

# An Introduction to Social Network Analysis With Gephi

January 20, 2021

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/](https://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>

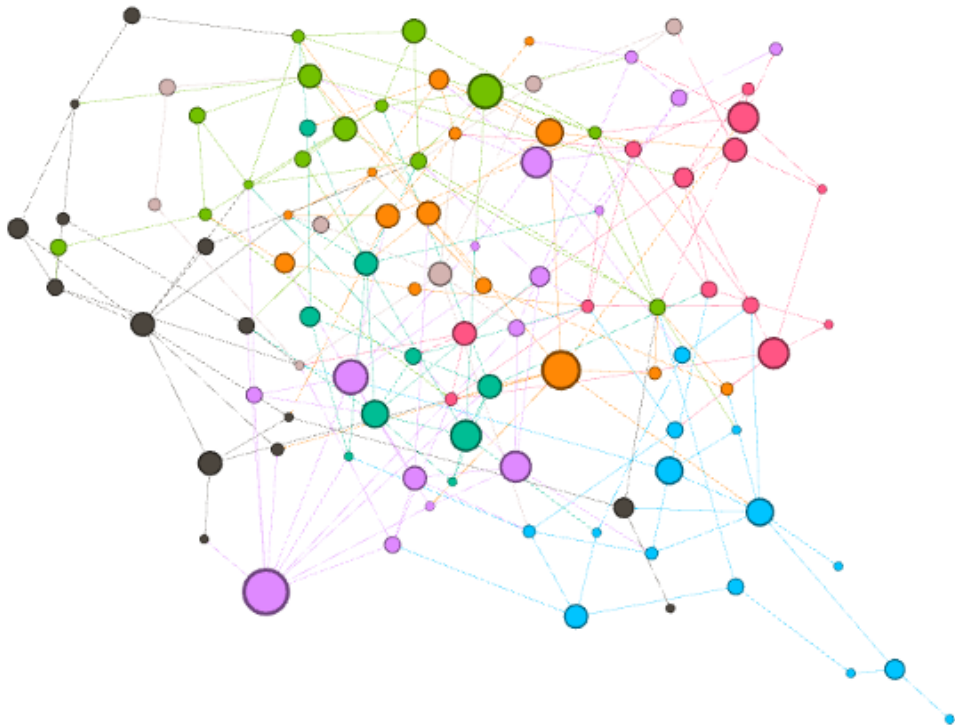
# On Social Network Analysis

*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



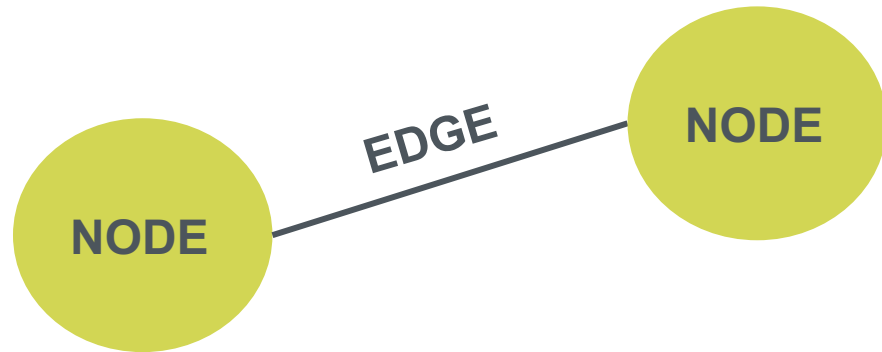


# Describing Relationships

**Node:** the 'actor' in the network

**Edge:** the relationship connecting actors

**Attribute:** features of the node or edge



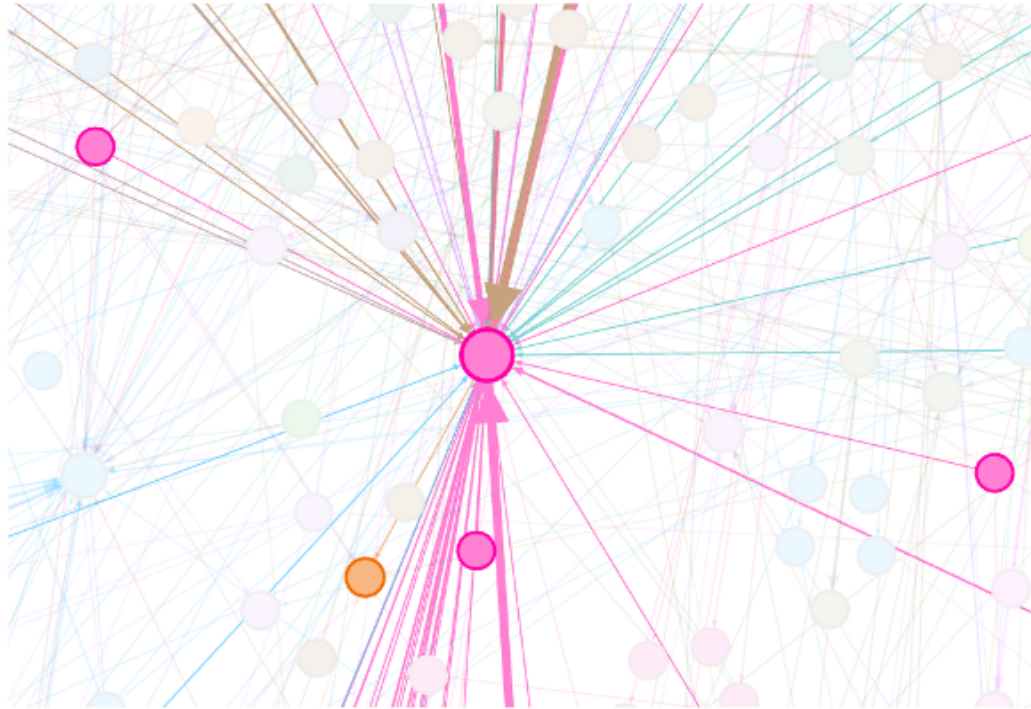
## Other terminology

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

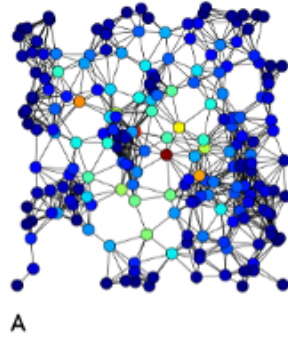
---



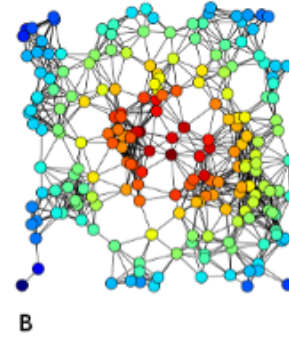
## Determining which nodes are the most important in the cluster or graph...

