



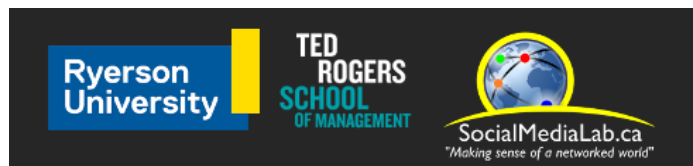
Session 5: Getting Started with Communalytic Social Network Analysis of Signed Networks

July 8, 2021

SocialMediaLab.ca



“Making Sense of a Networked World”



The Social Media Lab is a multidisciplinary research laboratory at Ryerson University in Toronto, Canada

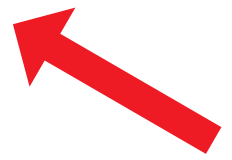
The lab studies how social media is changing the way people communicate, share information and form communities online, and how these changes impact society.

Videos and slides from past sessions are available online at: communalytic.com



FAQ TUTORIALS PUBLICATIONS

CSS BOOTCAMP



APR
07
2021

Social Media Lab's Computational Social Science (CSS) Bootcamp – Summer 2021

Off

By COMMUNALYTIC



What: CSS Bootcamp on Examining Online Discourse & Networks with Communalytic

When: 2nd and 4th Thursday of the month at 10 am -11:30 am (ET) between May and July, 2021.

**Where: Zoom (see details below)
Free Registration via Zoom**

Join the Commanalytic Discussion Group

<https://groups.google.com/u/1/g/commanalytic-community-group>

☰ Groups

- + New conversation
- 👤 My groups
- 🕒 Recent groups
- ▶ 📌 Favourite groups
- ☆ Starred conversations

Commanalytic Community Group

- 🗨️ **Conversations**
 - Approved
 - Pending
- 👤 People
 - Members

🔍 Conversations Search conversations within commanalytic-c...



☆ Commanalytic Community Group



CSS Bootcamp Schedule Summer 2021

Session #1	Getting Started with Communalytic: Data Collection from Reddit	May 13, 2021, 10:00- 11:30am (EDT)
Session #2	Toxicity Analysis with Reddit Data using Perspective API	May 27, 2021, 10:00- 11:30am (EDT)
Session #3	Getting Started with Communalytic: Data Collection from Twitter (Twitter Thread via API v2.0 and Twitter Academic Track)	June 10, 2021, 10:00- 11:30am (EDT)
Session #4	Toxicity Analysis of Twitter data using Perspective API	June 24, 2021, 10:00- 11:30am (EDT)
Session #5	Social Network Analysis of Signed Networks with Reddit and Twitter data	July 8, 2021, 10:00- 11:30am (EDT)
Session #6	Getting Started with Communalytic: Data Collection from Facebook & Instagram (via CrowdTangle API) + Social Network Analysis of Two-mode Semantic Networks with CrowdTangle data	July 22, 2021, 10:00- 11:30am (EDT)



Communalytic is a research tool for studying online communities and online discourse.

Communalytic can collect and analyze public data from social media platforms. It uses advanced text and social network analysis techniques to automatically pinpoint toxic and anti-social interactions, identify influencers, map shared interests and the spread of misinformation, and detect signs of possible coordination among seemingly disparate actors.



collect



extract





analyze



visualize



<https://communalytic.com/>

How to choose between Communalytic Edu and Pro.

	 communalytic EDU	 communalytic PRO
Account Type	Free	\$349/6-mo. to support site infrastructure (server-side data collection, storage, processing, analysis and visualization)
Designed For	Students and is ideal for teaching and learning about social media analytics	Academic researchers and is ideal for large scale academic research projects
Account Caps	≤ 30K records shared across 3 datasets	≤ 10M records shared across 50 datasets
Reddit	Live-collection* of public posts from any public subreddit for ≤ 7 consecutive days (Limit: ≤ 30K posts)	Live* & historical collection of public posts from any public subreddit for ≤ 31 consecutive days (Limit: Account Caps)
Twitter Threads (API ver.2) req. Twitter developer's account	Public replies to any public tweet posted within the previous 7 days (Limit: ≤ 30K tweets)	Public replies to any public tweet posted within the previous 7 days (Limit: ≤500K tweets/month)
Twitter Academic Track (API ver.2) req. Application to Twitter	Not supported	Full-archive historical-search of tweets back to 2006 (Limit: ≤ 10M tweets/month)
CrowdTangle (FB/IG) URL Search req. a CrowdTangle account	Public Facebook or Instagram posts that shared the same URL (Limit: ≤ 30K posts)	Public Facebook or Instagram posts that shared the same URL (Limit: Account Caps)

* Live-collection = the collection of posts/tweets posted on or after the date when you initiated the data collection.

How to choose between
Communalytic Edu and Pro.

	 communalytic EDU	 communalytic PRO
Exploratory Data Analysis (EDA)	<ul style="list-style-type: none"> • Emoji cloud (freq. used emojis) • Word cloud (freq. used words) • Time series (posts per day) • Top posters (top 10) 	<ul style="list-style-type: none"> • Emoji cloud (freq. used emojis) • Word cloud (freq. used words) • Time series (posts per day) • Top posters (top 10)
Text Analysis	<ul style="list-style-type: none"> • Toxicity analysis based on machine learning via Google's Perspective API 	<ul style="list-style-type: none"> • Toxicity analysis based on machine learning via Google's Perspective API
Social Network Analysis (SNA)	<ul style="list-style-type: none"> • Reply Network • Reply Network with toxicity scores • 2-mode Semantic Network (for CrowdTangle FB and IG data only) 	<ul style="list-style-type: none"> • Reply Network • Reply Network with toxicity scores • 2-mode Semantic Network (for CrowdTangle FB and IG data only)
Additional features	<ul style="list-style-type: none"> • Simultaneously Run Multiple Data Collectors: 1 Reddit, 1 Twitter & 1 CrowdTangle • Import existing datasets (CSV and Gzip-ed CSV) • Collaboration Friendly: Access to various team collaboration features 	<ul style="list-style-type: none"> • Simultaneously Run Multiple Data Collectors: 2 Reddit, 1 Twitter & 1 CrowdTangle • Import existing datasets (CSV and Gzip-ed CSV) • Collaboration Friendly: Access to various team collaboration features

Today's agenda



Introduction to Social Network Analysis (Philip Mai)



Analyzing “Signed” Networks in Communalytic (Anatoliy Gruzd)



Exporting and analyzing networks with 3rd-party programs like Gephi (Alyssa Saiphoo)

How Do We Collect Information About Social Networks?

Common approach for collecting social network data:

- Self-reported social network data
- Surveys or interviews



How Do We Collect Information About Social Networks?

A sample question about students' perceived social structures

Please indicate on a scale from [1] to [5],

YOUR FRIENDSHIP RELATIONSHIP WITH EACH STUDENT IN THE CLASS

- [1] - don't know this person
- [2] - just another member of class
- [3] - a slight friendship
- [4] - a friend
- [5] - a close friend

Alice D. [1] [2] [3] [4] [5]

...

Richard S. [1] [2] [3] [4] [5]

Source: C. Haythornthwaite, 1999

How Do We Collect Information About Social Networks?

Problems with surveys or interviews:

- Time-consuming
- Questions can be too sensitive
- Answers are subjective or incomplete
- Participant can forget people and interactions
- Different people perceive events and relationships differently



How to Make Sense of Social Media Data?

Method: Social Network Analysis (SNA)

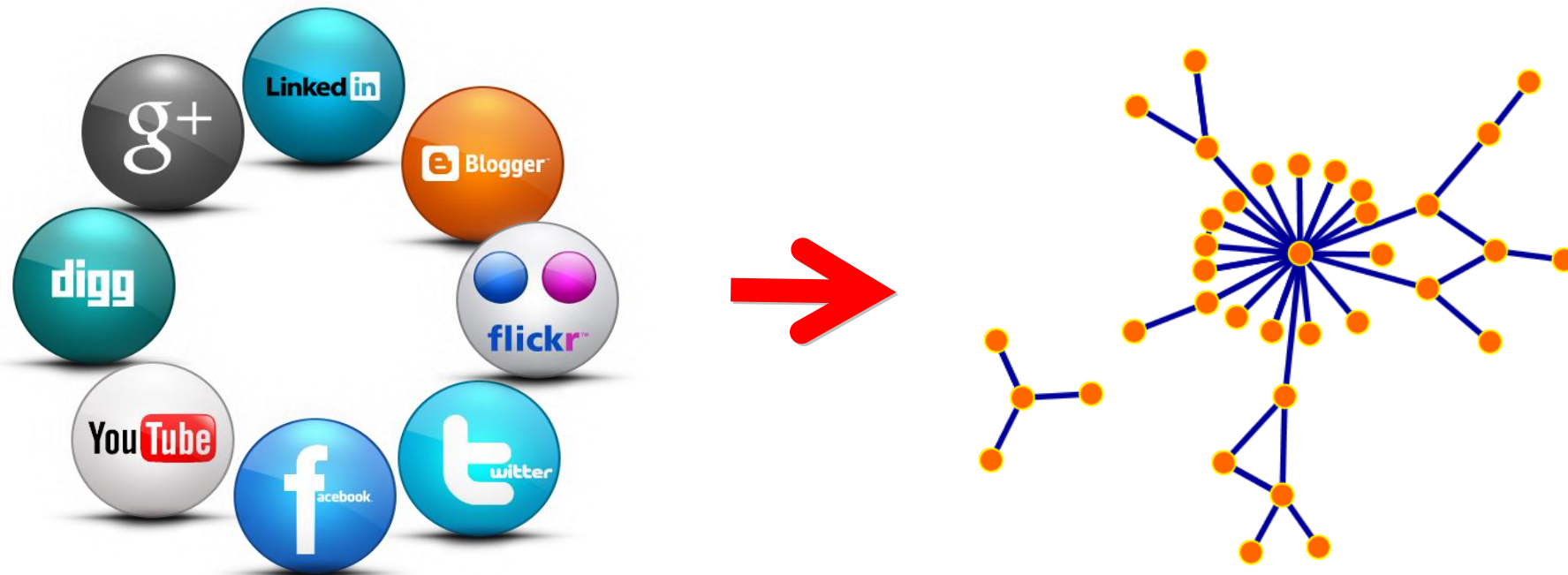
Nodes (dots) = Accounts (People, Companies, Gov, Org...)

Edges/Ties (lines) = Any Types of Relations. (Ex. of relations...)

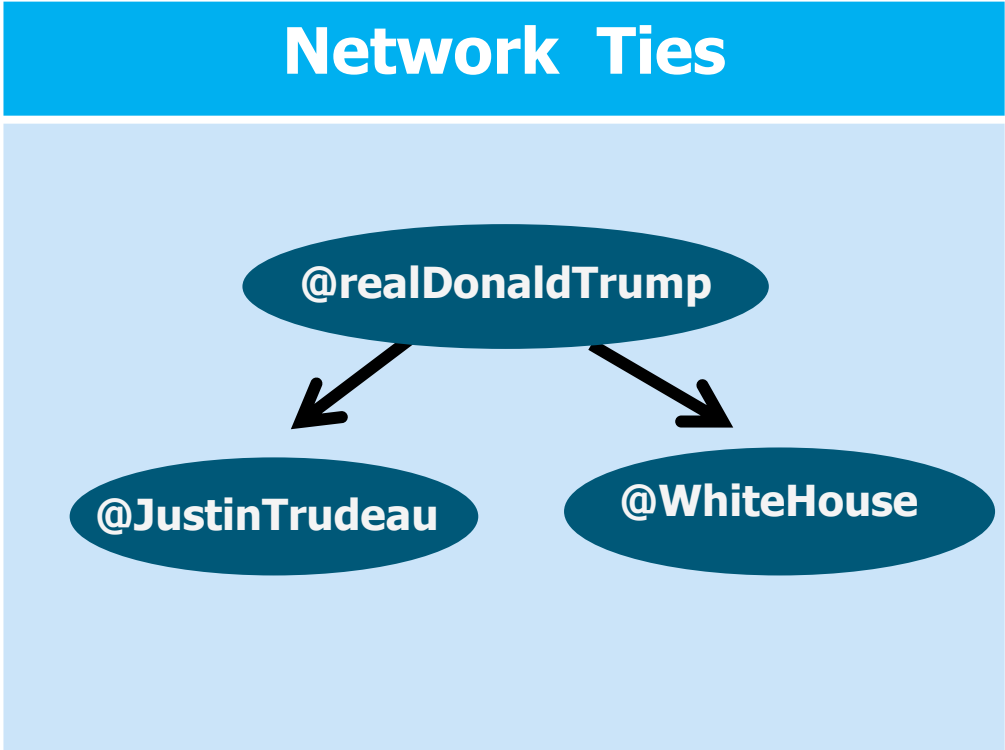
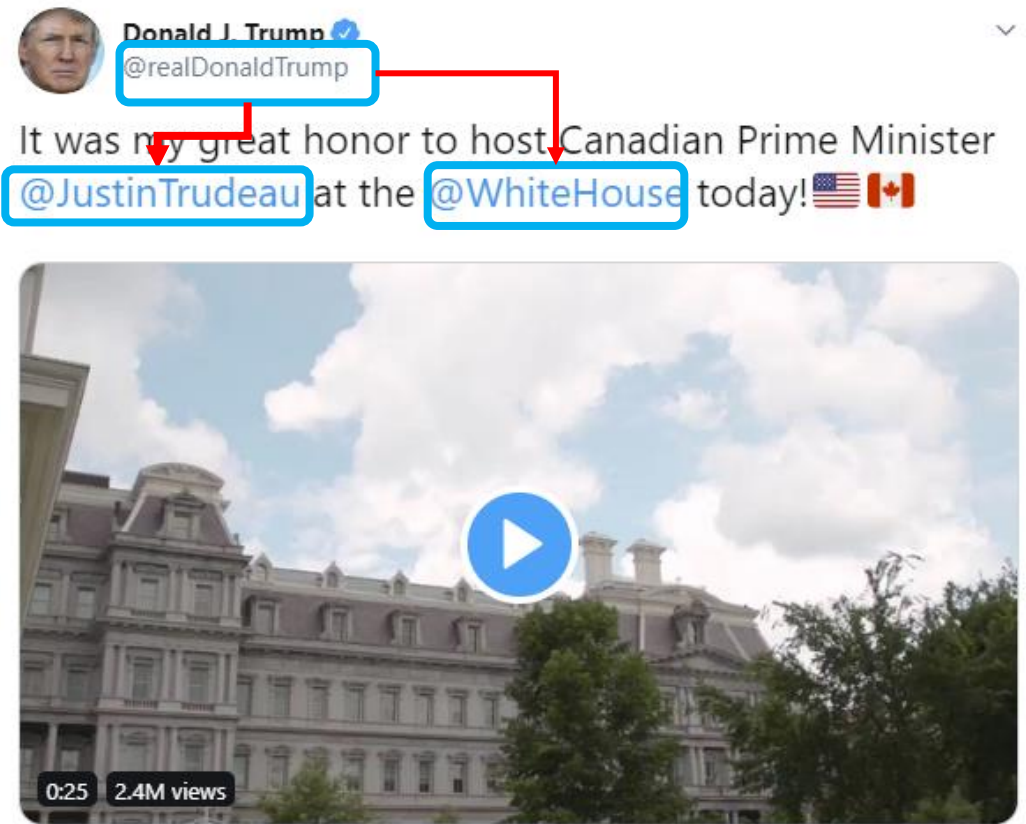
Offline: Who you are close to? Who do you talk to? Who do you work with?

Online: Who retweeted, shared, replied, mentioned whom, etc. ...

Tie Strength = The number of retweets, replies or mentions, etc...



