

An Introduction to Social Network Analysis With Gephi

February 2, 2022

Devon Mordell, Educational Developer
MacPherson Institute, McMaster University

Do More With Digital Scholarship Series



Image: [Mhsheikholeslami](#) (CC 4.0 BY-SA)

Session Recording and Privacy

This session is being recorded with the intention of being shared publicly via the web for future audiences.

In respect of your privacy, participant lists will not be shared outside of this session, nor will question or chat transcripts.

Questions asked via the chat box will be read by the facilitator without identifying you. Note that you may be identifiable when asking a question during the session in an audio or visual format.

Code of Conduct

The Sherman Centre and the McMaster University Library are committed to fostering a supportive and inclusive environment for its presenters and participants.

As a participant in this session, you agree to support and help cultivate an experience that is collaborative, respectful, and inclusive, as well as free of harassment, discrimination, and oppression. We reserve the right to remove participants who exhibit harassing, malicious, or persistently disruptive behaviour.

Please refer to our code of conduct webpage for more information:

scds.ca/events/code-of-conduct/



We're here to help!

Use **TH:** [query] in chat to
let the facilitators know
that you're having
technical issues

By the end of this workshop...

You'll be able to:

- Define key concepts in social network analysis
- Explain what phenomena can be observed through social network analysis
- Create a network visualization in Gephi



Have you downloaded Gephi?

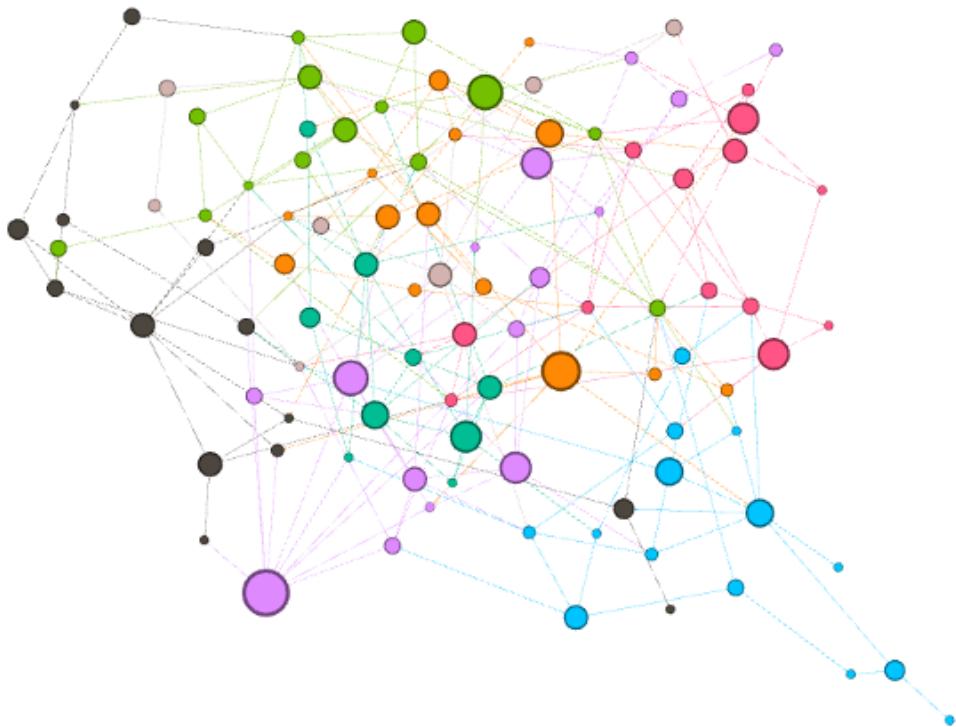
- Go right to the source: <https://gephi.org>

On Social Network Analysis

This is not a comprehensive course on social network analysis

Social network analysis is a methodological approach to representing the shape and characteristics of social structures.

- i.e. visualizing relationships between interdependent entities

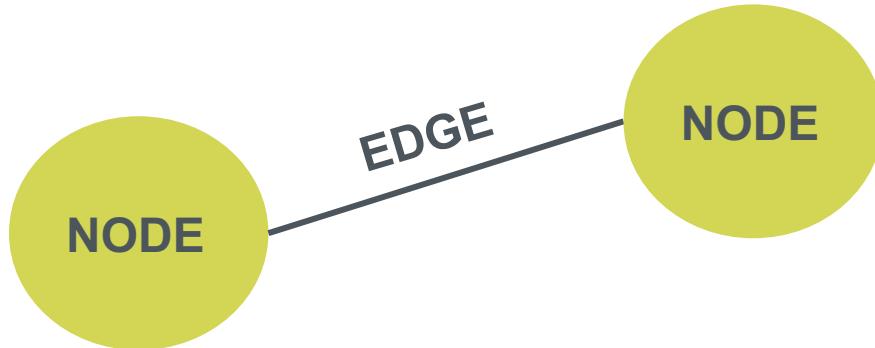


Describing Relationships

Node: the ‘actor’ in the network

Edge: the relationship connecting actors

Attribute: features of the node or edge



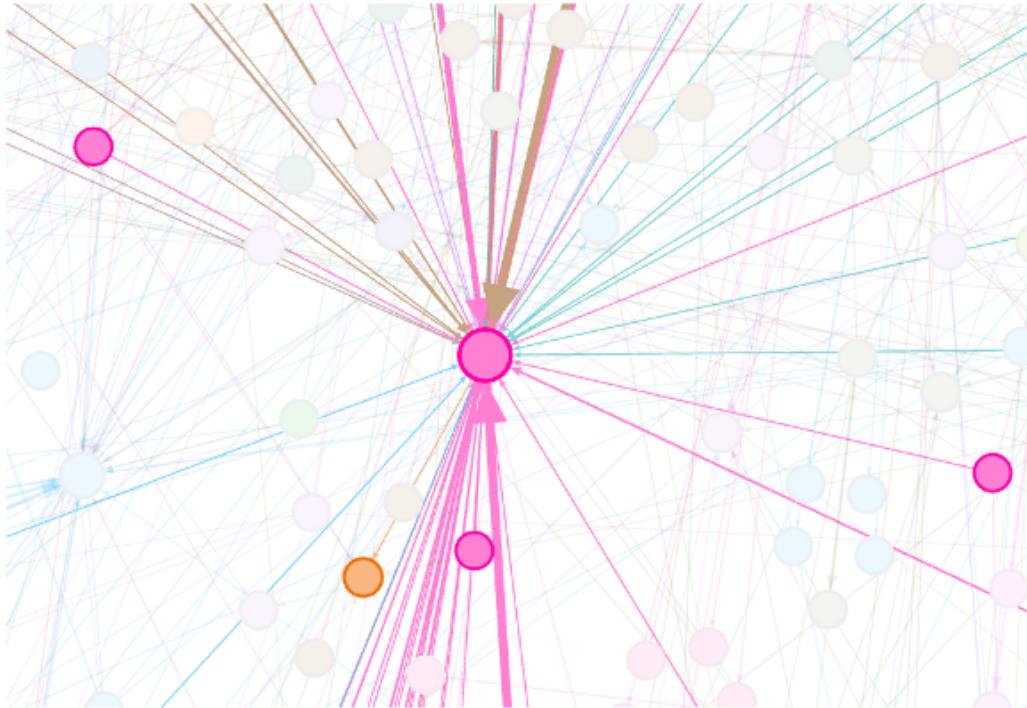
Other terminology

Edges can be **directed** or **undirected**.