- **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

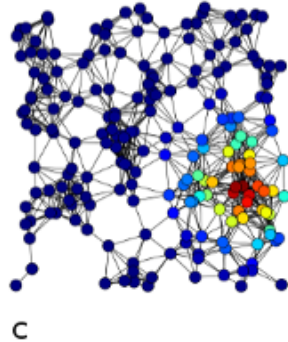
A. Betweenness



B. Closeness



C. Eigenvector



D. Degree

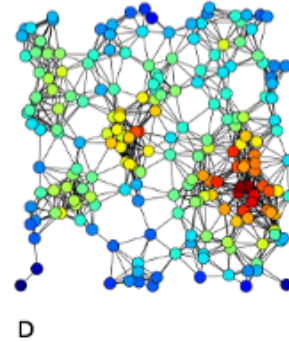
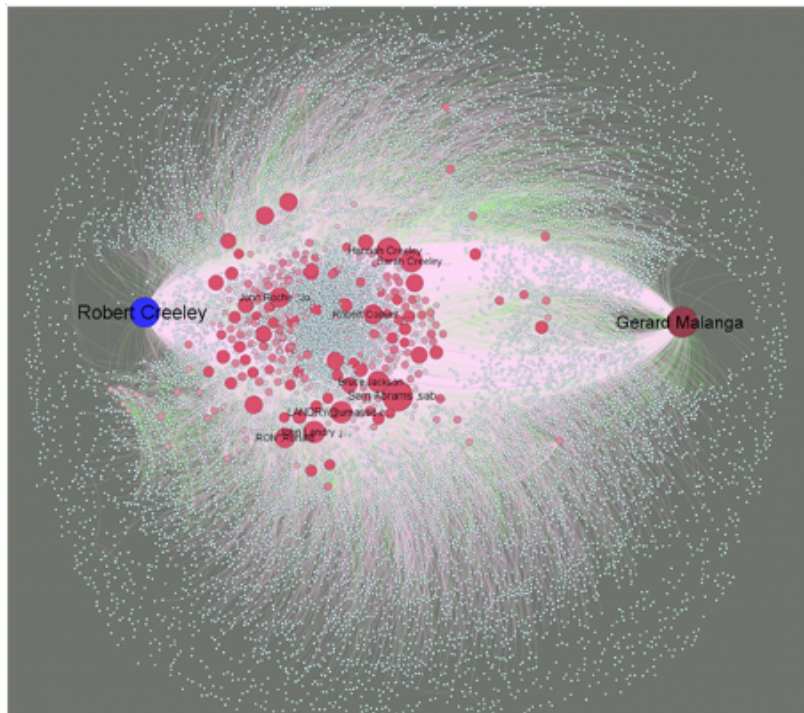


Image: [Tapiocozzo](#) (CC 4.0 BY-SA)

# 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)

---

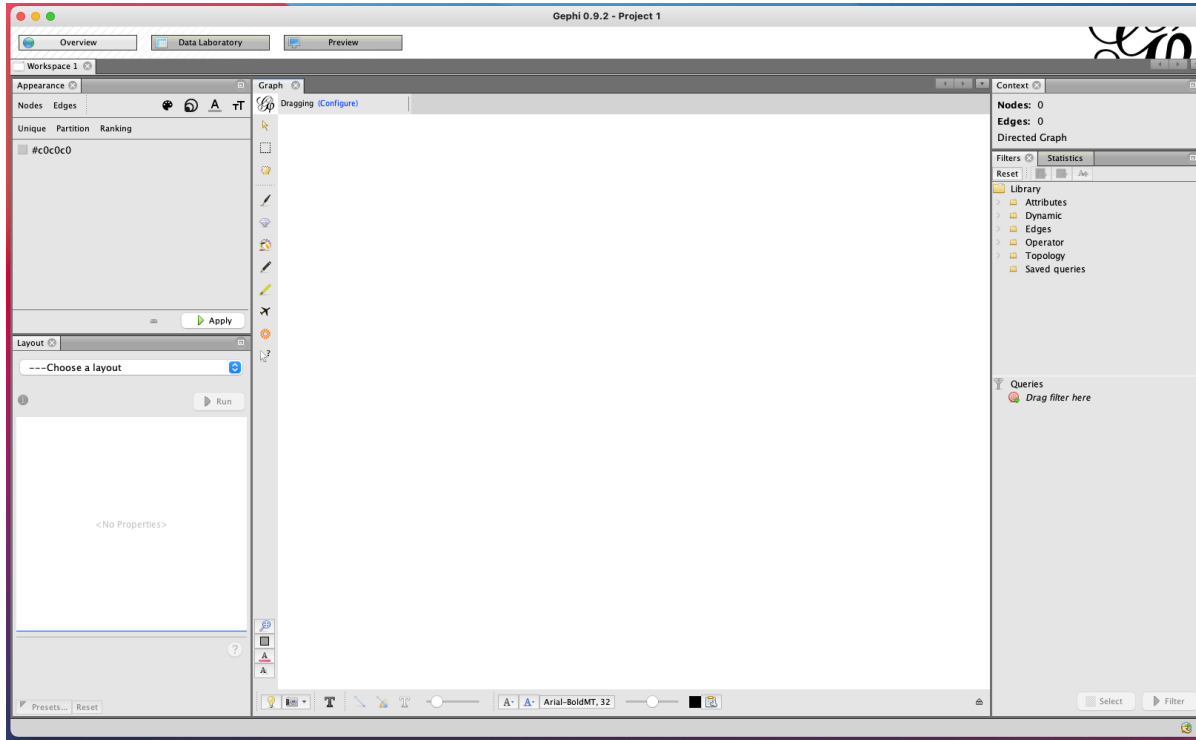




# The dataset

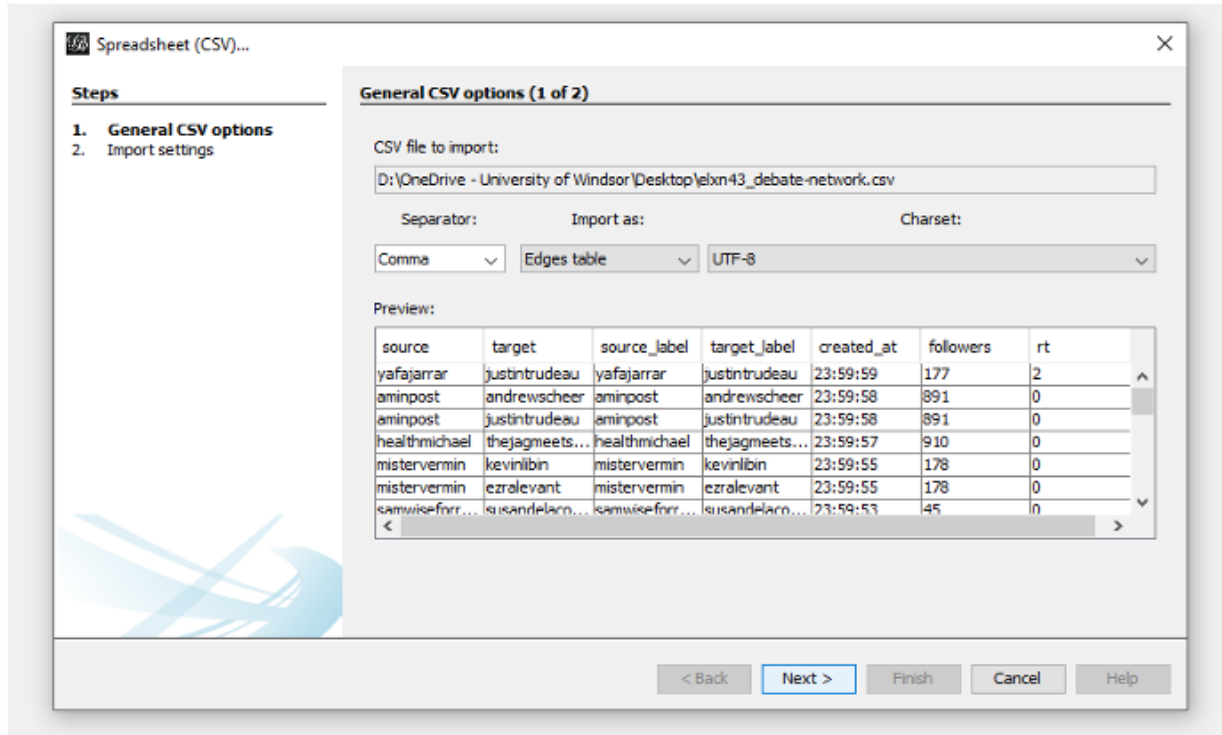
Download the [#elxn43 debate night dataset](#)

