edges represent relationships based on whether a user has retweeted, replied or mentioned; and "Semantic networks" where nodes represent unique concepts (terms/words extracted from the collection, hashtags and actors) and edges represent relationships between concepts, based on co-occurrence.

4 Text analysis with VOSON Dashboard

For a network with text data stored as either node or edge attribute, it is possible to conduct basic text analysis with `VOSON Dashboard`. At present, V`VOSON Dashboard` provides three types of text analysis: "Word frequency" bar charts, "Word clouds" and "Comparison clouds" (only when categorical data is available), and "Sentiment analysis".

It depends on the network type as to whether the text data is stored as an edge or node attribute. For the environmental activist "WWW hyperlink network" supplied as a demo network, the text data are stored as a node attribute. For Twitter/Reddit/YouTube networks created using `VOSON Dashboard`, the text data are stored as either edge attribute ("Actor networks"), or node attribute ("Activity networks").

Further note that for text data to be available for analysis in `VOSON Dashboard`, the name of the node/edge attribute where the text is stored must have "vosonTxt_" prepended to it.

It is possible that a user may, using R/igraph directly or other network analysis software (not via `VOSON Dashboard`), create new edge/node attributes containing text data. For example, one might create a node

attribute in a Twitter network that contains the profile data for Twitter users. In the situation where a network contains text data (with "vosonTxt_" prefix) stored as *both node and edge attributes*, then `VOSON Dashboard` will *only use the text data stored as a node attribute*. In future versions, we plan to allow the user to select which text data are used for text analysis.

Finally, for networks with node categorical attributes (e.g. "Type" in the demo environmental activist WWW hyperlink network), text analysis is also reported by category and a comparison word cloud can be generated. As noted above, it is presently not possible to conduct text analysis using groups obtained via categorical edge attributes.

4.1 Text filters

"Filters" allow you to manipulate the text before analysis(Figure 8):

- "Remove Standard Stopwords": this function removes stopwords (English) such as "and", "the", and "but".
- "User-Defined Stopwords": here, you can create your own list of stopwords (use commas or spaces between stopwords as you enter them).
- "Apply word stemming": this attempts to reduce words to their stems.
- For Twitter networks, two other options become available: "Remove Twitter hashtags" and "Remove Twitter Usernames".

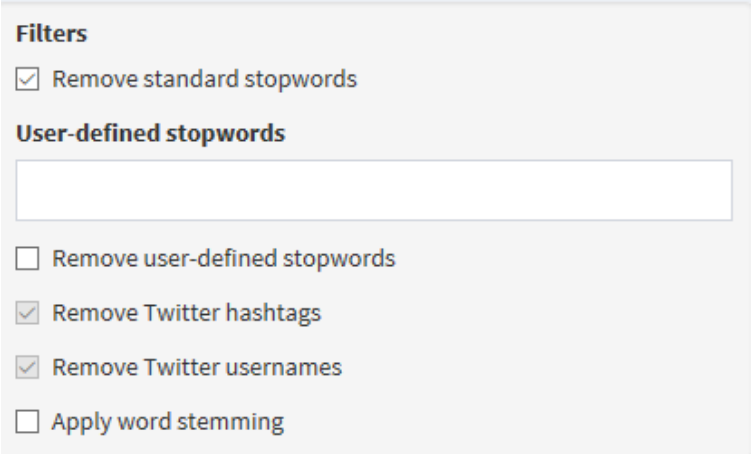


Figure 8: Text filters

4.2 Visualising word frequencies

Below the text analysis "Filters", there is a "Summary" of the results, including general text attributes, stopwords, and word counts, among others. Text analysis visualisations are presented in the main canvas. There are four types of text analysis visualisation available via tab:

- "Word Frequency": contains frequency bar chart(s) and, if applicable, a bar chart per category will be displayed (Figure 9). Word Frequency has two controls; "Display Top" is used to define the number of words that will appear in the visualisation (only the top 'X' words, based on frequency counts), and "Minimum Frequency", which determines how many times a word needs to have been used in order for it to feature in the visualisations.

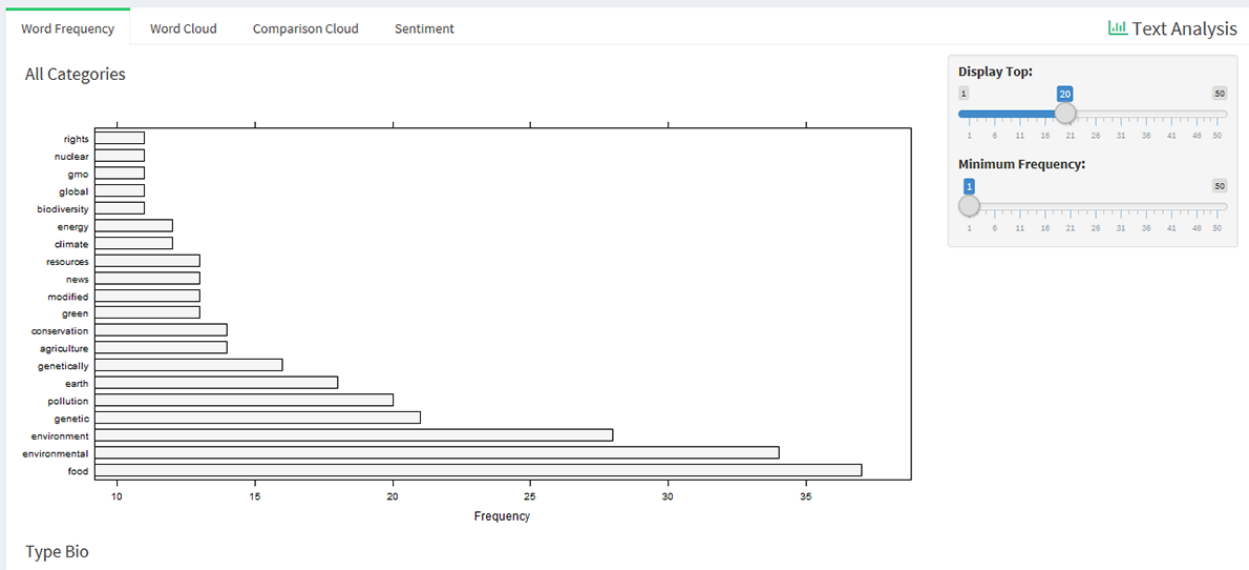


Figure 9: Word frequency bar chart

- "Word Cloud": where the size of a word represents its frequency. For a network with subgroups, word clouds per category will also be displayed (Figure 10). Word Cloud has two controls: "Minimum frequency" which determines how many times a word needs to have been used in order for it to feature in the visualisation and "Maximum words" to control for the number of words appearing in the graph.