Directed edges allow us to calculate **in-degree** and **out-degree**.

- In-degree: number of incoming directed edges
- Out-degree: number of outgoing directed edges

Directed edges can also have a relative **weight**.



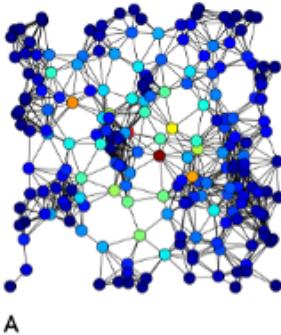
Network graph with weighted edges

Centrality

Determining which nodes are the most important in the cluster or graph...

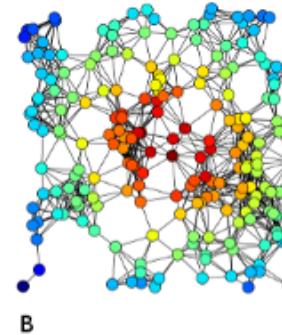
- **Degree**
 - nodes with the most connections (i.e. edges)
- **Closeness**
 - nodes closest to all other nodes (as a path)
- **Betweenness**
 - nodes which bridge the shortest paths
- **Eigenvector**
 - nodes that have a higher relative influence

A. Betweenness



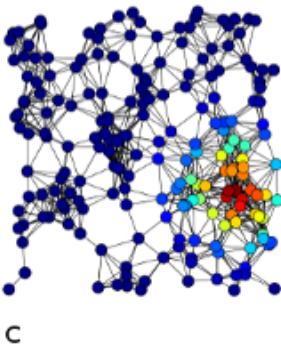
A

B. Closeness



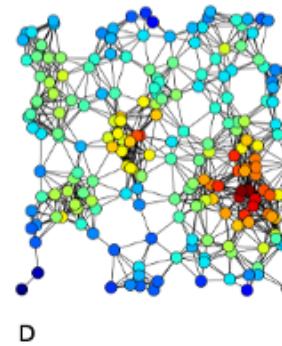
B

C. Eigenvector



C

D. Degree



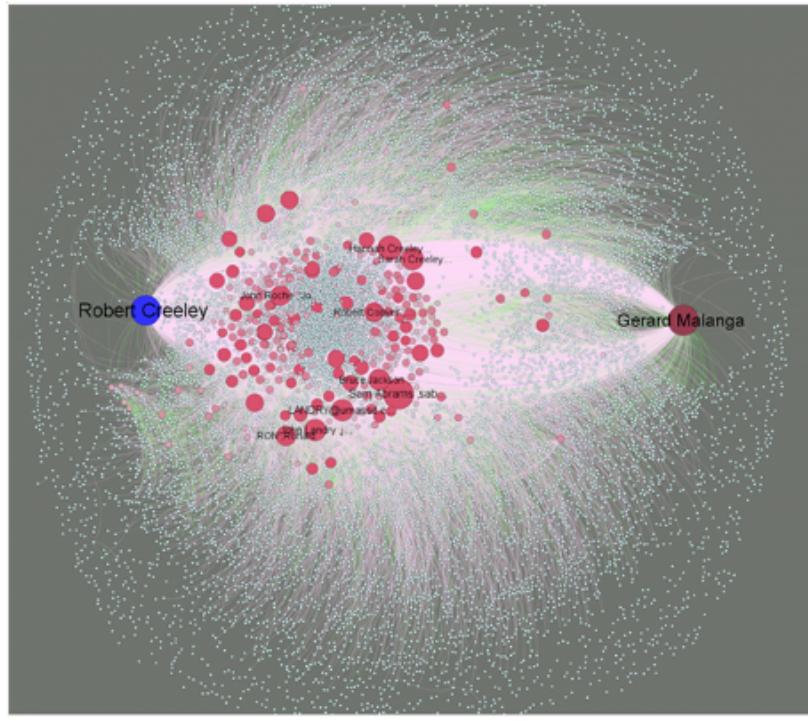
D

Image: [Tapiocozzo](#) (CC 4.0 BY-SA)

What Can SNA Be Used For?

Many applications in digital scholarship:

- Words that appear together often in a text (i.e. concordance)
- Correspondence between people (e.g. email or twitter)
- Communities in social networks (e.g. facebook friends)
- etc.



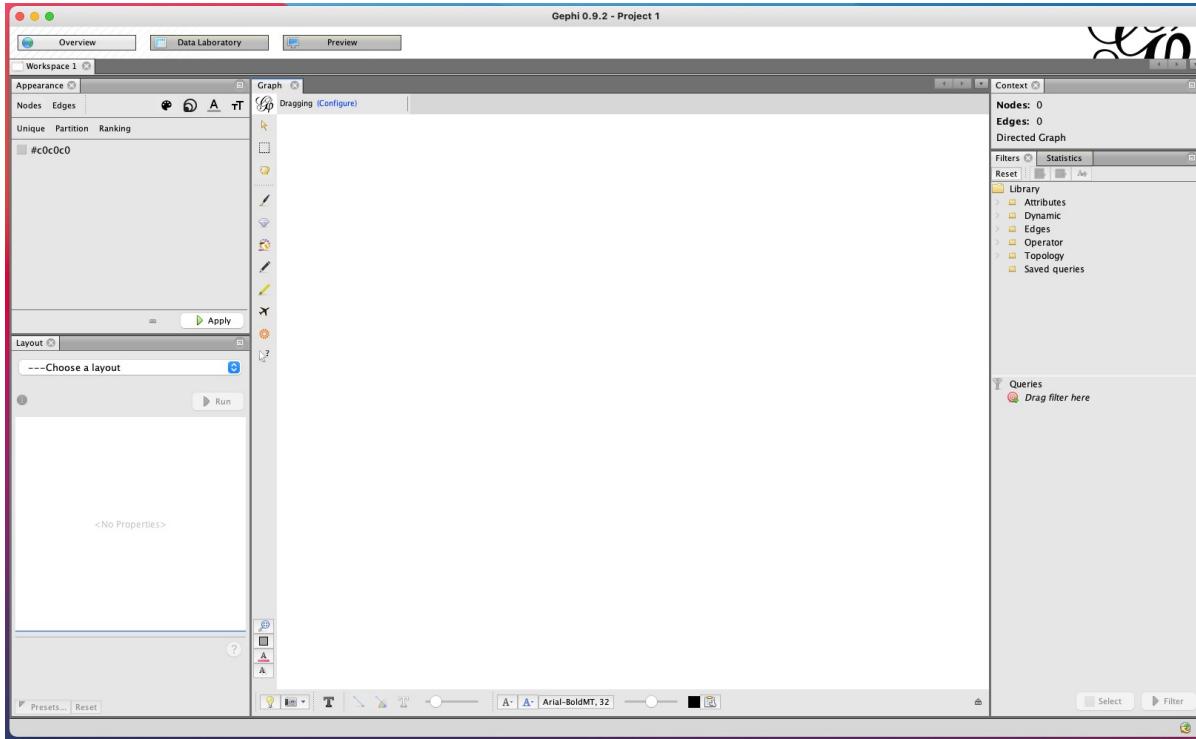
<https://dhs.stanford.edu/visualization/robert-creeley-e-mail-correspondence-network/>

Robert Creeley email archive (Stanford U)

The dataset

Download the [#elxn43 debate night dataset](#)

- Scraped from Twitter on Oct 7, 2019 → use of #elxn43 hashtag
- Pre-processed in OpenRefine
 - Isolated original tweets (excluded retweets)
 - Extracted mentions (@) from tweet text
 - Created source-target row for each mention



Gephi > New Project

Spreadsheet (CSV)...