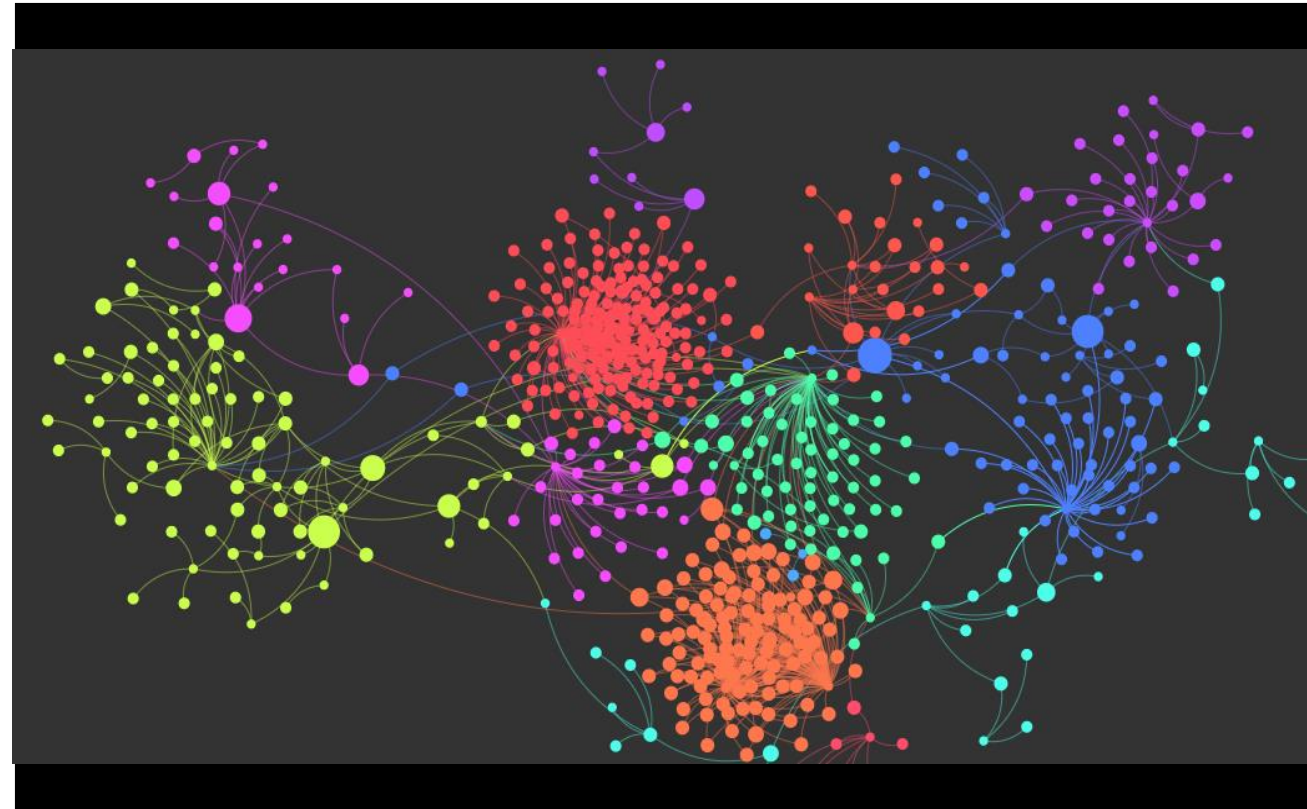
Discovering a Network Structure One Tweet at the time

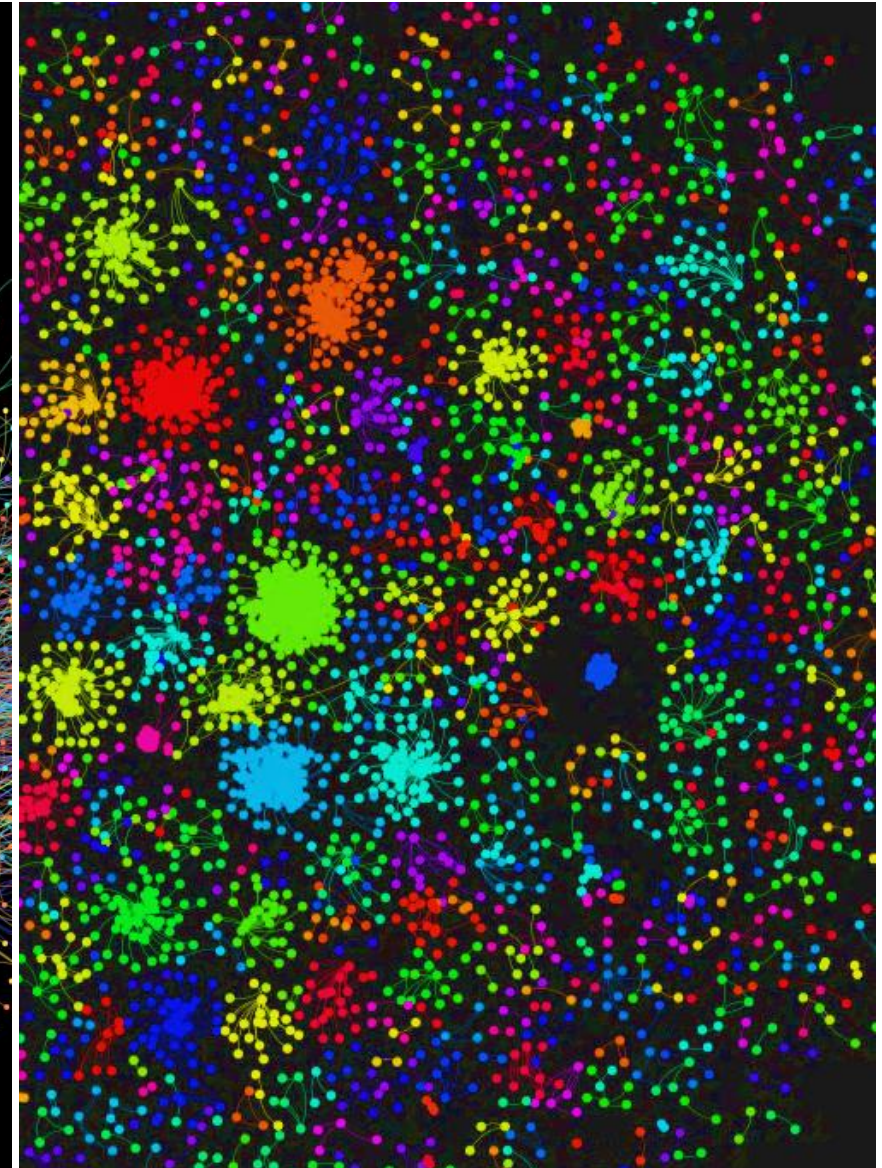
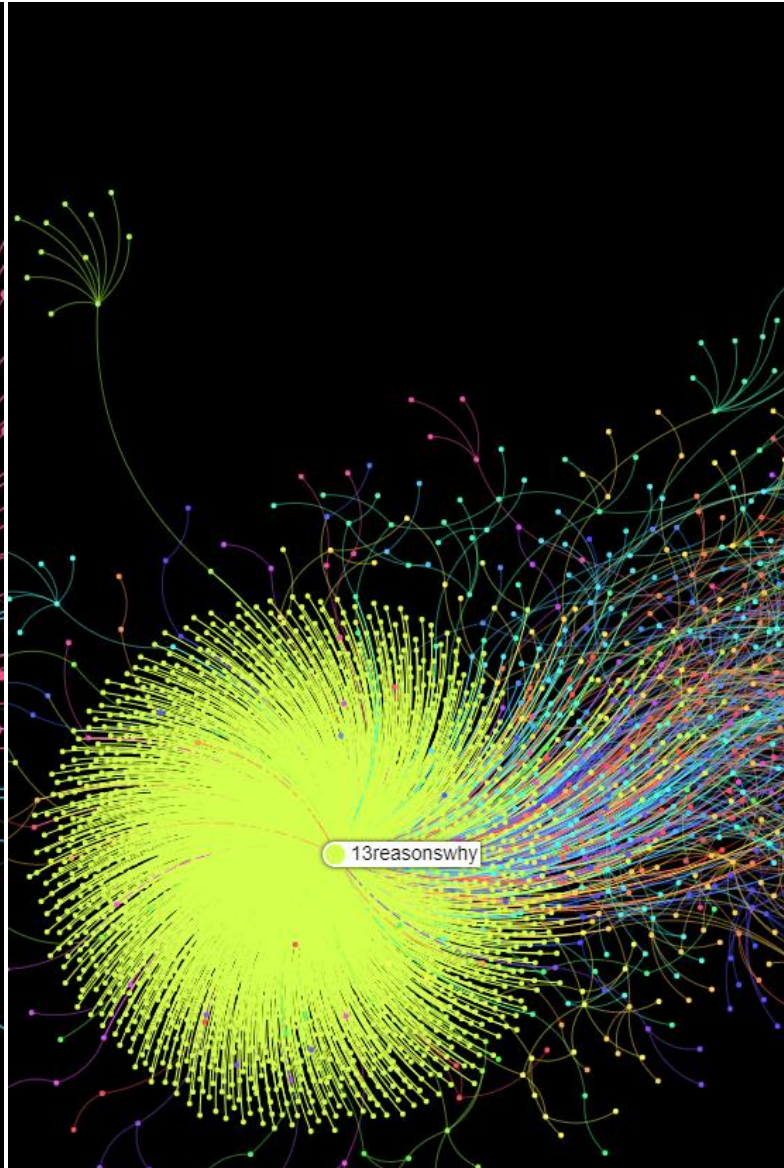
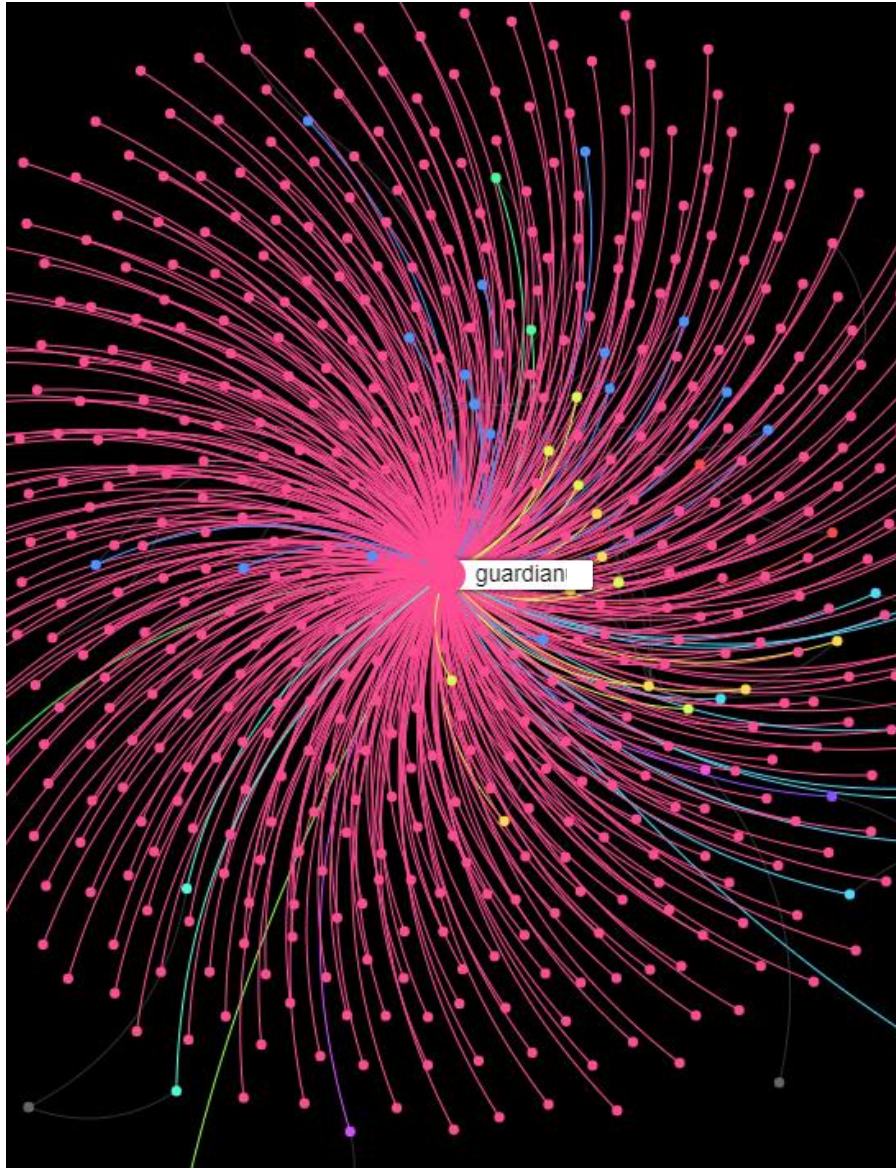


1:49 PM · Jun 20, 2019 · Twitter for iPhone

Advantages of Social Network Analysis (SNA)

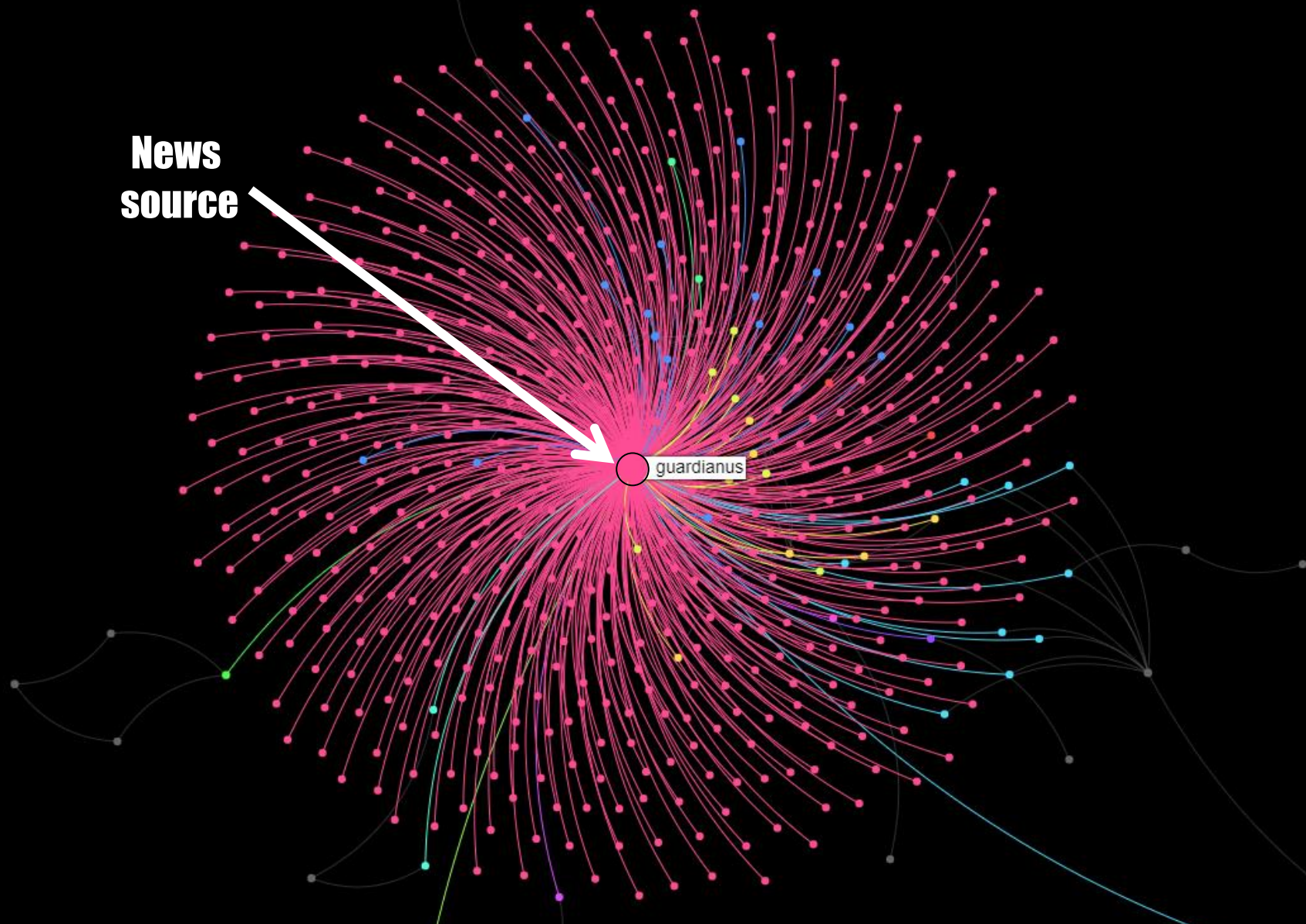
- Easier to understand what is going on in a large group
- Once the network is discovered, we can find out:
 - How do people interact with each other,
 - Who are the most/least active members,
 - Who is influential in a group,
 - Who is susceptible/open to being influenced, etc...





Examples of Different Information Propagation Models on Twitter

Star-shaped Network Configuration



Example 1

theguardian

News | Sport | Comment | Culture | Business | Money | Life & style | Travel |

News > World news > Ukraine

Ukraine police storm Kiev protest camp – live updates

LIVE

At least 18 killed and more than 200 seriously injured in clashes in Ukrainian capital

In pictures: violent clashes in Kiev

Ben Quinn and Kayla Epstein
theguardian.com, Tuesday 18 February 2014 23:45 GMT

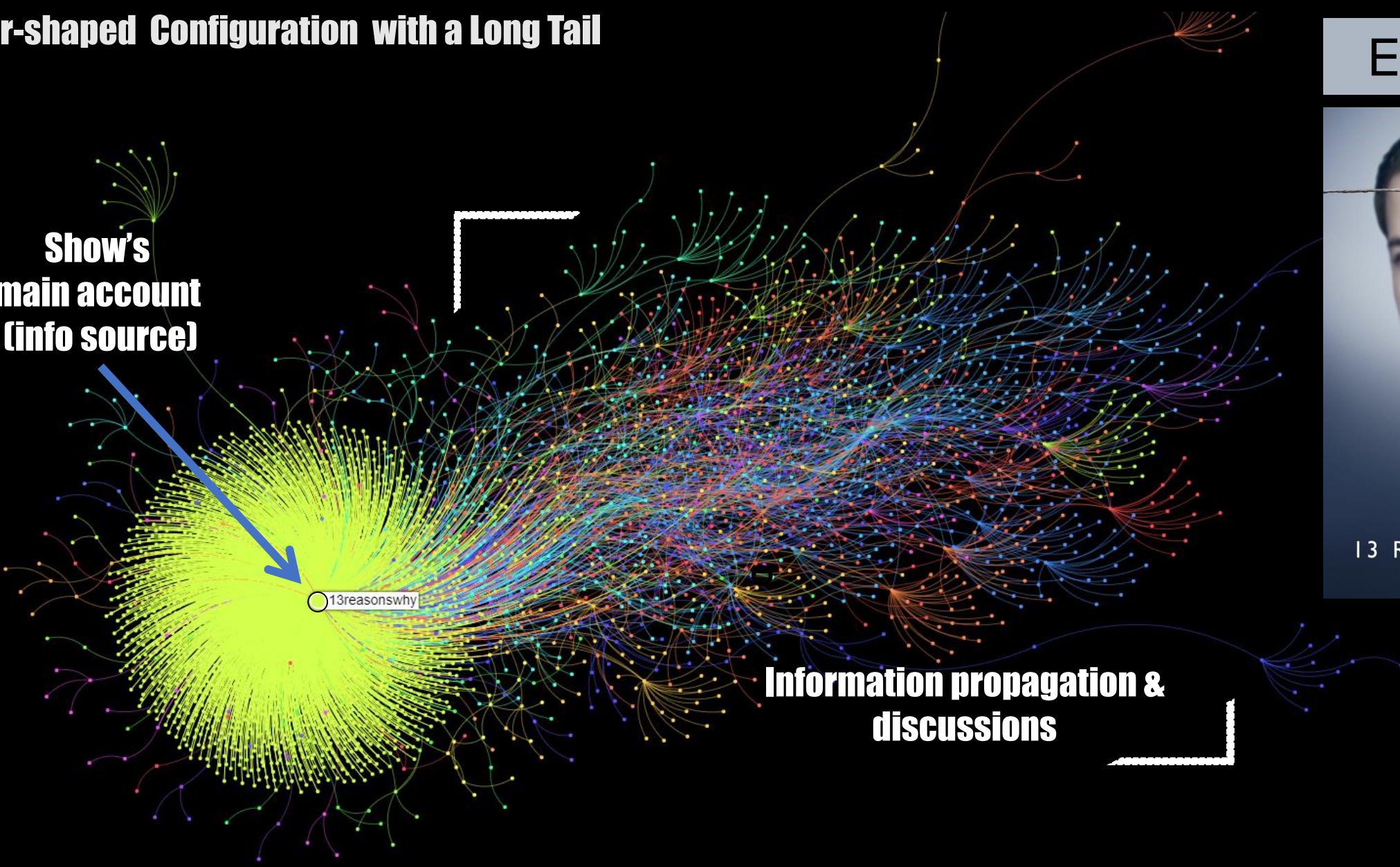
Jump to comments (498)

