

Data Collection and Visualization

An introduction to TAGS and Gephi
Tuja Khaund, COSMOS





Outline

1. Data Collection using TAGS
 - a. Background
 - b. How it collect data?
2. Data Visualization using Gephi
 - a. Background
 - b. Data
 - c. How to visualize data?



TAGS - an Introduction





TAGS - Background

- ❑ TAGS stands for **T**witter **A**rchiving **G**oogle **S**heets
- ❑ Developed as a 'hobby project' by Martin Hawksey in June 2010
- ❑ Six major versions have been released since its origin. Learn more about its history [here](#).



TAGS - How to collect data?

- ❑ Requirements:
 - A Twitter account, and a Google account for authentication
- ❑ Getting started:
 1. Visit the url <https://tags.hawksey.info/> to start
 2. Click on '**Get TAGS**' to proceed; then click on '**TAGS v6.1**'
 3. Click on '**Make a copy**' to get started
- ❑ The entire data collection process is also available [here](#).

TAGS - How to collect data?

- Steps 1, 2 and 3:

Copy of TAGS v6.1.9

File Edit View Insert Format Data Tools Add-ons Help TAGS 3

100% £ % .0 .00 123 Verdana 18 B I S A

TAGS v6.1.9

A B C

1 TAGS v6.1.9

2 <http://tags.hawksey.info> Created by mhawksey. Read more about this at:
<http://tags.hawksey.info>

3

4 With this spreadsheet you can: analytics GA

5 - automatically pull results from a Twitter Search into a Google Spreadsheet

6 Instructions:

7 1. If you've never run TAGS > Setup Twitter Access do so now (this should only need be done once for all your TAGS sheets)

8

9 2. Enter term you can use search operators like AND OR as well as from: and to: eg '#JobsNow AND from:BarackObama' (without quotes) 1

10

11 **Note:** Make a one off collection with TAGS > Run now! or set a trigger to collect every hour TAGS > Update archive every hour. To change the frequency open Tools > Script Editor then Triggers > Current script's triggers... and adjust

12 Advanced Settings:

13 Period	default	
15 Follower count filter	0	< if search term is being spammed you can set the minimum followers a person must have to be included in archive
16 Number of tweets	3000	< maximum varies based on the type of archive you are collecting
17 Type	search/tweets	< use a search term in step 3 above to get results from last 7 days

TAGS - How to collect data?

- Step 4, 5:
 - The output can be found in the 'Archive' sheet from bottom left

Rename file ☆ 📁

File Edit View Insert Format Data Tools Add-ons Help TAGS All changes saved in Drive

100% £ % .0 .00 123 Roboto

Advanced Settings:

A B

TAGS v6.1.9

Created by mhawkeye. Read more at <http://tags.hawkeye.info>

With this spreadsheet you can:

- automatically pull results from a Twitter Search into a Google Spreadsheet

Instructions:

1. If you've never run TAGS > Setup Twitter Access do so now (this should be done once)
2. Enter term <- you can use search operators like 'AND from:BarackObama' (without quotes)

Note: Make a one off collection with TAGS > Run now! or set a trigger to collect every hour TAGS > Update archive every hour. To change the frequency open Tools > Script Editor then Triggers > Current script's triggers... and adjust

Advanced Settings:

Period	default	<- if search term is being spammed you can set the minimum followers a person must have to be included in archive
Follower count filter	0	<- maximum varies based on the type of archive you are collecting
Number of tweets	500	<- use a search term in step 3 above to get results from last 7 days
Type	search/tweets	

Run Now! 4

Update archive every hour 5

Readme/Settings Archive

TAGS - How to collect data?

□ Data Snippet:

- The dataset is available to view [here](#).