Steps

1. General CSV options
2. Import settings

General CSV options (1 of 2)

CSV file to import:
D:\OneDrive - University of Windsor\Desktop\elxn43_debate-network.csv

Separator: Import as: Charset:

Comma Edges table UTF-8

Preview:

source	target	source_label	target_label	created_at	followers	rt
yafajarrar	justintrudeau	yafajarrar	justintrudeau	23:59:59	177	2
aminpost	andrewsheer	aminpost	andrewsheer	23:59:58	891	0
aminpost	justintrudeau	aminpost	justintrudeau	23:59:58	891	0
healthmichael	thejagmeets...	healthmichael	thejagmeets...	23:59:57	910	0
mistervermin	kevlinbin	mistervermin	kevlinbin	23:59:55	178	0
mistervermin	ezalevant	mistervermin	ezalevant	23:59:55	178	0
samwisefrmr...	susandlarn...	samwisefrmr...	susandlarn...	23:59:53	45	0

< Back Next > Finish Cancel Help

Import dataset (CSV) as: Edges table



Spreadsheet (CSV)...

Steps

1. General CSV options
2. Import settings

Import settings (2 of 2)

Time representation
Intervals

Imported columns:

source

target

source_label
String

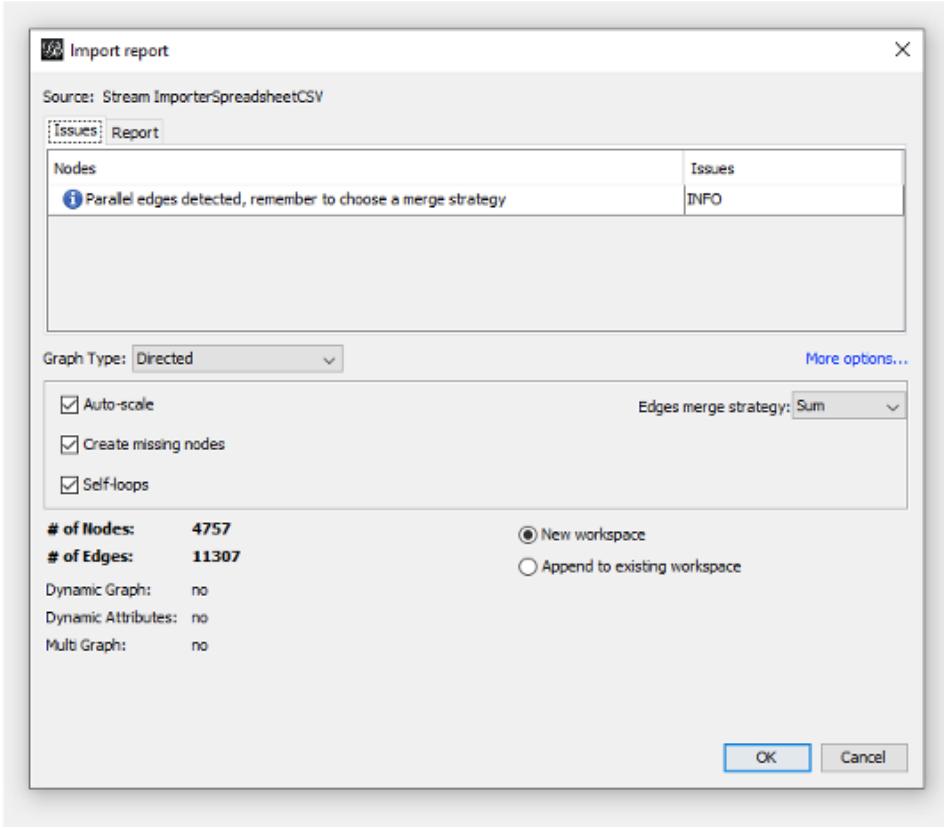
target_label
String

created_at
String

< Back Next > **Finish** Cancel Help

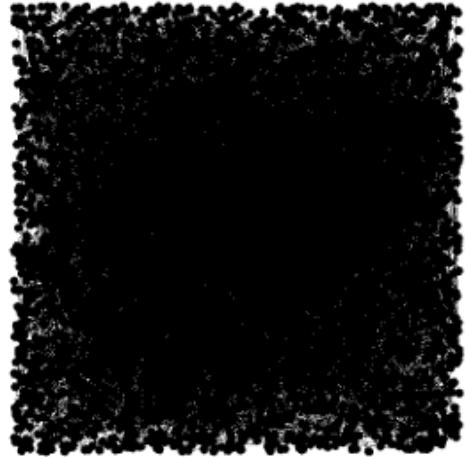
Note: name columns “source” and “target”



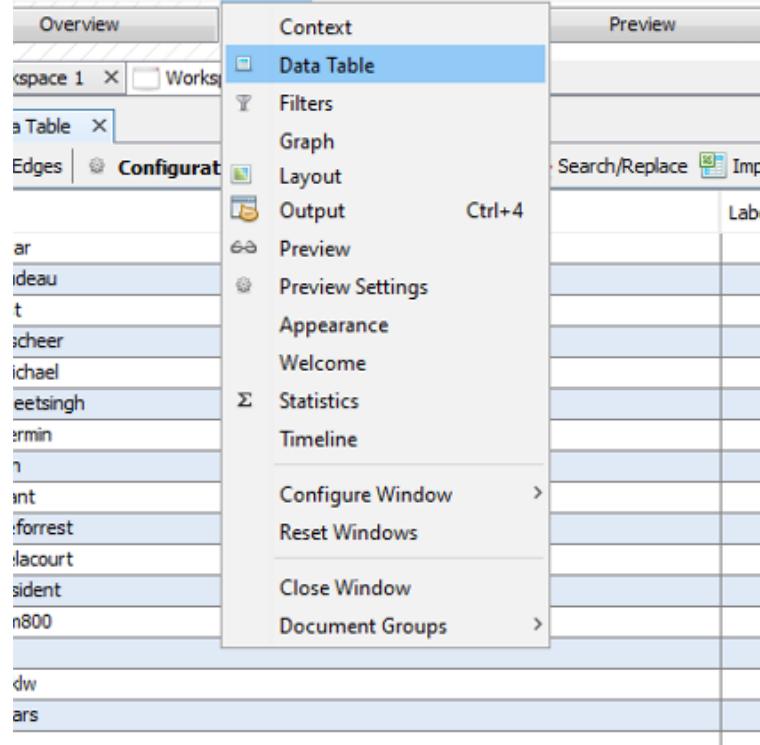


Almost there... directed graph?