Protesters gather in Independence Square in Kiev. Photograph: EPA/Igor Kovalenko

Sharing a news story about the 2014 Euromaidan revolution in Ukraine

Star-shaped Configuration with a Long Tail

Show's main account (info source)



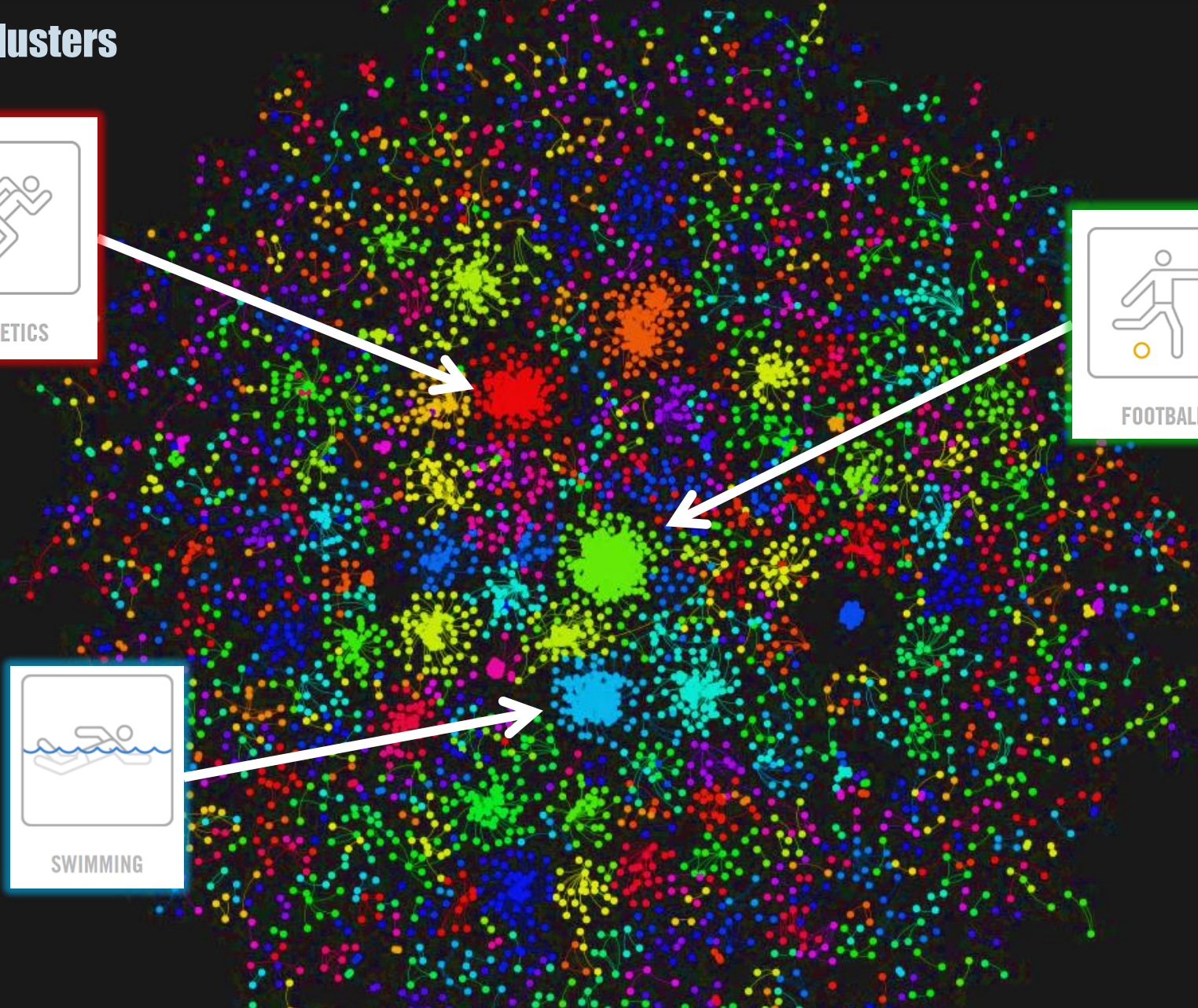
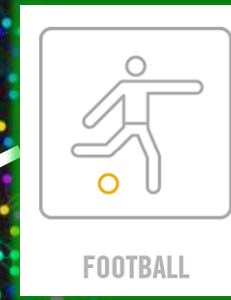
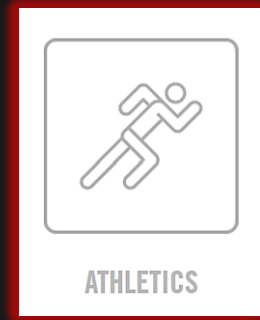
Information propagation & discussions

Example 2



Information sharing and conversations about a Netflix show called "13 Reasons Why"

Multiple Clusters



Example 3



Twitter chatter about different sporting events during the 2012 Olympics in London

SNA Measures

Micro-level

In-degree centrality

Out-degree centrality

Betweenness centrality

**Other centrality measures
(e.g., closeness, eigenvector)**



Today's focus

Macro-level

Density

Diameter

Reciprocity

Centralization

Modularity

...

Sample Twitter Dataset: #HongKong

3557 tweets (Dec 3, 2015)

#HongKong

Top | Live | News | Photos | Videos | More options ▾

Top news [View all](#)

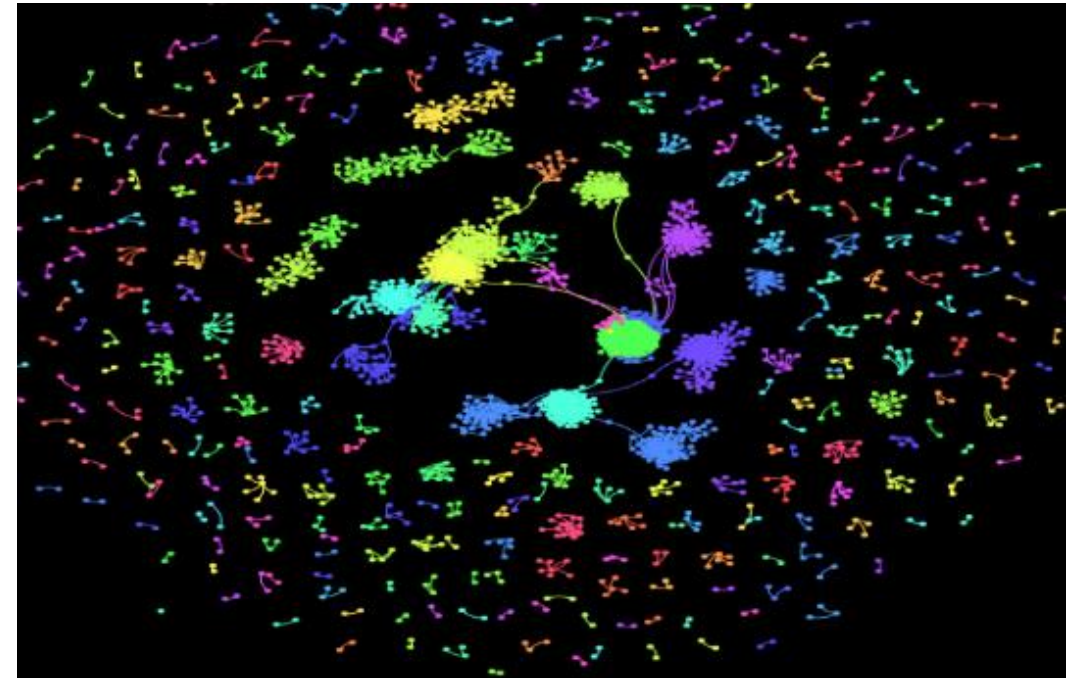


Hong Kong's best kept secrets: 1960s-'70s style postcards and quirky gifts in historical Wan Chai

Paper dolls, vintage lamps, cassette players and kitsch Hong Kong postcards among finds at Hong Kong House of Stories in 1920s tong lau

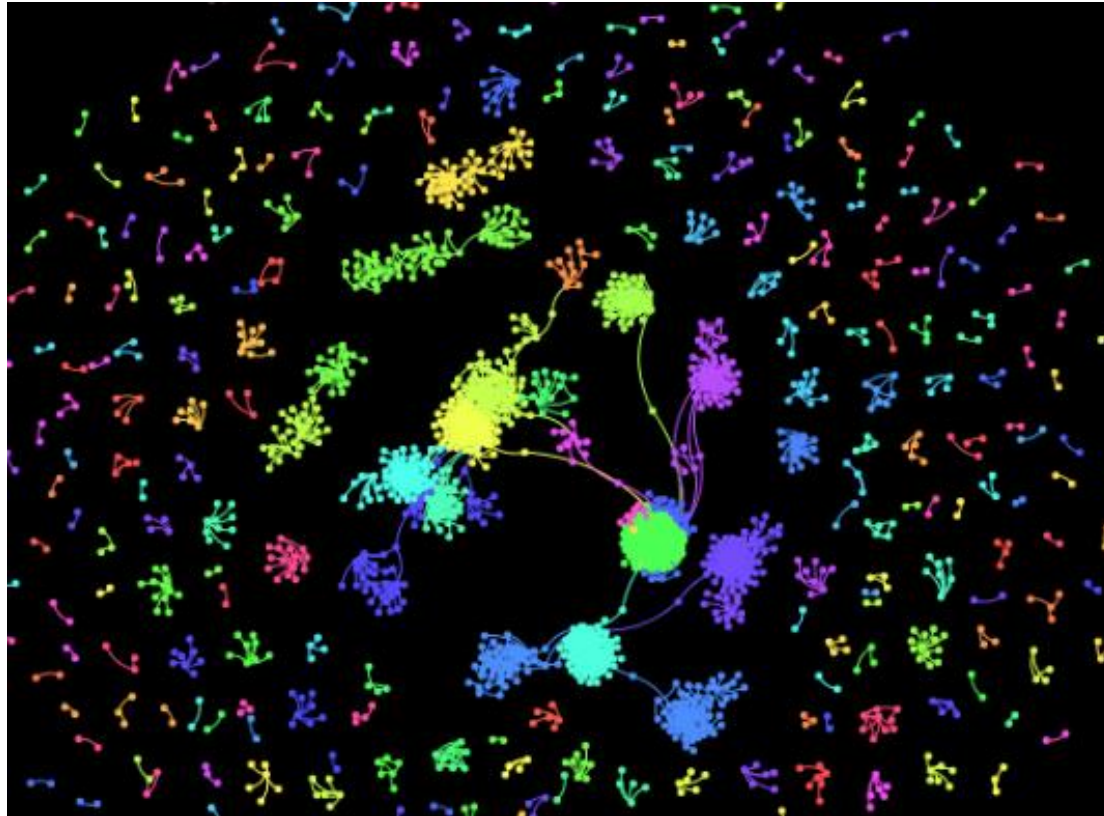
SCMP News Dec 02, 2015

Network Representation



#HongKong Twitter Network

What does this network visualizations tell us?



SNA Measures

Micro-level

In-degree centrality

Out-degree centrality

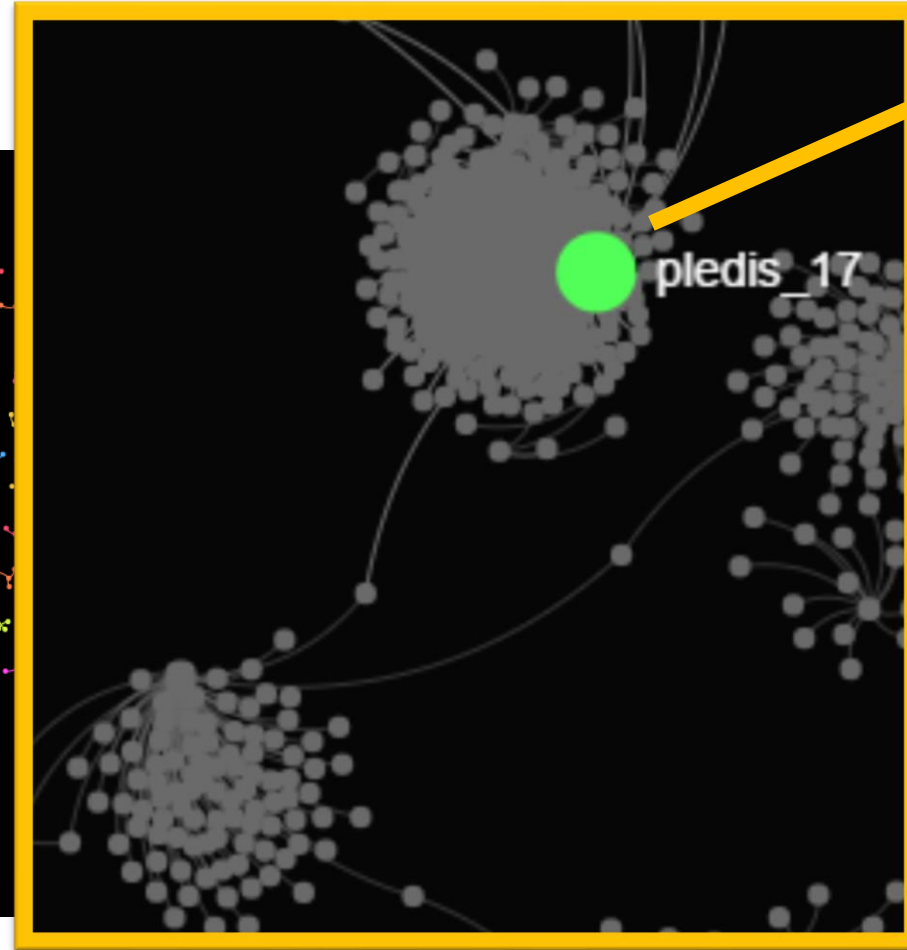
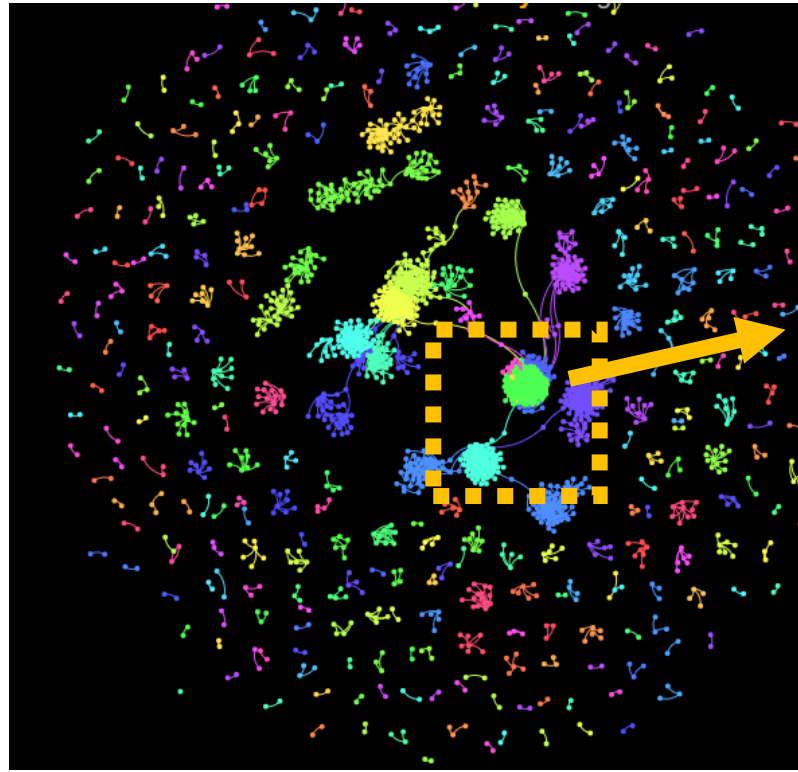
Betweenness centrality

**Other centrality measures
(e.g., closeness, eigenvector)**

- **In-degree** suggests “prestige” highlighting frequently mentioned or replied Twitter users

In-degree centrality #HongKong Twitter network

SEVENTEEN or SVT is a S.Korean boy group formed by Pledis Entertainment



세븐틴(SEVENTEEN)
@pledis_17 Follow

#SEVENTEEN's Today Thank you #Hongkong Thank you #2015MAMA
[instagram.com/p/-y7HgfoUU1/](https://www.instagram.com/p/-y7HgfoUU1/)

RETWEETS 5,454 LIKES 6,711

8:12 AM - 2 Dec 2015

권순영닷컴 @19960615com · 16h
@pledis_17 너무 벅차고 행복하다. 이만큼 즐거라고 전엔 상상 못 했는데 상상 못 한게 너무 미안할 정도로 너희는 늘 최고야! 이제는 그 잠재력이 대체 어디까지인지 상상이 안 돼. 그러니까 꼭 함께할게, 오늘보다 더 빛날 세븐틴을 위해서! ❤️

부른다 @980116_net · 16h
@pledis_17 해외에서 정말 큰무대에서 멋지게 무대하느라 수고많았어요 폭쉬고 콘서트 준비도 무리하지말고 했으면!! 오늘 세상에서 제일 멋졌어요 최고로 멋진 내가수를 덕분에 매일매일 행복해.. ❤️