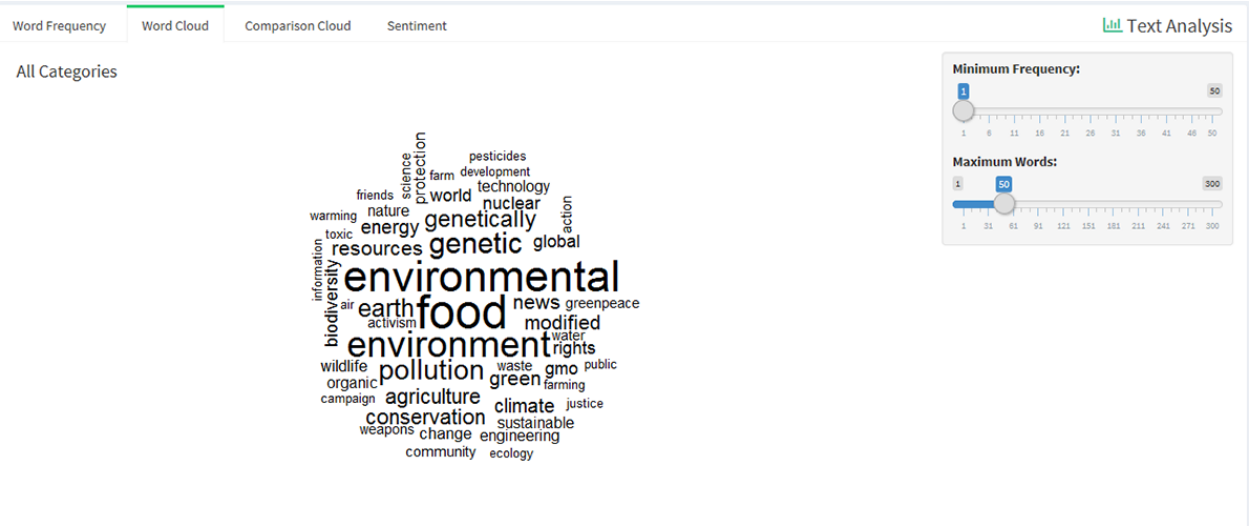


Figure 10: Word cloud

- "Comparison Cloud" (only for datasets with categorical node attributes): this graph displays the terms that are identified with the nodes in each group or category and hence can be used to discern differences between the groups of network nodes (Figure 11). Note that in a comparison cloud a term does not need to be solely used by nodes a particular group, for it to be identified with that group. Comparison cloud has one control: "Maximum words" to control for the number of words appearing in the graph.
- "Sentiment Analysis": This function uses the syuzhet package. There are bar charts with the words classified based on the NRC Emotion Lexicon, which is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and two

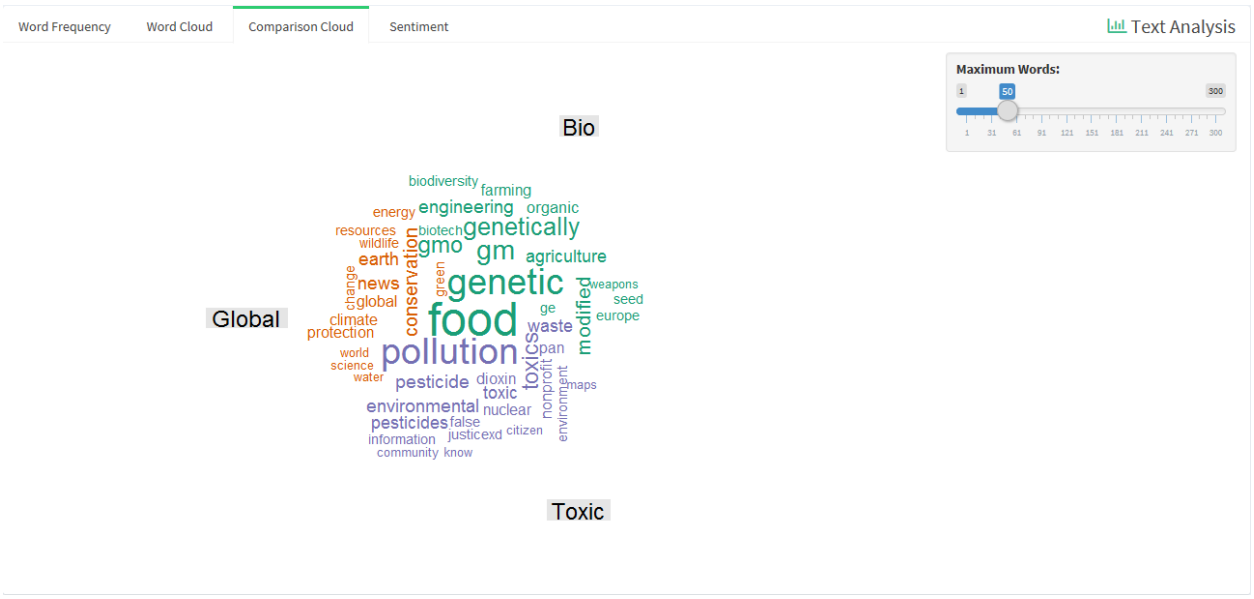


Figure 11: Comparison cloud

sentiments (negative and positive) (Figure 12).

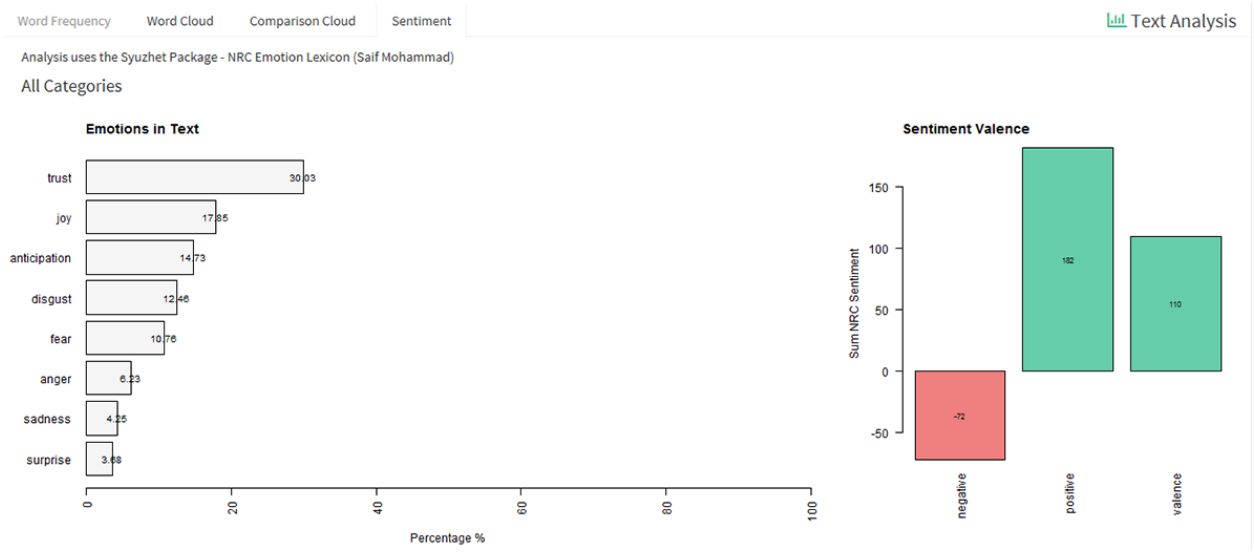


Figure 12: Sentiment Analysis

To save text analysis visualisations, right-click on "Save Image" and input the file name and file type.

5 Collecting data with VOSON Dashboard

5.1 API Keys

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

be out of date.

At present, you do not need an API key to collect data from Reddit, as `VOSON Dashboard` provides access to the Reddit public API.