- 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

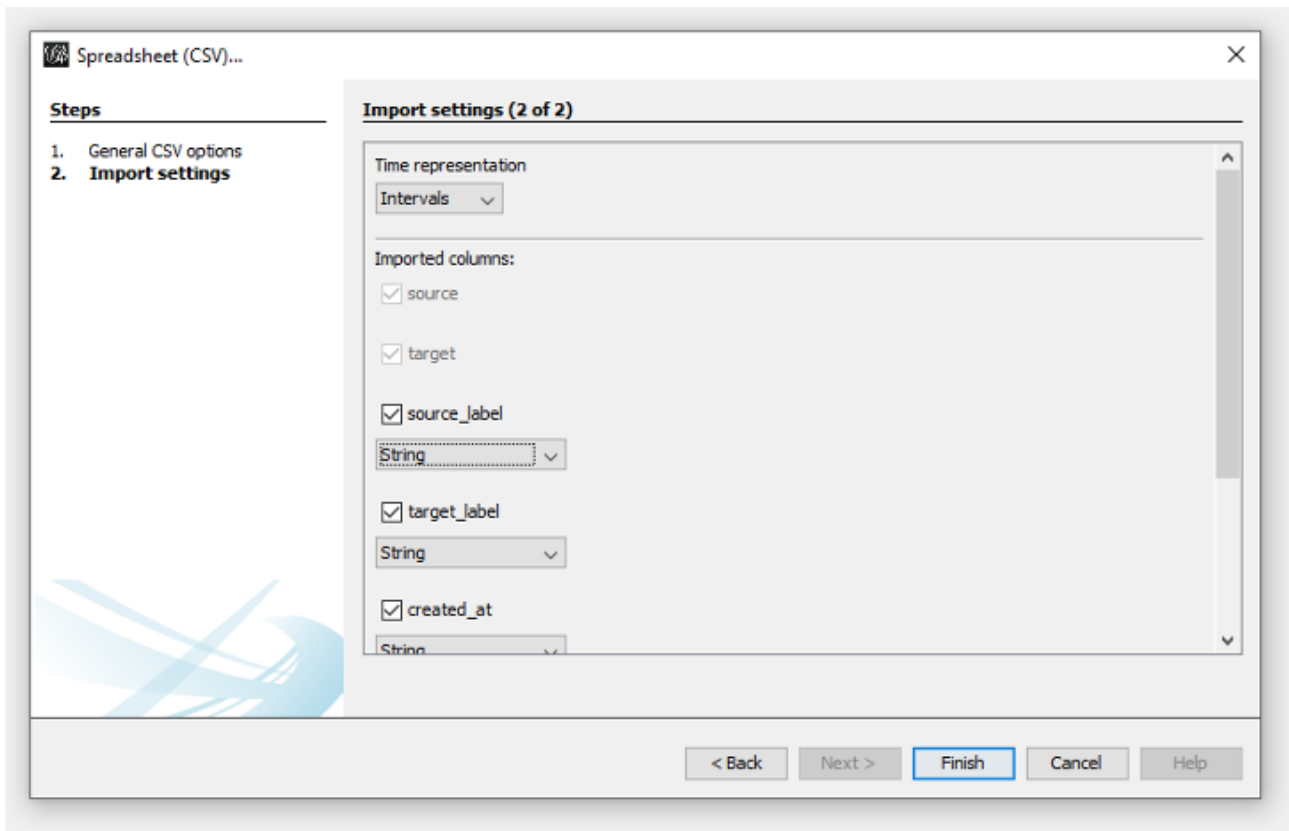


## Gephi > New Project

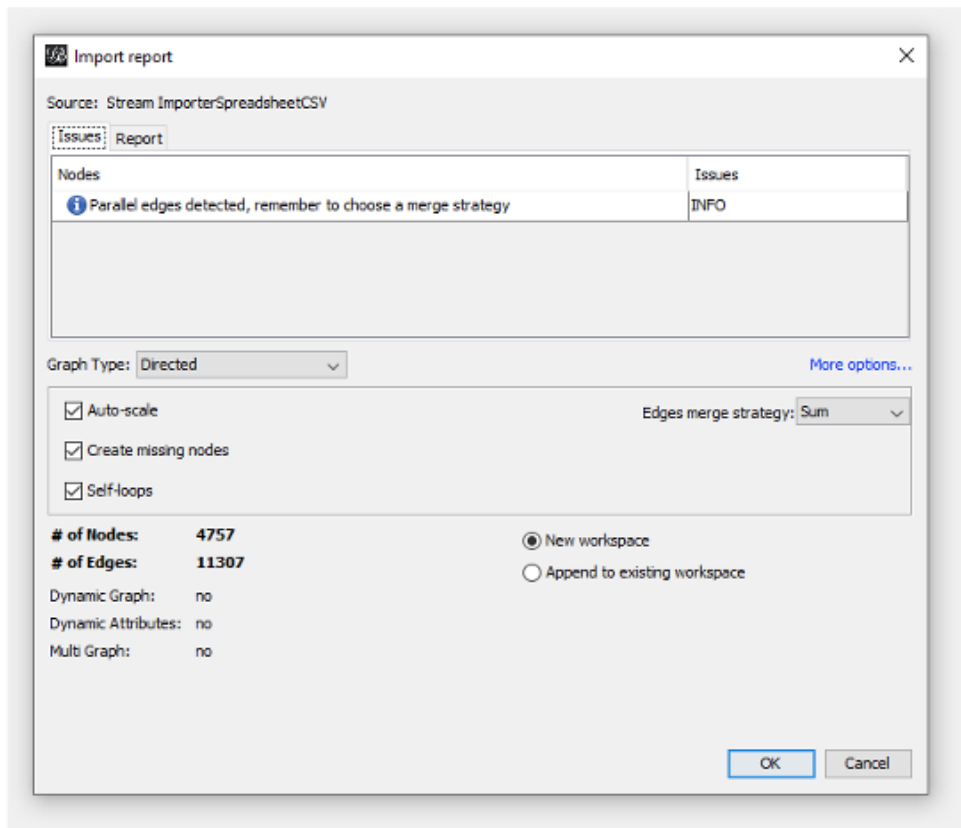
---



Import dataset (CSV) as: Edges table



**Note:** name columns “source” and “target”

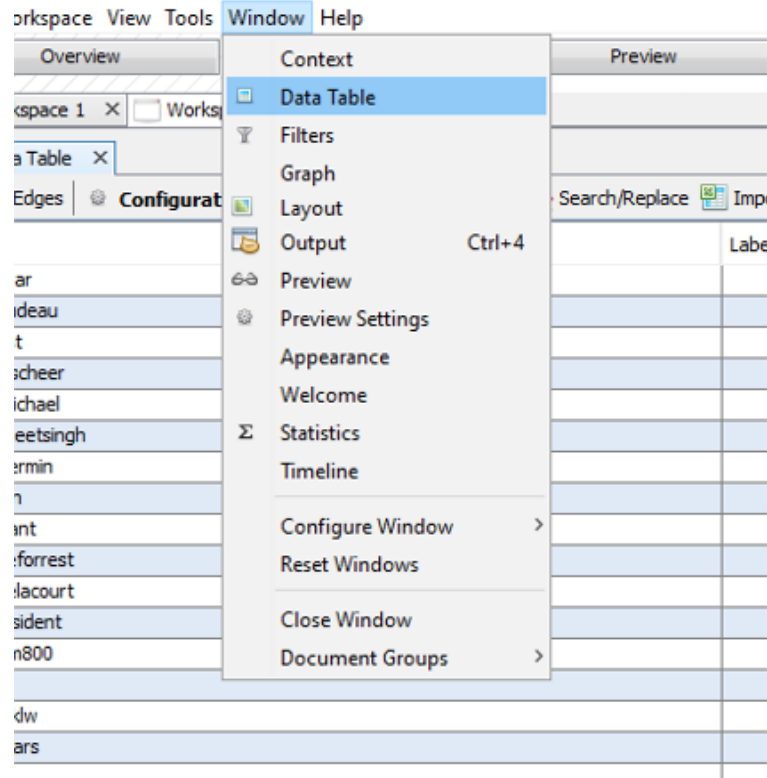


Almost there... directed graph?

