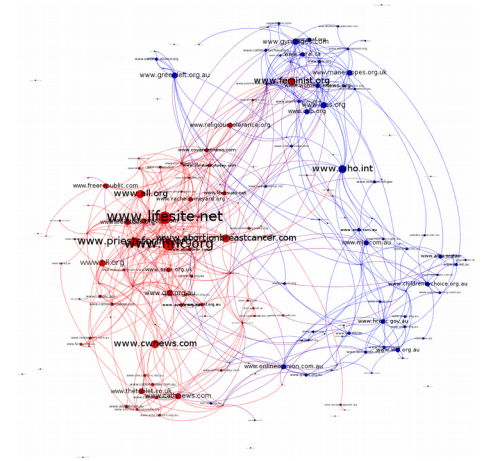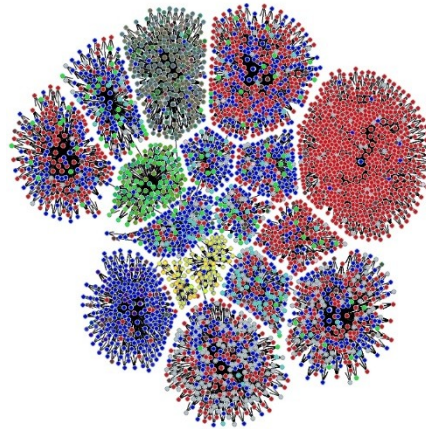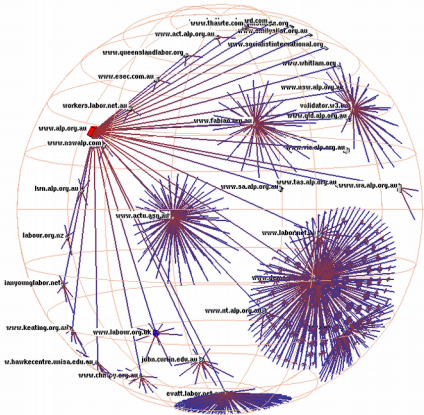# Big Data Analysis for Social Scientists

Workshop held at Social Informatics 2018, St Petersburg, Russia
25th September 2018

Dr Robert Ackland
School of Sociology
Virtual Observatory for the Study of Online Networks (VOSON) Lab
Research School of Social Sciences
Australian National University

E: robert.ackland@anu.edu.au
T: @RobAckland
W: http://vosonlab.net

VOSON Lab
Virtual Observatory for the Study of Online Networks

Welcome to the

Virtual Observatory
for the Study of
Online Networks

VOSON Lab!

(Artist's representation of our Lab... Not drawn to scale or reality).

A/Prof Robert Ackland

Dr Tim Graham

A/Prof Mathieu O'neil

Francisca Borquez

Xiaolan Cai

Dr Jamsheed Shorish

Bryan Gertzel

Conducting **research**, **teaching** and **tool development** in areas of web (social) science, network science, computational social science, social science of the internet.

Australian Research Council funded research

- DP0452051 "New Methods for Researching the Existence and Impact of Political Networks on the WWW" – Ackland and Gibson (2004-2006)
- SR0567298 "Virtual Observatory for the Study of Online Networks (VOSON)" – Ackland, Gibson, O'Neil, Buchhorn, Bimber, Ward (2005)
- LP0990974 "The role of online social networks in successful ageing: benefitting from 'who you know' at older ages" – Booth, Ackland, Windsor (2009-2012)
- DP110100446 "The institutional structure of e-government: a cross-policy, cross-country comparison" – Henman, Ackland, Margetts (2011-2013)
- DP140103688 "Understanding online attention and user-generated content creation: An information consumption and production perspective" – Ackland (2014-2016)

Research tools
- **VOSON software** for hyperlink network construction & analysis (publicly available since 2006, over 3000 user accounts)
  - Commercialised via Uberlink (http://www.uberlink.com) uberlink
- **R packages**:
  - **vosonSML ("social media lab")** (with Tim Graham) – released on CRAN Nov 2015
    - collects (via free APIs) from: Twitter, YouTube [working on Reddit]
    - creates network and text datasets
  - **VOSON Dashboard** (with Bryan Gertzel)
    - R/Shiny web app for network/text data collection (via vosonSML) and analysis

Teaching

- Social Science of the Internet - masters (undergrad) since 2008 (2017)
- Online Research Methods - masters (undergrad) since 2009 (2018)
- Economic Analysis of the Digital Economy – masters and undergraduate starting 2019

ACSPRI
Australian Consortium for Social & Political Research Inc.

vosonSML R package

- Aims to be the "Swiss army knife" for collecting social media data via free APIs and constructing datasets for network and text analysis

- **Tim Graham** (ANU, @TimothyJGraham)
  – lead developer & maintainer

- **Rob Ackland** (ANU, @RobAckland)