5.1.1 Loading API keys

The "API Keys" window (Figure 13) provides all the fields to be entered in order to access social media data.

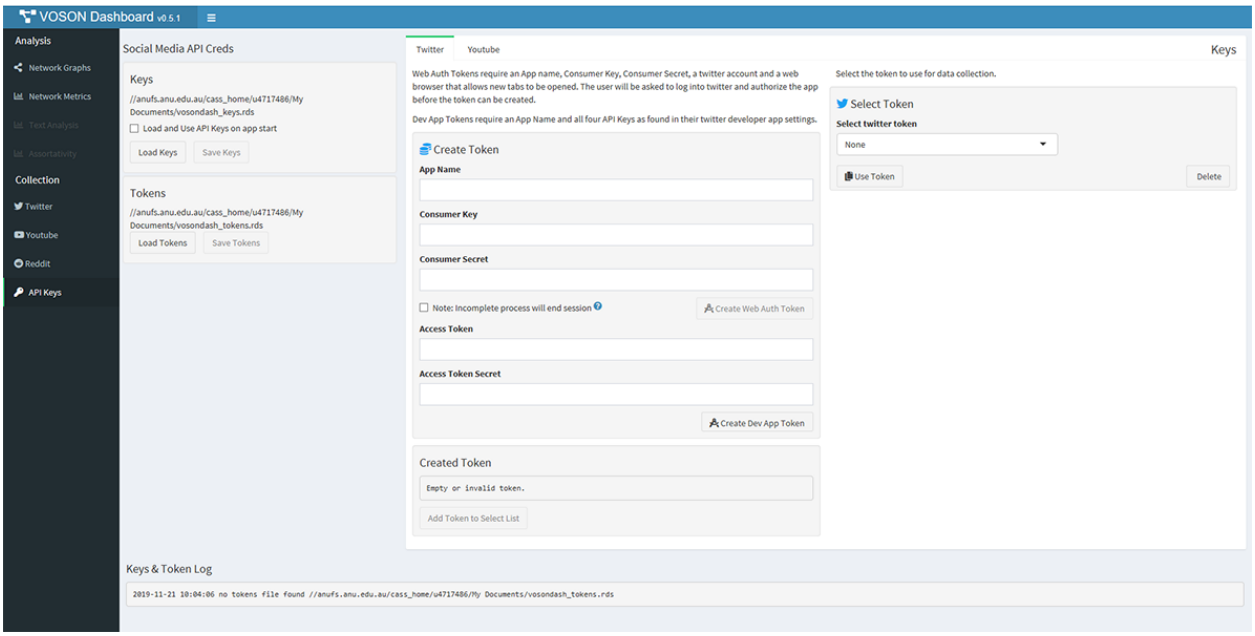


Figure 13: Input, save and load API keys in the "API Keys" window

For Twitter, if you have your own Twitter developer account, enter the fields accordingly and then click on the button "Create Dev App Token". A box below will display details of the token. The second step is to save it by clicking on the "Add Token to select list" button. That action will populate the "Select token" window in the top right. Finally, click on the "Use Token" button. You will be able to check the steps in the "Keys and Token Log" box.

In a classroom setting, it may be desirable for the instructor (who has a Twitter developer account) to create an app and provide the app name, and the consumer key and secret to students, who can then generate a "web auth" token. This will allow the students to undertake Twitter data collections via `VOSON Dashboard`, even if they do not have their own Twitter developer account (Twitter developer accounts are increasingly difficult to obtain, and there can be a lengthy approval process). After filling in the app name and consumer key and secret, check the box "Experimental (aborting will end session)" and click on the "Create Web Auth Token" button - this will open up a web page where you need to sign into your Twitter account and approve the app to have access to your Twitter account. A token will be created and then you have to click the "Add Token to Select List" button down the bottom of the panel in `VOSON Dashboard`. Finally, select the token in the select box in the upper RHS, and press the "Use Token" button.

Please note that Twitter no longer creates an API authentication token for each twitter collection. Instead Twitter tokens are now created and managed in the API Keys tab, therefore they should be saved and re-used. API keys and tokens are saved to files in the users home directory as specified by the system environment variable HOME. This location can be found in the `VOSON Dashboard` start up information, in the apps API Keys tab or by using the following R function in RStudio:

```
Sys.getenv(HOME)
```

The keys and tokens files are named 'vosondash_keys.rds' and 'vosondash_tokens.rds' respectively.

If you use a public accessible computer, please make sure to delete your keys and tokens files.

For YouTube, simply enter the API key field and click on "Use Key" (Figure 14).

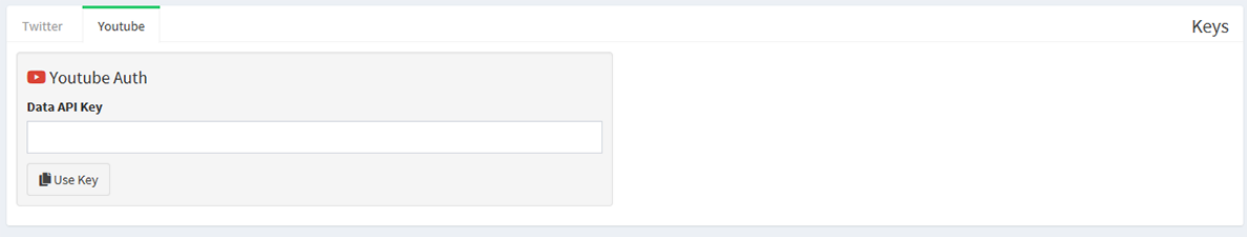


Figure 14: Input, save and load API keys in the "API Keys" window

5.2 Collecting data from Twitter

5.2.1 Searching for tweets

There are two major ways you can search for tweets and -hence- construct networks resulting from those tweets:

1. Search for terms - this includes hashtags, account names and words. This is achieved by writing the search phrase in the "Search Terms" text box. For more information on search rules (syntax), refer to the Twitter's Rules and Filtering documentation.
2. Search for tweets authored by or directed to a particular user. This is achieved by selecting the "Additional Filters" box and then writing the Twitter username in the "From" and/or "To" text box. "From" will collect the tweets and re-tweet authored by the account, and "To" will collect tweets and re-tweets directed to the account (Figure 15).

Note that it is possible to undertake both of the types of searches for tweets above e.g. one can search for tweets containing a particular hashtag **and** authored by a particular Twitter user.

