Using Social Media Data in Research

WebDataRA from WSI

Prof Leslie Carr

Web Data Research Assistant

- *Scrapes Twitter, Facebook and Google data into a spreadsheet*

- *Uniquely allows free historic data capture*

- *No programming required*

- *Browser extension, one-click install*

Overview of Use

- The Web Data RA will capture Twitter, Facebook and Google data from a browser and allow you to paste a table of information directly into a spreadsheet. This tutorial focuses on its use with Twitter.

1. Visit **bit.ly/WebDataRA** in Chrome, click on the blue "+ Add to Chrome" button. The small green icon will appear in the top right of the browser window, next to the URL bar.

2. Go to **twitter.com** and create a Twitter search or display a timeline

3. Click on the **WebDataRA** icon to start collecting tweets.
   - Every 5 secs the browser will automatically scroll to the bottom of the page to make Twitter load the next batch of results and add the updates to the clipboard.

4. When you have collected enough results, paste the data into an Excel spreadsheet.

5. Use Excel to analyse data, or export to other programs such as Gephi or Voyant for other kinds of analysis.

WebDataRA Tables

The tweet data, with author, mentions, hashtags, text and counts of retweets, replies and likes broken out in separate columns.	Account occurrence summary, a count of the number of times that each Twitter account appears in the dataset as author or a mention (including retweets).	Counts of the appearances of each hashtag.	A table of edges of the conversational network, i.e. the number of times each pair of accounts communicate with each other.

Using The Tweet Data Table

- The tweet data (gray) contains the basic data about each tweet: what was said, when, by who and to whom.

- Use this data to form a general overview of the communication over time and identify the most significant tweets.

- Examine specific tweets and their context by referring back to the Twitter site using each tweet's URL.

Pivot Table Visual Twitter Timeline

- Click on any gray cell in the Tweet Data table

- Choose "Pivot Table" from the Insert ribbon.

- In the Pivot Table builder
  - drag "Author" from the Field Name panel into the "Rows" panel
  - drag "Timestamp" into the "Columns" panel
  - drag "Author" (again) into the "Values" panel (it will automatically turn into "Count of Author").

- Reformat to create a helpful Timeline summary of contributors (vertical axis) by days (horizontal axis).
  - narrow the columns, slant the column headings, change the angle of the text to 60°
  - use the "Row Labels" control to sort by the author count
  - show only the rows where the total author count is greater than a chosen threshold.
  - use conditional formatting to highlight the most extreme values.
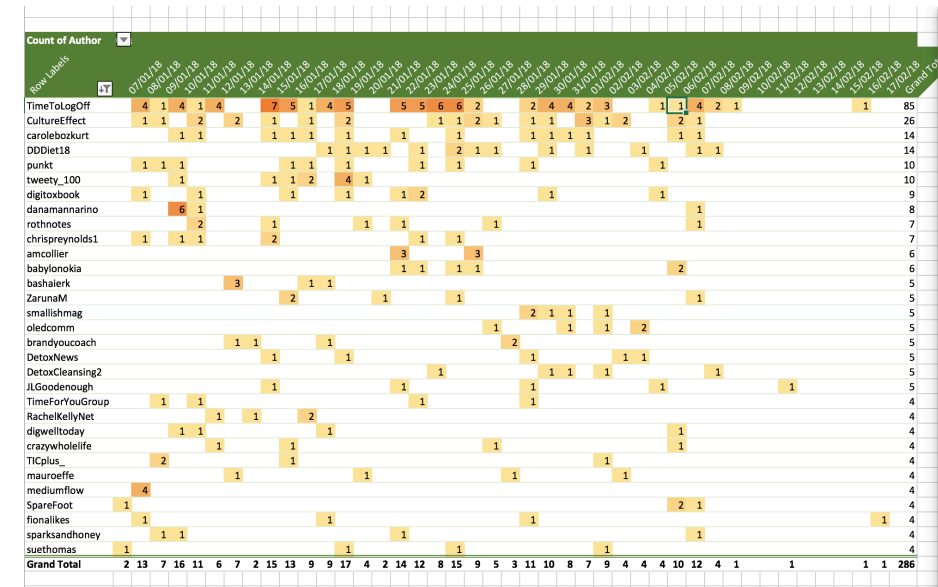
Other Questions to ask of the Data