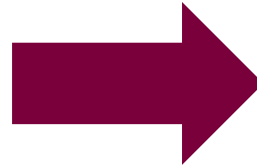
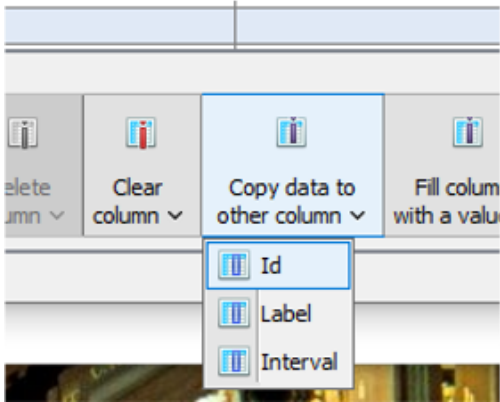


Overview → HAIRBALL!!!!

---



## Data Laboratory > Data Table view



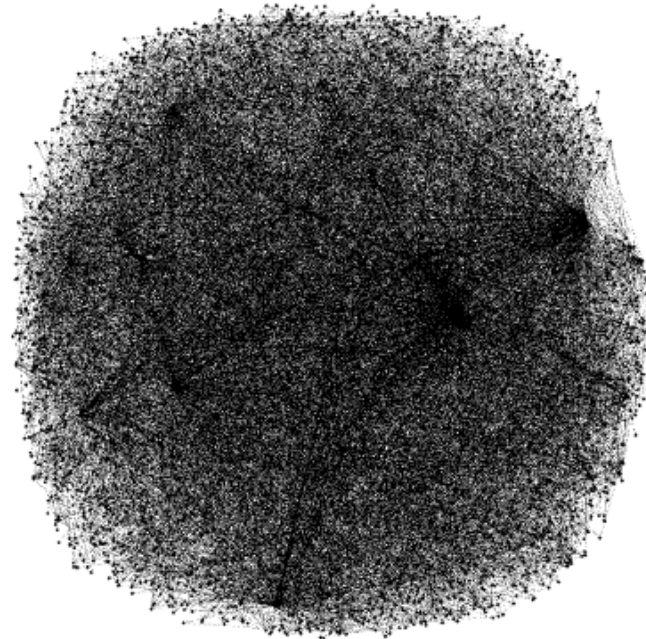
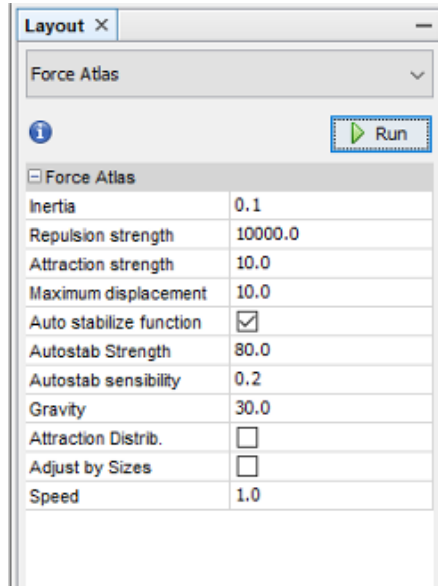
Data Table		
Nodes	Edges	Configuratio
Id	Label	Inte
yafajarrar	yafajarrar	
justintrudeau	justintrudeau	
aminpost	aminpost	
andrewscheer	andrewscheer	
healthmichael	healthmichael	
thejagmeets...	thejagmeets...	
mistervermin	mistervermin	
kevinlibin	kevinlibin	
ezrelevant	ezrelevant	
samwiseforr...	samwiseforr...	
susandelaco...	susandelaco...	
cfnpresident	cfnpresident	
baconam800	baconam800	
ctv	ctv	
am800ckdw	am800ckdw	
girlfrmmars	girlfrmmars	
rgl007	rgl007	
elizabethmay	elizabethmay	
yfblanchet	yfblanchet	
globalnews	globalnews	

Copy data from Id to Label (in Node view)