There are additional search parameters that can be enabled (Figure 16):

- By clicking the box "Re-tweets" you can define the type of Twitter activities to be collected.
- By clicking the "Retry on rate limit" box, the collection will automatically pause when the Twitter API rate limit is reached, and then restart.
- The "Count" select box is designed to control the number of collected tweets.
- The "Language" text box is designed to select the language of tweets to be collected. This search parameter accepts two-letter language codes as per the ISO 639-1 standard.
- The "Date Until" function 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.*
- Each tweet is assigned a unique sequential ID number. By selecting the "Tweet ID range" box, you can use "Since ID" to collect only those tweets with IDs *greater* than ID number (e.g. tweets authored *after* the time the tweet with that ID was authored). For example, say you are collecting tweets including the #auspol (Australian politics hashtag) and the most recently authored tweet in the collection you undertook yesterday had an ID of x, then by putting x in the "Since ID" text box, you would ensure that your latest collection would only collect those tweets featuring #auspol and authored after the

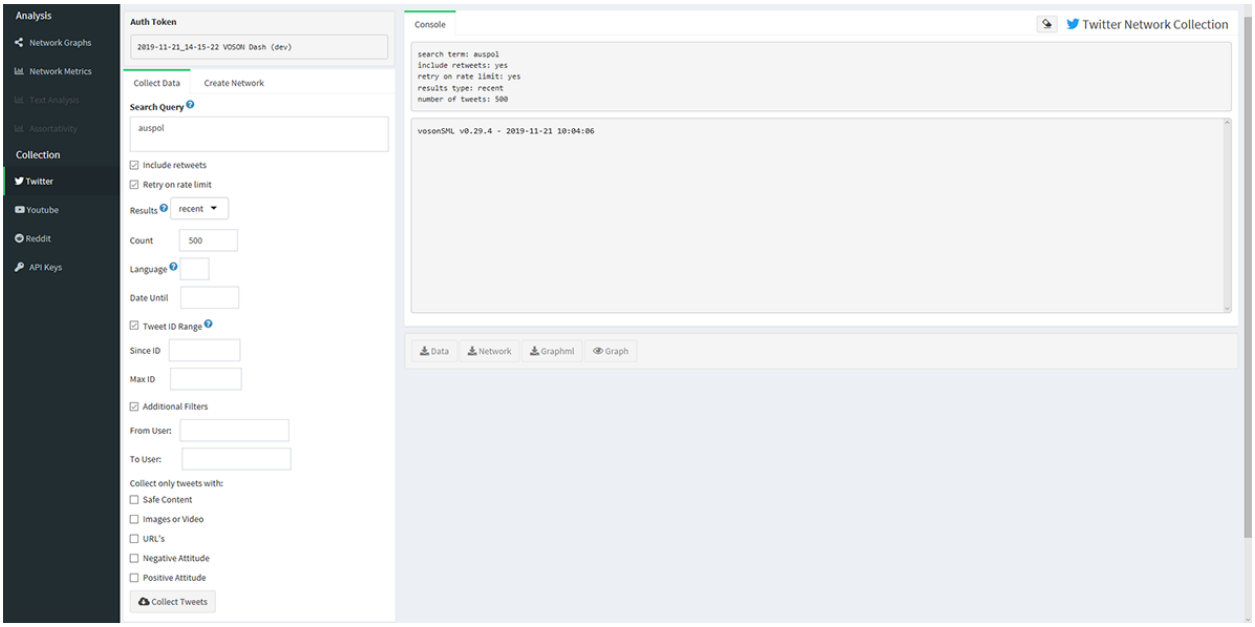


Figure 15: Twitter collection window

tweet with ID x. Similarly, "Max ID" can be used to collect only those tweets that were authored *before* the tweet with a particular ID number. But remember that the Twitter Free API only allows you to collect tweets authored in the past 7 days.

- You can further filter your collection through checking the boxes: "Safe Content", "Images or Videos", "URLs", "Negative attitude" and "Positive attitude". Note that these attributes are assigned by Twitter, so the filter applies as a collection parameter. More information on parameters for Twitter collection can be found in this link.

5.2.2 Navigating the database

Once your collection has run, a table will list the data at the bottom of the "Console". Basic attributes of each tweets are listed in this table, including: text, favourite and favourite count, IDs of tweets and re-tweets, time and etc. By clicking the "Truncate text" on the top left of the table, smaller cells will be displayed as comments are summarised. If you un-click the "Truncate text", the row will display the full comment, in the text column. The Search box is on the top right of the table, where you can enter text to find matched records (Figure 17).

5.2.3 Creating Twitter networks

Once your collection has run, VOSON Dashboard provides the option to create different types of networks via the "Create Network" tab.

- Activity network: In activity networks, nodes represent tweets and edge types are: retweets, replies, or quote retweets. Click the "Add text" checkbox, to include text data (e.g. Twitter payload) to your network (Figure 18).
- Actor Network: In an actor network, nodes are users who have tweeted, mentioned or retweeted users, and edges represent interactions ('mention', 'reply', 'retweet', 'Quote tweet' or 'Self loop) between Twitter users. Click the "Add text" checkbox, to include text data (e.g. Twitter payload) to your network. The "Lookup User Data" option will retrieve profile information for users that became nodes