티얼스드림 @TearsDream_WW · 14h
@pledis_17 수고많았어~ 늘 고마워 사랑해 ❤️

SNA Measures

Micro-level

In-degree centrality

Out-degree centrality

Betweenness centrality

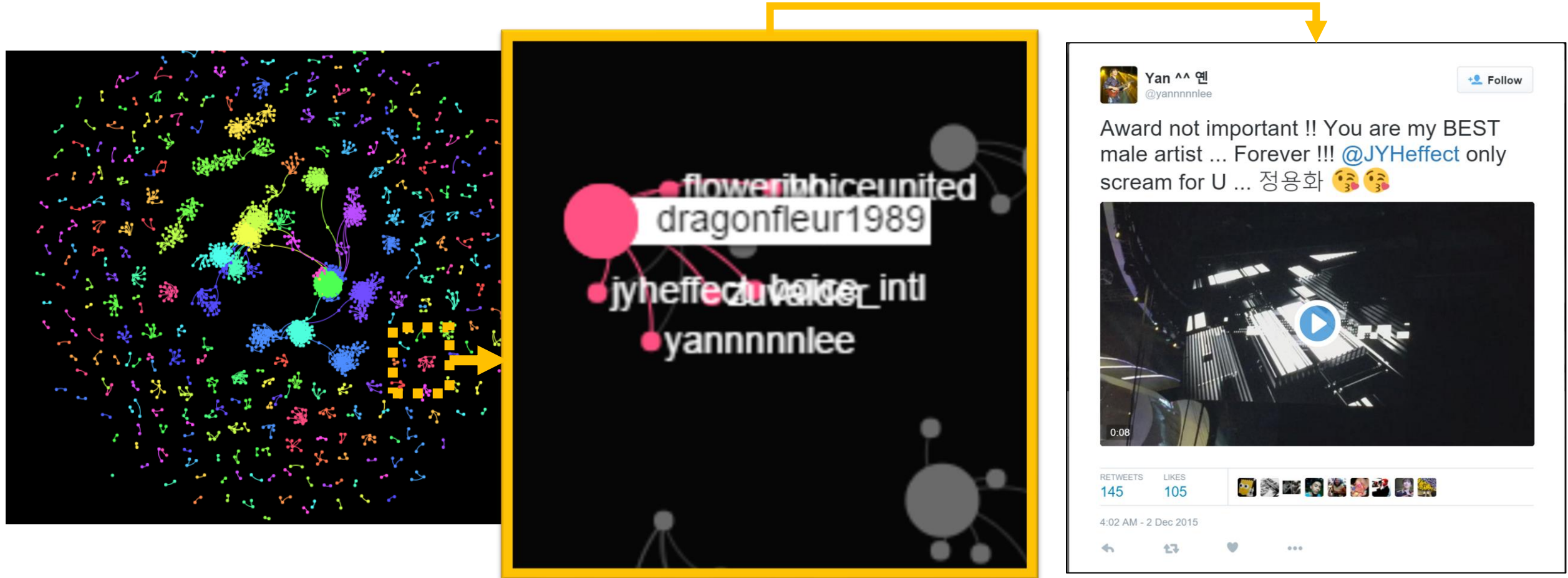
**Other centrality measures
(e.g., closeness, eigenvector)**

- **Out-degree** reveals active Twitter users with a good awareness of others in the network

Out-degree centrality

#HongKong Twitter network

Note: A music fan (many retweets & replies to others)



SNA Measures

Micro-level

In-degree centrality

Out-degree centrality

Betweenness centrality

**Other centrality measures
(e.g., closeness, eigenvector)**

- **Betweenness** shows actors who are *located on the greatest number of information paths* and who often *connect* different groups of users in the network

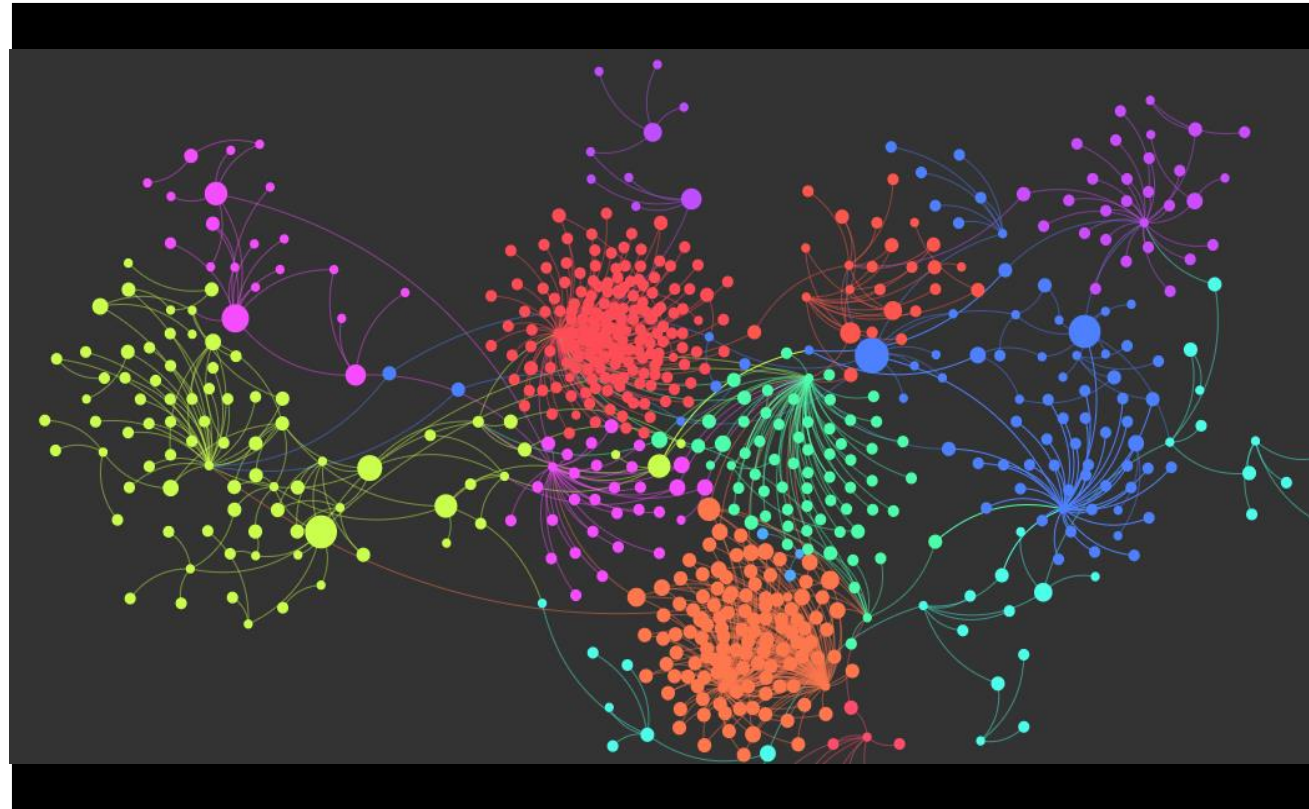
Betweenness centrality #HongKong Twitter network

Note: A fan (retweets/replies to messages from two different fan communities/sites)



Learn More About Social Network Analysis (SNA)

- Social Network Analysis: An Introduction
By Alexandra Marin and Barry Wellman
http://www.alexandramarin.ca/uploads/8/5/7/4/85749/marin_and_wellman_-_2010_-_social_network_analysis_an_introduction.pdf
- How To Get Started with Social Network Analysis
A Complete Beginner's Guide to Getting Up and Running Making Beautiful Network Graphs
<https://towardsdatascience.com/how-to-get-started-with-social-network-analysis-6d527685d374>



Today's agenda



Introduction to Social Network Analysis (Philip Mai)



Analyzing “Signed” Networks in Communalystic (Anatoliy Gruzd)



Exporting and analyzing networks with 3rd-party programs like Gephi (Alyssa Saiphoo)



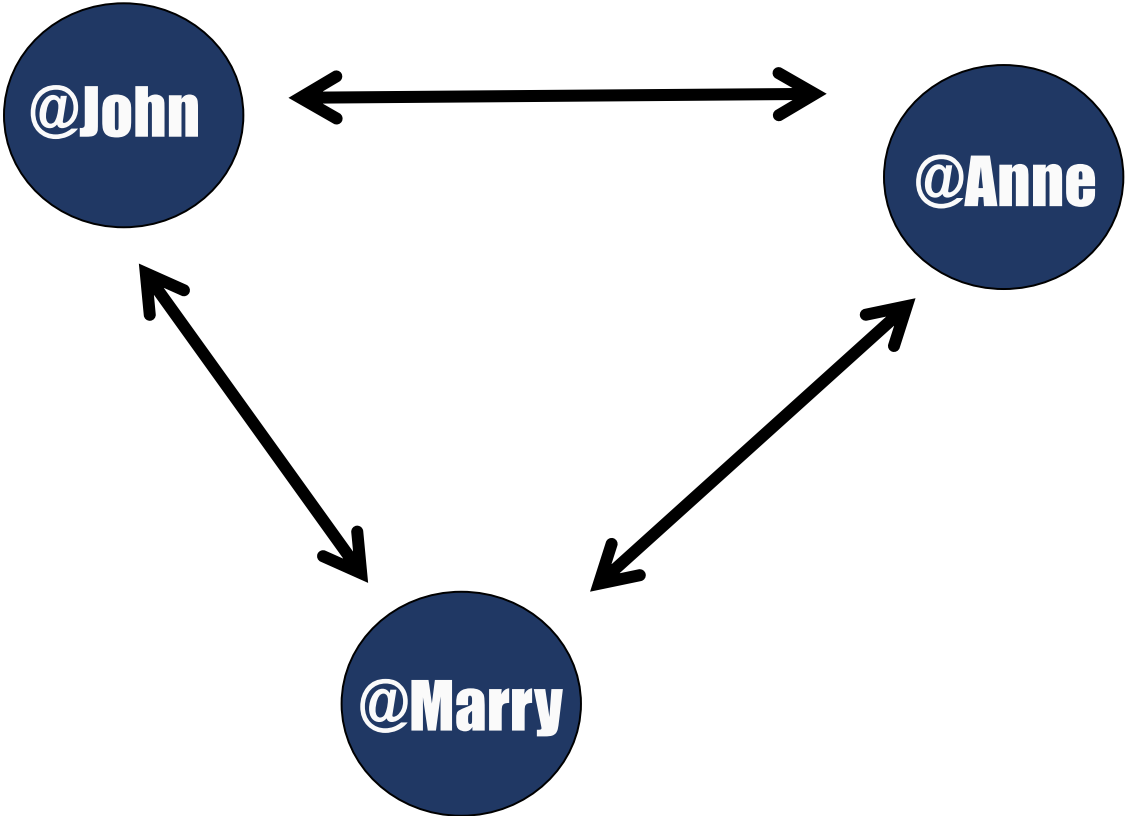
Signed Networks

- ‘**Signed**’ networks are networks with positive and negative weights assigned to each tie/connection.
- A sign indicates the extent to which interaction(s) between two nodes are either **positive (+)** or **negative (-)**.
- A **signed network** is helpful because we can judge the level of hostility at the individual and group-levels
 - Ex. Ties between subgroups tend to be more negative than internal ties within a subgroup (Harrigan and Yap, 2017)

Communication Network (non-signed)

● Node = User account

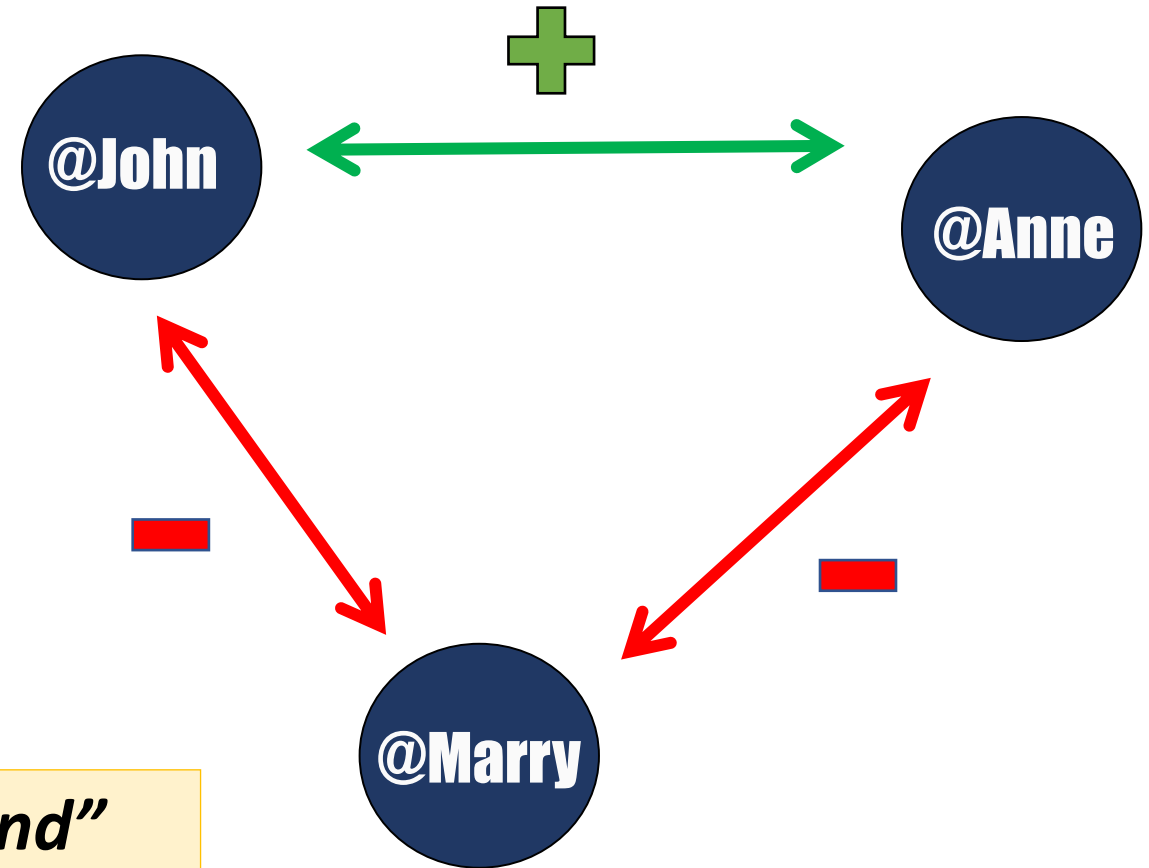
↕ Tie = Interaction



“Signed” Communication Network

● Node = User account

↕ Tie = Interaction (may carry a positive or negative weight)



“The enemy of my enemy is my friend”

Two individuals that have positive ties with each other should maintain similar negative ties in order to achieve **global structural balance** (Traag and Bruggeman, 2009)

How to Discover Positive/Negative Ties?

Previous Work

Aiello et.al., 2012; Song and Meyer, 2015

- Predicting positive/friendship ties in a non-signed network

Leskovec et.al., 2010

- Predicting future positive or negative ties with known initial positive or negative ties

Tang et.al., 2015

- Predicting future negative ties based on the presence of positive ties and content or user interaction features

Our Approach

- Previous approaches require either information about pre-existing positive or negative ties or information about a friendship network.
- However, when working with publicly available social media data, usually only interaction data (in particular, communication data) is available.
- To transform a communication network into a signed network, we can use toxicity scores to label exchanges as positive or negative.

Using Toxicity Scores to Discover Negative Ties

POST



Sample tweet attacking Jagmeet Singh

REPLY

Khalistani Sikh extremist Jagmeet Singh wants to turn Canada into Pakistan by allowing in thousands of Pakistanis.

CANADA BEWARE!