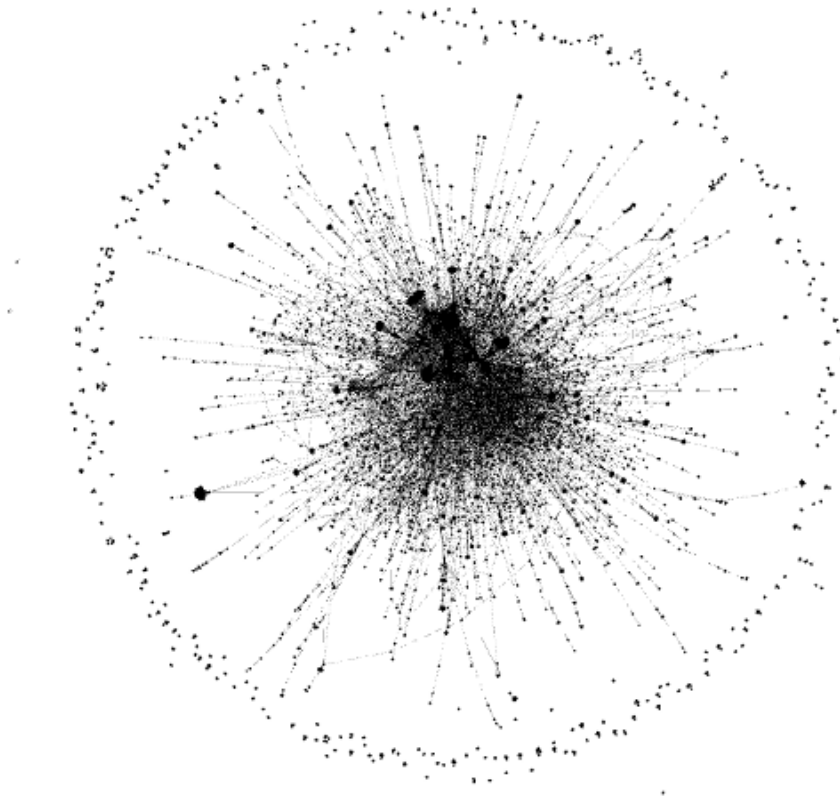
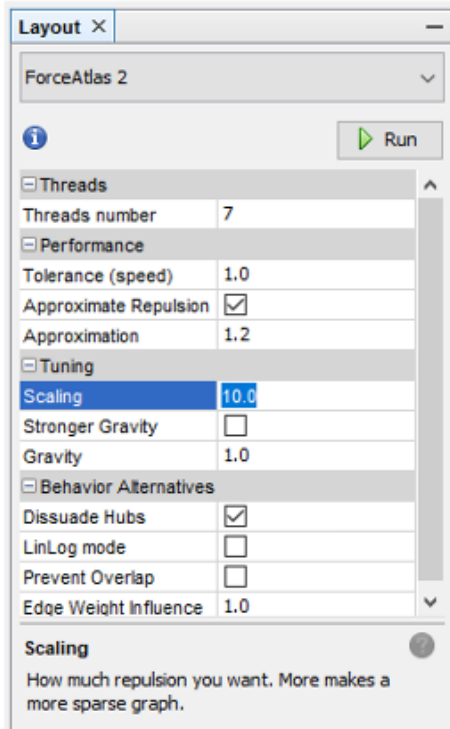
Source	Target	Type	Id	Label	Interval	Weight	followers	rt
yefajamar	justin Trudeau	Directed	0			3.0	177	0
amipost	andrewscheer	Directed	1			1.0	891	0
amipost	justin Trudeau	Directed	2			1.0	891	0
healthchannel	thejagmeetsingh	Directed	3			1.0	910	0
mistervermin	kevinblair	Directed	4			1.0	178	0
mistervermin	sarahelavirt	Directed	5			1.0	178	0
sarahelavirt	susandelaacourt	Directed	6			1.0	45	0
cfupresident	andrewscheer	Directed	7			4.0	4190	6
bacanam800	clv	Directed	8			1.0	1570	0
bacanam800	am800dhw	Directed	9			1.0	1520	0
griffmairs	andrewscheer	Directed	10			1.0	174	0
rg107	thejagmeetsingh	Directed	11			1.0	50	2
rg107	klizabithenay	Directed	12			1.0	50	2
rg107	yfbianchet	Directed	13			1.0	50	2
rg107	justin Trudeau	Directed	14			1.0	50	2
rg107	andrewscheer	Directed	15			1.0	50	2
globalnews	alhilara	Directed	16			7.0	358424	1
globalnews	susandelaacourt	Directed	17			1.0	358419	4
emile_widham	thejagmeetsingh	Directed	18			1.0	147	0
emile_widham	andrewscheer	Directed	19			2.0	147	0
vandongene	thejagmeetsingh	Directed	20			1.0	2330	3
clngaman	thejagmeetsingh	Directed	21			1.0	2	2
clngaman	klizabithenay	Directed	22			1.0	2	2
clngaman	yfbianchet	Directed	23			1.0	2	2
clngaman	justin Trudeau	Directed	24			1.0	2	2
clngaman	andrewscheer	Directed	25			1.0	2	2
mbrianodgers	andrewscheer	Directed	26			1.0	20	0
opencoe	danyalraza	Directed	27			2.0	3899	2
angemedc	thejagmeetsingh	Directed	28			1.0	58	0
the_terraviva	justin Trudeau	Directed	29			1.0	286	0
tzqwanuhina	thejagmeetsingh	Directed	30			1.0	838	0
sofelindgrly	justin Trudeau	Directed	31			1.0	18	0

What does 'weight' refer to...?

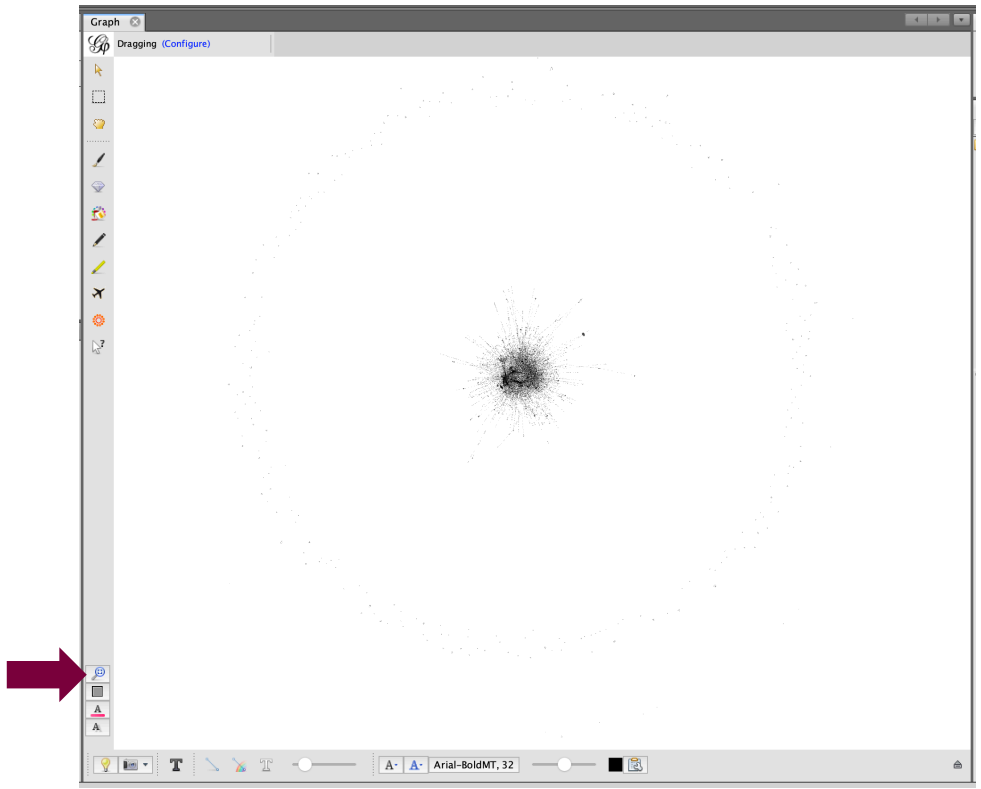


## Layout (begin with Force Atlas)

---

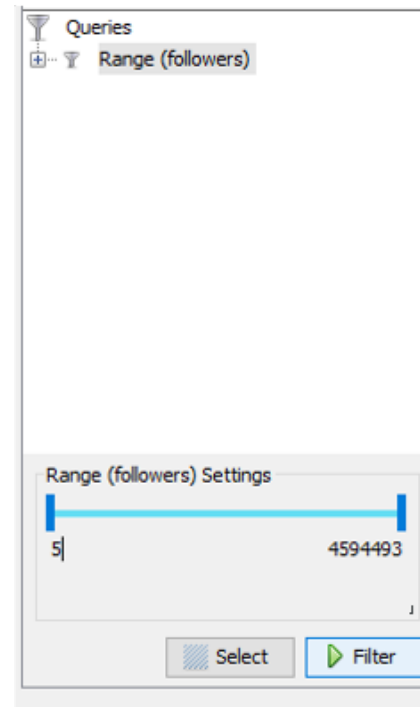
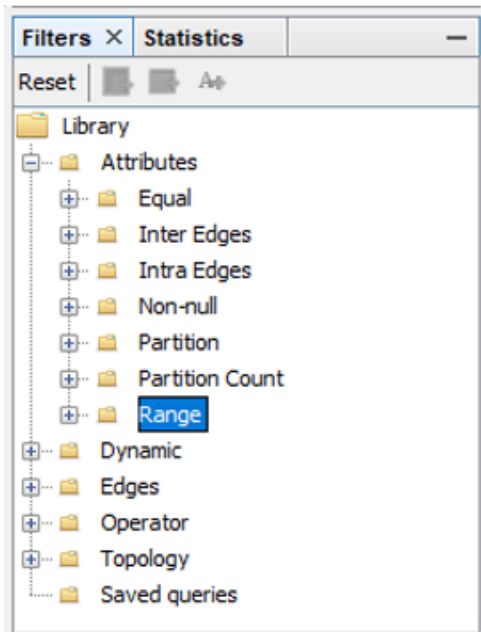


Try ForceAtlas 2...