- All kinds of summaries and analyses are possible using Excel on this data, including:
  - Showing the distribution of the tweet sample through time
  - Identifying the most prolific and/or popular actors, and showing their activity through time
  - Showing the use of individual hashtags (this might be useful in a big conversation, or one that evolves over a longer period)
  - Comparing the relative proportion of contributions from different actors / hashtags

Using The Account Data Table

- The account table (green) shows
  - the most active tweeters,
  - the most frequent repliers,
  - the most retweeted users.
- This shows the key actors in a conversation, and the main roles that they take.
- Get detailed information by clicking on the account names (linked) to see the account bios and the relevant timelines of these actors in the Twitter website.
- Understand whether they are corporate accounts, private individuals, bots or trolls.

Account	Author Count	...Inc RTs	Mention Count	...Inc RTs
1234margs	1	3		
2202Gayle	1	1		
25Score	1	2		
50Folds	1	3		
90_mile_beach	1	1		
949powerfm	1	2		
ABRAMSbooks	1	5		
ADJBlog	1	2		
AFARmedia			2	6
ALLITAustralia	1	2		
ASMRLittleMelon	1	1		
ASTsupportAAli	1	2		
AbhaReach	1	2		
AdamtheBullFAN			1	1
AdeldMeyer			1	1
AdrianChen			1	2
AgenceW	1	2		
AlanSee	1	3		
AlexRWilliams	1	4		
AlishaValerie	1	2		
Aloe_Vera_FLP	1	2		
AmlyBotanicals			1	6
Andrea_MacE	1	2		
AngelaInBold			1	1
Anitapratapsing	1	2		
AnnaBell_writes			1	5
Anonfighter5			1	1
AppliedIG	1	4		
Argentine58			2	2
AstraLifeID	1	5		
AvonworthEA	1	2		
BBC			1	1
BBCJonSopel	1	1		
BBCWomansHour			1	2

Inspecting Twitter Accounts

Account	#	Bio
ItsTimeToLogOff	30	Time To Log Off is the home of digital detox. We're spearheading the movement to disconnect regularly from digital devices and reconnect with the world offline. We do this through collecting facts on the need for digital detox, running campaigns to get everyone off their screens and hosting retreats, events and workshops.
DinnerTableMBA	9	A commercial organisation working together to help families become more confident, successful, and self-empowered
SpareFoot	8	A storage company. We make it easy to move and store your stuff. Reserve storage for free and get your mind out of the clutter.
CultureEffect	5	Author of Digitox: How to Find a Healthy Balance for your Family's Digital Diet

The account names in the account "author and mentions" (green) table are clickable, and open the page of the account profile in your default web browser.

Following the account hyperlinks for the most prolific authors in the green table, we see that they are all commercial or institutional actors to one extent or another.

Using The Hashtag Data Table

- The hashtag table (blue) shows you the most frequently used hashtags. This can help you extend your data gathering to look for more tweets relevant to your research question.

Hashtags	Count	...Inc RTs
##skincare	2	2
#124;	1	1
#365daysofself care.	1	1
#AI.	1	1
#AageSeRight	1	4
#AbuDhabi	1	2
#Apps	1	2
#Argyll	1	15
#BCNMI	1	1
#BLOGOSPHE RECHAT	1	2
#BabyZakTime	1	2
#Bachelor	1	3
#BeWellDoWe ll	1	5
#BienvenueChe zNous	1	2
#BlogosphereC hat	15	27
#BlogosphereC hat.	1	1
#CBCconf18	1	3
#Cafepress	2	3
#CaribbeanSea!	1	7
#CellPhone	1	7
#Cheshire	2	19
#Children	15	51

Using The Edge Data Table

- The edge table (yellow) will help you to see the interactions between actors, and help you to understand groupings of actors, and the pattern of their interaction.
  - Is a key account dominating a conversation and talking to many others?
  - Are they responding or just being passive recipients of marketing messages?
  - Is there a group of equals having a balanced conversation with equal participation?