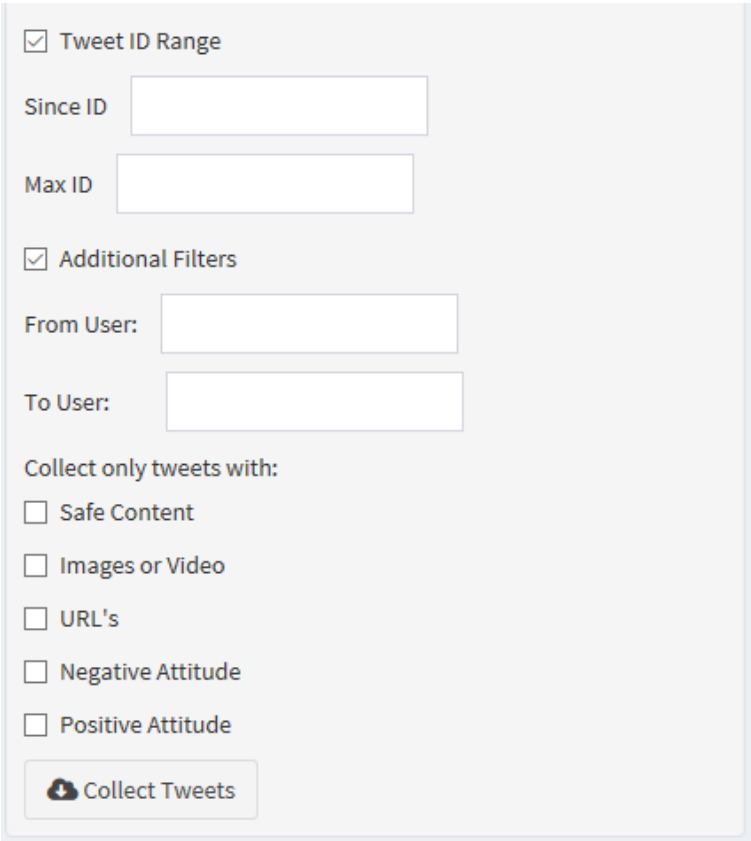


Figure 16: Additional filters for Twitter collection

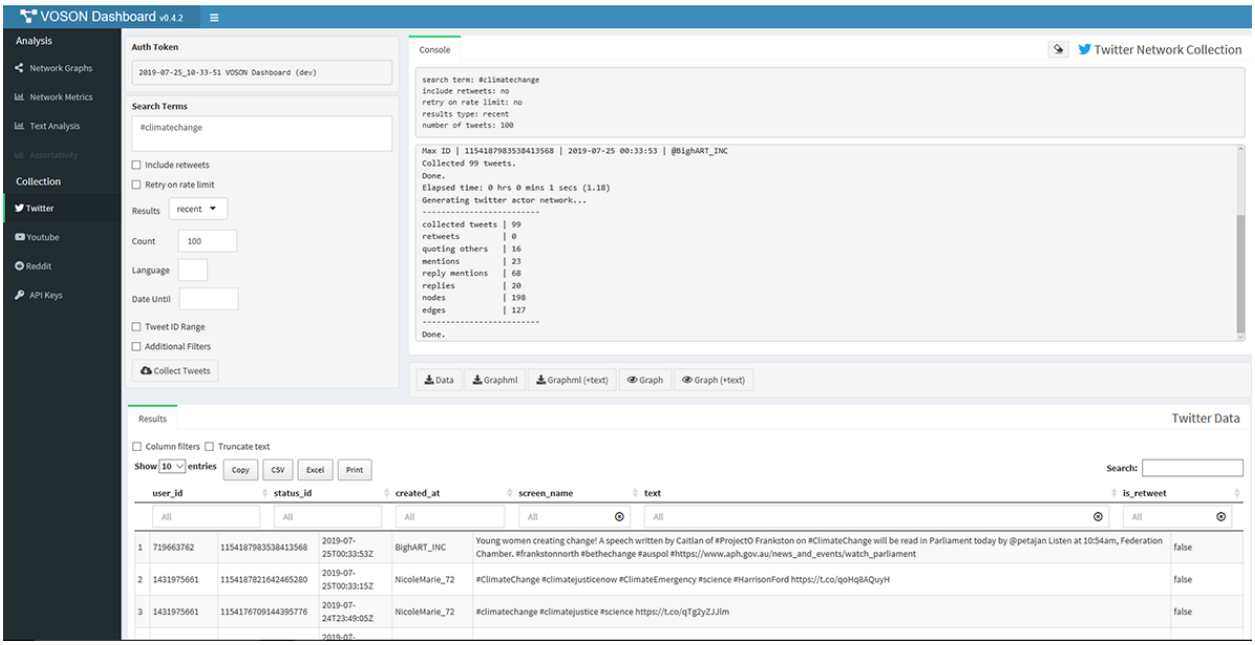


Figure 17: Twitter collection view

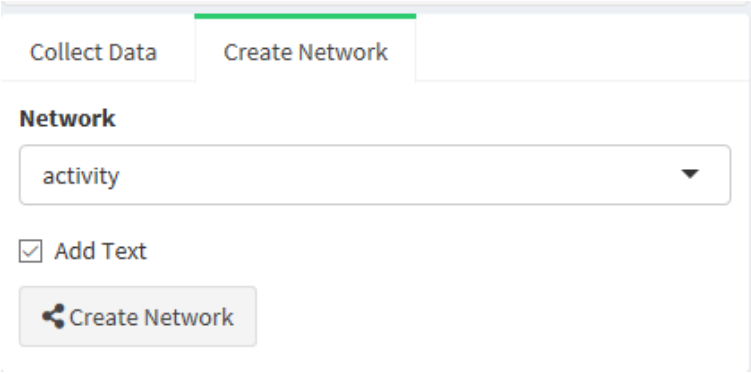


Figure 18: Activity network,controls

during network creation but were not authors of tweets. Their profile information was missing most likely because they were referenced in tweets but none of their tweets were collected in the search. In a Twitter search, profile information is only returned for the authors of tweets captured in the search (Figure 19).

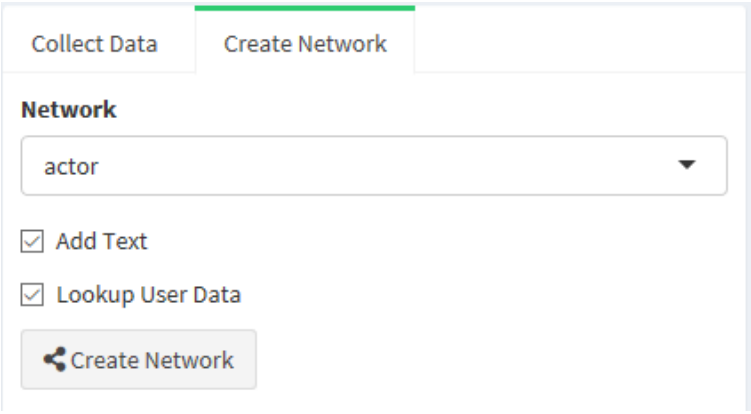


Figure 19: Actor network, controls

- Two-mode network: In a Two-mode network, nodes are actors (Twitter users) and hashtags. Edges represent relationships based on the usage of hashtag in tweets or reference to other users in tweets. Edges are directed. To illustrate, if we collect the latest 100 tweets including #auspol (a predominant hashtag used in Australian politics) and we look at the two-mode network, the #auspol hashtag will be a predominant node, because all the observations contain that hashtag (collection parameter). We can filter out that term from the network by entering it in the text box "Remove terms". This box accepts a comma delimited list of terms, that can be actors and hashtags (Figure 20).
- Semantic network: In Semantic networks, nodes represent unique concepts (terms/words extracted from the collection, hashtags and actors), and edges represent relationships between concepts, based on co-occurrence. Edges are undirected and weighted. If we use the example of collection proposed in Two-mode networks, the Twitter semantic network of a collection of 100 tweets including the Auspol hashtag, nodes represent either hashtags (e.g. "#auspol") or single terms (e.g. "politics"), or actors (e.g. if a Twitter user was mentioned). If there are 100 tweets in the data set (i.e. 100 observations), and the term "#auspol" and the term "politics" appear together in every tweet, then this would be represented by an edge with weight equal to 100. Click the "Add text" checkbox, to include text data (e.g. Twitter payload) to your network. A term can be filtered out from the network by entering it in the text box "Remove terms". This box accepts a comma delimited list of terms, actors and hashtags. The semantic network currently uses the default `vosonSML` options of only including the 5 percent of the

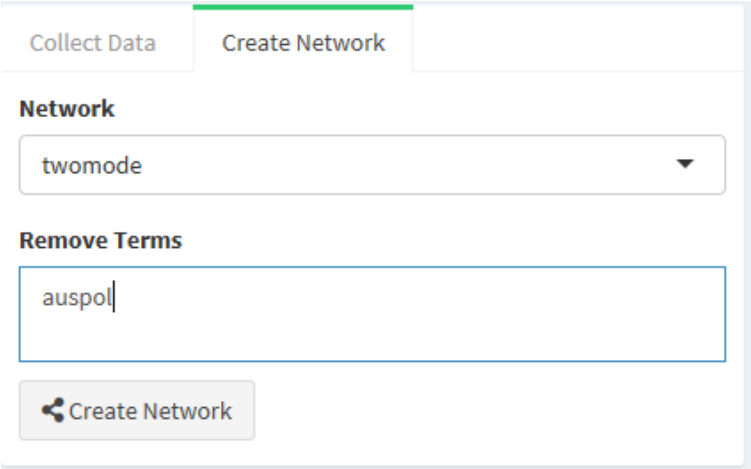


Figure 20: Two-mode network, controls

most frequent terms and 50 percent of the most frequently occurring hashtags in the network, metrics which can be adjusted using the dials (Figure 21).

5.2.4 Working with Twitter data

Once your collection has run and the network has been created, you can either download the raw data ("Data" button); save the network you just created including all applied filters; save the Graphml file; or open it (eye icon). Note that text fields (e.g. tweet content) will only be displayed if the "Add Text" option was checked at the network creation stage.