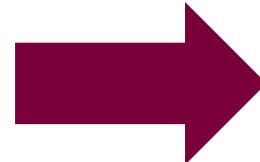
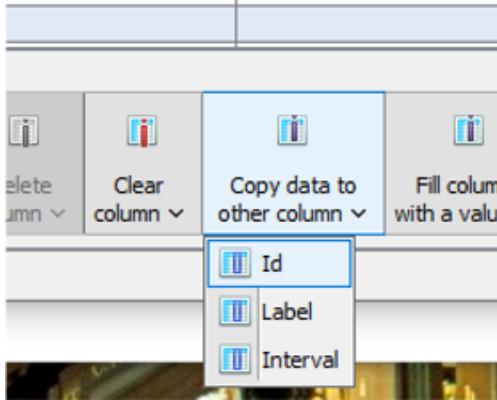


Overview → HAIRBALL!!!!



Data Laboratory > Data Table view





Data Table		
Nodes	Edges	Configuration
Id	Label	Inte
yafajarrar	yafajarrar	
justintrudeau	justintrudeau	
aminpost	aminpost	
andrewscheer	andrewscheer	
healthmichael	healthmichael	
thejagmeets...	thejagmeets...	
mistervermin	mistervermin	
kevinlibin	kevinlibin	
ezralevant	ezralevant	
samwiseforr...	samwiseforr...	
susandelaco...	susandelaco...	
cfnupresident	cfnupresident	
baconam800	baconam800	
ctv	ctv	
am800cklw	am800cklw	
girfrmars	girfrmars	
rgl007	rgl007	
elizabethmay	elizabethmay	
yfblanchet	yfblanchet	
globalnews	globalnews	

Copy data from Id to Label (in Node view)

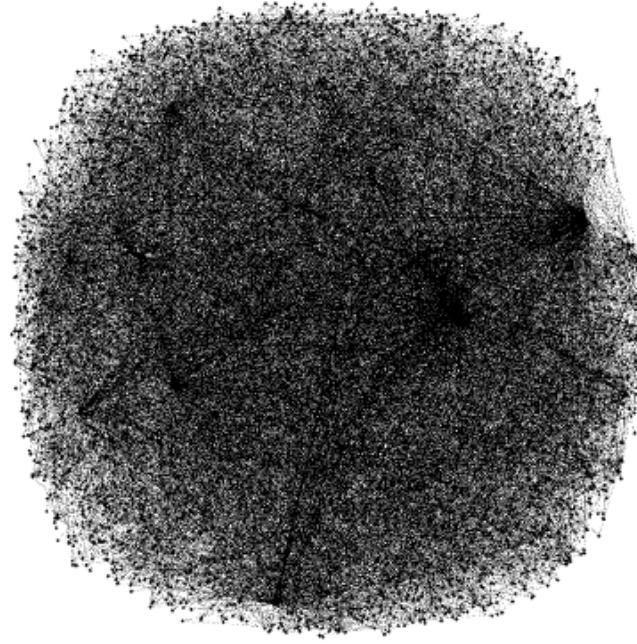
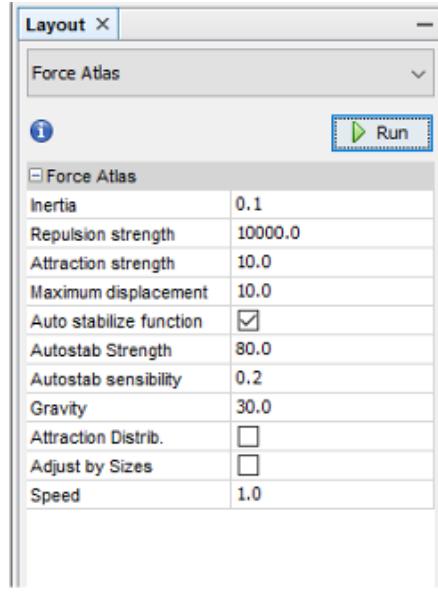
Data Table X

Nodes Edges Configuration Add node Add edge Search/Replace Import Spreadsheet Export table More actions Filter Source

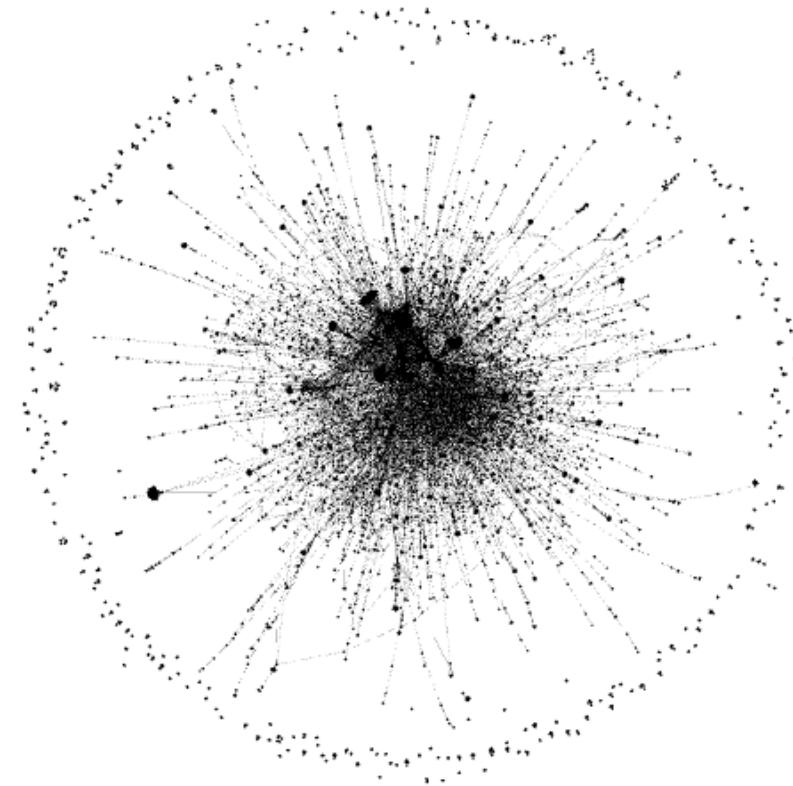
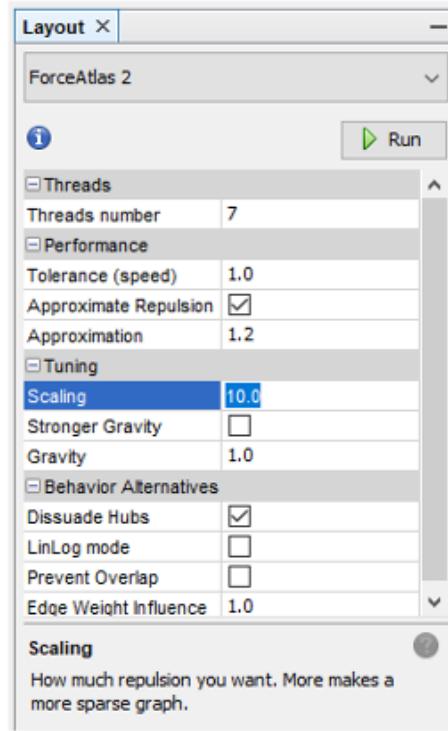
Source	Target	Type	Id	Label	Interval	Weight	follower	rt
yefajarr	justintrudeau	Directed	0			3.0	177	0
anipost	andrewshear	Directed	1			1.0	891	0
anipost	justintrudeau	Directed	2			1.0	891	0
healthncheel	theyagmeetsingh	Directed	3			1.0	910	0
mistervermen	kevnlion	Directed	4			1.0	178	0
mistervermen	carallevant	Directed	5			1.0	178	0
samwiseforest	susandelacourt	Directed	6			1.0	45	0
cfrupresident	andrewshear	Directed	7			4.0	4190	0
baconwm800	clyw	Directed	8			1.0	1520	0
baconwm800	am80oddw	Directed	9			1.0	1520	0
griffmarrs	andrewshear	Directed	10			1.0	574	0
rg007	theyagmeetsingh	Directed	11			1.0	50	2
rg007	elizabethmray	Directed	12			1.0	50	2
rg007	yblanchet	Directed	13			1.0	50	2
rg007	justintrudeau	Directed	14			1.0	50	2
rg007	andrewshear	Directed	15			1.0	50	2
globalnews	alihara	Directed	16			7.0	358424	1
globalnews	susandelacourt	Directed	17			1.0	358419	4
emile_wickham	theyagmeetsingh	Directed	18			1.0	147	0
emile_wickham	andrewshear	Directed	19			2.0	147	0
vandongene	theyagmeetsingh	Directed	20			1.0	2330	3
dingman	theyagmeetsingh	Directed	21			1.0	2	2
dingman	elizabethmray	Directed	22			1.0	2	2
dingman	yblanchet	Directed	23			1.0	2	2
dingman	justintrudeau	Directed	24			1.0	2	2
dingman	andrewshear	Directed	25			1.0	2	2
mbrianadams	andrewshear	Directed	26			1.0	20	0
opencoce	danyairazza	Directed	27			2.0	3899	2
ang_english	theyagmeetsingh	Directed	28			1.0	58	0
the_bearristol	justintrudeau	Directed	29			1.0	286	0
taqwamuhina	theyagmeetsingh	Directed	30			1.0	938	0
lofekindigly	justintrudeau	Directed	31			1.0	18	0