Source	Target	Weight
1234margs	YouTube	1
2202Gayle	BlogosphereM	1
2202Gayle	ccfest	1
2202Gayle	edfringe	1
90_mile_beach	sundarakarma	1
ABRAMSbooks	tanyagoodin	1
ASMRLittleMelon	Anonfighter5	1
Andrea_MacE	iamwellandgood	1
AppliedIG	DanConnors16	1
BBCJonSopel	BBC	1
BBCJonSopel	OilPaul	1
BBCJonSopel	SebGorka	1
BVG_Kampagne	adidas	1
BVG_Kampagne	betablogr	1
BarryEdwardsJr	LloydEdwards91	1
BellyDanceRav	YouTube	2

Inspecting the Conversation Network

- Copy and paste the yellow table *into a separate spreadsheet* and save it as a CSV file (call it *edgetable.csv* or similar).

- Load up the network visualisation program "Gephi", and start a new project.

- In the "Data Laboratory", choose "Import Spreadsheet" and load up the CSV data as an *edge table*.
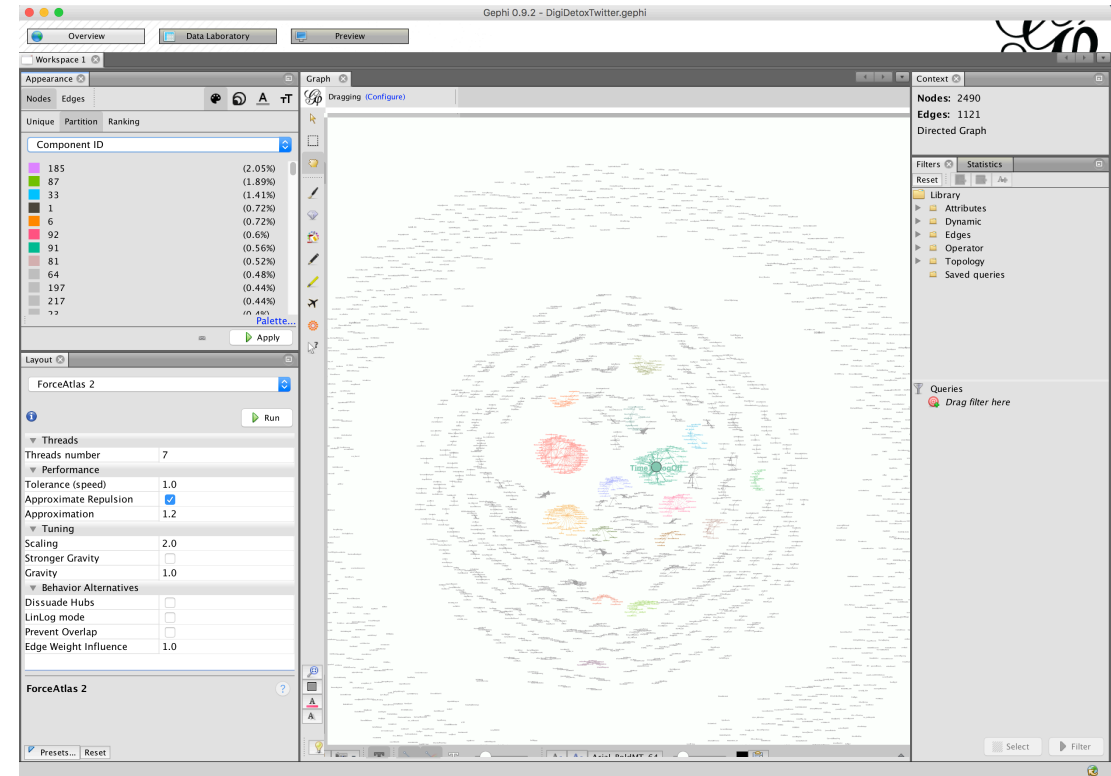
You can then apply a variety of network layout algorithms in the "Overview" pane.