The Results area displays a tabulation of the data, reflecting either node's or edge's attributes. As mentioned, for Twitter collection, text data (i.e. tweet payload) can be stored either as an edge (Actor, Two-mode, or Semantic networks) or as a node attribute (Activity networks), and can be accessed via the "vertex" and "edges" tables, respectively.

Again, the "Truncate text" option on the top left of the data table is available to display smaller cells (summarised text content), so more entries can be seen per page.

At the top right of the table there is a Search box; by entering text, all the matching records will be shown on the table. Data can be sorted by different fields, through clicking on the heading of a given column.

5.2.5 Saving the database

Data can be downloaded via the "Twitter Collection" (CSV, Graphml) view or the "Network Graph" view. In the "Network Graph" view, simply click the buttons below the graph window, or download the data in CSV or Excel formats via the buttons in the "Results" area. Please note that in the "Network Graph" option, the downloaded files will only contain full text data if that was the chosen option when you created the network.

Additionally, the results area provides the options to "Copy" or "Print" the data.

*****Important note: if you do not explicitly save your data (including graph files) to a local folder, the system will not store the data after the end of the session.*****

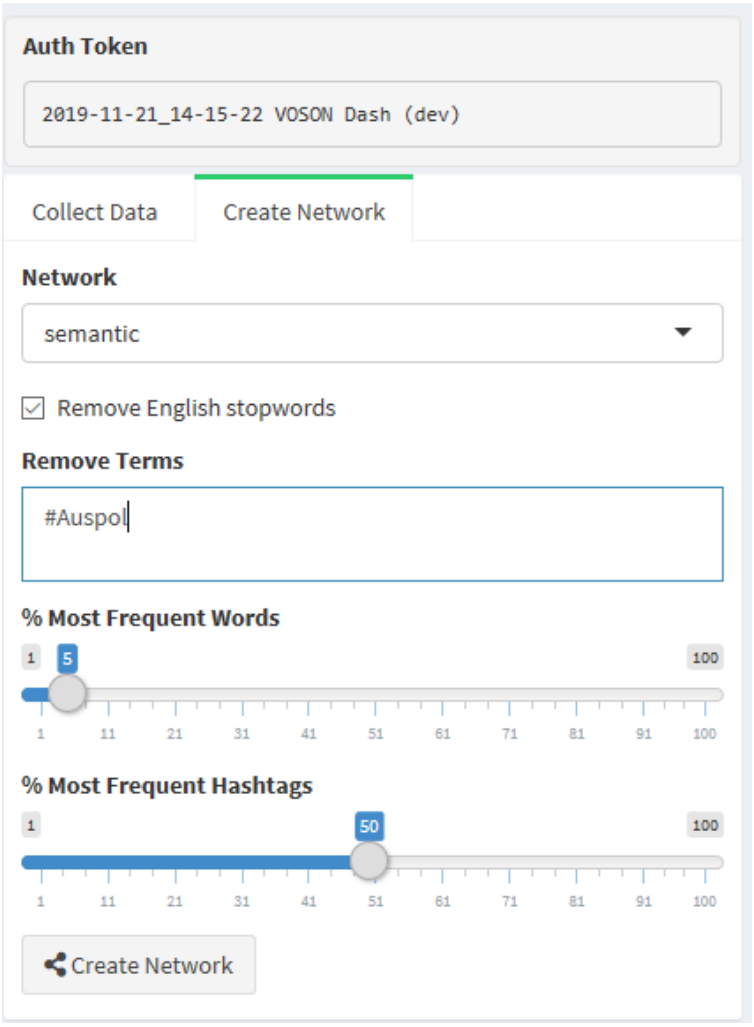


Figure 21: Semantic network, controls

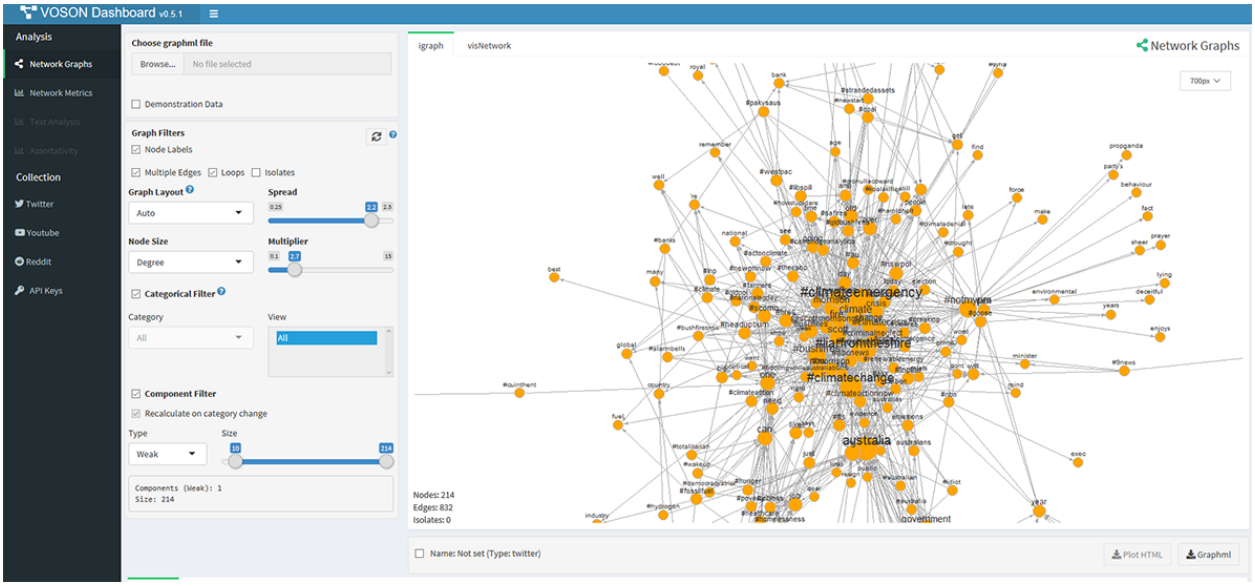


Figure 22: Example network visualisation: Semantic network

5.3 Collecting data from YouTube

5.3.1 Search term

YouTube comments can be collected by either using the full URL or the video ID, which is the part after "=" in the YouTube URL (Figure 23). Copy the URL into the "Add YouTube URL" box and click on "Add". This action will populate the collection box (below) with the video ID. You may add or remove videos for your collection. Click on "Collect comments" to proceed with data collection. When more than one video ID is applied to the collection, the comments of each video will be collected via the API.

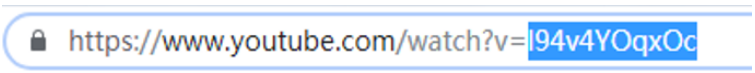


Figure 23: YouTube Video ID

5.3.2 Creating YouTube networks

Once you have collected YouTube comments, two types of networks can be created: "Activity networks" and "Actor networks".

- Activity network: In an Activity network, nodes are either users who have commented (either top-level comment or reply to top-level comment), or videos (as videos represent an starting comment). There are three types of edges in YouTube activity networks, representing commenting activity: "Top-level comment" (comments directed to video creator), "Reply comment" (comments directed to user comments), and "Self loops" (reply to self) (Figure 24).
- Actor Network: an Actor network maps the relationships of YouTube users who have interacted with each other in the comments section for particular videos (i.e. user i has replied to user j, or mentioned user j in a comment, or reply to self ("Self loop"). Videos are considered nodes too, as they represent the user who posted them (Figure 25).

