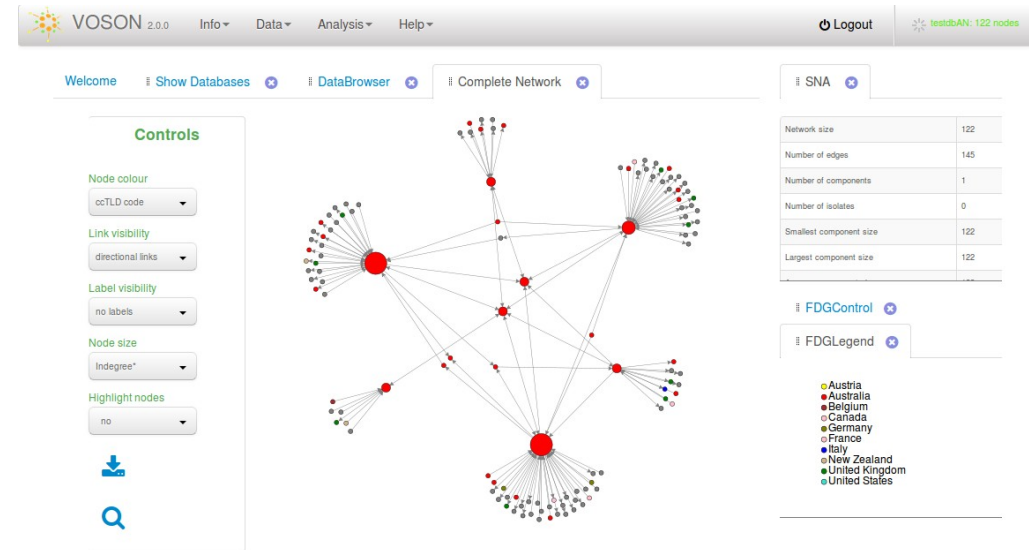


Introduction

- R is free and open source software for statistical computing and graphics – www.r-project.org
- The R syntax and approach presents a bit of a learning curve (especially for social scientists) but benefits are worth the effort:
 - incredibly wide range of R packages/libraries are available
 - promotes scientific collaboration via open source code and data
- R is hard to pass up if you want to do social media research and (intermediate-to-advanced) teaching
- Three R packages being developed by the VOSON Lab (and associates):
 - **vosonR** – client for VOSON software for WWW hyperlink network and website text content collection and analysis
 - **SocialMediaLab** – collection and analysis of social media data
 - **AdaptiveSampling** – network sampling (not discussed here)

VOSON

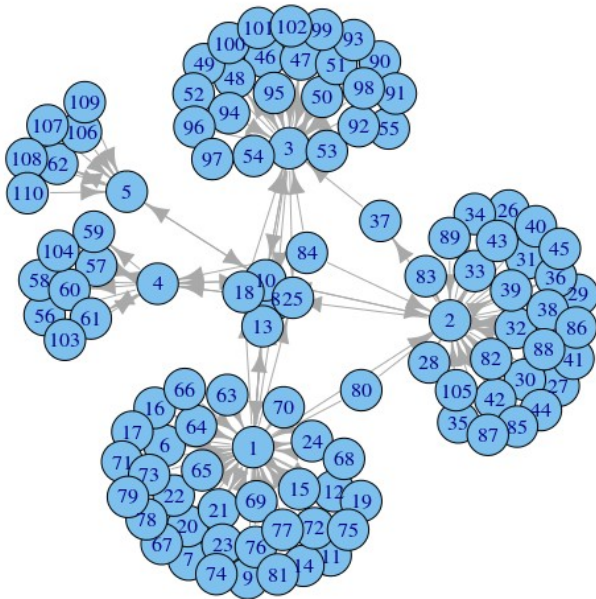
- VOSON – collection and analysis of WWW hyperlink and text content
- VOSON 2.0 web interface works with Firefox, Chrome, Safari, iPad
- Available for academics/students since 2006, for general use since 2010