Add column Merge columns Delete column Clear column Copy data to other column Fill column with a value Duplicate column Create a boolean column from regex match Create column with list of regex matching groups

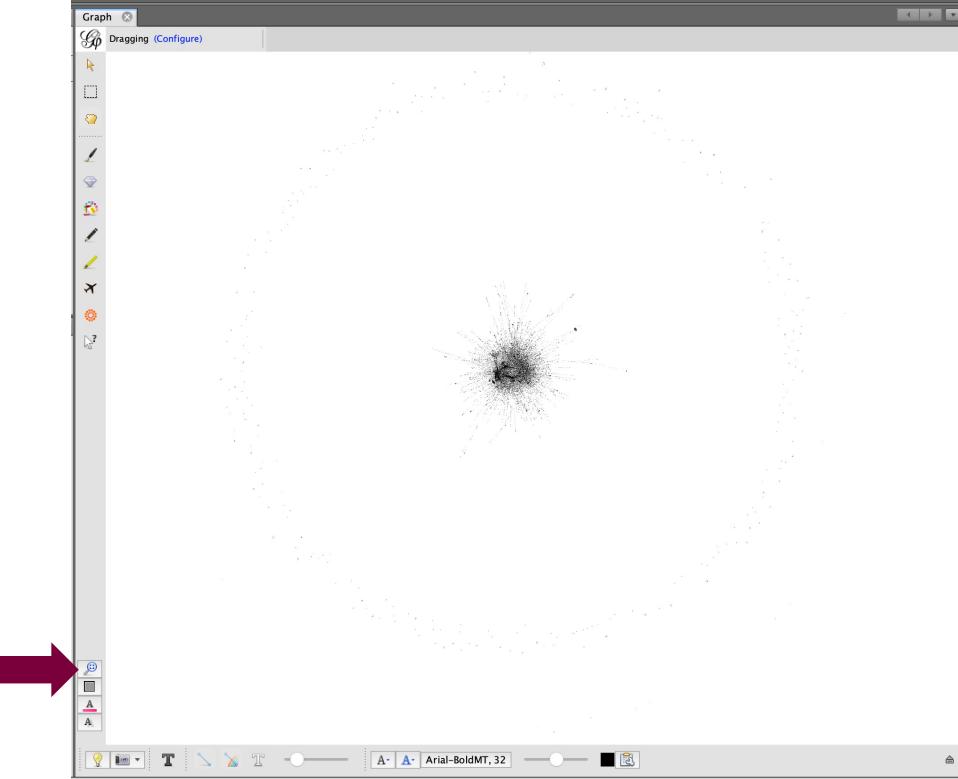
What does ‘weight’ refer to...?



Layout (begin with Force Atlas)



Try ForceAtlas 2...



Getting around the canvas

The screenshot shows two panels of a software interface. The left panel is a 'Filters' library with a tree view of filter types: Library, Attributes, Dynamic, Edges, Operator, Topology, and Saved queries. Under 'Attributes', several sub-options are listed: Equal, Inter Edges, Intra Edges, Non-null, Partition, Partition Count, and Range. The 'Range' option is highlighted with a blue selection bar. The right panel shows a 'Queries' section with a single entry: 'Range (followers)'. Below it is a 'Range (followers) Settings' dialog box containing a horizontal slider with values '5' and '4594493'. At the bottom of the dialog are 'Select' and 'Filter' buttons, with 'Filter' being the active button.

Filters to omit nodes (e.g. bot accounts)

Context X

Nodes: 4757
Edges: 8381
Directed Graph

Filters Statistics X

Settings

Network Overview

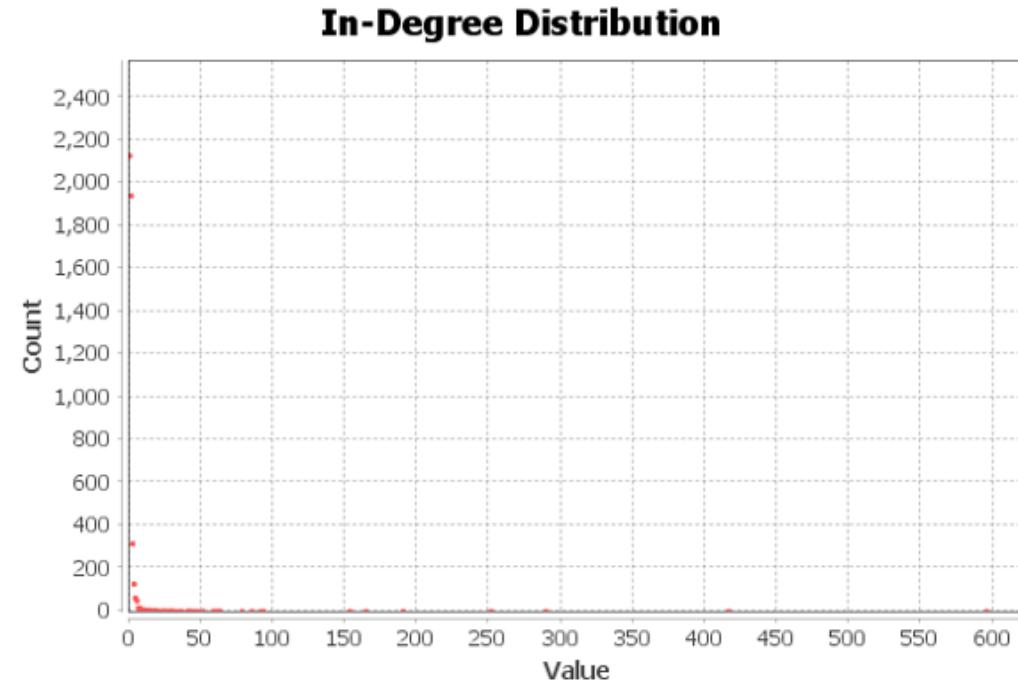
- Average Degree Run
- Avg. Weighted Degree Run
- Network Diameter Run
- Graph Density Run
- HITS Run
- Modularity Run
- PageRank Run
- Connected Components Run