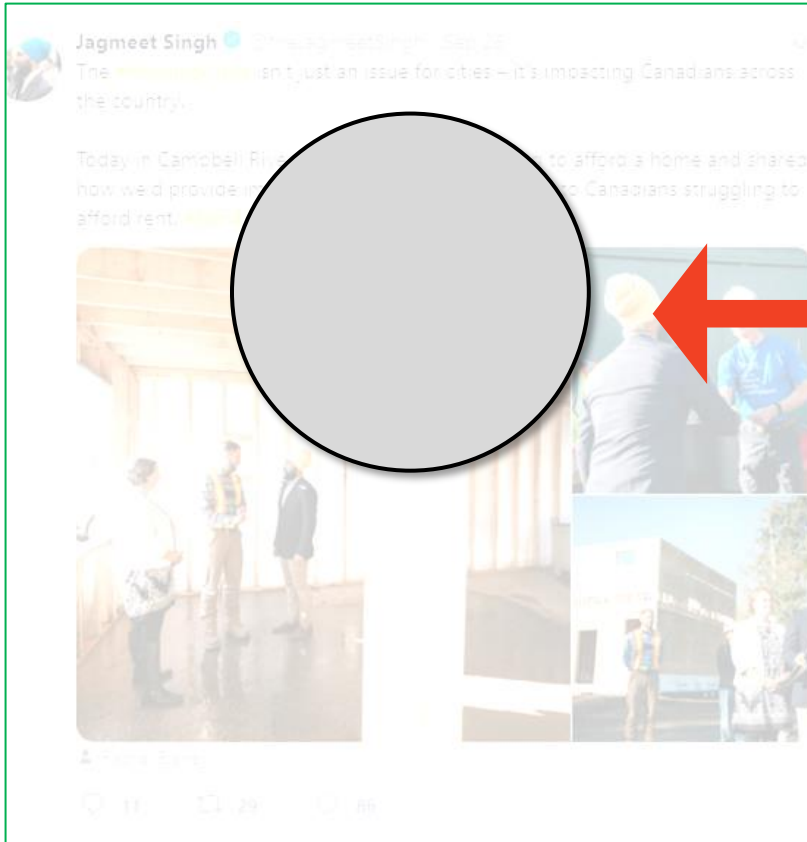
#cdnpoli #elxn43 #askcanadiansikhs #jagmeetsingh #elxn2019 #NDP #onpoli #InItForYou #JS4PM #ElectionsCanada #worldsikhorg #jagmeet #elxn19



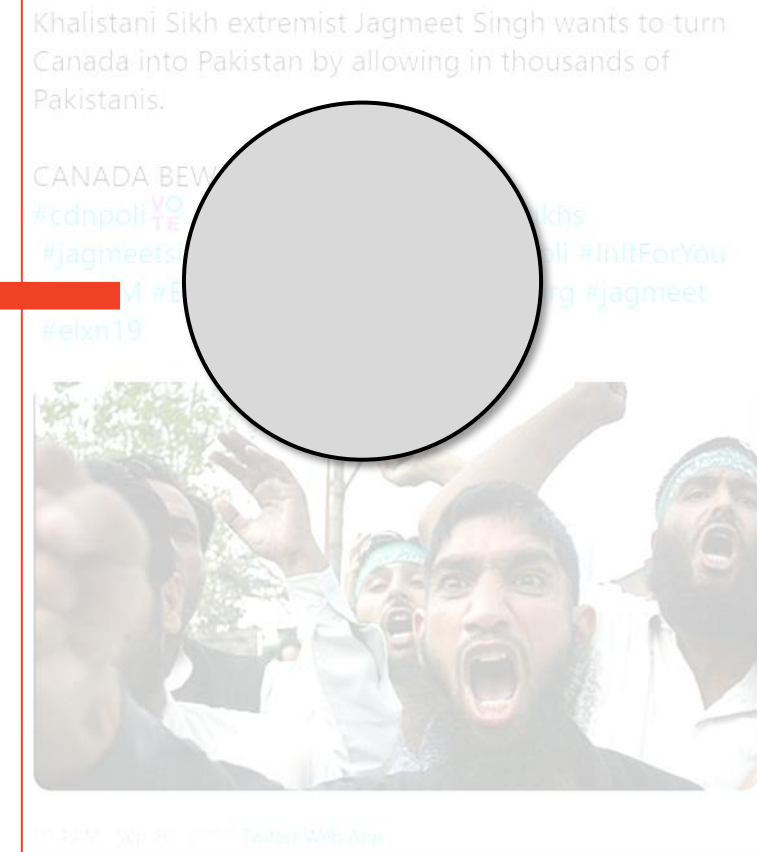
(toxicity score=0.8)

Using Toxicity Scores to Discover Negative Ties

POST

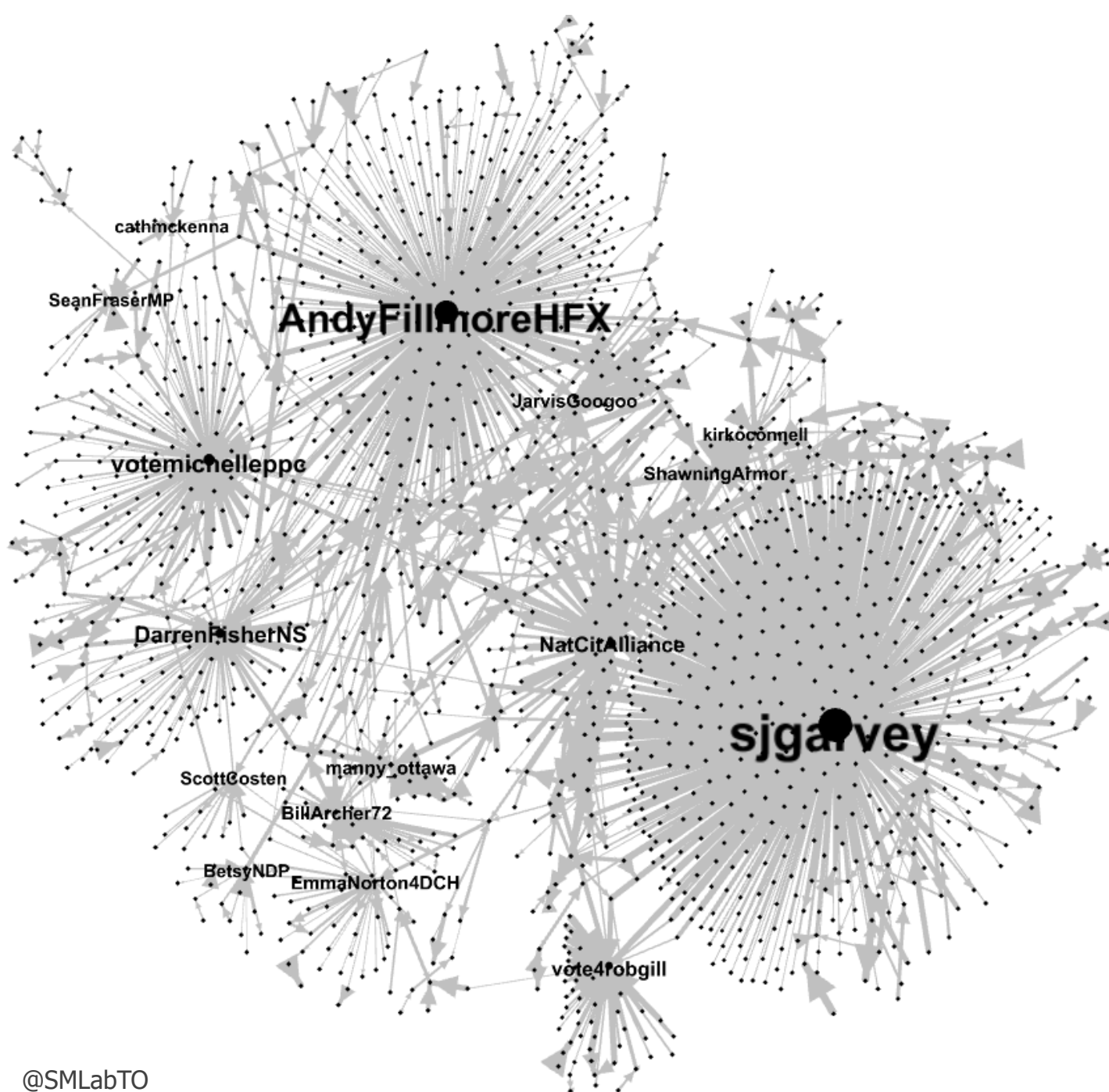


REPLY

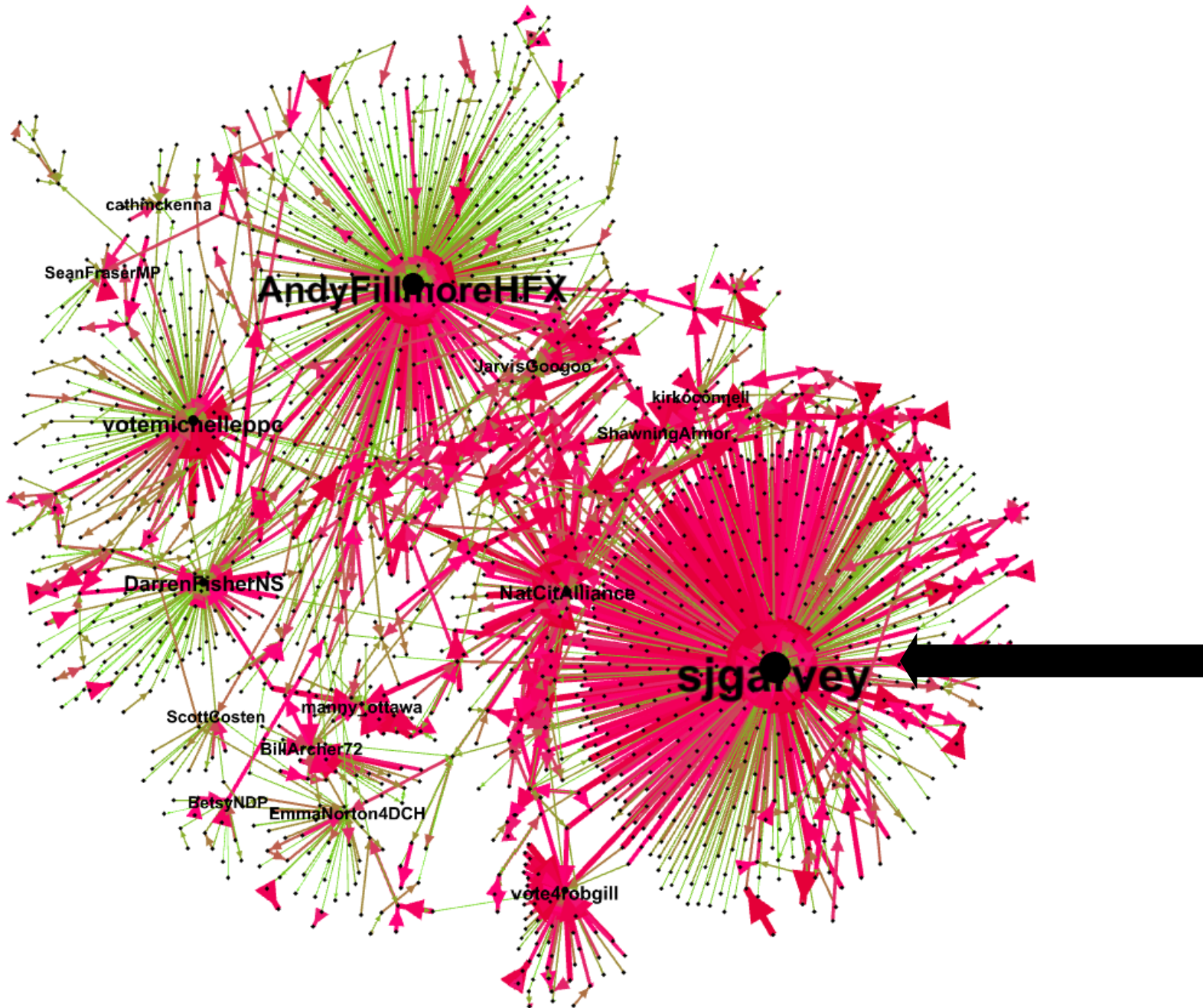


A negative tie
(toxicity score ≥ 0.8)

Tweets from/to a Group of Candidates during the 2019 Federal Election in Canada



Candidates as Toxicity "Magnets"



CANADIANS FIRST



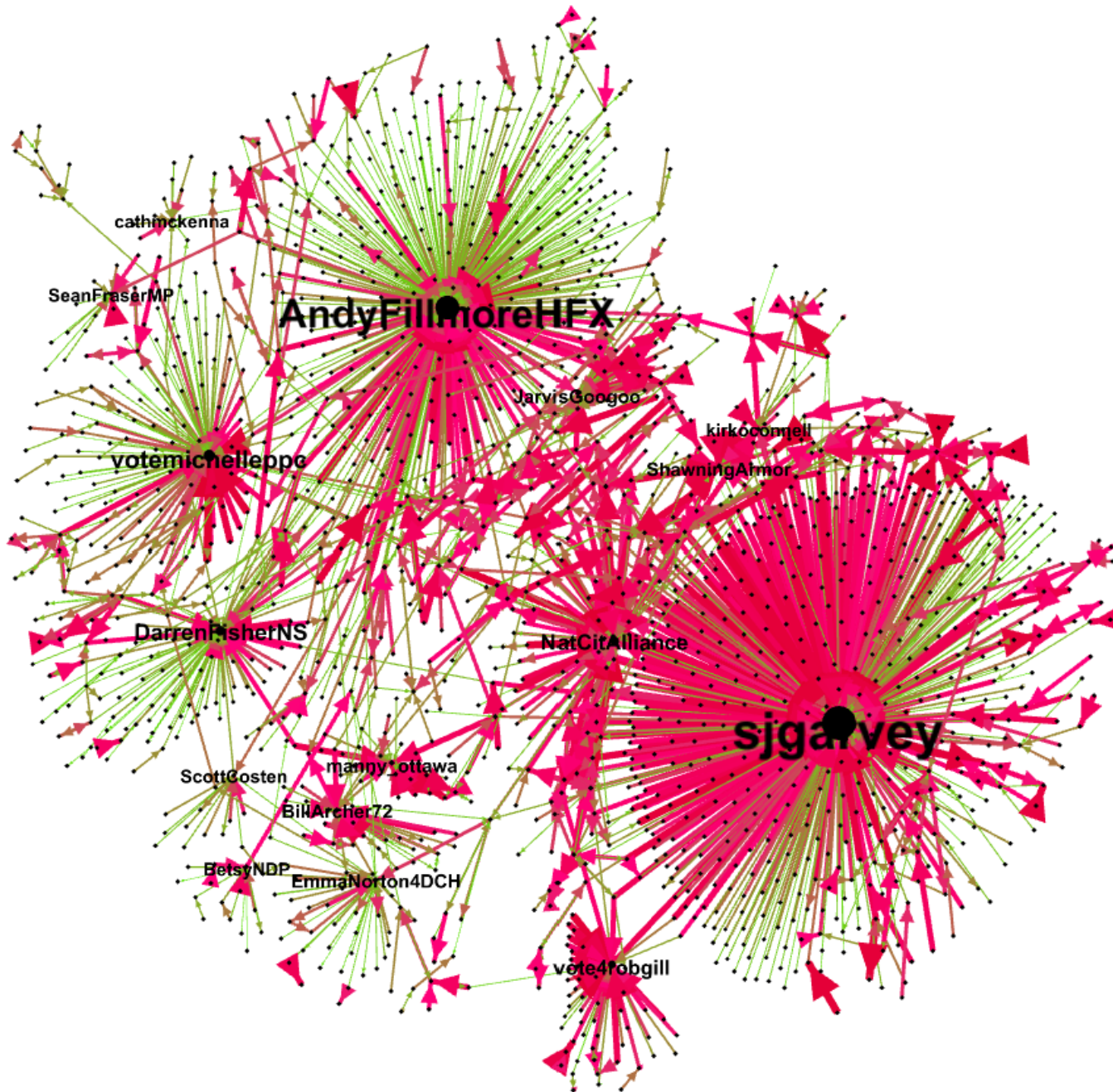
No Fed Income Tax \$22,000 or less, Middle Income 15% Fed Tax
CPP - 3x Increase in Annual Benefits
50,000 or Less Total Immigration per yr

Follow

Stephen Garvey 🇨🇦
@sjgarvey
Founder & Leader of National Citizens Alliance (NCA); pro-Canada/anti-globalist; no PC; zero tolerance for radical extremist ideologies; Canadian people FIRST
📍 Canada 🌐 nationalcitizensalliance.ca 📅 Joined April 2011
66 Following 1,050 Followers

The **National Citizens Alliance** of Canada (NCA) led by Stephen J Garvey is an officially registered federal party.

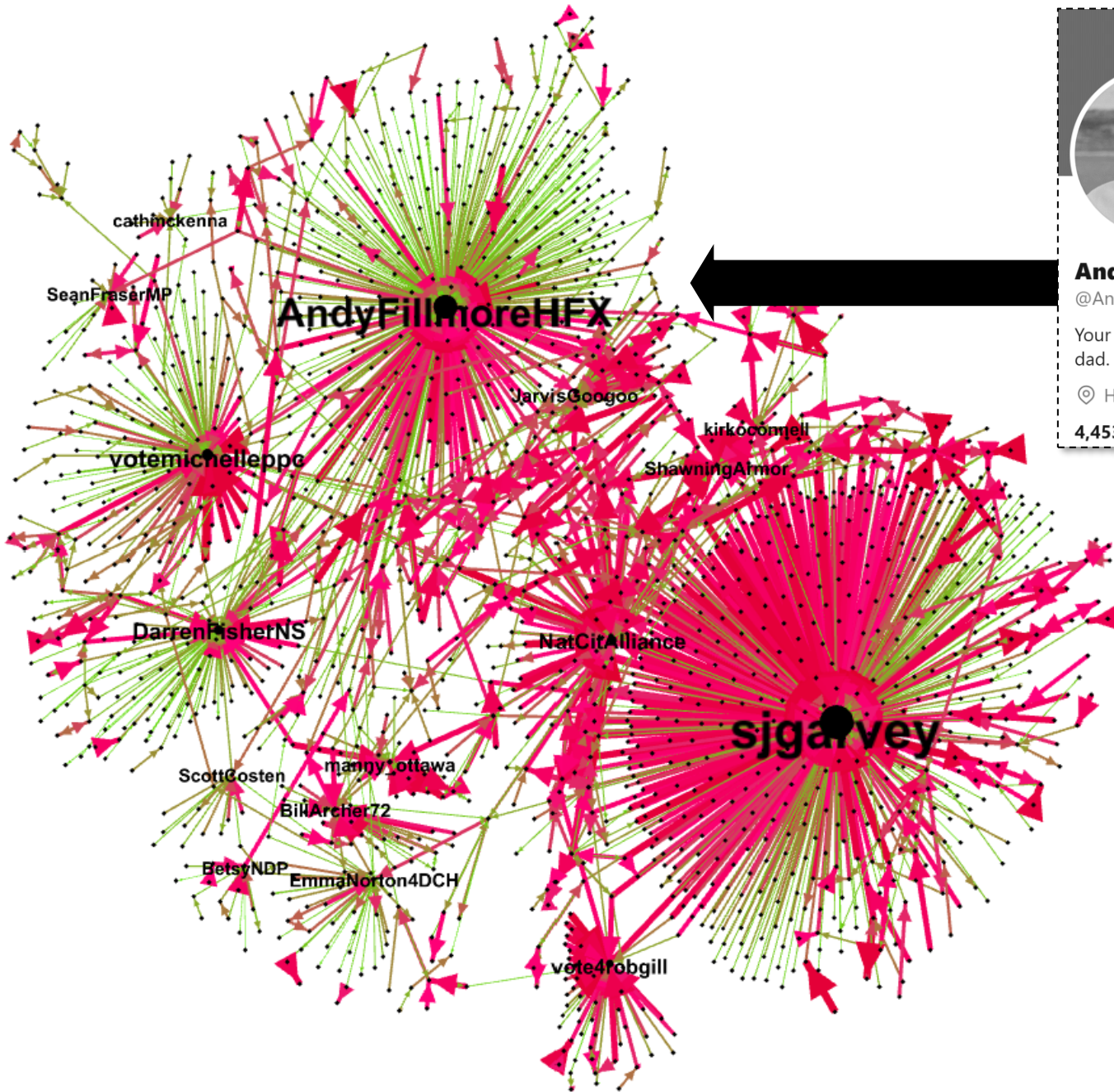
Sample Tweets Targeting the National Citizens Alliance (NCA) Candidate