id_str	from_user	text	created_at	time	geo_coordinates	user_lang	in_reply_to_user_id_	in_reply_to_screen_n	from_user_id_str	in_reply_to_status_id	source	profile_image_url																																												
109512	toogoodyy	RT @royallymegtruth: Meghan getting a shout out from the BBC commentator at the #BAFTAs about her fashion choice with sustainable designers and companies. The Meghan impact is real. My girl ain't even there. #DuchessofSussex	Tue Feb 12 00:43:48	12/02/2019 00:43:48		en			1594657132		<a href="http://twitter: <a 82fsuqxddy"="" href="http://obs.twimg.com</td><td></td></tr><tr><td>109512</td><td>perez_lorenzo1</td><td>RT @peopletv: So sweet! #TaylorSwift skips #GrammyAwards, attends #BAFTAs afterparty with boyfriend #JoeAlwyn. More on #PeopleNow: https://t.co/82FsuqxDDY https://t.co/YTauQT1VJA	Tue Feb 12 00:43:31	12/02/2019 00:43:31		es			941414241449283584		<a href="http://twitter: <a #baftas="" <a="" club.="" fan="" href="https://t.co/anmpyCMHgh" jaw="" left="" madden's="" muscle"="" richard="">https://t.co/anmpyCMHgh	Tue Feb 12 00:43:17	12/02/2019 00:43:17		en			1029870067083886115		<a href="http://twitter: <a href="http://obs.twimg.com</td><td></td></tr><tr><td>109512</td><td>louisaharriet</td><td>@MRodOfficial Thank you for taking the time to have a photo with me at the BAFTA awards in London this weekend. You are so amazing and I am honoured to have met you ☺ excuse my face, I was so overwhelmed. Couldn't hold my tears of joy back. #Bafas2019 #bafas https://t.co/wqHSyAerSO	Tue Feb 12 00:43:16	12/02/2019 00:43:16		en	51219631	MRodOfficial	50399071		<a href="http://twitter: <a 0zeqh0bu7z"="" href="http://obs.twimg.com</td><td></td></tr><tr><td>109512</td><td>lizandersoo</td><td>RT @billboard: Taylor Swift steps out at #BAFTAs afterparty with Joe Alwyn: see pic https://t.co/0Zeqh0BU7Z https://t.co/WmHtBsUQsS	Tue Feb 12 00:40:43	12/02/2019 00:40:43		es			1559182224		<a href="http://twitter: <a 8sgooiuqps"="" href="http://obs.twimg.com</td><td></td></tr><tr><td>109512</td><td>selenativist</td><td>RT @JenLawrence_BR: Jennifer Lawrence no Bafta de 2019 é algo que merece ser lembrado sempre! Auge da beleza. #EBAFTAs #BAFTA #BAFTAs https://t.co/8sgOOIUQpS	Tue Feb 12 00:40:42	12/02/2019 00:40:42		pt			3289935290		<a href="http://twitter:

TAGS - Visualization

Walkthrough:

- **Important!** - Make sure the file sharing is set to **public**
- You will need to publish your work on the web
- Click on 'TAG Explorer' to visualize your data
- You can view a sample visualization [here](#).

The image shows a Google Sheets spreadsheet titled "BAFTA Awards 2019 TAGS v6.1.9". The spreadsheet contains a list of tweets related to the BAFTA Awards. A context menu is open over the spreadsheet, and the option "Publish to the web..." is highlighted with a red circle. Below the spreadsheet, the "Publish" dialog is visible, showing options to "Link" or "Embed" the document, with "Entire Document" and "Web page" selected. A green "Publish" button is visible at the bottom of the dialog.

BAFTA Awards 2019 TAGS v6.1.9

File Edit View Insert Format Data Tools Add-ons Help TAGS

Share... .0 .00 123 Arial 8

New

Open... Ctrl+O

Import...

Make a copy...

Download as

Email as attachment...

Version history

Rename...

Move to...

Move to trash

Publish to the web...

Email collaborators...

Document details...

Spreadsheet settings...

Print Ctrl+P

Readme/Settings Archive

Link Embed

Entire Document Web page

Publish

Published content & settings



TAGS - Visualization

- ❑ Pros:
 - Visualization can be done with just one click
 - Can get better insights without data preprocessing
- ❑ Cons:
 - Limited to only three types of networks, reply, mention and retweets
 - Cannot make any visual changes
 - Hard to visualize big datasets

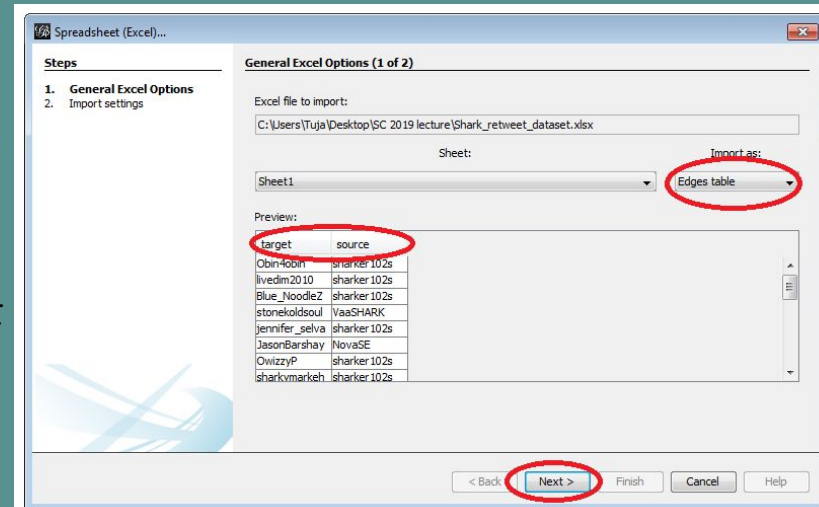


Gephi - a graph visualization tool



Gephi - Install and Import

- ❑ Step 1: Get set up
 - Download [Gephi](#), install it, open it up and start a new project.
- ❑ Step 2: Import a Spreadsheet
 - Input file needs two columns, “Source” and “Target.”
 - Click *File > Import spreadsheet*
 - Select ‘directed’ from the next screen