## Getting around the canvas

---



**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

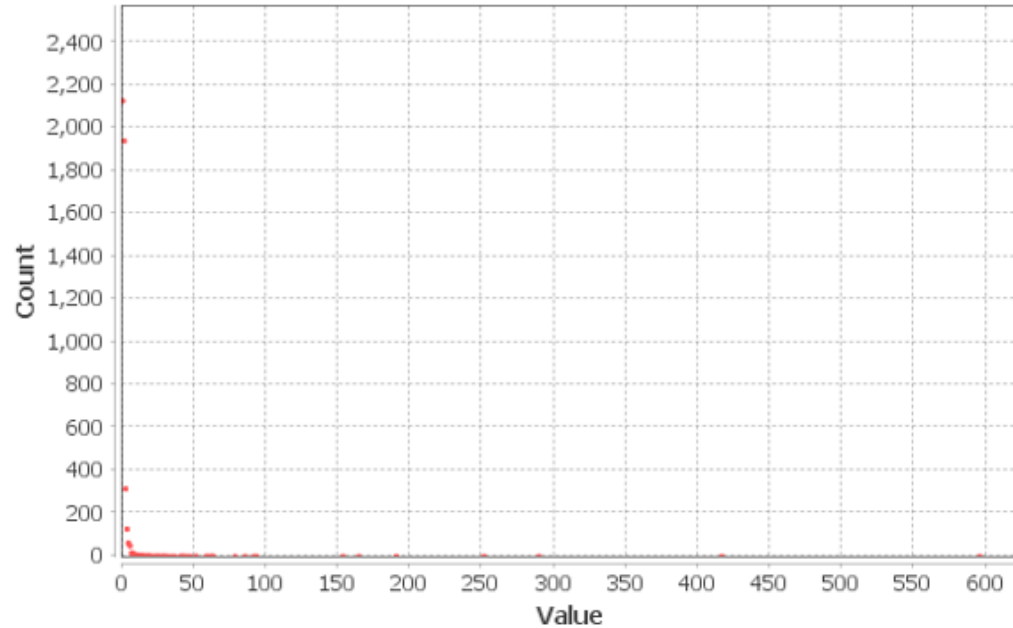
Avg. Clustering Coefficient Run ●

Eigenvector Centrality Run ●

Edge Overview

Avg. Path Length Run ●

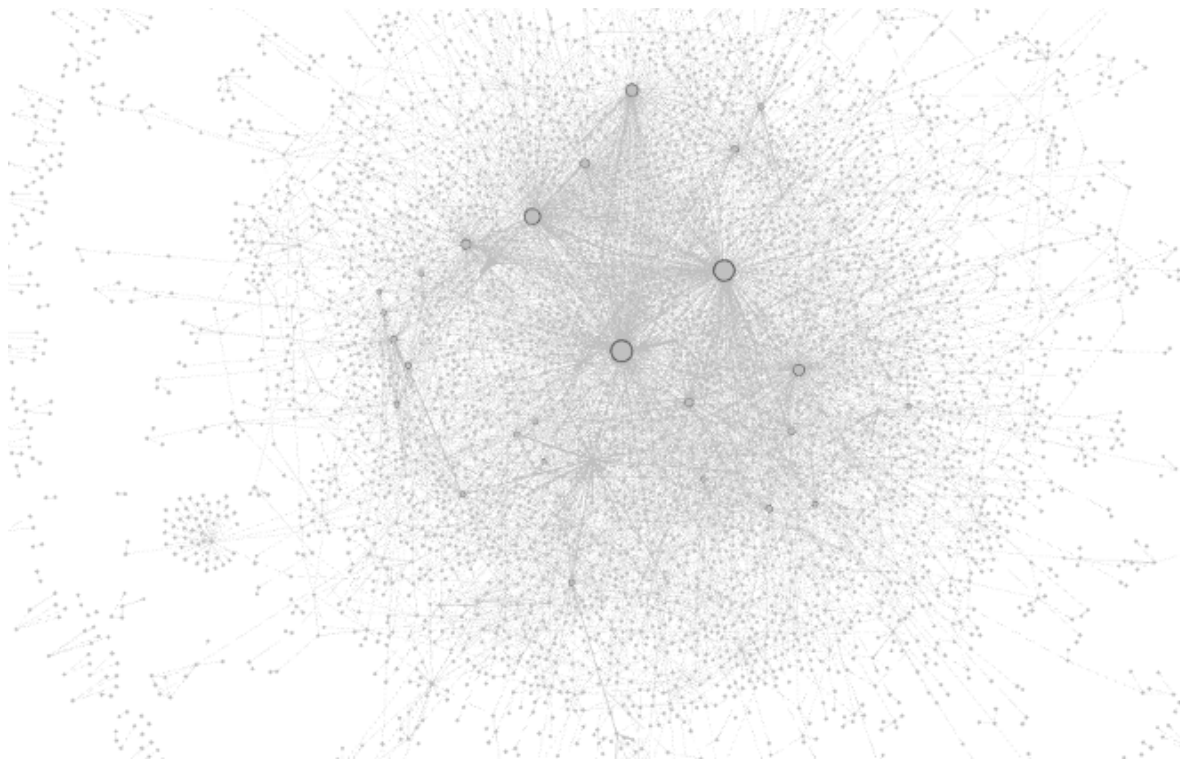
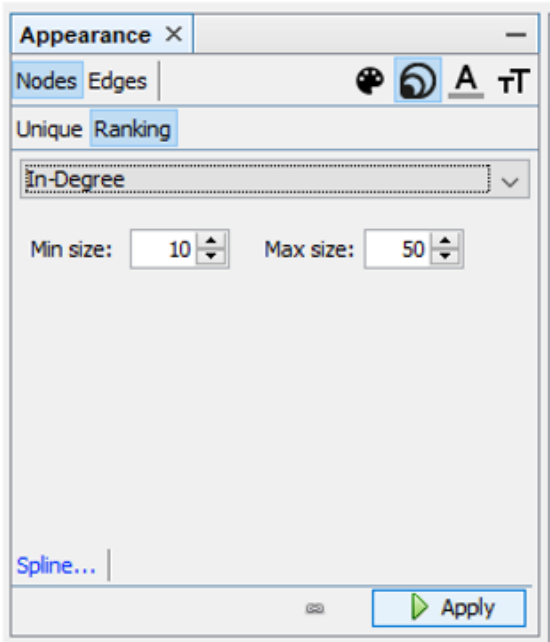
## In-Degree Distribution



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
lisalafiamectv	lisalafiamectv		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
dauidakin	dauidakin		45	10	55
ctvnews	ctvnews		43	7	50
peoplespca	peoplespca		42	0	42