“ you’re a disgusting piece of sh*t. ”

“ you're a racist little bitch and u ugly as f*ck. ”

“ you are a horrible piece of sh*t. ”







For Halifax

Andy Fillmore ✓
 @AndyFillmoreHFX

Your Liberal Party candidate running for re-election in Halifax. City planner and proud dad. Endorsed by GreenPAC and Unifor. 🇨🇦

📍 Halifax, Nova Scotia 🌐 andyfillmore.ca 📅 Joined June 2010

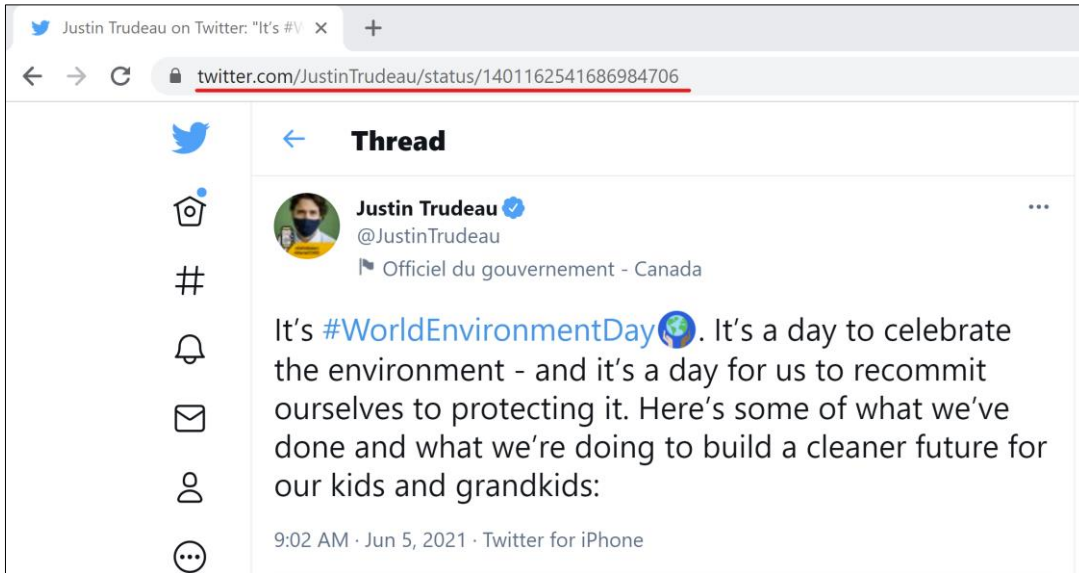
4,453 Following 11.9K Followers

Liberal Party candidate running for re-election in Halifax



How to discover and visualize
networks with toxicity scores

Example: Twitter Thread Data Collection



<https://twitter.com/JustinTrudeau/status/1401162541686984706>



The screenshot shows a web interface titled "My Datasets". At the top, it displays "Server Time: June 06, 2021 18:36 UTC". Below this is a section "Collect data from" with four buttons: "Reddit (LIVE)", "Twitter Thread" (highlighted in blue), "CrowdTangle", and "CSV File". An orange arrow with the number "2" points from this section down to the "Twitter Thread Data Collection" form. The form includes a "Back to My Datasets" button, a title "Twitter Thread Data Collection", and the instruction "Please enter the following details to begin collection". There are two input fields: the first contains "It's #WorldEnvironmentDay" and the second contains "https://twitter.com/JustinTrudeau/status/1401162541686984706". Below the fields is a checkbox for "Email me once job completes". A yellow information box contains text about the collector's capabilities and requirements. At the bottom is a blue "Start Data Collection" button.

My Datasets

Server Time: July 06, 2021 18:15 UTC

Collect data from

- Reddit (LIVE)
- Twitter Thread
- CrowdTangle
- CSV File

Search dataset list...

My Datasets Shared With Me

Total records in the account: 977 / 30000

3%

Dataset Name	Number of Records	Toxicity Analysis	Network Analysis	Download Dataset	Collaborators	Delete
Justin Trudeau: It's #WorldEnvironmentDay. It's a day to celebrate the environment - and it's a day for us to recommit ourselves to protecting it twitter : https://twitter.com/JustinTrudeau/status/1401162541686984706	Records: 977					



← Back to My Datasets

Overview

Dataset Name: **Justin Trudeau: It's #WorldEnvironmentDay. It's a day to celebrate the environment - and it's a day for us to recommit ourselves to protecting it**

Platform: **twitter**

Search Criteria: **https://twitter.com/JustinTrudeau/status/1401162541686984706**

Collection started: **2021-06-06 18:26**

Collection ended: **2021-06-06 18:26**

Records: **977**

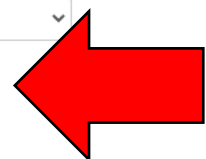
- Posts Per Day
- Word Cloud
- Emoji Cloud
- Top Ten Posters
- Toxicity Analysis
- Network Analysis

Toxicity Analysis

The Toxicity analysis can only run with one language mode at a time. Please select the primary language of your dataset using the drop-down menu below.

English

Start Analysis



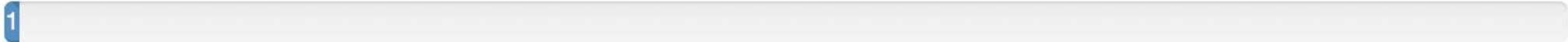
- Toxicity Analysis is based on a machine learning API called [Perspective API](#) by Google. If you already have an API key for Perspective, it can be added to your account under [My Profile](#); otherwise, see [this guide](#) on how to generate a new Perspective API.
- Due to the post length restriction imposed by the API, Communalytic can only analyze the first 3000 characters of each post. Posts longer than 3000 characters will be automatically truncated.
- Link's and URLs will be removed before text is sent to API for analysis.
- The resulting toxicity scores will be added to the export files and available for download via the "Export Posts" and "Export Network" options.
- The API currently supports the following languages: English, French, German, Italian, Portuguese, Spanish and Russian.



Toxicity Analysis

Analysis in Progress ...

You may close this window and visit it later.



Progress: 59 / 4451

Estimated Time Left: 1h 20m

Check progress in **43** sec

Cancel Analysis

1. Toxicity
Analysis



2. Network
Discovery



3. Network
Visualization



4. Network
Export

Toxicity Scores from Perspective API

	Definition	Sample post
Toxicity	Rude, disrespectful, or unreasonable post	“This is one of stupidest things I've read but fuck me I laughed at the second line”
Severe toxicity	Very hateful, aggressive, disrespectful post. This score is less sensitive to posts that include positive uses of curse words	“Fuck off pathetic loser, no one cares about your worthless opinion”
Insult	Insulting, inflammatory, or negative post toward an individual or a group	“How fucking stupid is [Name]? That is pretty fucking stupid. What's next - a deep fake having him say racist things as a "social experiment"?”
Identity attack	Negative post attacking someone because of their identity (including race, gender, sexual orientation, ideology, religion, nationality, etc.)	“You people are a bunch of fags. And I voted for [Political Party Name]”
Profanity	Post with swear words or other obscene language	“Why vote for the [Political Party Name] when you know they won't win shit.”
Threat	Post with an intention to inflict pain, injury, or violence against an individual or group	“Shoot all yellow vests! We have to kill all Nazis!”

<https://support.perspectiveapi.com/s/about-the-api-attributes-and-languages>

1. Toxicity Analysis



2. Network Discovery



3. Network Visualization



4. Network Export

Toxicity Analysis

There were **977** comments analyzed in **English**

Print

Reset Analysis

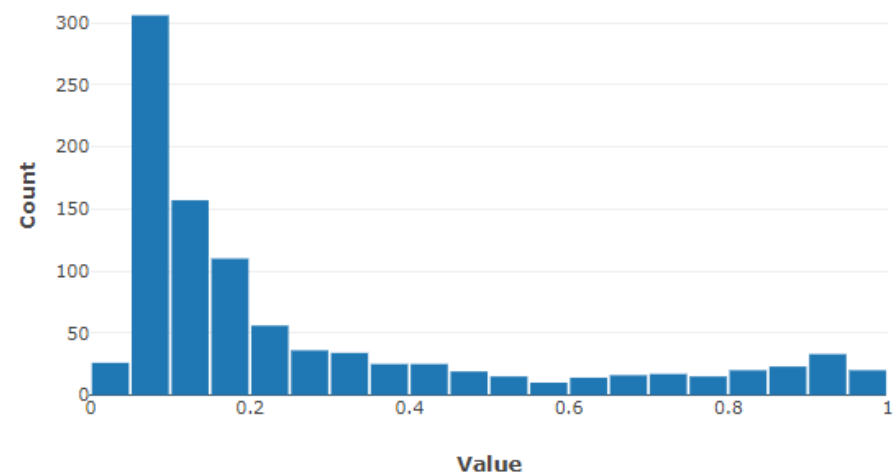
⚠ The Toxicity analysis can only run with one language mode at a time. To change the primary language selection for this dataset, click on the Reset Analysis button and rerun the analysis as needed.

Click on the highest/lowest values to see the top posts for each category

Download the Toxicity analysis results as a [CSV file](#).

	Average for dataset	Highest value	Lowest value
Toxicity ⓘ	0.27	0.99	0.00
Severe toxicity ⓘ	0.15	0.95	0.00
Identity attack ⓘ	0.16	0.93	0.00
Insult ⓘ	0.27	0.98	0.00
Profanity ⓘ	0.18	0.99	0.00
Threat ⓘ	0.16	0.93	0.00

Toxicity Values Distribution Chart



1. Toxicity Analysis



2. Network Discovery



3. Network Visualization



4. Network Export

My Datasets

Server Time: July 06, 2021 18:15 UTC

Collect data from

- Reddit (LIVE)
- Twitter Thread
- CrowdTangle
- CSV File

Search dataset list...

My Datasets Shared With Me

Total records in the account: 977 / 30000

3%

Dataset Name	Number of Records	Toxicity Analysis	Network Analysis	Download Dataset	Collaborators	Delete
Justin Trudeau: It's #WorldEnvironmentDay. It's a day to celebrate the environment - and it's a day for us to recommit ourselves to protecting it twitter : https://twitter.com/JustinTrudeau/status/1401162541686984706	Records: 977					





← Back to My Datasets

Overview

Dataset Name: Justin Trudeau: It's #WorldEnvironmentDay. It's a day to celebrate the environment - and it's a day for us to recommit ourselves to protecting it

Platform: twitter

Search Criteria: <https://twitter.com/JustinTrudeau/status/1401162541686984706>

Collection started: 2021-06-06 18:26

Collection ended: 2021-06-06 18:26

Records: 977

Posts Per Day

Word Cloud

Emoji Cloud

Top Ten Posters

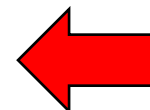
Toxicity Analysis

Network Analysis

Network Analysis

i This step automatically creates a 'who replies to whom' communication network from your Twitter dataset. If you are new to Social Network Analysis (SNA), check out [this SNA primer](#) to learn the fundamentals of Network Analysis.

↓ Generate Network



1. Toxicity Analysis



2. Network Discovery



3. Network Visualization



4. Network Export



← Back to My Datasets

Overview

Dataset Name: Justin Trudeau: It's #WorldEnvironmentDay. It's a day to celebrate the environment - and it's a day for us to recommit ourselves to protecting it

Platform: twitter

Search Criteria: https://twitter.com/JustinTrudeau/status/1401162541686984706

Collection started: 2021-06-06 18:26

Collection ended: 2021-06-06 18:26

Records: 977

Posts Per Day

Word Cloud

Emoji Cloud

Top Ten Posters

Toxicity Analysis

Network Analysis

Network Analysis

This step automatically creates a 'who replies to whom' communication network from your Twitter dataset. If you are new to Social Network Analysis (SNA), check out [this SNA primer](#) to learn the fundamentals of Network Analysis.

Visualize Network

Download Network (GraphML)

Download Network (Gephi)

Reset Network File



1. Toxicity Analysis



2. Network Discovery



3. Network Visualization

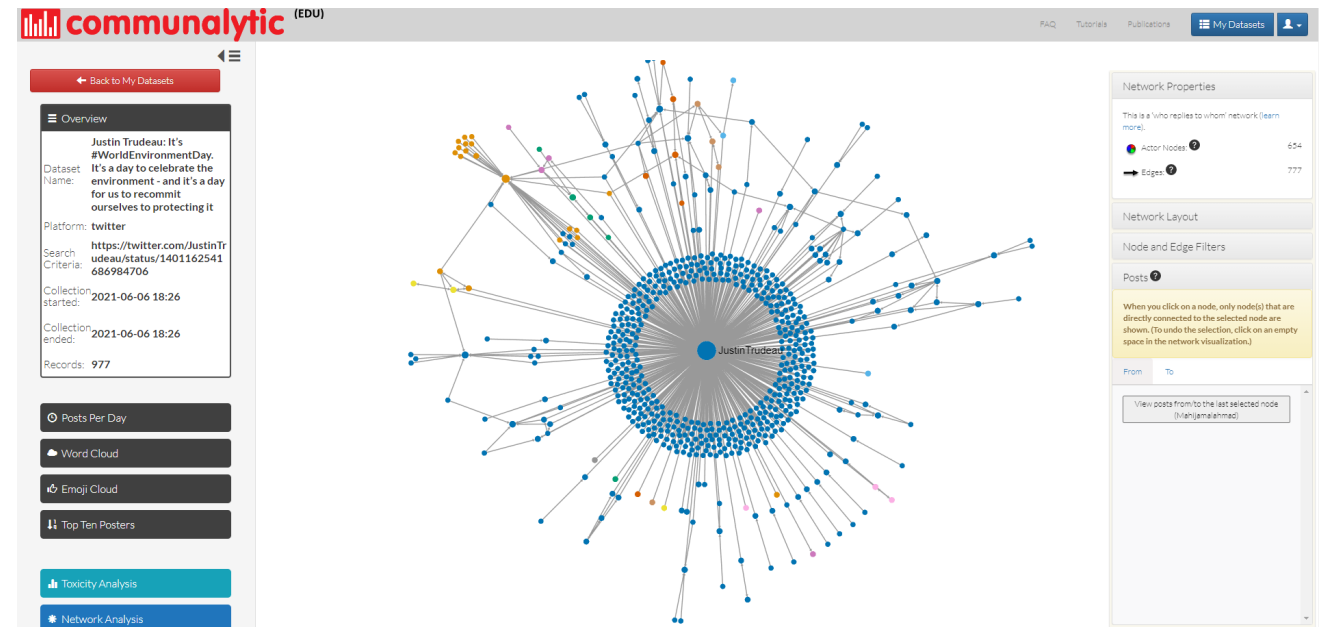
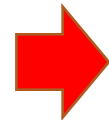


4. Network Export

Example: Network Visualization of a Twitter Thread

The screenshot shows a Twitter thread from Justin Trudeau (@JustinTrudeau) on June 5, 2021. The main tweet discusses World Environment Day and environmental protection, with 335 retweets, 74 quote tweets, and 2,099 likes. Below it are two replies, the first of which mentions that 14% of marine and coastal areas are now protected, up from 1% in 2015.

Network Analysis in Communalytic



1. Toxicity Analysis



2. Network Discovery

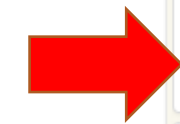


3. Network Visualization



4. Network Export

Example: Network Visualization of a Twitter Thread



Network Properties

This is a 'who replies to whom' network ([learn more](#)).

Actor Nodes: 654

Edges: 777

Network Layout

The default network layout is based on the [SFDP algorithm](#).