Node Overview

- Avg. Clustering Coefficient Run
- Eigenvector Centrality Run

Edge Overview

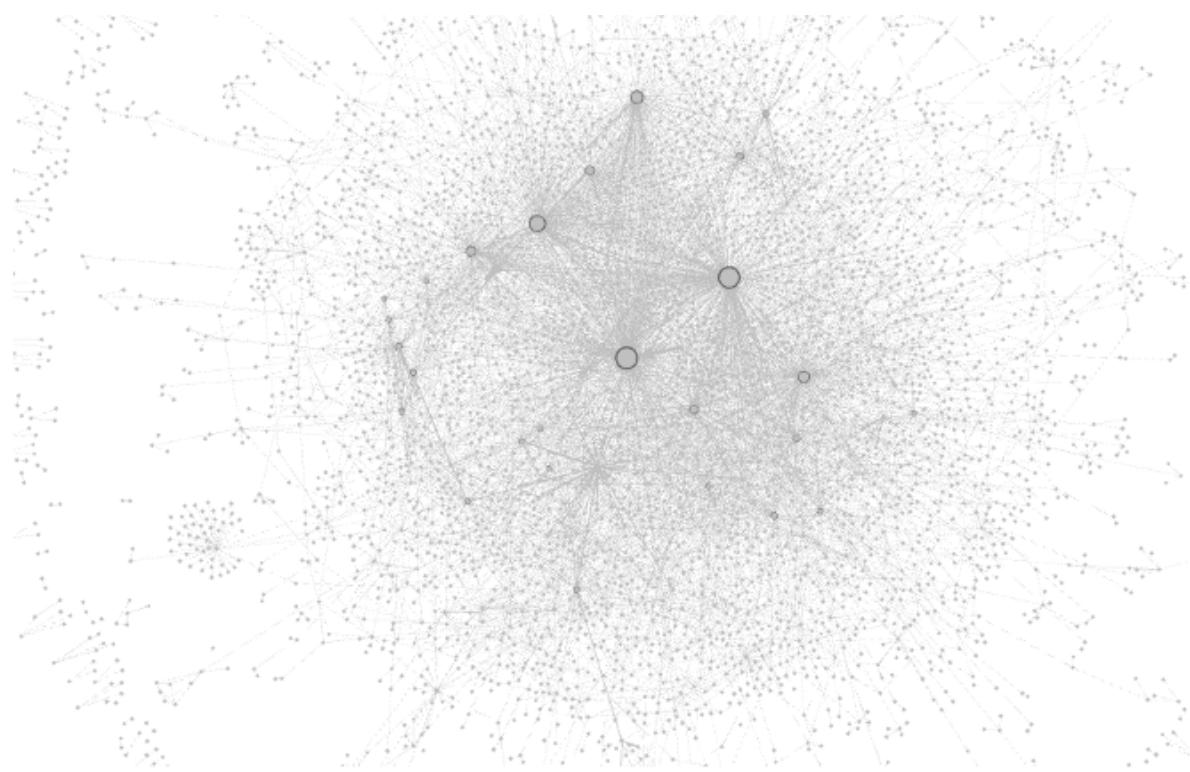
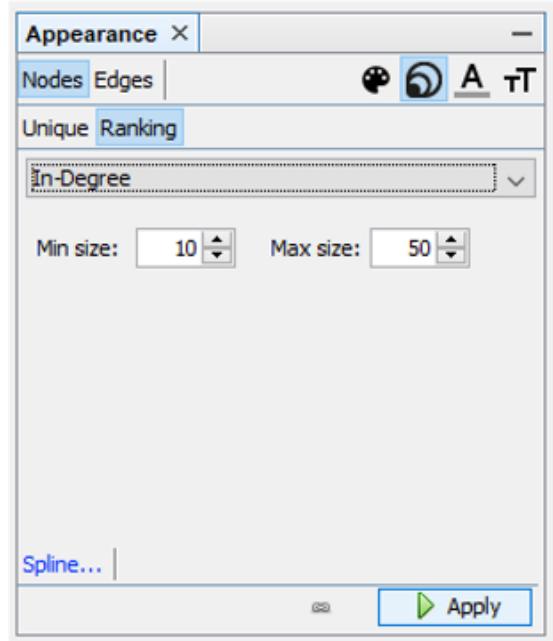
- Avg. Path Length Run



Run Statistics for more attributes...

