

<--data visualization for the humanities-->

# INTRODUCTION TO SOCIAL NETWORK ANALYSIS WITH GEPHI

---

March 5, 2020

Devon Mordell, Digital Scholarship and Archiving Librarian





## 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

Follow along at <https://bit.ly/leddylib-gephi>

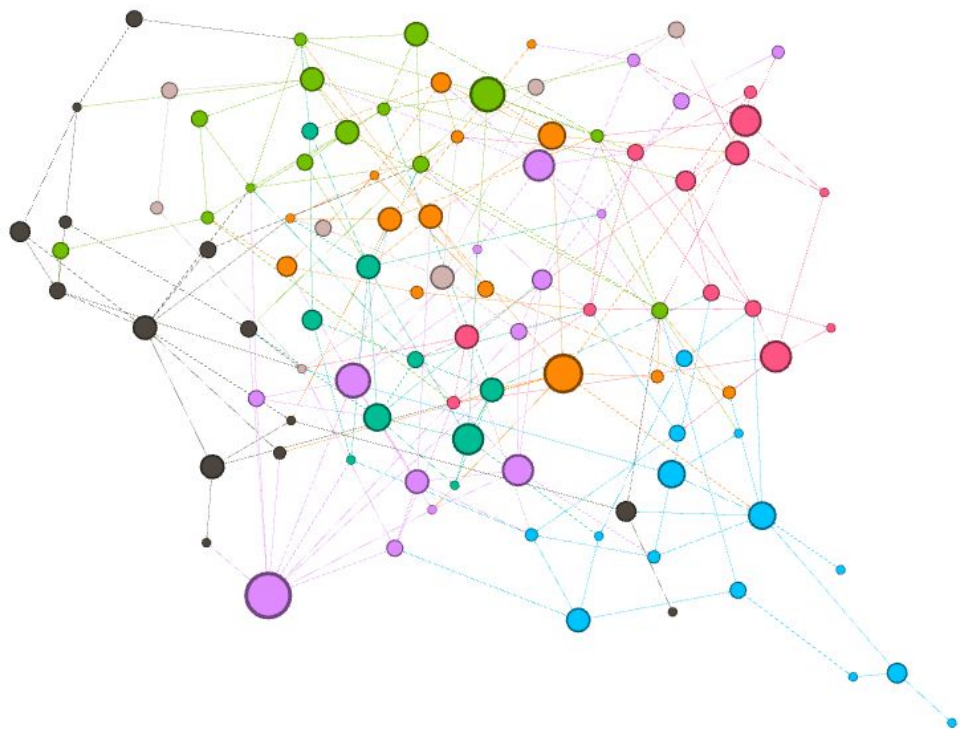


# ON SOCIAL NETWORK ANALYSIS

\*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

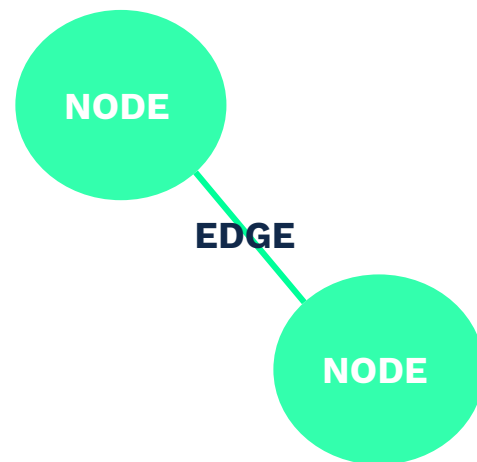


# DESCRIBING RELATIONSHIPS

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

**Edge:** the relationship connecting actors

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





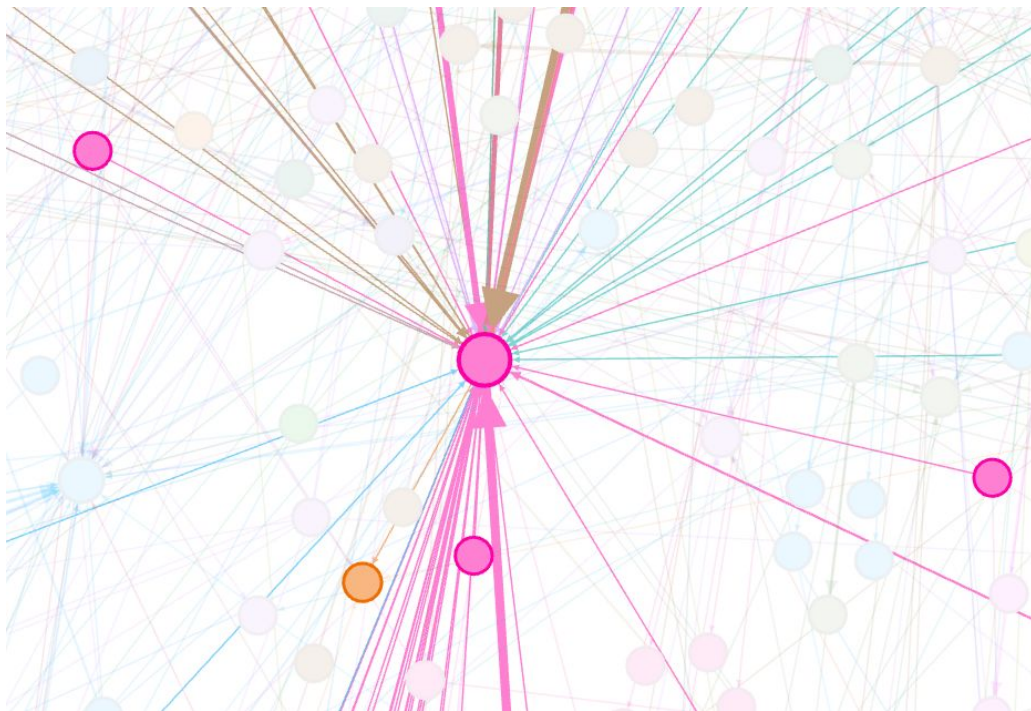
# OTHER TERMINOLOGY

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

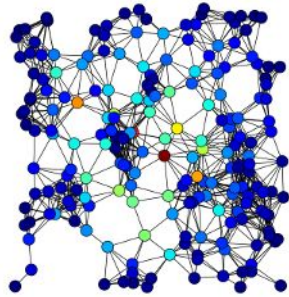


# CENTRALITY

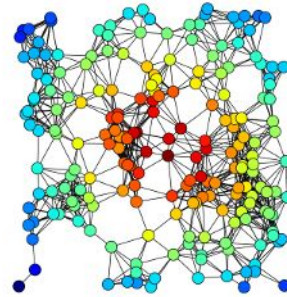
I.e. which nodes are the most important in the cluster or graph?