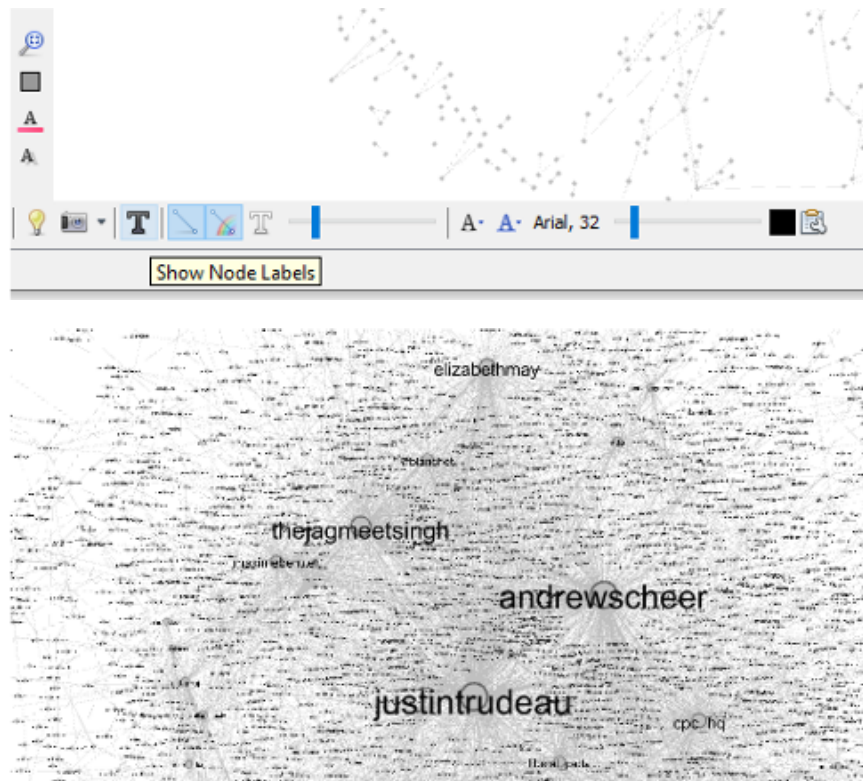
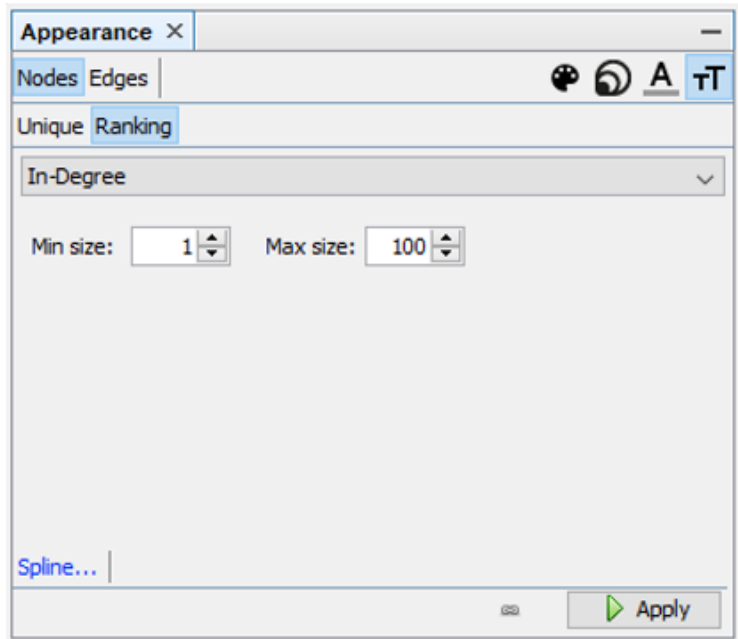
Statistics are also populated in the Data Table



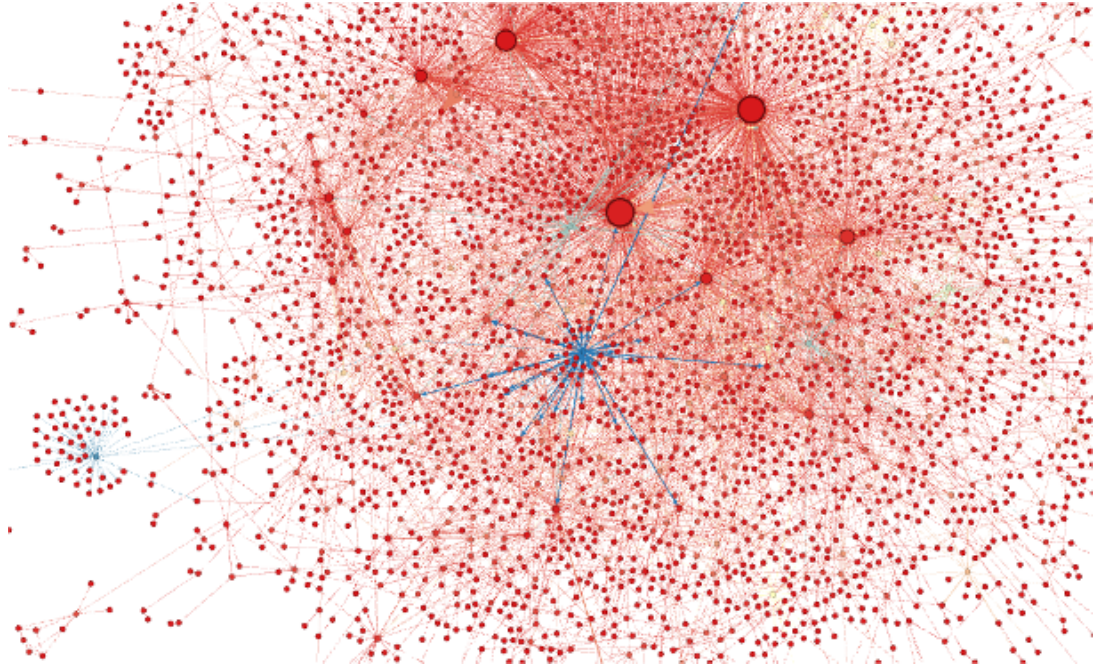
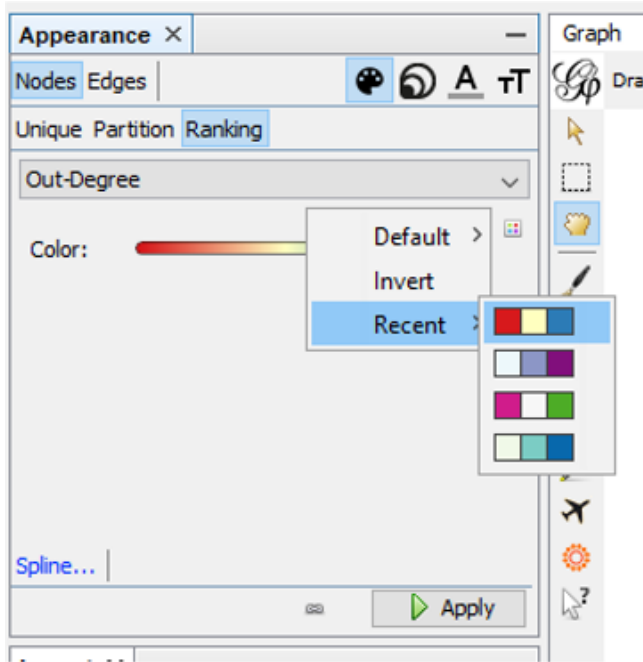
Who is being tweeted *at* the most?