Understanding the Conversation Network

- Many summaries and analyses are possible using Gephi's network visualisations.
- Showing the interaction of the network actors
- Identifying the communities and active participant subgroups within the larger sample
- Identifying the roles of different actors in the communications network
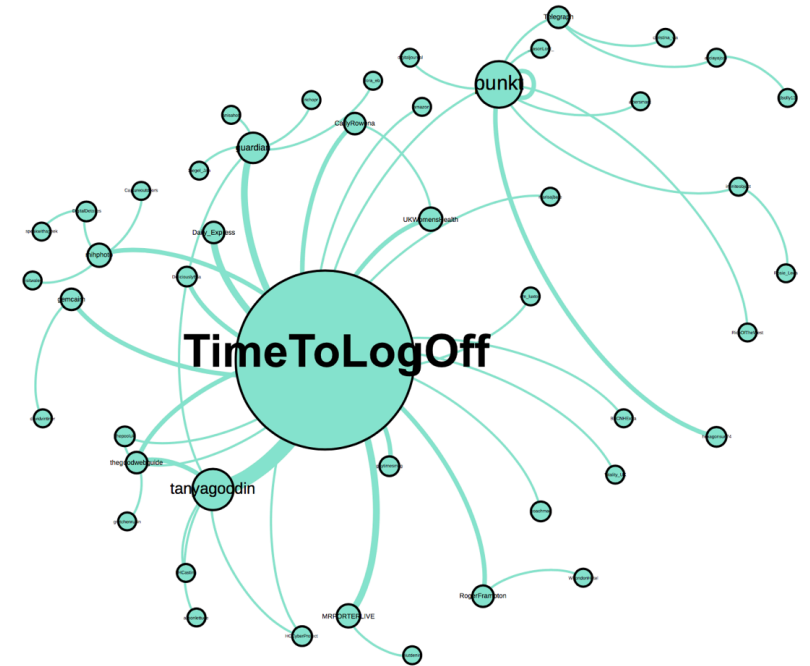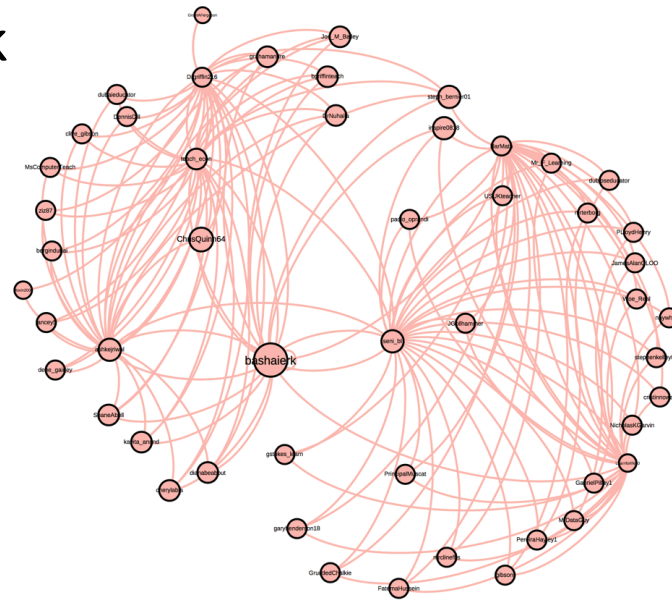
Textual Analyses of the Social Conversation

- In the gray table, copy the "Sanitised Text" column.
  - This contains the text of all the texts, but with all the Twitter features (@names, #hashtags, URLs) removed to leave only the English text.
- Go to the Voyant-Tools.org website
  - *Voyant Tools is a textual corpus analyser. It considers a Twitter conversation as a single document & individual tweets as individual sentences.*
- Paste the text into the textbox
- Press the "Reveal" button.
  - You will see a screen with several panels that help you explore the text of the tweets in different ways.