- 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

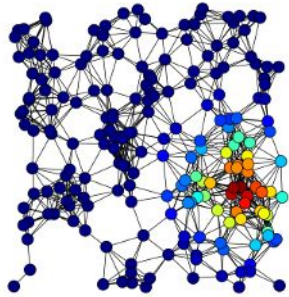




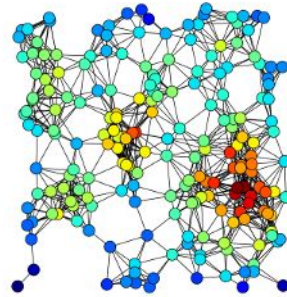
A



B



C



D

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

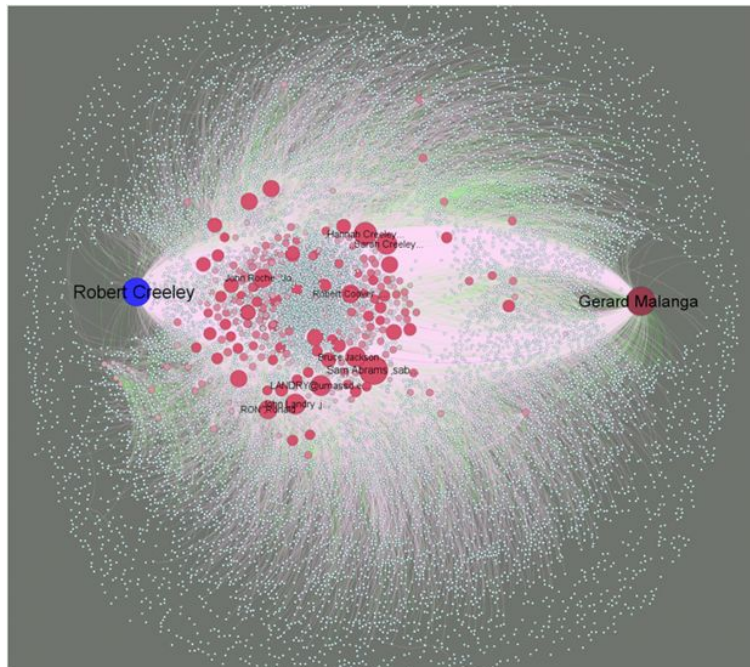
**Centrality:** A. Betweenness; B. Closeness; C. Eigenvector; D. Degree.



# 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's Email Archive** at Stanford University



# THE DATASET

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

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

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

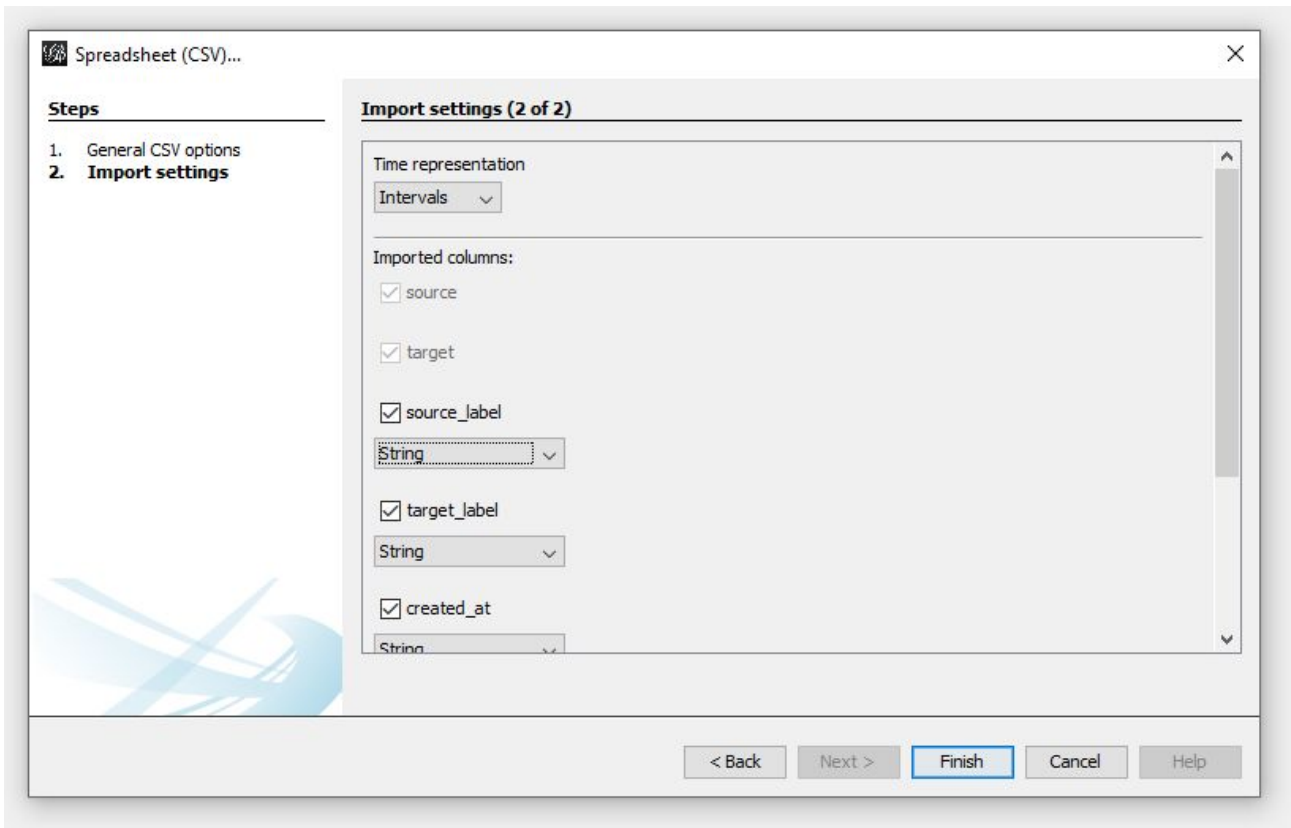
Separator: Comma Import as: Edges table Charset: UTF-8

Preview:

source	target	source_label	target_label	created_at	followers	rt
yafajarrar	justintrudeau	yafajarrar	justintrudeau	23:59:59	177	2
aminpost	andrewscheer	aminpost	andrewscheer	23:59:58	891	0
aminpost	justintrudeau	aminpost	justintrudeau	23:59:58	891	0
healthmichael	thejagmeets...	healthmichael	thejagmeets...	23:59:57	910	0
mistervermin	kevinlibin	mistervermin	kevinlibin	23:59:55	178	0
mistervermin	ezrelevant	mistervermin	ezrelevant	23:59:55	178	0
samwise forr...	isandelaro...	samwise forr...	isandelaro...	23:59:53	45	0