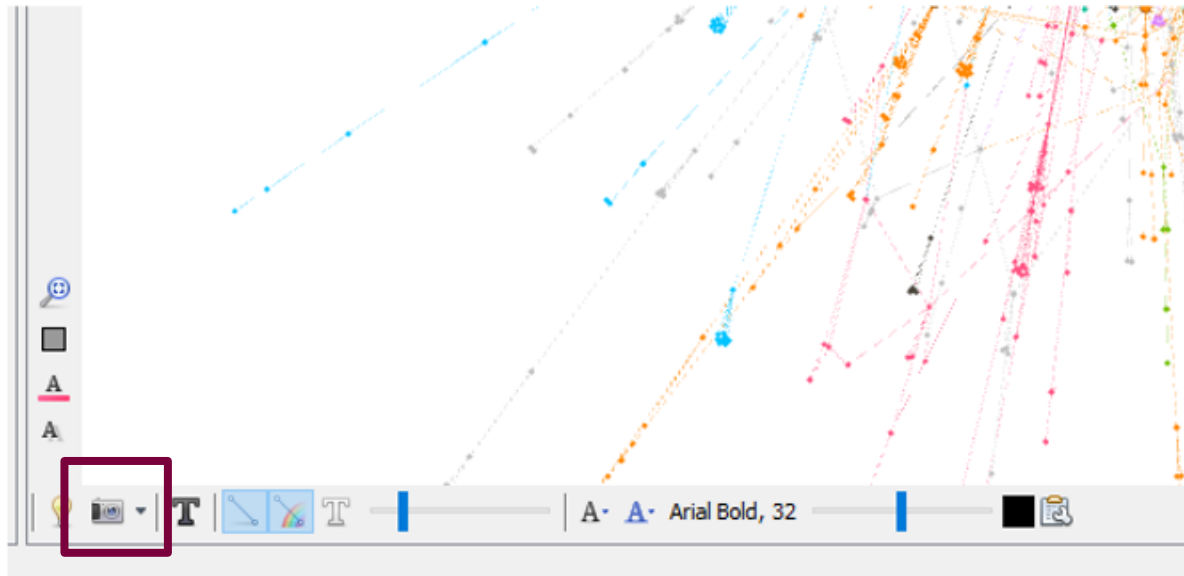


Adjust label size to identify account names



Who is doing most of the tweeting?

---



TAKE A SCREENSHOT... recall **provenance**

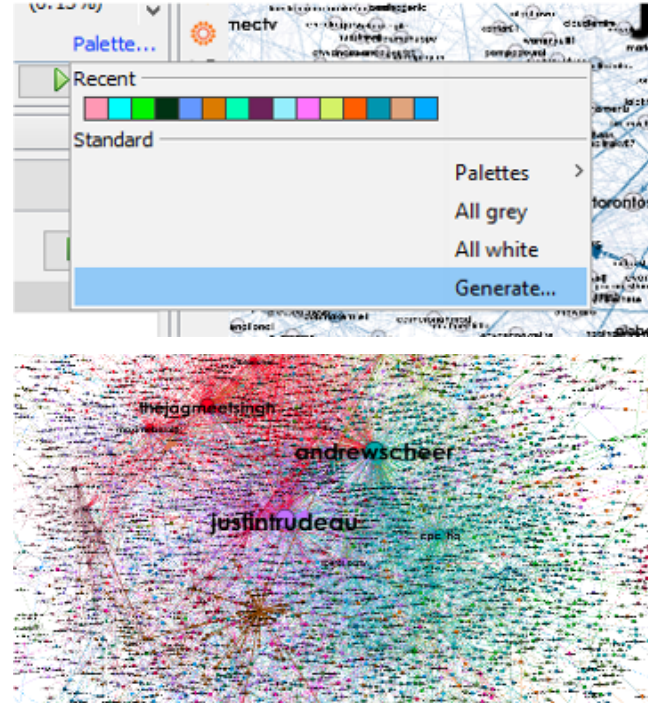
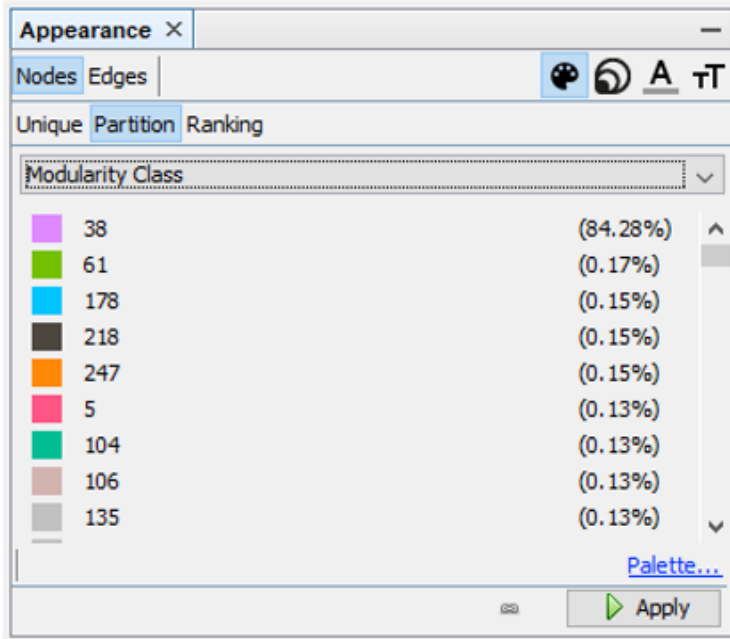
---



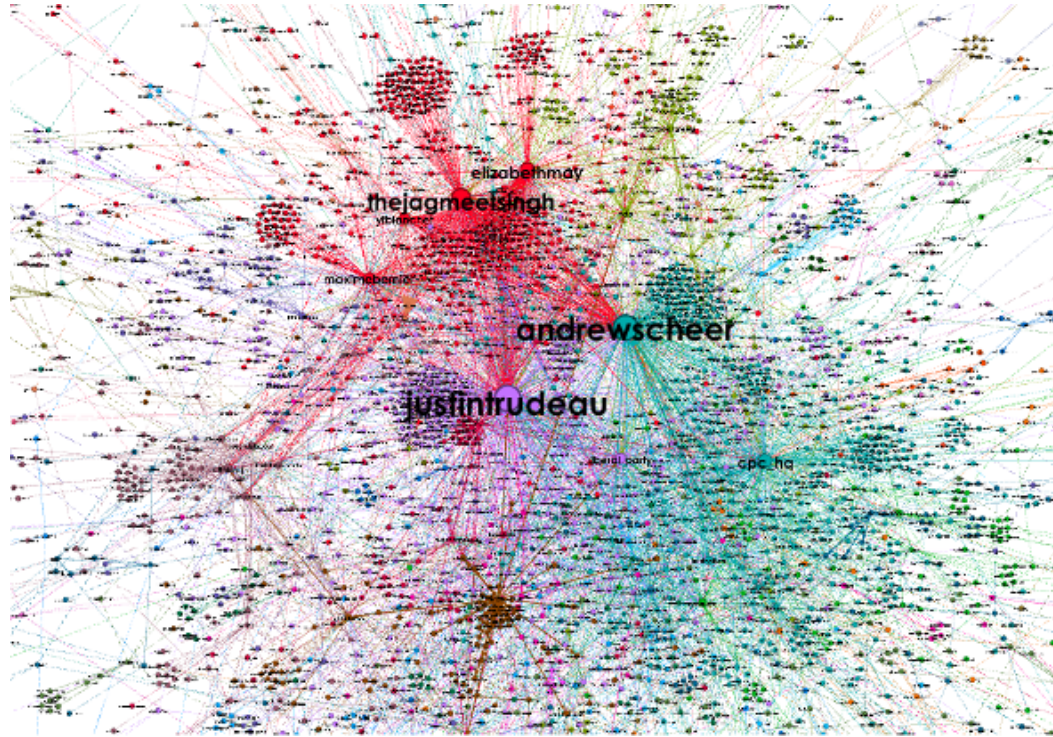
Your turn...

## 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](mailto:mordelld@mcmaster.ca)