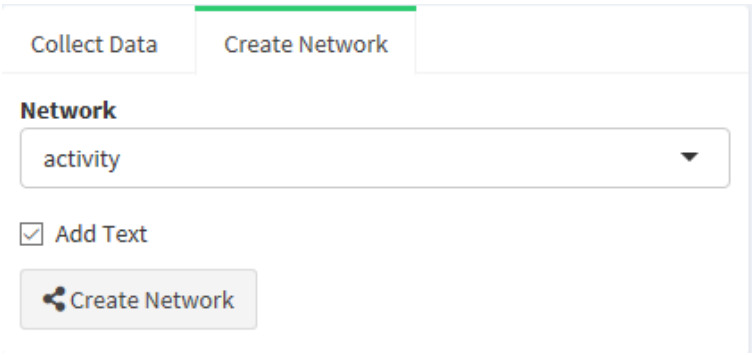


Figure 24: Youtube Activity network,controls

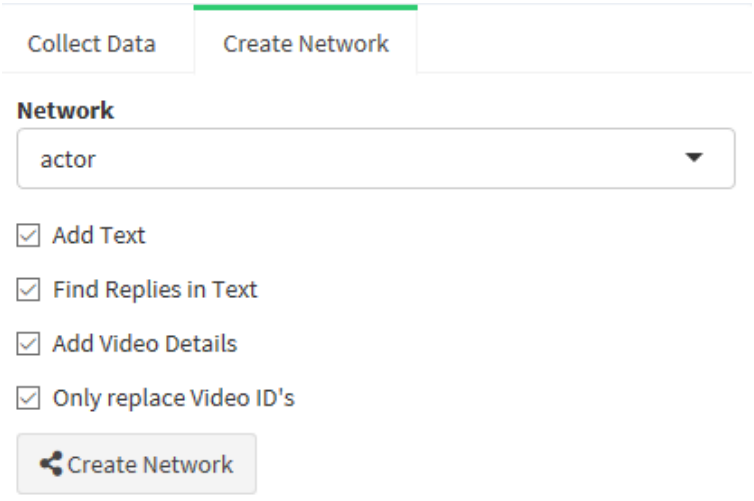


Figure 25: Youtube Actor network, controls

5.3.3 Navigating the database

Results of the collected YouTube data are presented on a table, at the bottom of the "Console". The main attributes of the collected comments are listed in this table, including the user who post the comment, the text content of comment, reply to other users, etc. As with Twitter data, you may chose to download Row data; save the network (including the applied filters); save the Graphml file; or view (eye icon) the network via the "Network Graph" window. This can be done by clicking on the respective button.

Once in the "Network Graphs" window, you can visualise the network of YouTube comments and access to the "Graph data" section to observe nodes and edges attributes. Similar to Twitter data, by clicking the "Truncate text" on the top left of the table, smaller cells will be displayed as comments are summarised. If you un-click the "Truncate text", the row will display the full comment, in the text column. The Search box is on the top right of the table, where you can enter text and matched records will be shown. Data can be sorted by different fields, through clicking on the heading of the column (Figure 26).

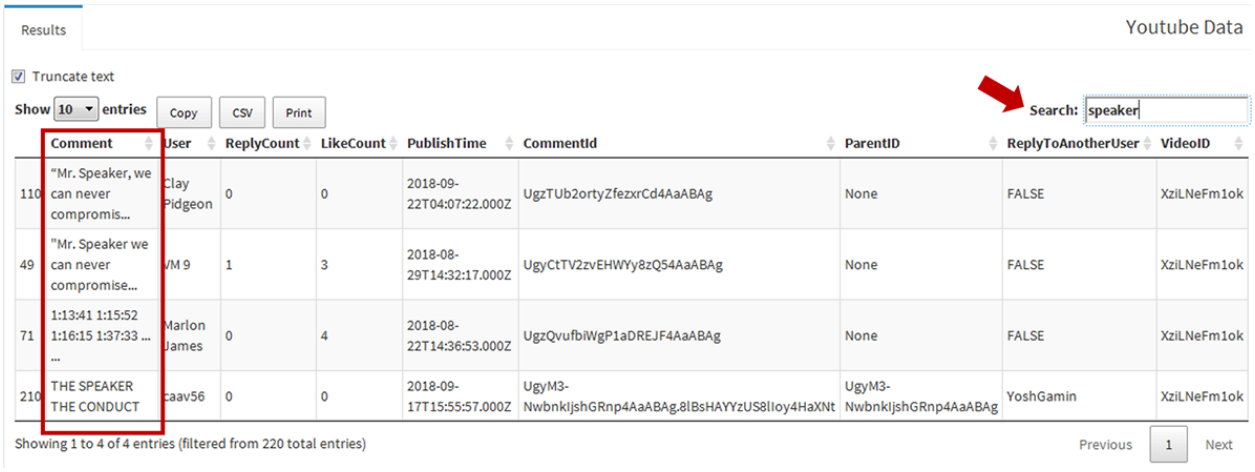


Figure 26: Searching terms in YouTube data

5.3.4 Saving the database

From the Youtube collection view you can download CSV, Graphml files directly. To download the Graphml data in the "Network Graph" view, click the buttons below the graph window or download the data in CSV or Excel formats via the buttons in the "Results" area. Please note that in the "Network Graph" option,the downloaded files will only contain full text data if that was the chosen option when you created the network.Additionally, the results area provides the options to "Copy" or "Print" the data.

Again, the downloaded files will contain full text data, if that was that option was chosen at the network creation stage.

*****Important note: if you do not explicitly save your data (including graph files) to a local folder, the system will not store the data after the end of the session.*****

5.4 Collecting data from Reddit

5.4.1 Reddit API

At present, `VOSON Dashboard` accesses the Reddit API via `vosonSML`, based on an approach found in the RedditExtractoR package; therefore, there is no need to load the API keys for this data source before

collection. Due to API limitations, the number of comments available for retrieval is limited to 500 per thread.

5.4.2 Search term

Reddit comments can be collected by entering the URL of a particular thread in the "Add Reddit URL" window. When more than one thread is applied to the search, `VOSON Dashboard` will collect the comments of each thread. Click on "Collect Threads" to start the collection (Figure 27).

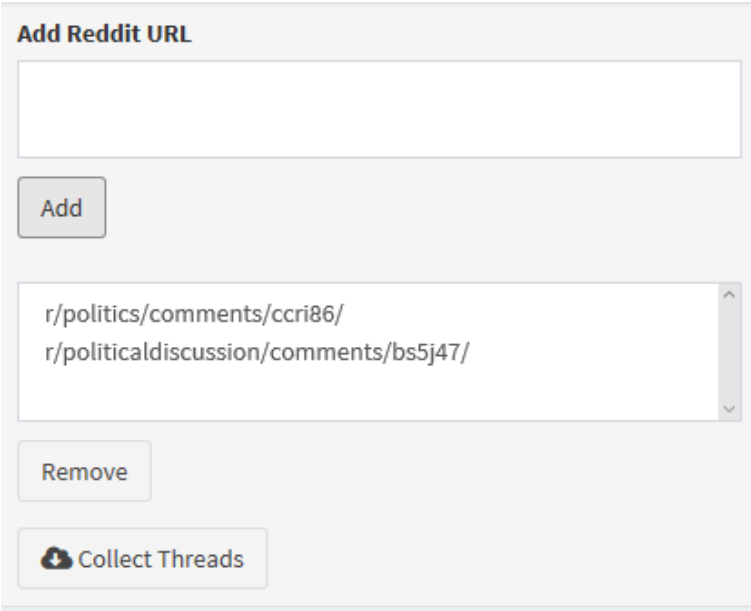


Figure 27: Reddit collection parameters

5.4.3 Navigating the database

Once your collection has run, a table will list the data at the bottom of the "Console" (Figure 28). The main attributes of the Reddit comments are listed in this table, including subreddit, author, number of comments, text content, score, controversiality, among others.