< Back Next > Finish Cancel Help

Import dataset (CSV) as: Edges table



(Gephi recognizes “source” and “target” columns as nodes, & creates edges)

Import report

Source: Stream ImporterSpreadsheetCSV

Issues Report

Nodes	Issues
<i>i</i> Parallel edges detected, remember to choose a merge strategy	INFO

Graph Type: Directed [More options...](#)

☒ Auto-scale ☐ Create missing nodes ☒ Self-loops

Edges merge strategy: Sum

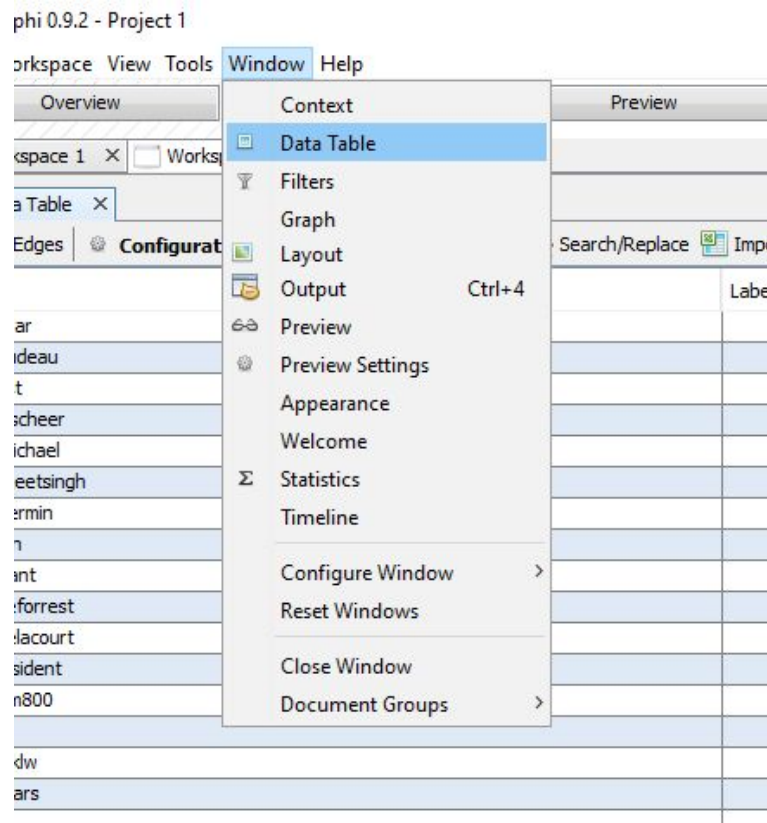
# of Nodes: 4757  
# of Edges: 11307

Dynamic Graph: no  
Dynamic Attributes: no  
Multi Graph: no

☒ New workspace  
☐ Append to existing workspace

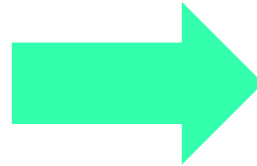
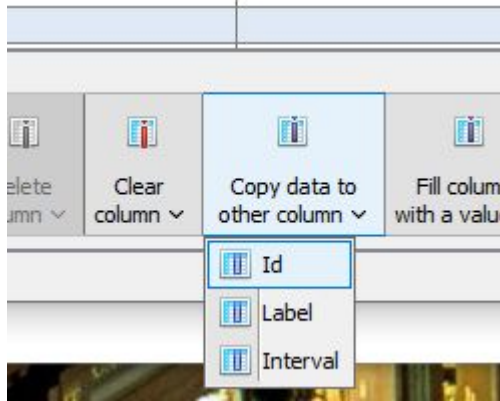
OK Cancel

Verify graph type is “Directed” and number of edges



Open up Data Laboratory > **Data Table** view if not already visible



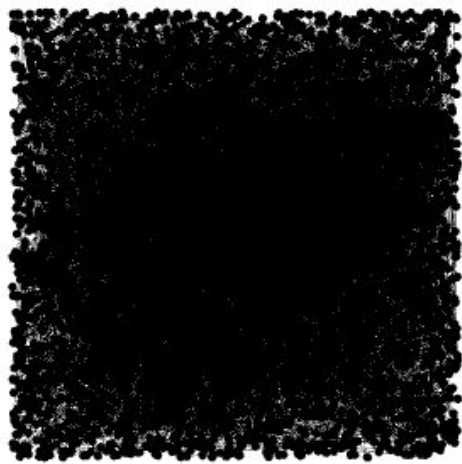


Data Table		
Nodes		
Configuration		
Id	Label	Inte
yafajarrar	yafajarrar	
justintrudeau	justintrudeau	
aminpost	aminpost	
andrewscheer	andrewscheer	
healthmichael	healthmichael	
thejagmeets...	thejagmeets...	
mistervermin	mistervermin	
kevinlibin	kevinlibin	
ezrelevant	ezrelevant	
samwiseforr...	samwiseforr...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

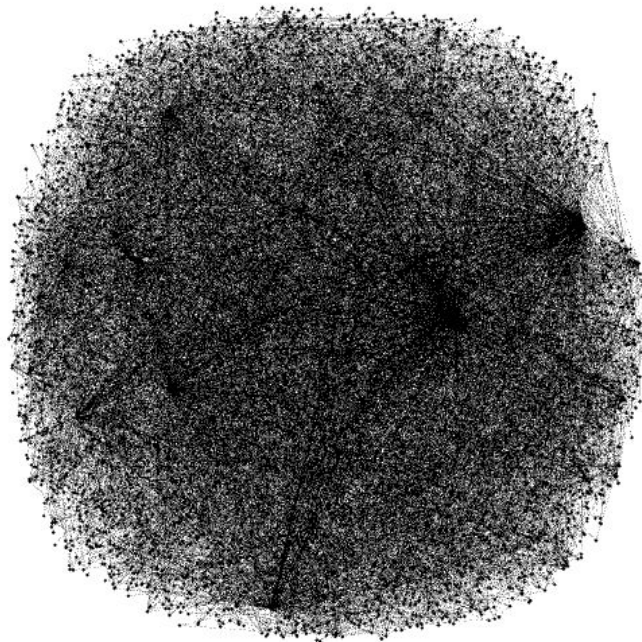
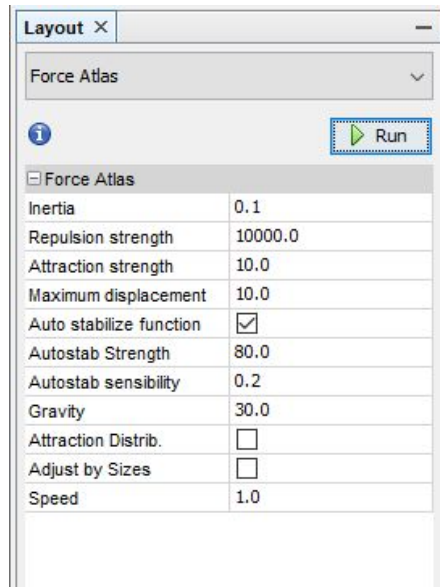
Create labels for the nodes by copying data from Id to Label (Node view)