Run for 3 sec Stop Reset

Node and Edge Filters

Community Clusters (>=10 nodes): All nodes

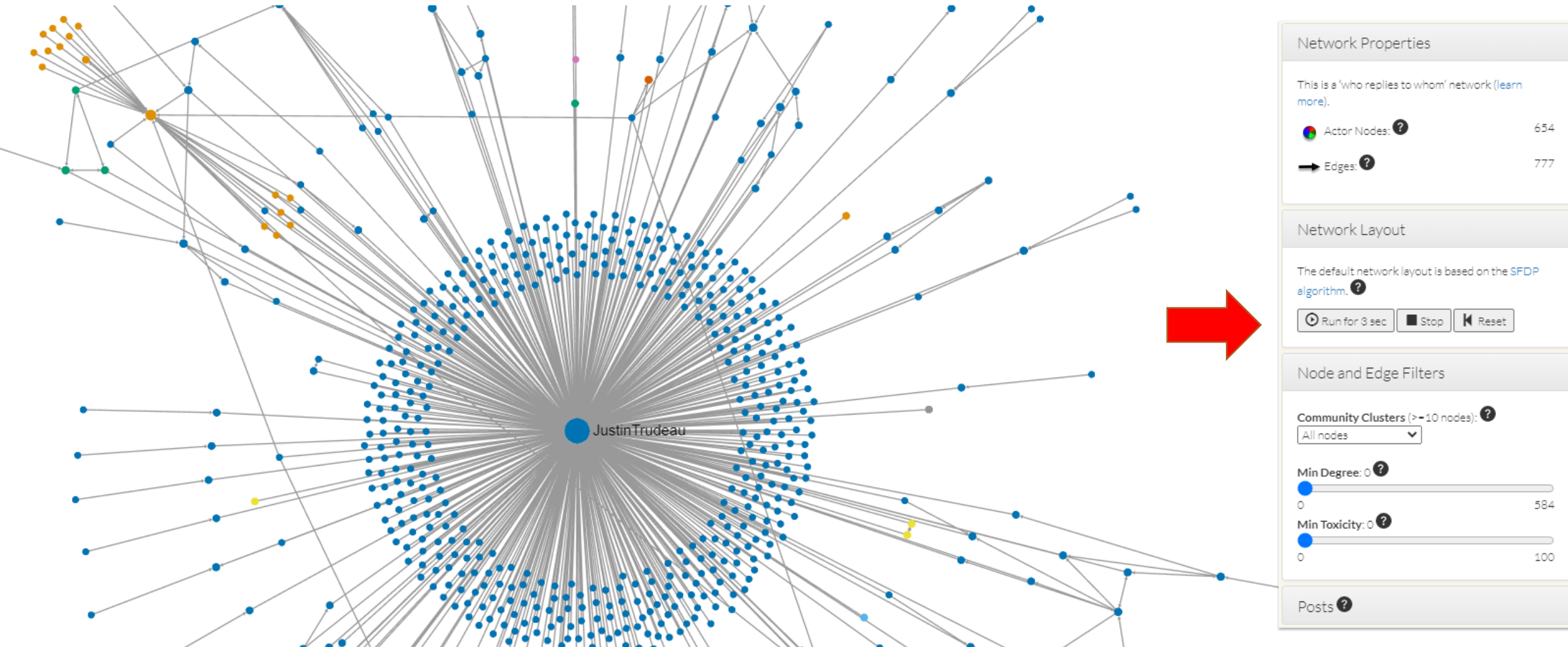
Min Degree: 0 584

Min Toxicity: 0 100

Posts



Example: Network Visualization of a Twitter Thread



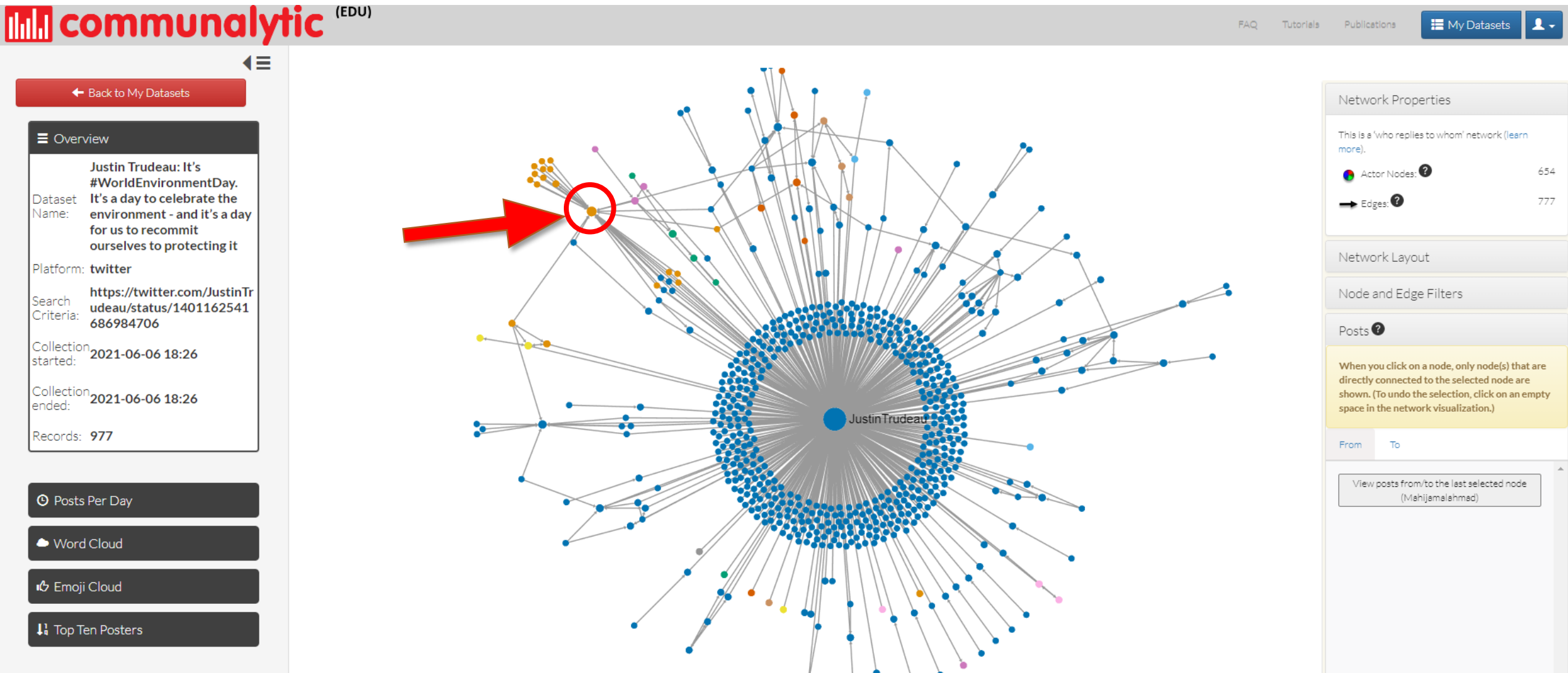
1. Toxicity Analysis

2. Network Discovery

3. Network Visualization

4. Network Export

Example: Network Visualization of a Twitter Thread



1. Toxicity Analysis



2. Network Discovery

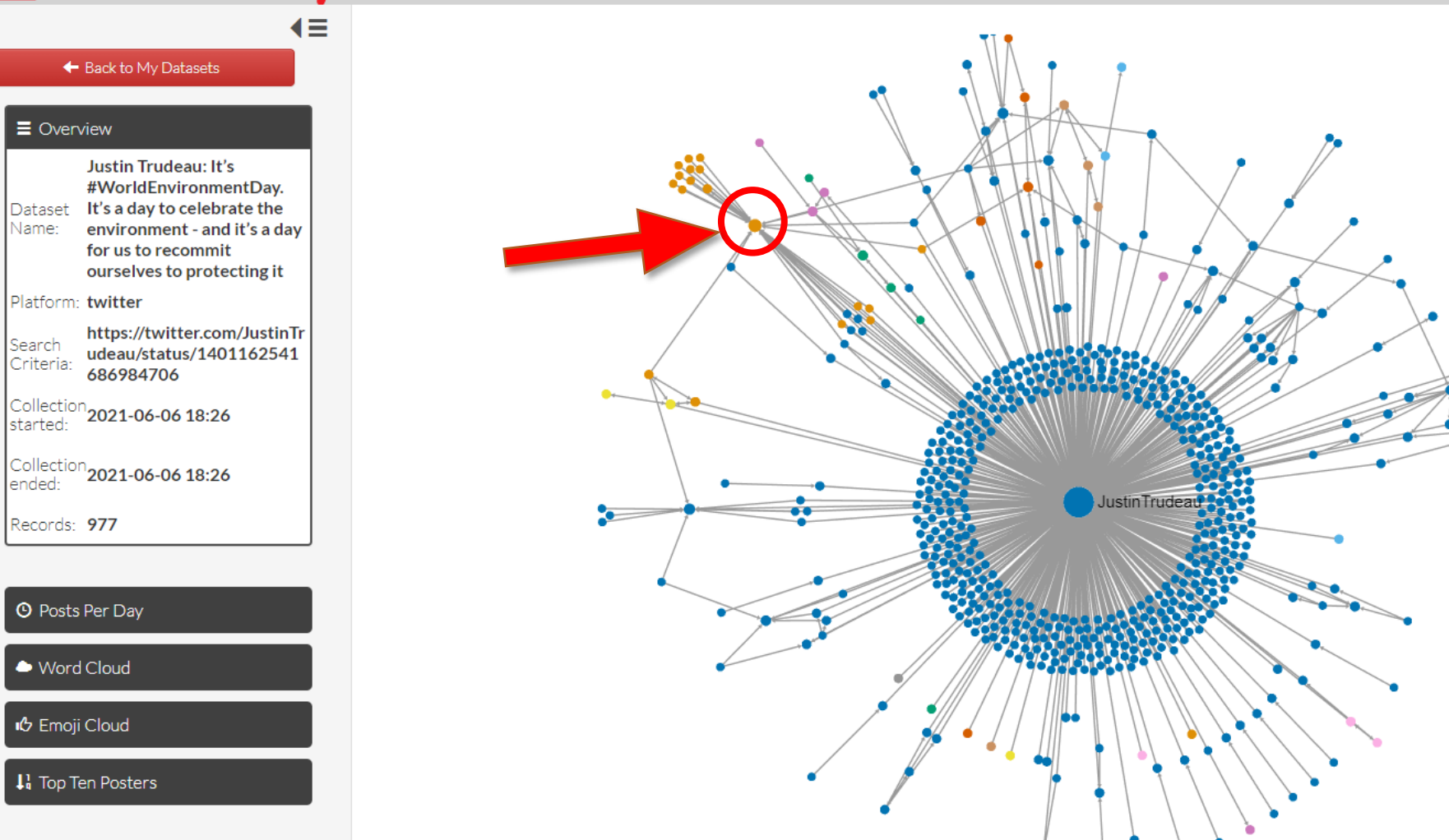


3. Network Visualization



4. Network Export

Example: Network Visualization of a Twitter Thread



Posts ?

When you click on a node, only node(s) that are directly connected to the selected node are shown. (To undo the selection, click on an empty space in the network visualization.)

From 2 To 25

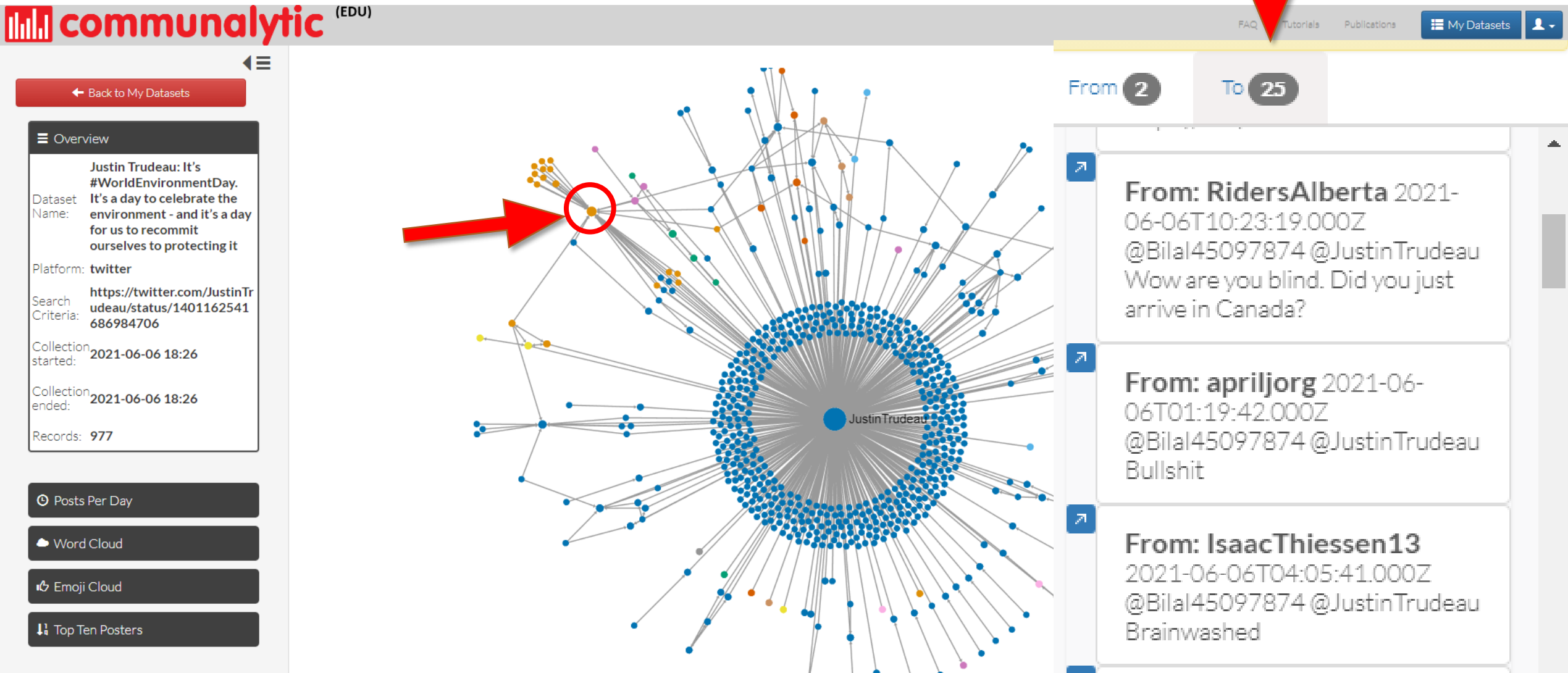
The following are posts from **Bilal45097874**:

To: strgIndigenous2021-06-06T14:56:50.000Z
@strgIndigenous @JustinTrudeau
Actually he is a honest person and take caring of nation such as his family.. <https://t.co/s6df3F3vcj>

To: JustinTrudeau2021-06-05T13:10:02.000Z
@JustinTrudeau You are always think about people's and you are good person for every religion those are living in your country. I salute you sir 🙌🙌🙌



Example: Network Visualization of a Twitter Thread



1. Toxicity Analysis



2. Network Discovery

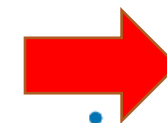
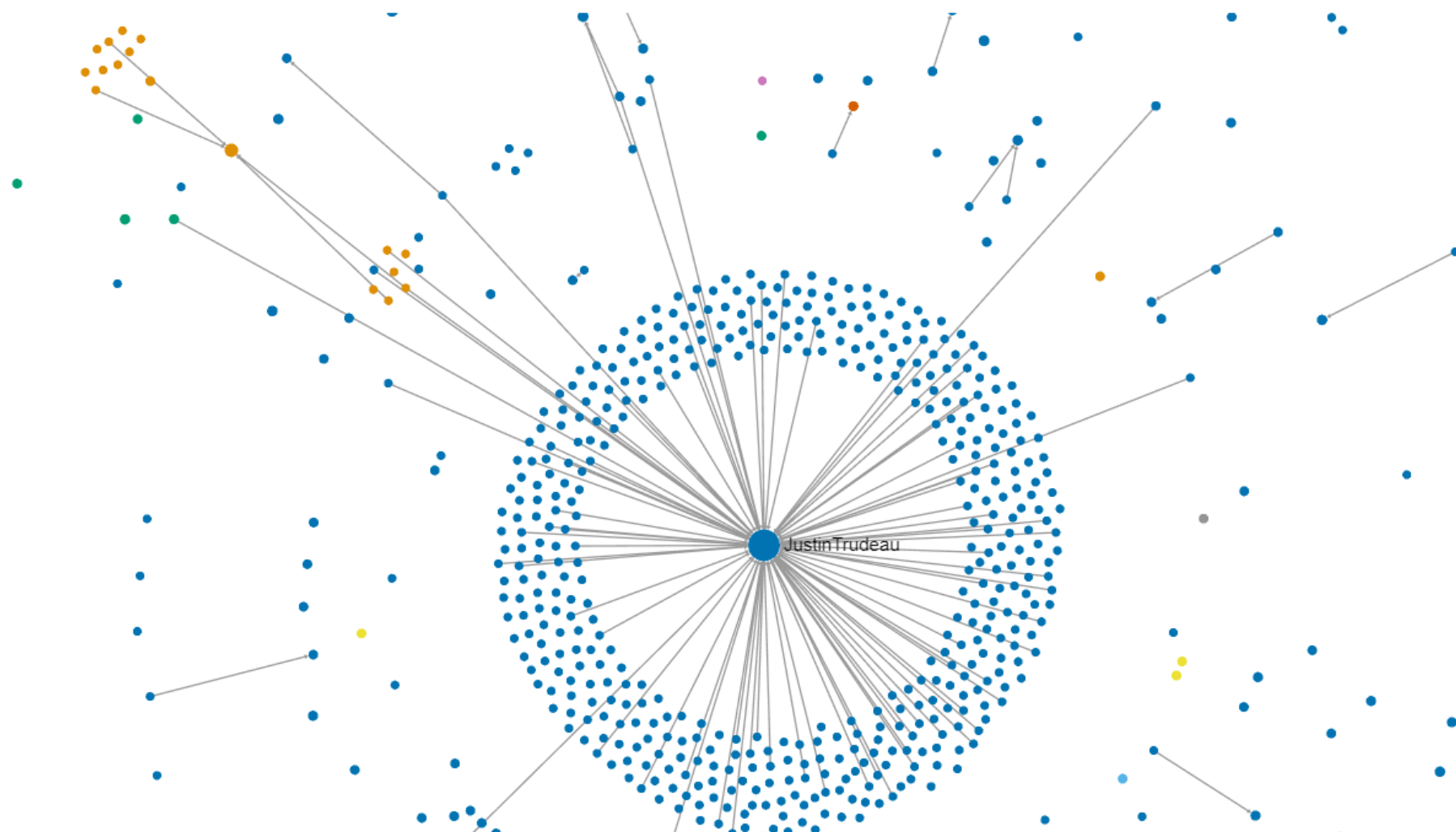


3. Network Visualization



4. Network Export

Example: Filter edges based on Toxicity Scores (≥ 0.7)



Network Properties

This is a 'who replies to whom' network ([learn more](#)).

Actor Nodes: ? 654

Edges: ? 777

Network Layout

The default network layout is based on the SFDP algorithm. ?

Run for 3 sec Stop Reset

Node and Edge Filters (Applied)

Community Clusters (>=10 nodes): ?

All nodes

Min Degree: 0 ?

0 584

Min Toxicity: 70 ?

0 100

Posts ?