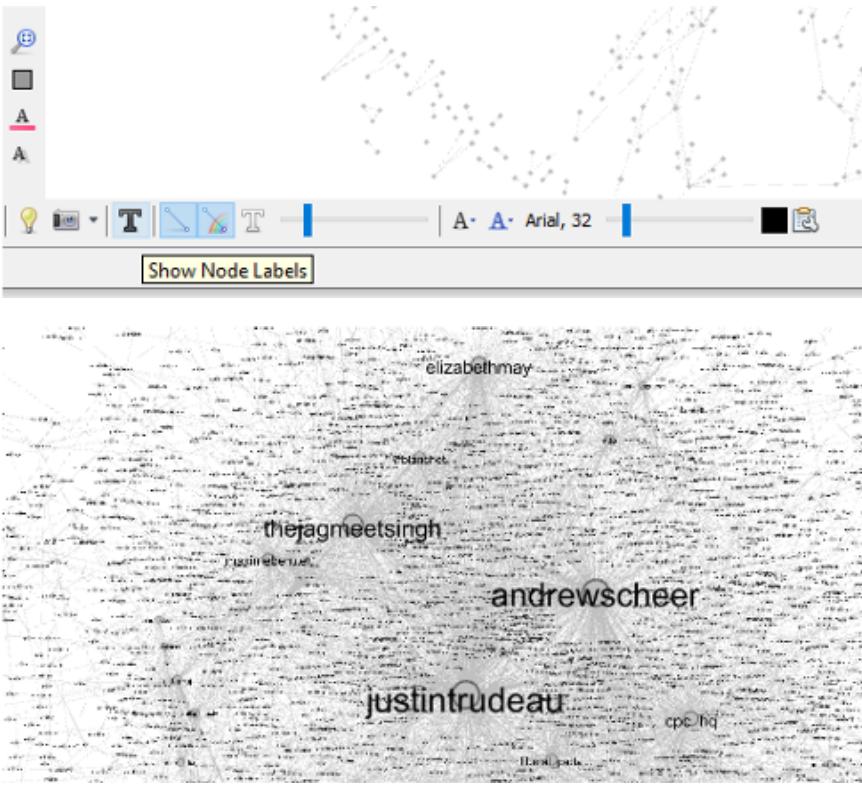
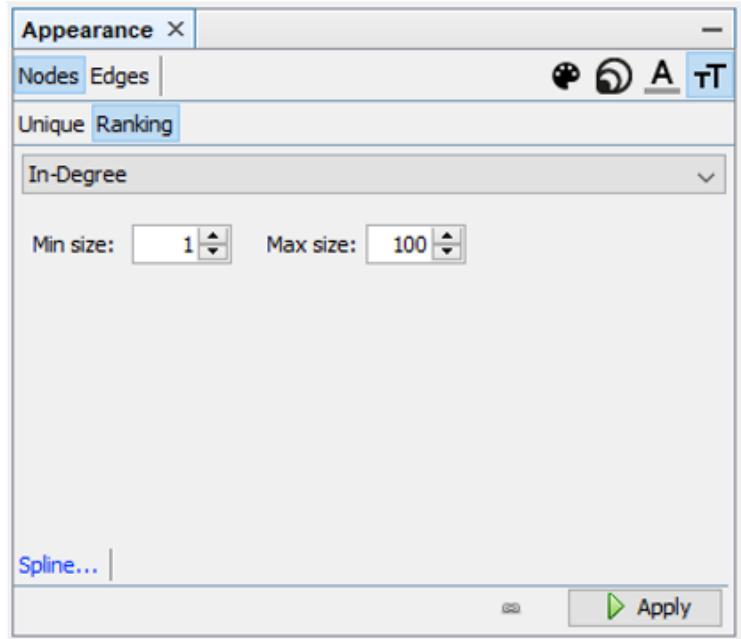
Id	Label	Interval	In-Degree	Out-Degree	Degree
justintrudeau	justintrudeau		621	1	622
andrewscheer	andrewscheer		595	0	595
thejagmeets...	thejagmeets...		416	0	416
elizabethmay	elizabethmay		289	0	289
cpc_hq	cpc_hq		251	3	254
maximebernier	maximebernier		190	0	190
yfblanchet	yfblanchet		164	0	164
liberal_party	liberal_party		153	1	154
althiaraj	althiaraj		93	0	93
ndp	ndp		91	1	92
gmbutts	gmbutts		85	4	89
fordnation	fordnation		78	0	78
canadiangre...	canadiangre...		63	1	64
cbcnews	cbcnews		62	1	63
lisalaflammeectv	lisalaflammeectv		61	0	61
susandelaco...	susandelaco...		59	2	61
rosiebarton	rosiebarton		58	0	58
torontostar	torontostar		51	0	51
jkenney	jkenney		51	0	51
globalnews	globalnews		47	5	52
davidakin	davidakin		45	10	55
ctvnews	ctvnews		43	7	50
peoplespca	peoplespca		42	0	42

Statistics are also populated in the Data Table

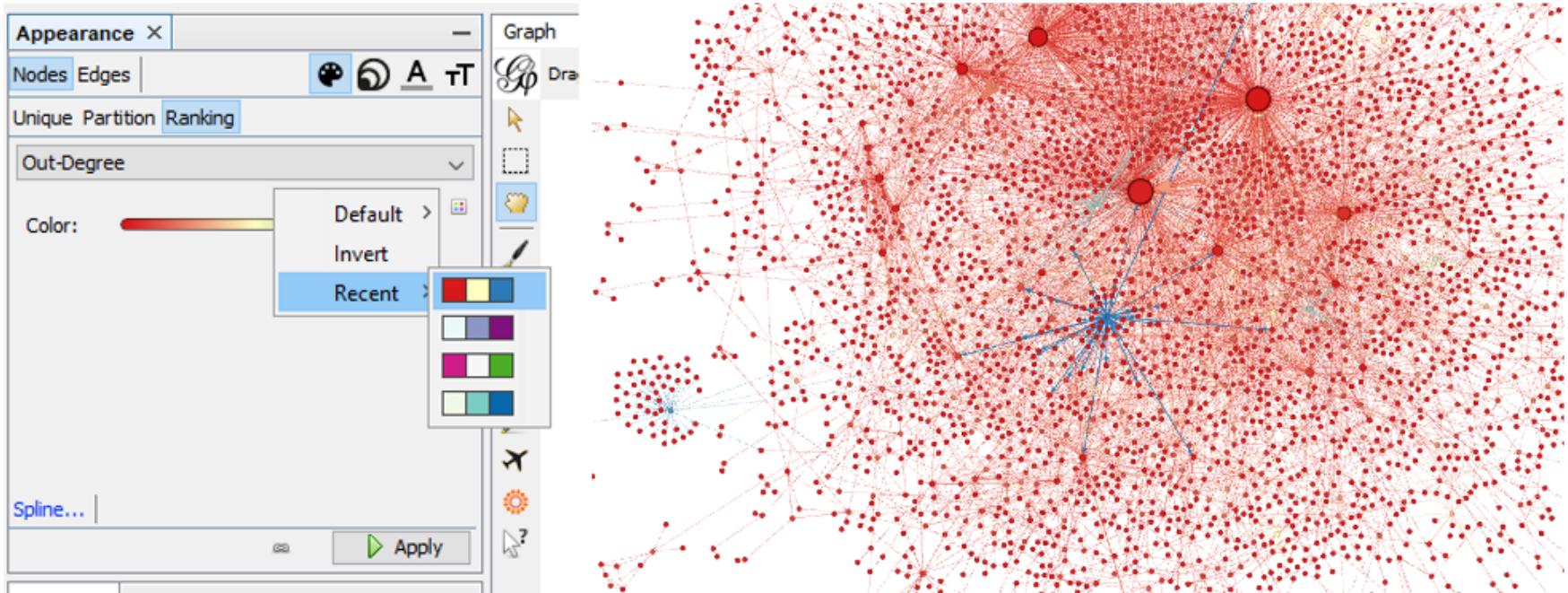


Who is being tweeted at the most?

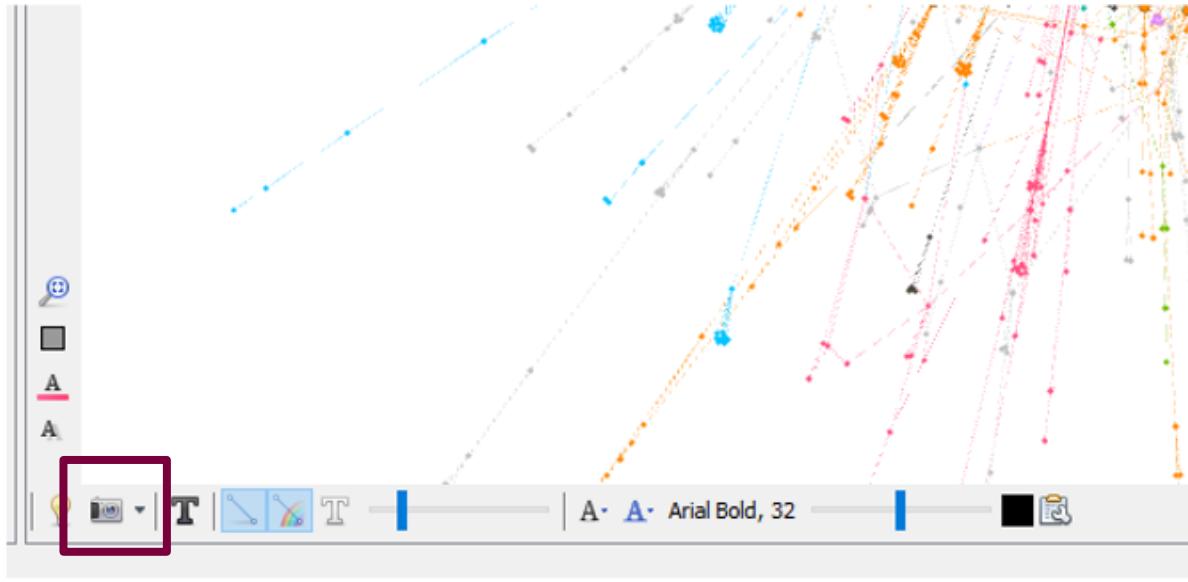




Adjust label size to identify account names



Who is doing most of the tweeting?