Source	Target	Type	Id	Label	Interval	Weight	followers	rt
yafajarrar	justintrudeau	Directed	0			3.0	177	0
aminpost	andrewscheer	Directed	1			1.0	891	0
aminpost	justintrudeau	Directed	2			1.0	891	0
healthmichael	thejagmeetsingh	Directed	3			1.0	910	0
mistervtermin	kevinlibin	Directed	4			1.0	178	0
mistervtermin	ezralevant	Directed	5			1.0	178	0
samwise Forrest	susandelacourt	Directed	6			1.0	45	0
cfnpresident	andrewscheer	Directed	7			4.0	4190	6
baconam800	ctv	Directed	8			1.0	1520	0
baconam800	am800dlw	Directed	9			1.0	1520	0
griffmarmars	andrewscheer	Directed	10			1.0	574	0
rgl007	thejagmeetsingh	Directed	11			1.0	50	2
rgl007	elizabethmay	Directed	12			1.0	50	2
rgl007	yfblandhet	Directed	13			1.0	50	2
rgl007	justintrudeau	Directed	14			1.0	50	2
rgl007	andrewscheer	Directed	15			1.0	50	2
globalnews	althiaraj	Directed	16			7.0	358424	1
globalnews	susandelacourt	Directed	17			1.0	358419	4
emile_wickham	thejagmeetsingh	Directed	18			1.0	147	0
emile_wickham	andrewscheer	Directed	19			2.0	147	0
vandongene	thejagmeetsingh	Directed	20			1.0	2330	3
clingaman	thejagmeetsingh	Directed	21			1.0	2	2
clingaman	elizabethmay	Directed	22			1.0	2	2
clingaman	yfblandhet	Directed	23			1.0	2	2
clingaman	justintrudeau	Directed	24			1.0	2	2
clingaman	andrewscheer	Directed	25			1.0	2	2
mrbrianrodgers	andrewscheer	Directed	26			1.0	20	0
openice	danyaalraza	Directed	27			2.0	3899	2
angiemdec	thejagmeetsingh	Directed	28			1.0	58	0
the_terrorist	justintrudeau	Directed	29			1.0	286	0
taqwanuhsina	thejagmeetsingh	Directed	30			1.0	938	0
tofakeindgily	justinpjtrudeau	Directed	31			1.0	18	0

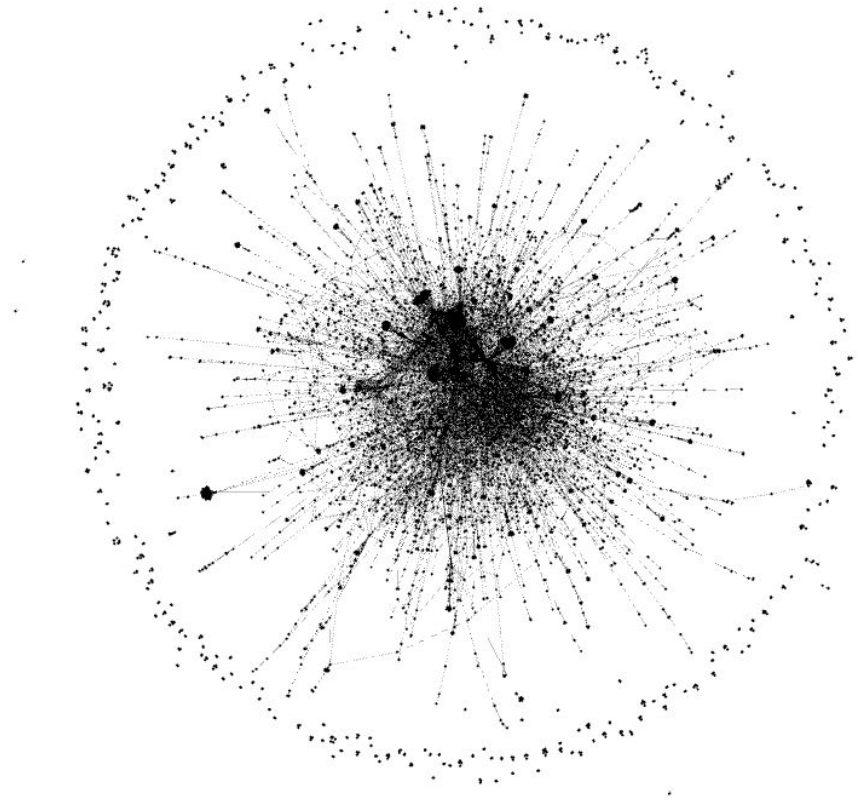
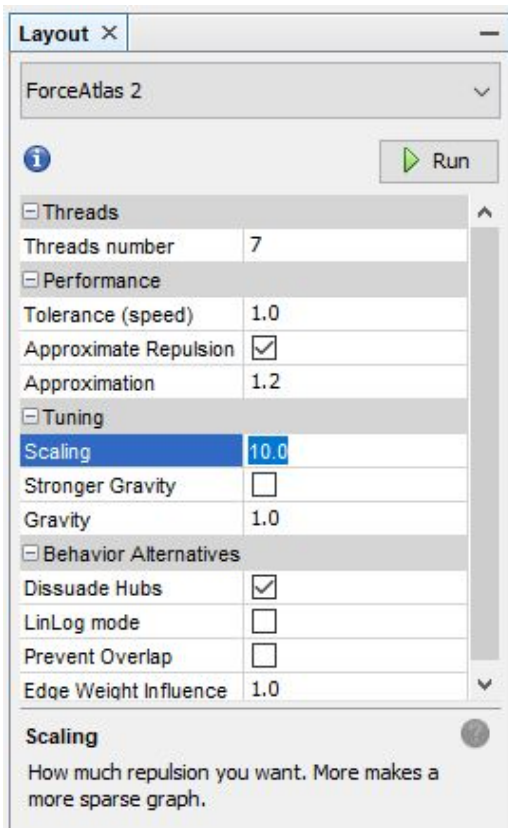
Review our data (nodes and edges) → what does ‘weight’ refer to?