- VOSON is **not free and open source** (free version and paid versions)
- For VOSON accounts: <http://www.uberlink.com>

vostonR

- R client for VOSON software (via web service)
- With vostonR, you can schedule a crawl and import networks into R (as igraph graph objects). Also exports graphml.
- ICA2015 alpha version for download: <http://voston.anu.edu.au/tools>



```
> library(vostonR)
> SetAuthCred("rob","xxxxxx")
[1] "Successful connection to VOSON API."
> ShowDatabases()
  Name Project Author Type Rows Last_modified Parent
1 a1AN rob rob voston-analysis 112 2015-04-13_12:49:53 a1
2 a1 rob rob voston 247 2015-04-13_12:49:53 n.a.
3 bb rob rob voston 5 2015-05-19_01:34:58 n.a.
4 testdbAN tutorial rob voston-analysis 123 2013-06-09_17:02:38 testdb
5 testdb tutorial tutorial voston 196 2014-06-27_22:10:02 n.a.
  Tie_indicator Node_type Comment
1 hyperlink pagegroup eA==
2 hyperlink x eA==
3 hyperlink x dGhpcyBpcyBhIGNvbW1lbnQ=
4 hyperlink pagegroup x
5 hyperlink x x
> g <- GetNetwork("testdbAN", "tutorial")
> g
IGRAPH D--- 110 129 --
+ attr: Image File (v/c), Ringset (v/n), Country Code TLD (v/n),
Generic TLD (v/n), id (v/c), Width (e/n)
> plot(g)
```

- Beta version will be released soon
- Contributors:
 - Robert Ackland (vostonR, VOSON core development team)
 - Tim Graham (vostonR)
 - Jamsheed Shorish (VOSON core development team)
- Email me if would like to know more...

SocialMediaLab

- There are many R packages that can be used for collection and analysis of social media data
- However some of these packages require advanced programming skills...
- **SocialMediaLab** is an R package being developed by Tim Graham (Univ. Queensland) and Robert Ackland (ANU)
- Provides a “virtual lab” for the collection and analysis of social media data
- A wrapper around many other R packages, making it easier for non-programmers to do social media analytics research and teaching in R
- The package is being developed for use in “Big Data for Social Scientists” ACSPRI short course (29 Jun - 3 Jul, Brisbane):
<http://voston.anu.edu.au/big-data-analysis>
- We will release alpha version of SocialMediaLab package and associated open source courseware by 29 June

SocialMediaLab – data collection

- Twitter – for supplied search term (e.g. hashtag), produces:
 - networks (graphml files, igraph graph objects): actor network (users who tweeted term or mentioned in these tweets), bimodal network (actors and extracted terms), semantic network of extracted terms
 - csv file containing tweet payloads
- Youtube – for supplied video id, produces:
 - networks: actor network (thread network from video comments), bimodal network (actors and extracted terms), semantic network (extracted terms)
 - csv file containing comments
- Facebook fan pages – for supplied fan page, produces:
 - networks: actor network (based on co-comments, co-likes), bimodal network (actors and posts), semantic network of extracted terms
 - csv file containing fan page posts and comments
- WWW hyperlink networks (via vosonR)

SocialMediaLab & course – analysis

- Network analysis (core): Intro to SNA, covering main network-level and node-level metrics, and clustering (‘community detection’)
- Visualisation (core): Network visualisations, interactive networks (on the web), word clouds and dendrograms from network data.
- Text analysis (core): Supervised machine learning (e.g. support vector machines), unsupervised machine learning (topic modelling), sentiment analysis, hierarchical clustering, and descriptive analytics.

SocialMediaLab & course – analysis

- Temporal analysis (advanced): Analysing networks and text over time, identifying significant changes in behaviour of individual nodes, clusters or entire networks.
- Filtering and sampling (advanced): Techniques for targeting data collection and analysis on a particular set of actors. Reducing the scale of datasets via sampling.
- Scaling up to very large datasets (advanced): What if your dataset is too large for your computer?

R packages used in SocialMediaLab (planned)

- igraph (network analysis and visualisation)
- twitterR (for collecting Twitter data)
- tm (text mining)
- RTextTools (machine learning package for automatic text classification)
- RCurl (collecting WWW data)
- XML (reading and creating XML documents)
- R.utils (programming utilities)
- wordcloud (text word clouds)
- ape and dendextend (dendograms, hierarchical clustering)
- FactoMineR and homals (multiple correspondence analysis)
- plyr and stringr (text sentiment analysis)