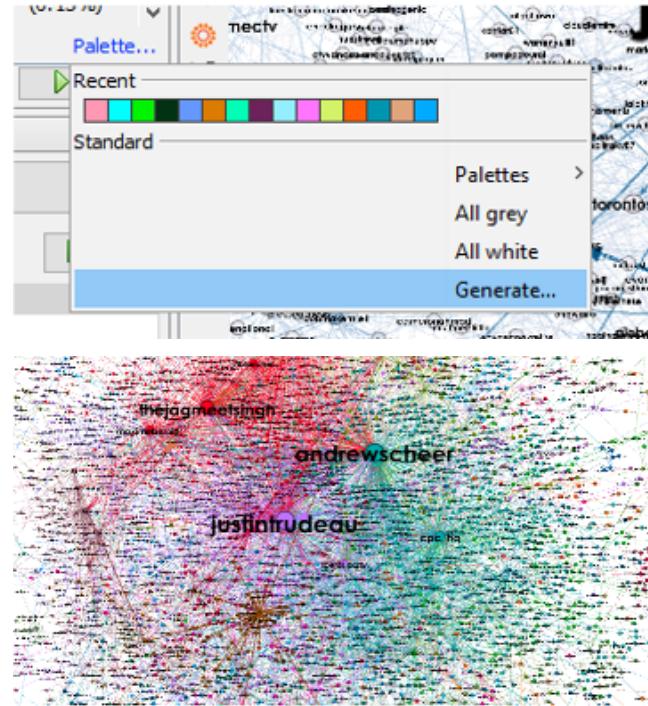
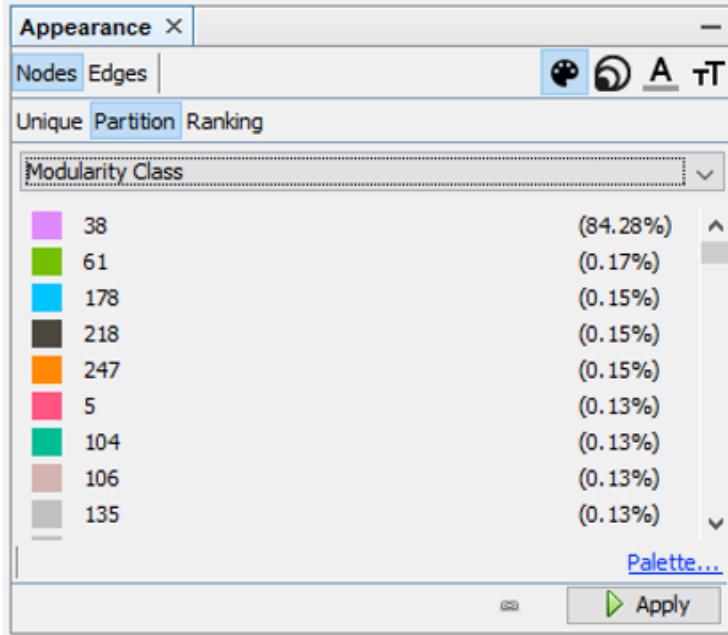
TAKE A SCREENSHOT... recall provenance



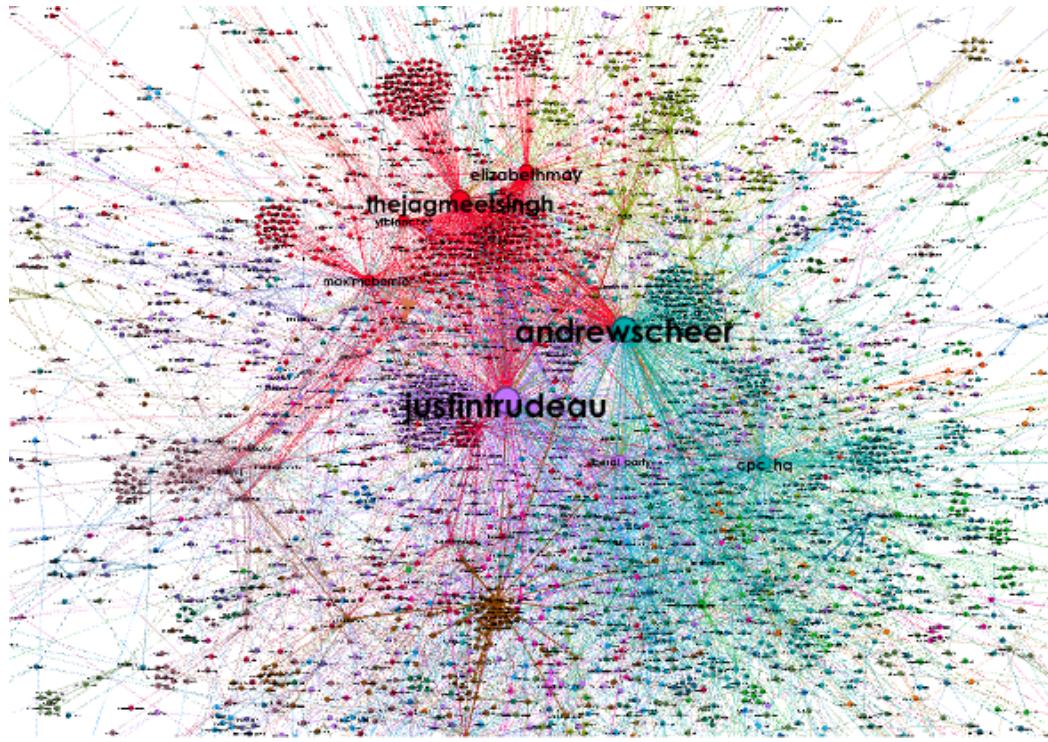
Your turn...

Start making sense of the data

- Use attributes...
 - Calculated: In-degree, out-degree, modularity, etc.
 - Supplied: followers, retweets
- ...to modify:
 - Node & edge colour
 - Node size
 - Label size & colour



Community Detection → Modularity Class



Finesse: adjust layout, font, etc.

Show Off Your Viz...

Export as graphic to preserve layout
(consider privacy, though!)

- PDF
- PNG
- SVG

Thanks for coming!

Questions: mordelld@mcmaster.ca