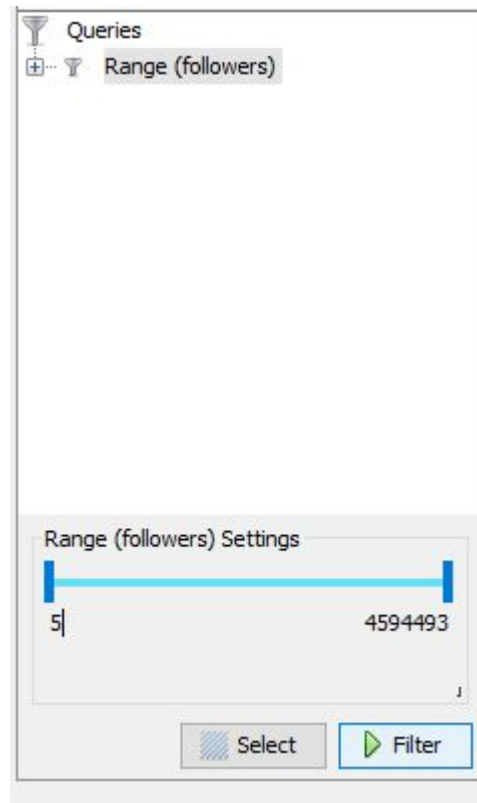
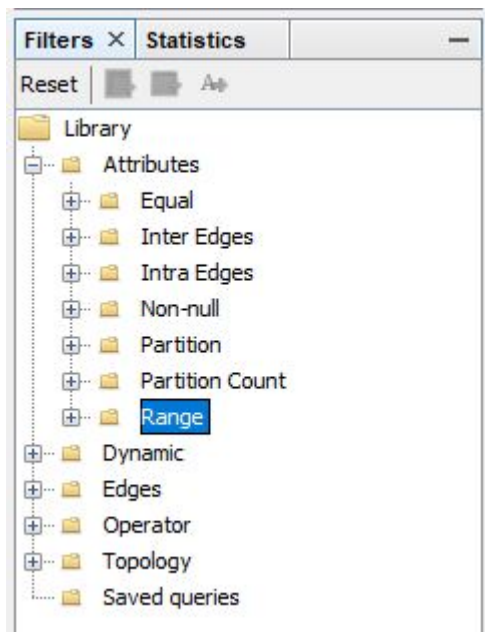
Switch to **Overview** → HAIRBALL!!!!



Un-teasing the hairball: **Layout** (begin with **Force Atlas**)



Try **ForceAtlas 2**...



Optional: use **Filters** to omit potential bot accounts (when tied to nodes)

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 ●

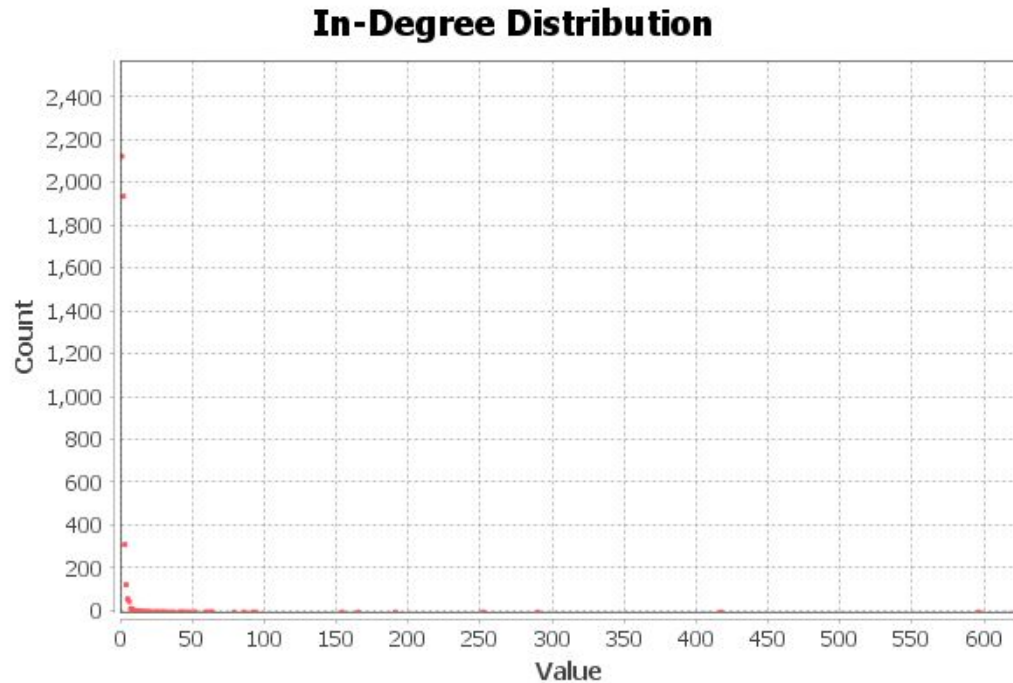
☒ **Node Overview**

Avg. Clustering Coefficient Run ●

Eigenvector Centrality Run ●

☒ **Edge Overview**

Avg. Path Length Run ●



Run **Statistics** so that we have some additional data to work with...



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
lisalaflammetv	lisalaflammetv		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