Similar to Twitter and YouTube results, by clicking the "Truncate text" on the top left of the table, smaller cells will be displayed as comments are summarised. If you un-click the "Truncate text", the row will display the full comment, in the text column. The Search box is on the top right of the table, where you can enter text and matched records will be shown.

5.4.4 Creating Reddit networks

Once you have collected Reddit comments, two types of networks can be created: "Activity networks" and "Actor networks".

- Activity network: In a Reddit Activity network, nodes are either comments and/or initial thread posts and the edges represent replies to comments (Figure 29).
- Actor Network: where nodes are users who have either: posted original posts, or who have posted comments and edges are the comments linking users.

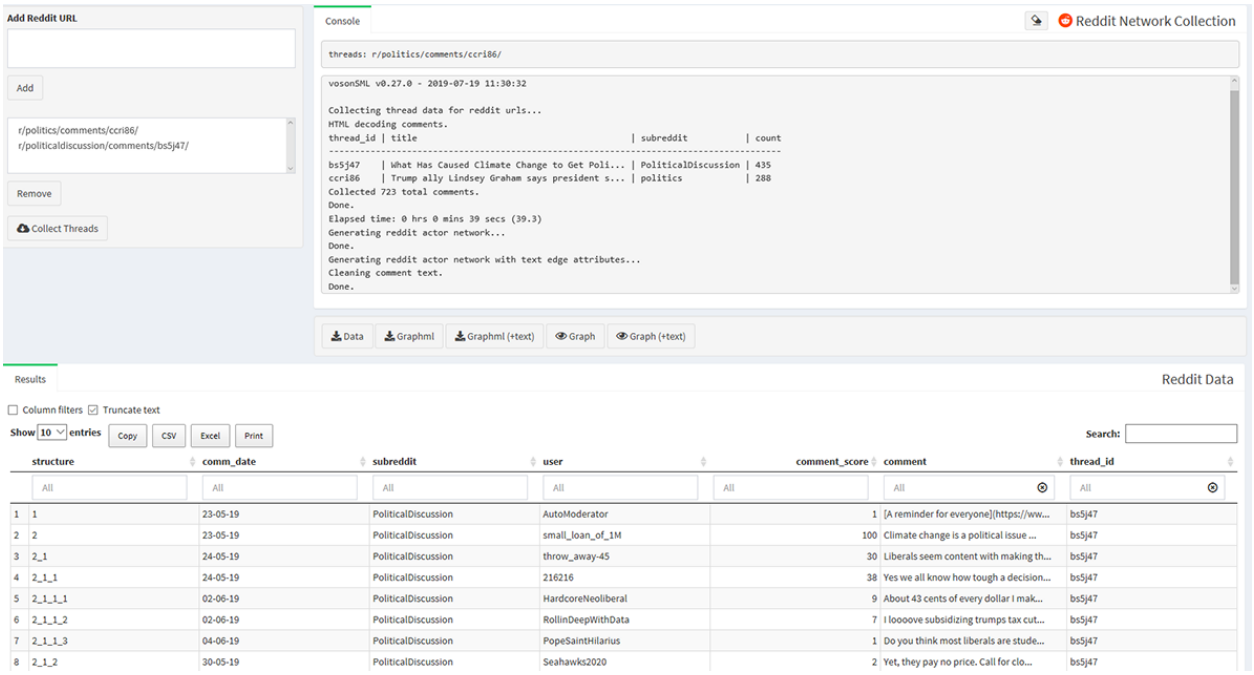


Figure 28: Reddit collection, data

Once the network has been created, you can open it in the "Network Graph Window" by clicking on the "Graph" button (eye icon).

5.4.5 Saving database

The Reddit database you just collected can be downloaded as: row data (Data), the "network" containing all the filters you have applied and Graphml files directly. From the results table, its possible to download in CSV and Excel formats. This menu also enables the options to copy the data or print it.

Please note that in the "Network Graph"option,the downloaded files will only contain full text data if that option was chosen, at the network creation stage.

*****Important note: if you do not explicitly save your data (including graph files) to a local folder, the system will not store the data after the end of the session.*****

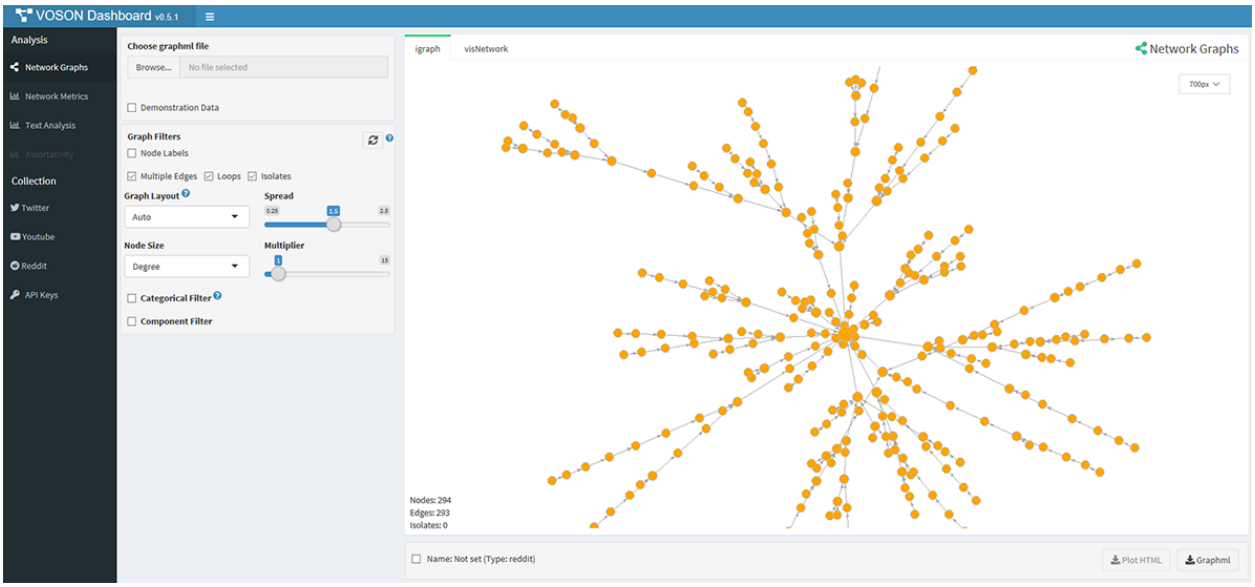


Figure 29: Reddit Activity network, controls