- **Chung-hong Chan** (Univ. of Mannheim, @chainsawriot) – *new UI using maggritr*
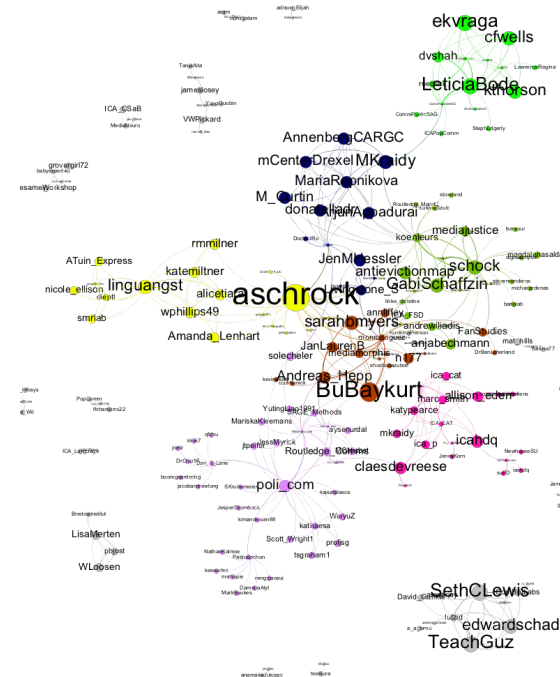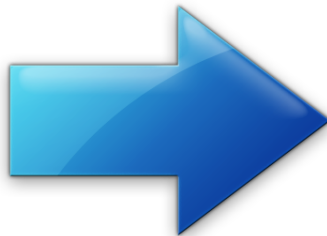
- **Bryan Gertzel** (ANU)

Tutorials / help: http://vosonlab.net/SocialMediaLab

vosonSML: example workflow

Collect 500 latest tweets from **#ica17** and construct an "actor" network showing replies+mentions+retweets between users (visualisation using Gephi)

vosonSML data typology

VOSON SocialMediaLab – Data Typology (5th April 2016, version 0.20.1)

	Facebook	Twitter	Instagram	Instagram – Ego	YouTube
Data collection	Manually created list of Facebook fan pages	Search on terms (usernames, words, hashtags)	Search on terms in caption OR geographical search location of person posting... i.e. uploader)	Manually created list of users (who may or may not have...	Manually created list of videos
Actor(s)	Users Facebook fan pages p... (comments)	Users	Users Captions... photo)	Users	Users
Network(s)	"bi-modal" - directed ties from user to post based on like and comments "actor network (users)" - undirected from user to user via... (not yet implemented in package ...graph) "actor network" – undirected ties from posts to... via project (not yet implemented)	"bi-modal" - directed ties from user to word/hashtag "actor network" - directed ties from user to user based on @mention, @reply, RT	"bi-modal"... ties from user to ... (photo) based... comments. Not... author of caption as vertex attribute.	...ties from user to ...based on follows	Directed ties from user to user based on mention or reply ("affiliation network")
Semantic network	...o	Yes – words and hashtags are different actor types, edges are co-occurrence in tweet payload	No	No	No
Dynamic network	Yes	No	N...	No	No
Text content	Post and comment text Usernames	Tweet payload Usernames	Comment and caption text Usernames	Usernames	Comment text Usernames

VOSON Dashboard R/Shiny App

- Lead Developer: Bryan Gertzel (ANU)

- Developer & Documentation: Rob Ackland (ANU)

- Documentation: Xiaolan Cai (ANU)

- Shiny GUI for collection of social media network/text data (via R/vosonSML) and analysis of networks (via R/igraph) and text (currently via R/tm and R/wordcloud)

VOSON Dashboard (28Aug18*)

Home
Network Graphs
Network Metrics
Text Analysis
Assortativity
Twitter
Youtube
API Keys

Choose graphml file
Browse... twitter_a
Upload complete

Filters
Node Names
Multiple Edges
Loops Isolates

Graph Layout   Spread
Auto   1   2.5

Node Size   Multiplier
None   1   10

Categorical Attribute   View
All   All

Connected Components
Type   Size
Wea   1   475

Plot   D3 Force   D3 Simple   Network Graphs

Name: twitter_auspol_600.graphml (Type: twitter)
Description: Network loaded from file
Save Graph

Youtube
API Keys

Remove Twitter usernames
Apply word stemming

Summary
All Categories
Components (Weak): 4
Nodes: 161
Edges: 1444

Selected Categories
Components (Weak): 4
Nodes: 161
Edges: 1444

quency   Word Cloud   Comparison Cloud   Text Analysis

Bio

Global

Toxic

VOSON for hyperlink network construction & analysis

Created at ANU (Ackland), now developed and hosted by Uberlink Corp

Uberlink Corp

- Uberlink VOSON development team:

  - **Rob Ackland** (ANU, Founder & CEO, Location: Canberra)

  - **Jamsheed Shorish** (CTO, Location: Brussels)

  - **Francisca Borquez** (Communication Officer & Research Assistant, Location: Canberra)

- **VOSON 2.5** released 6 June 2017

  - Improved user interface/workflow

  - More flexbility with database naming (e.g. special characters)

  - Collect **Twitter data** from the real-time stream of tweets matching your search criteria (e.g. hashtag use) over a scheduled time period.

WEB SOCIAL SCIENCE

CONCEPTS, DATA AND TOOLS FOR
SOCIAL SCIENTISTS IN THE DIGITAL AGE

ROBERT ACKLAND

- Part I – Methods
  - Ch 1 – Introduction - Web Primer and Perspectives
  - Ch 2 – Online Research Methods
  - Ch 3 – Social Media Networks
  - Ch 4 – Hyperlink Networks
- Part II – Examples
  - Ch 5 – Friendship Formation and Social Influence
  - Ch 6 – Organisational Collective Behaviour
  - Ch 7 – Politics and Participation
  - Ch 8 – Government and Public Policy
  - Ch 9 – Production and Collaboration
  - Ch 10 – Commerce and Marketing