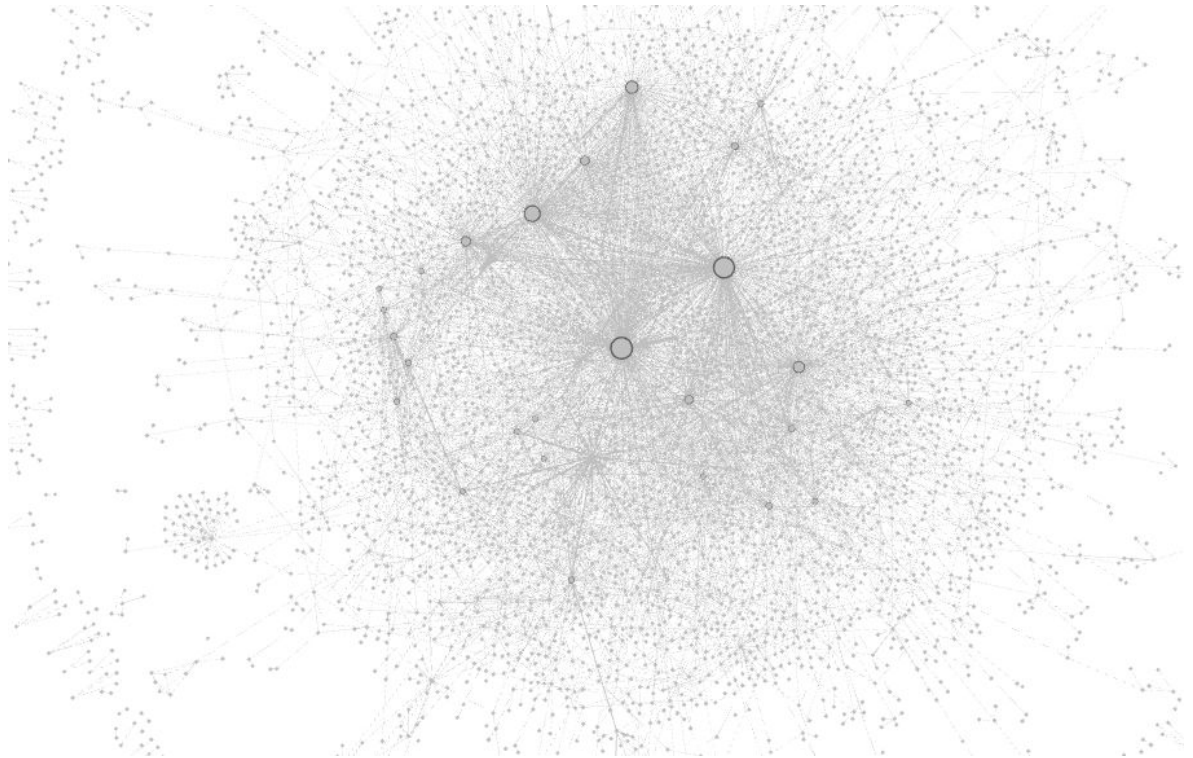
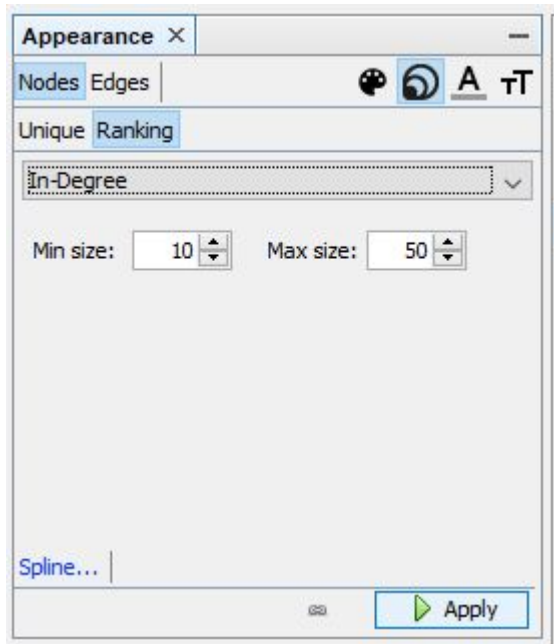
Statistics are also populated in the Data Table (e.g. sort by **In-Degree**)



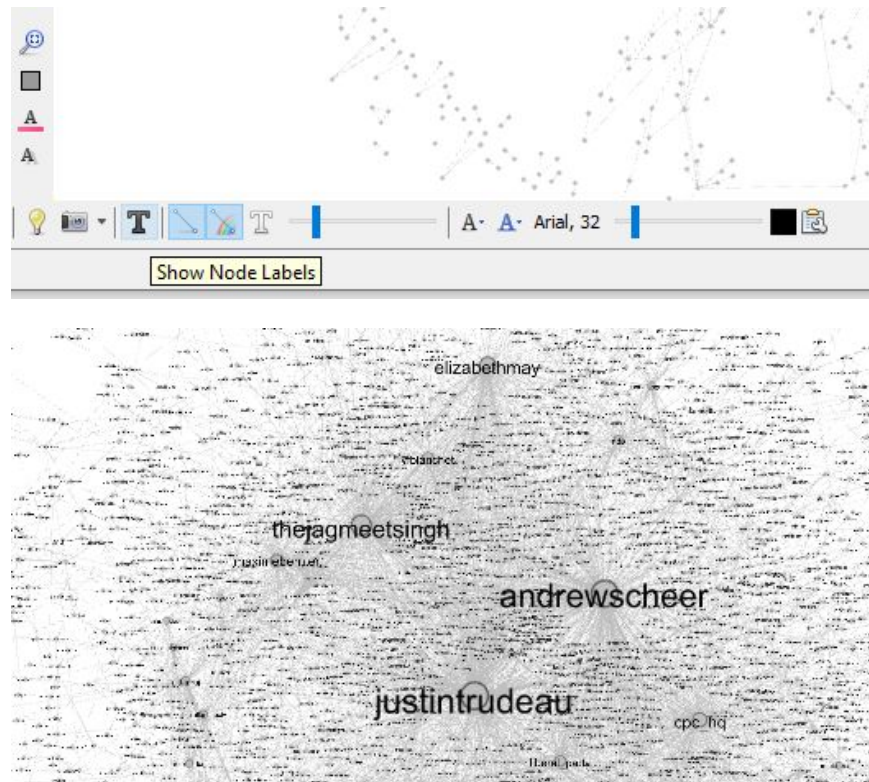
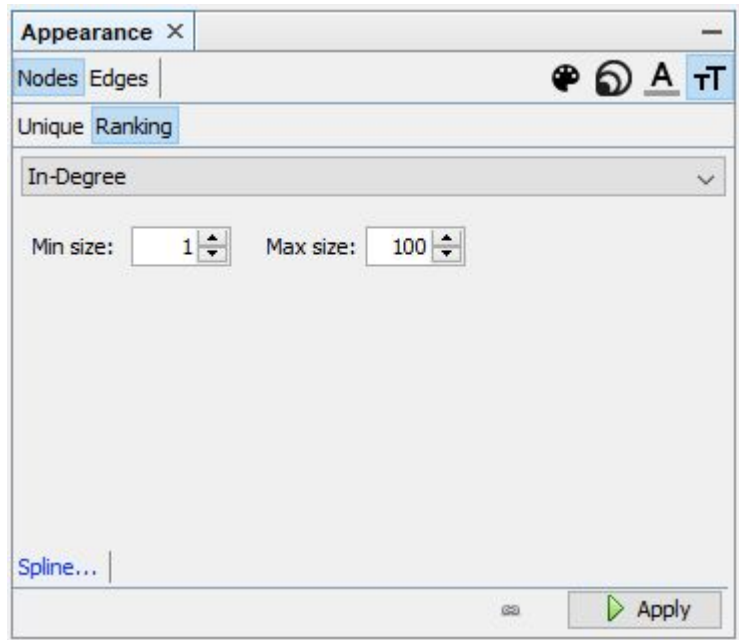