Gephi - understand the User Interface

The screenshot displays the Gephi 0.9.2 software interface. The main window shows a network graph with 812 nodes and 753 edges, visualized as a dense, radial cluster of black nodes and edges. The interface is divided into several panels:

- Appearance:** Located on the left, it includes a color selection tool (labeled '1') and an 'Apply' button.
- Layout:** Below Appearance, it shows the 'ForceAtlas 2' layout algorithm selected, with a 'Run' button (labeled '2') and a list of configuration parameters such as 'Threads number' (3), 'Tolerance (speed)' (1.0), and 'Scaling' (2.0).
- Graph:** The central workspace where the network graph is displayed, with a red '3' indicating the graph area.
- Context:** On the right, it provides summary statistics: 'Nodes: 812', 'Edges: 753', and 'Directed Graph'. A red '4' is placed next to the 'Edges' value.
- Filters and Statistics:** Below Context, there are tabs for 'Filters' and 'Statistics'. The 'Statistics' tab is active, showing a 'Network Overview' section with various metrics and their corresponding 'Run' buttons. A red '5' is placed next to the 'HTS' metric.

The bottom status bar shows the font settings: 'Arial Bold, 32'.

Gephi - creating labels

The screenshot shows the Gephi 0.9.2 interface with a data table and a dialog box. The data table has the following columns: Id, Label, Interval, and Modularity Class. The dialog box is titled 'Copy data to other column' and shows 'Copy data from 'Id'' and 'Copy to: Label'.

Id	Label	Interval	Modularity Class
sharker_102s			2
Clain4obin			2
livedm2010			2
Blue_jloodbz			2
VaaSHARK			0
stankaldesul			0
jennifer_selva			2
NovaSE			1
JasonBarshay			1
OnizzyP			2
sharkymarkah			2
NSUMDCollege			1
TeeZsharp			2
RiptideTheShark			3
edwardlevy10			3
MakoPeggyHughes			4
JuneauBug2011			4
nadner_weiner			4
Jonharding77			4
OCEARCH			4
tundraawolfqueen			4
jm_allen74			4
HazCenter			5
MaryElizabethE9			5
Bizarroluthor			6
TheNightyHtrae			6
BlackFishRich_			7
PicklesBottom			7
EricaVerrillo			8
GatzTdaMax			8
TheSharkDaymond			10

Dialog box content:

Copy data to other column

Copy data from 'Id'

Copy to: Label

Ok Cancel

Bottom toolbar buttons:

- Add column
- Merge columns
- Delete column
- Clear column
- Copy data to other column (circled in red)
- Fill column with a value
- Duplicate column
- Create a boolean column from regex match
- Create column with list of regex matching groups
- Negate boolean values
- Convert column to dynamic

Gephi - How to visualize?

The screenshot displays the Gephi 0.9.2 interface with a network graph visualization. The graph is a large, dense, circular cluster of nodes, primarily orange, with several smaller, distinct clusters of different colors (pink, blue, green, purple, yellow) scattered around the main cluster. A red pentagon is drawn around the entire graph. The interface includes several panels and toolbars:

- Appearance Panel (Left):** Shows the 'Modularity Class' dropdown menu with a red box around it. Below it, a list of classes is shown with their respective counts and percentages. A red '2' is next to the class with 64 nodes. The 'Apply' button is highlighted with a red box.
- Layout Panel (Left):** Shows the 'ForceAtlas 2' layout algorithm selected, with a red box around it and a red '3' next to it. The 'Run' button is also highlighted with a red box.
- Context Panel (Right):** Shows network statistics: Nodes: 812, Edges: 753, Directed Graph. The 'Modularity' value is 0.652, highlighted with a red box and a red '1' next to it.
- Graph Panel (Center):** Shows the network visualization with a red '4' next to it.

Modularity Class	Count	Percentage
29	52.96%	(52.96%)
40	6.77%	(6.77%)
64	4.19%	(4.19%)
17	3.33%	(3.33%)
28	3.2%	(3.2%)
44	2.59%	(2.59%)
16	2.46%	(2.46%)
66	1.6%	(1.6%)

Network Overview	Value	Run
Average Degree		Run
Avg. Weighted Degree		Run
Network Diameter		Run
Graph Density		Run
HTS	1	Run
Modularity	0.652	Run
PageRank		Run
Connected Components		Run

Node Overview	Run
Avg. Clustering Coefficient	Run
Eigenvector Centrality	Run

Edge Overview	Run
Avg. Path Length	Run

Dynamic	Run
# Nodes	Run
# Edges	Run
Degree	Run
Clustering Coefficient	Run

Gephi - How to visualize?