Textual Analyses

- Voyant includes a variety of textual analysis components
  - Word cloud
  - Trend analyser
  - Concordance
  - Summary
  - Vocabulary cluster analysis
  - Dimensional Reductions
  - Co-occurrence Network
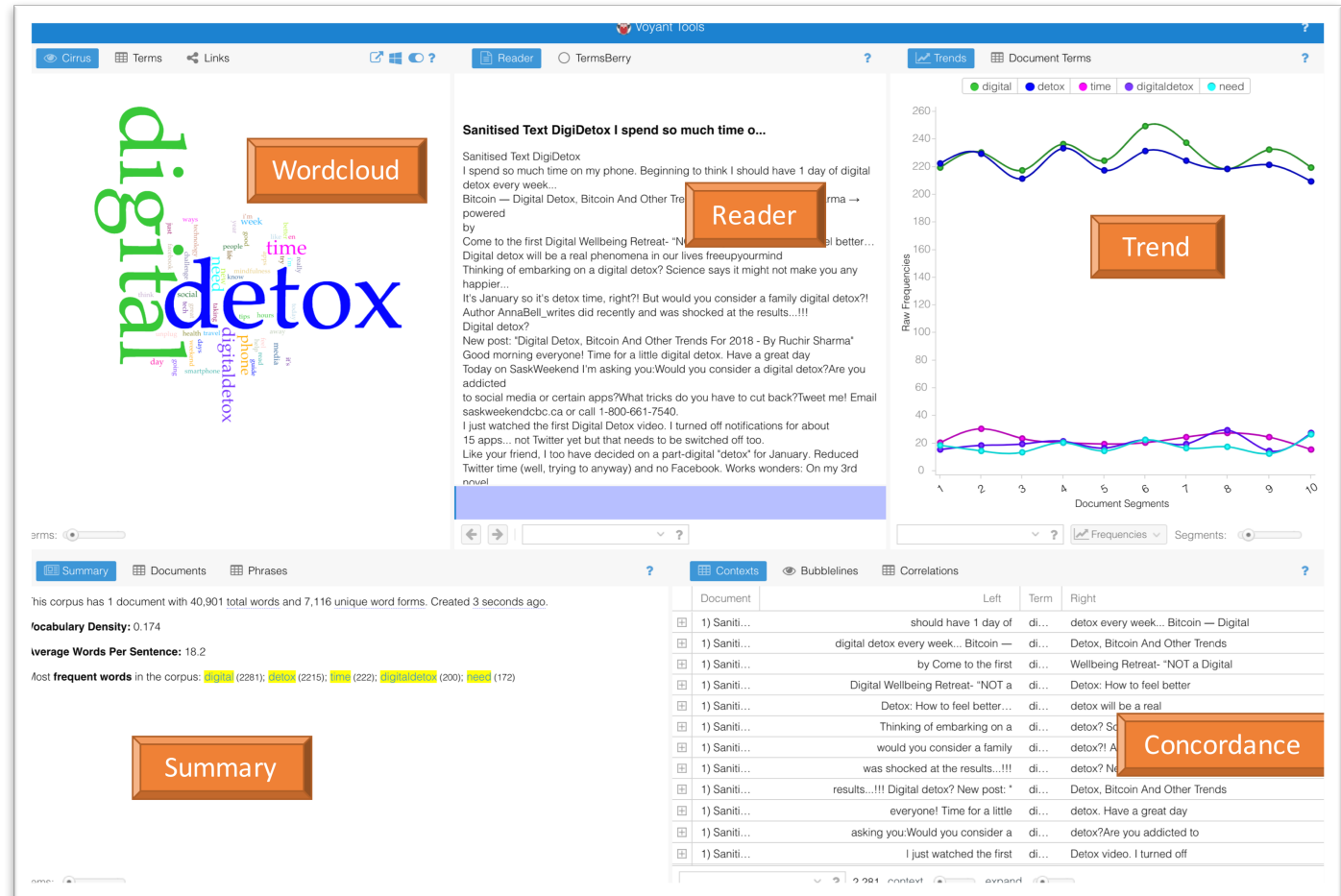
Sentiment Analysis

- Sentiment analysis can help you identify positive or negative comments in your sample.
  - This is a popular method in industry, especially with brand management companies. However it is academically contested, and does not have a high degree of transparency in the lexical processing.

- Paste the "Sanitised Text" column into **sentigem.com**.

- Consider to what extent the results seem accurate to you? How well does it identify positive and negative 'sentiment' in a tweet?
  - What kinds of inaccuracies can you see?

- Does it help you to identify any points of interest in your data for more thorough investigation?