# MAKING SENSE OF THE DATA

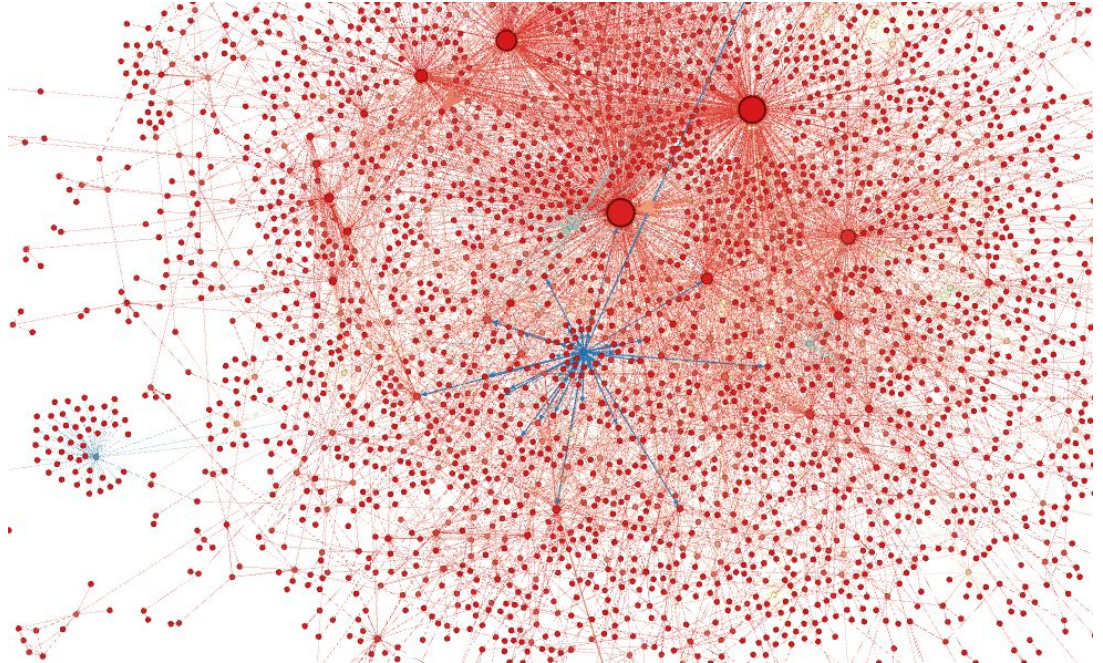
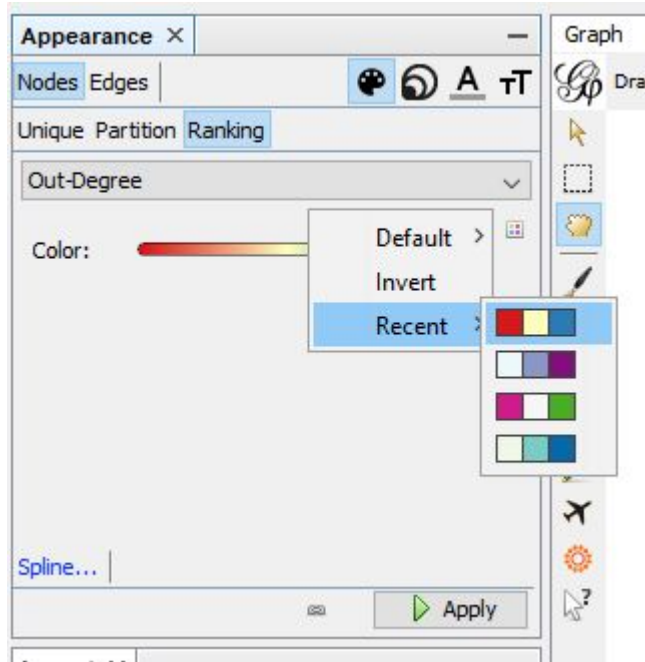
- Use attributes...
  - Calculated: In-degree, out-degree, modularity, etc.
  - Supplied: followers, retweets
- ...to modify:
  - Node & edge colour
  - Node size
  - Label size & colour



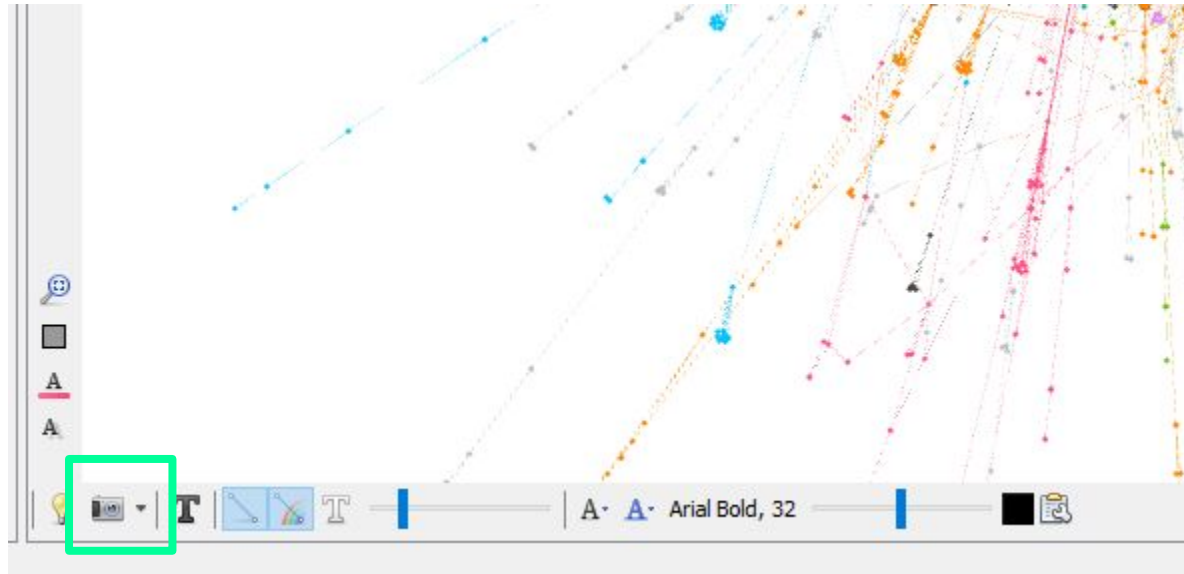
Who is being tweeted at the most?