The screenshot displays the Gephi 0.9.2 software interface. The main window shows a network visualization with a central node labeled "mopage19" and numerous surrounding nodes of various colors (orange, blue, green, pink, purple, cyan) and sizes, connected by edges. The interface includes a menu bar (File, Workspace, View, Tools, Window, Help) and a toolbar with buttons for Overview, Data Laboratory, and Preview. On the left, the "Preview Settings" panel is open, showing various configuration options for nodes and edges. At the bottom left, the "Export" button is circled in red, with the text "SVG/PDF/PNG" next to it. The "Preview" window on the right shows the current visualization at 100% zoom.

Gephi 0.9.2 - Project 1

File Workspace View Tools Window Help

Overview Data Laboratory Preview

Workspace 1

Preview Settings

Presets

Default

Settings Manage renderers

Nodes

Border Width 0.0

Border Color custom [0,0,0]

Opacity 80.0

Per-Node Opacity

Node Labels

Show Labels

Font

Proportional size

Color custom [0,0,0]

Shorten label

Max characters 30

Outline size 0.0

Outline color custom [255,255,255]

Outline opacity 80.0

Box

Box color parent

Box opacity 100.0

Edges

Show Edges

Thickness 5.0

Rescale weight

Min. rescaled weight 0.1

Max. rescaled weight 1.0

Color target

Opacity 100.0

Curved

Radius 0.0

Edge Arrows

Size 3.0

Preview ratio: 100%

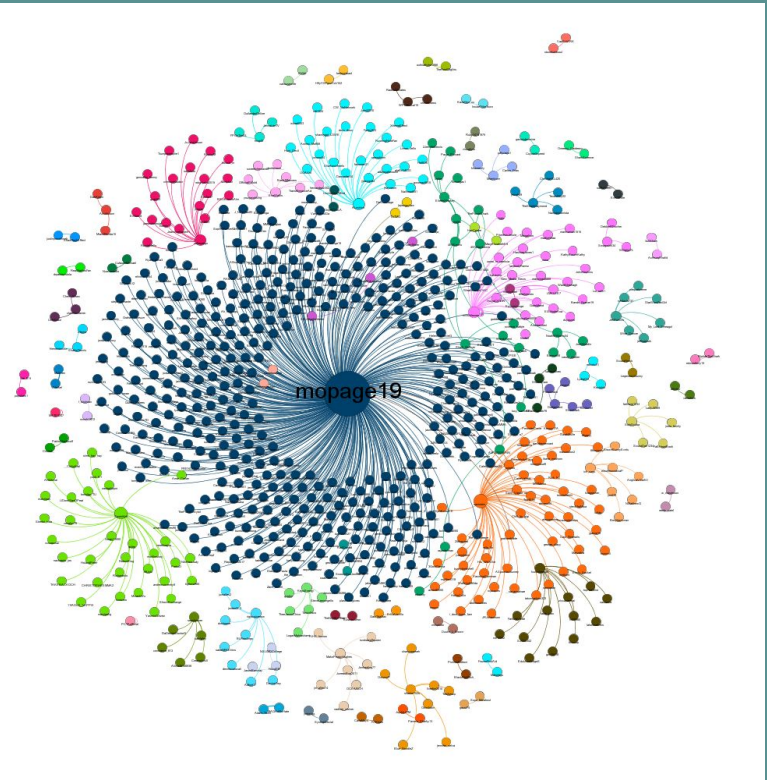
Refresh

Export: SVG/PDF/PNG

Background Reset zoom - +

Sample Data and Visualization

	A	B	C	D
1	target	source		
2	Obin4obin	sharker102s		
3	livedim2010	sharker102s		
4	Blue_NoodleZ	sharker102s		
5	stonekoldsoul	VaaSHARK		
6	jennifer_selva	sharker102s		
7	JasonBarshay	NovaSE		
8	OwizzyP	sharker102s		
9	sharkymarkeh	sharker102s		
10	NSUMDCollege	NovaSE		
11	Tee2sharp	sharker102s		
12	edwardlevy10	RiptideTheShark		
13	JuneauBug2011	MakoPeggyHughes		
14	nadiner_weiner	MakoPeggyHughes		
15	Jonharding77	MakoPeggyHughes		
16	OCEARCH	MakoPeggyHughes		
17	tundrawolfqueen	MakoPeggyHughes		
18	jm_allen74	MakoPeggyHughes		
19	MaryElizabethE9	HazCenter		
20	TheNightlyHtrae	BizarroLuthor		



Thank You!

Questions?