Adjust label size to identify account names

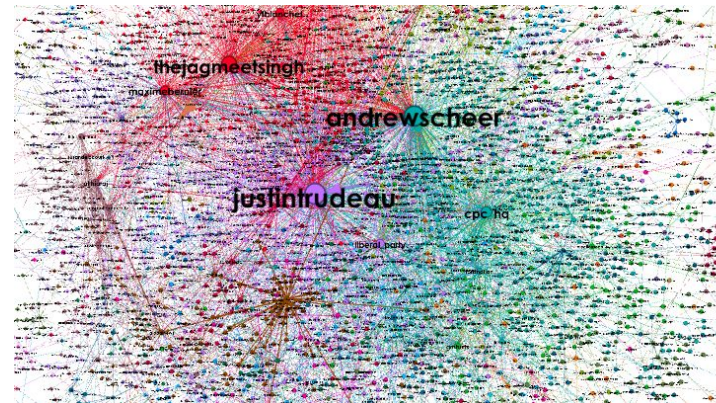
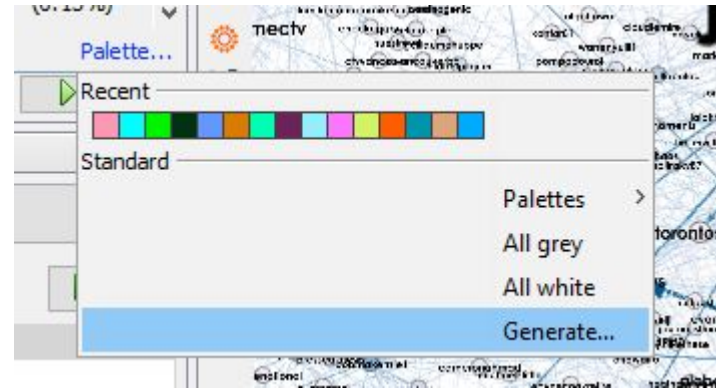
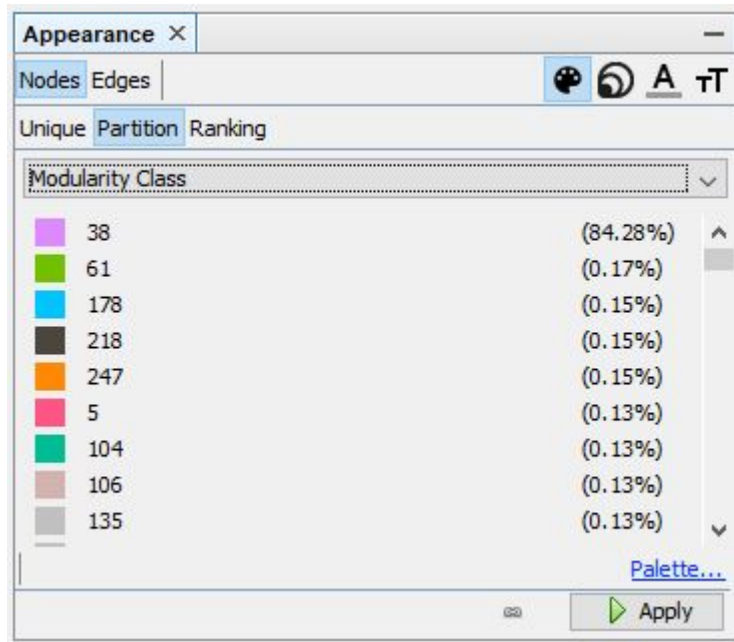


Who is doing most of the tweeting?

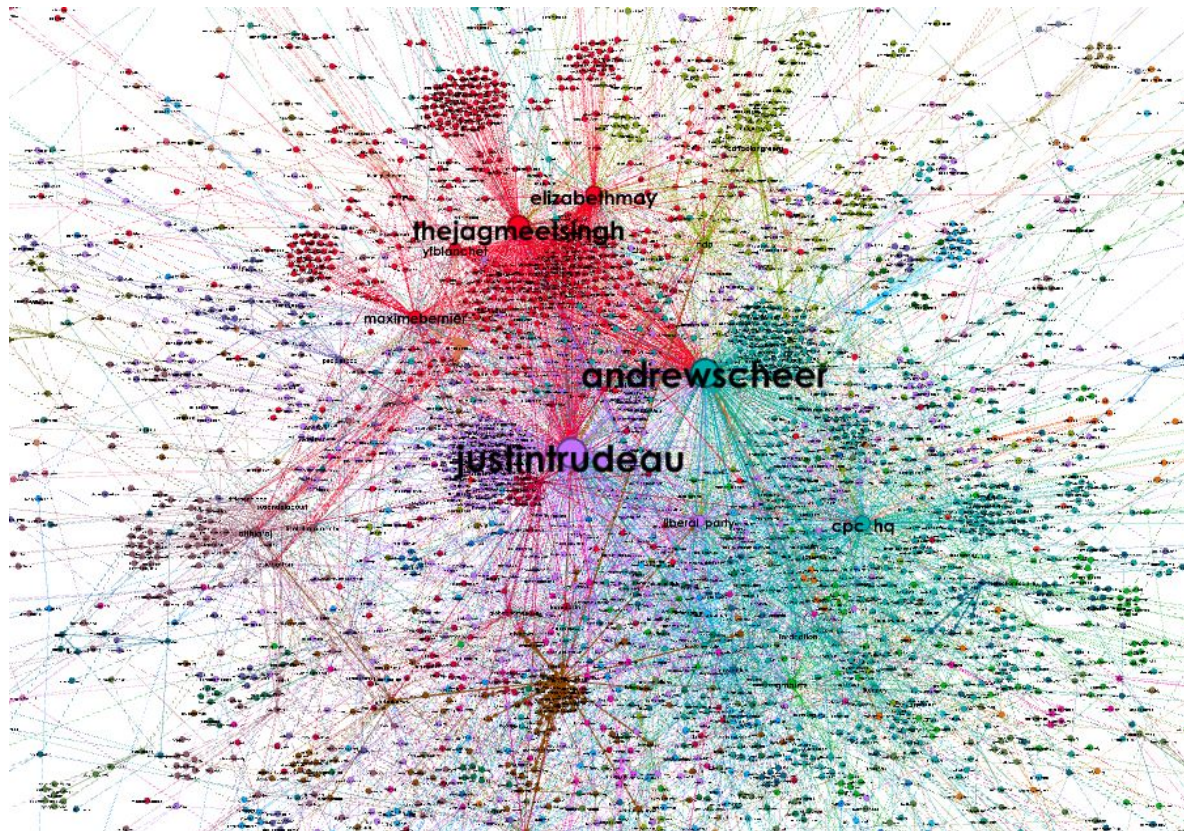


Good time to TAKE A SCREENSHOT... recall **provenance**





Community Detection → **Modularity Class**



Finessing the visualization further - adjust layout, font, etc.



# SHOW OFF YOUR VIZ

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

- PDF
- PNG
- SVG





# Thanks for coming!

Questions: [dmordell@uwindSOR.ca](mailto:dmordell@uwindSOR.ca)