1. Toxicity Analysis



2. Network Discovery

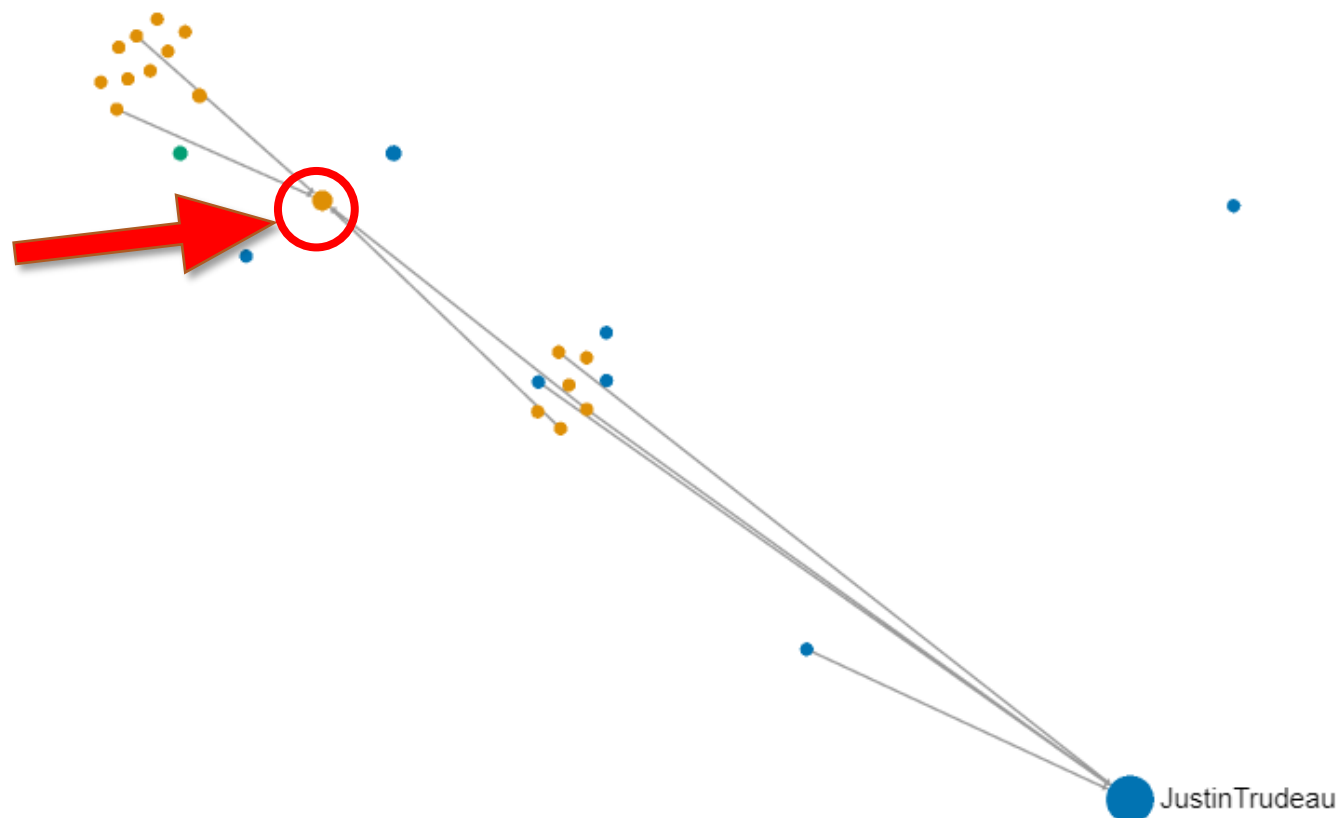


3. Network Visualization



4. Network Export

Example: Filter edges based on Toxicity Scores (≥ 0.7)



From 0 To 4

The following are replies to **Bilal45097874**:

➤ **From: AshFloyd3** 2021-06-06T02:27:47.000Z
@Bilal45097874 @JustinTrudeau
He is a woke idiot that is destroying this country.

➤ **From: coronado_cortez** 2021-06-05T17:44:44.000Z
@Bilal45097874 @JustinTrudeau
Ur a idiot

➤ **From: DazEtheGreat2** 2021-06-05T13:38:09.000Z
@Bilal45097874 @JustinTrudeau
What a moron. You forgot your

1. Toxicity Analysis



2. Network Discovery



3. Network Visualization



4. Network Export



← Back to My Datasets

Overview

Dataset Name: Justin Trudeau: It's #WorldEnvironmentDay. It's a day to celebrate the environment - and it's a day for us to recommit ourselves to protecting it

Platform: twitter

Search Criteria: https://twitter.com/JustinTrudeau/status/1401162541686984706

Collection started: 2021-06-06 18:26

Collection ended: 2021-06-06 18:26

Records: 977

Posts Per Day

Word Cloud

Emoji Cloud

Top Ten Posters

Toxicity Analysis

Network Analysis

Network Analysis

This step automatically creates a 'who replies to whom' communication network from your Twitter dataset. If you are new to Social Network Analysis (SNA), check out [this SNA primer](#) to learn the fundamentals of Network Analysis.

Visualize Network Download Network (GraphML) Download Network (Gephi) Reset Network File



1. Toxicity Analysis



2. Network Discovery

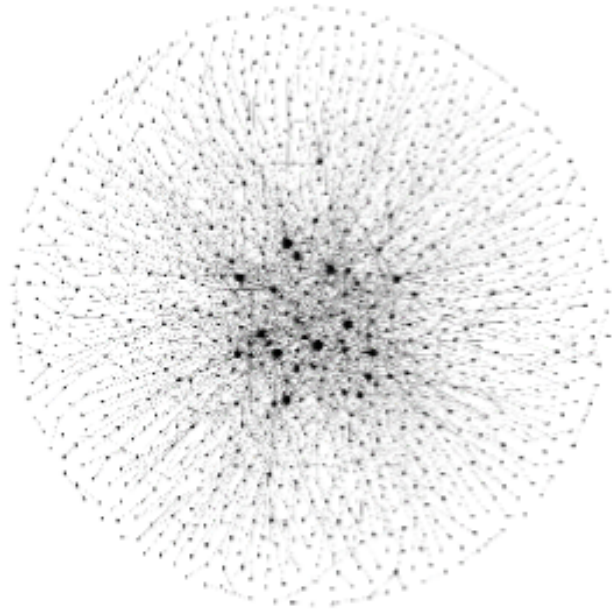


3. Network Visualization

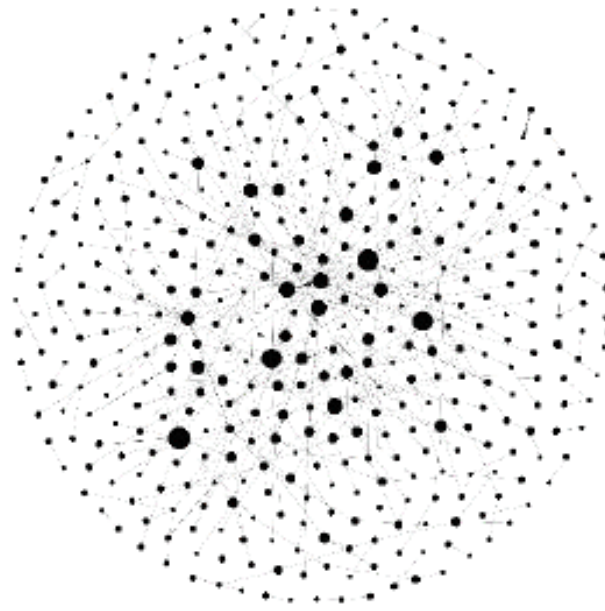


4. Network Export

Advantage of Exporting Network Data: Ability to examine additional toxicity scores and calculate additional SNA metrics



(a) Toxicity Sub-networks



(b) Identity Attack Sub-networks

Example: Network visualizations of anti-social interactions in **r/metacanada** with a threshold of 0.8 (Node size = Out-degree centrality)

Gruzd, Mai, & Vahedi. (2020). Studying Anti-Social Behaviour on Reddit with Communalytic. *Advance. Preprint.* <https://doi.org/10.31124/advance.12453749.v1>

1. Toxicity Analysis



2. Network Discovery

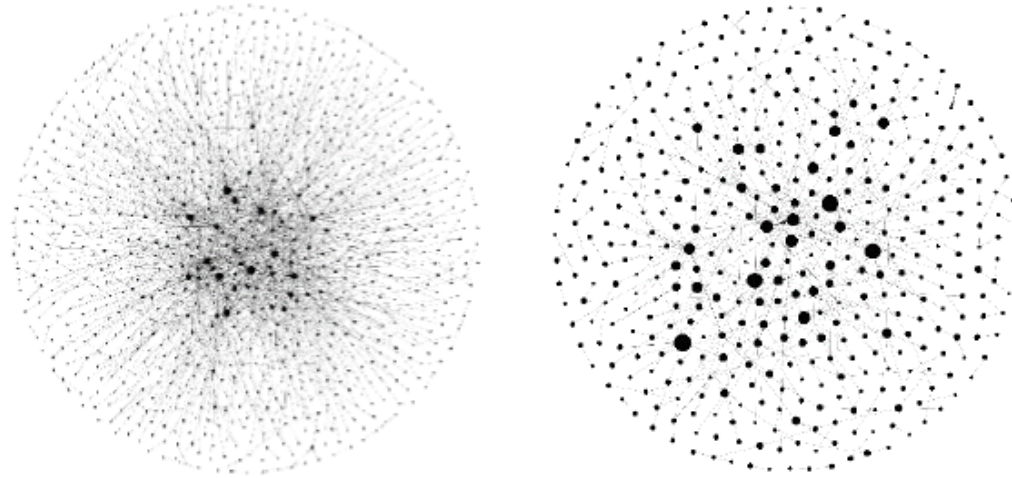


3. Network Visualization



4. Network Export

Advantage of Exporting Network Data: Ability to examine additional toxicity scores and calculate additional SNA metrics



Example: Network visualizations of anti-social interactions in **r/metacanada** with a threshold of 0.8 (Node size = Out-degree centrality)

	Toxicity	Identity Attack
Number of Nodes	987	435
Number of Edges	1979	485
Density	0.002	0.003
Clusters	35	49
Reciprocity	0.142	0.067
Modularity (based on label propagation community)	0.134	0.712

1. Toxicity Analysis



2. Network Discovery



3. Network Visualization



4. Network Export

Today's agenda



Introduction to Social Network Analysis (Philip Mai)



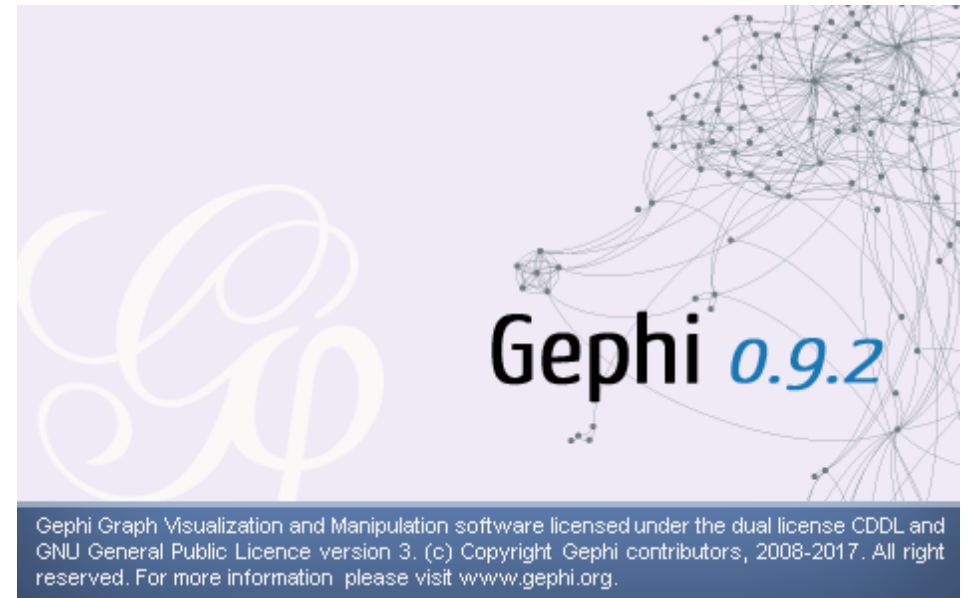
Analyzing “Signed” Networks in Communalytic (Anatoliy Gruzd)



Exporting and analyzing networks with 3rd-party programs like Gephi (Alyssa Saiphoo)

Gephi

- Free to [download](#)
- [Tutorials](#)
- Some of the same tools as the built in visualizer
- Data laboratory feature



Case Study

Public replies to COVID-19 related tweets from

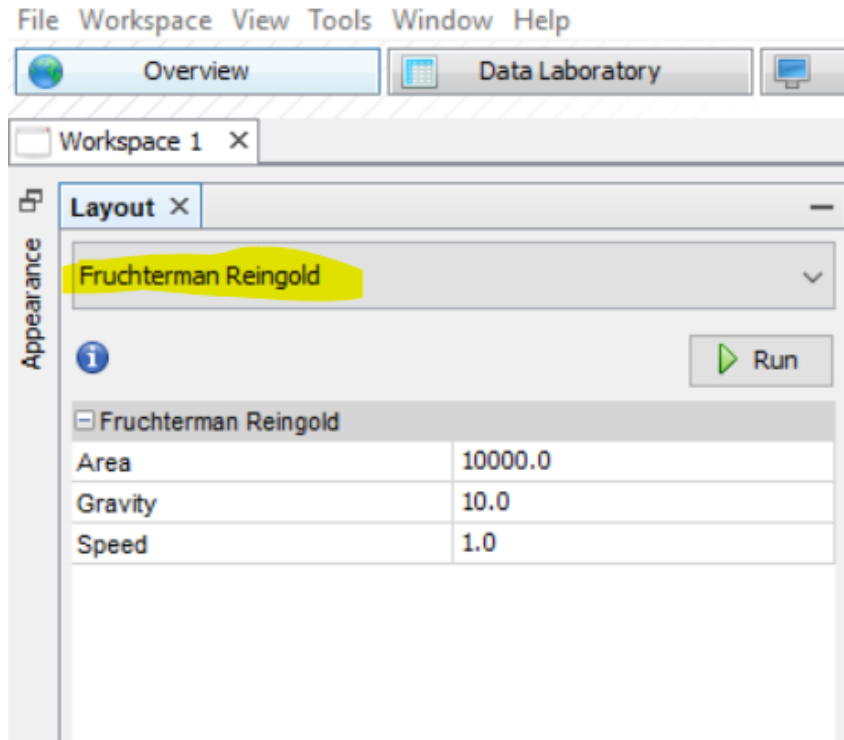
Donald Trump (Dataset 1, n = 5k) and Joe Biden (Dataset 2, n = 5k)



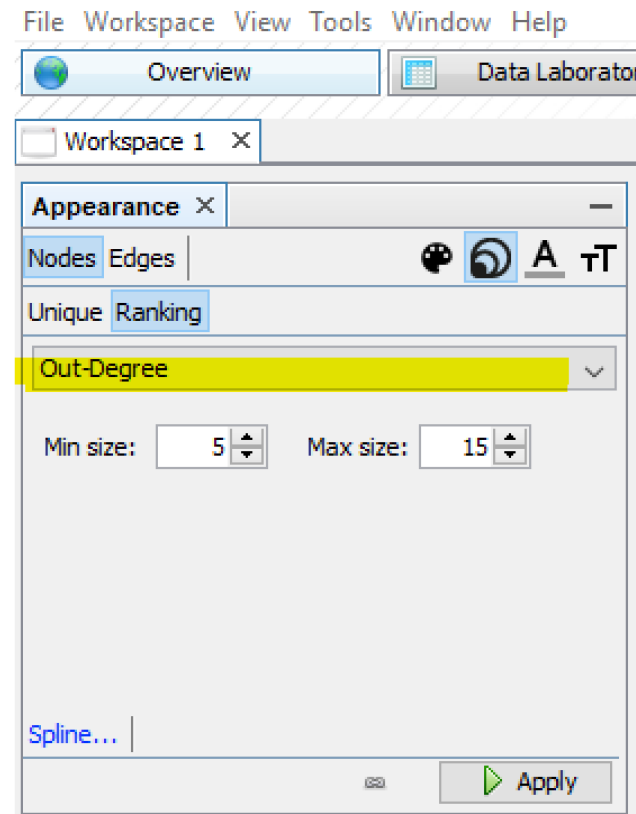
<https://commanalytic.com/2021/04/27/commanalytic-how-to-guide-exploratory-network-and-toxic-analysis/>

Visualizing Networks with Gephi

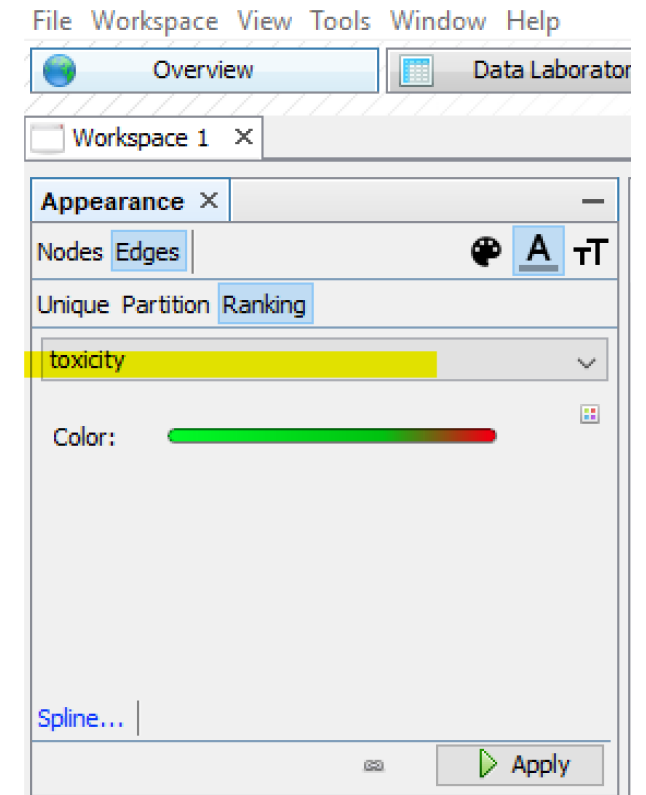
1. Apply a layout algorithm



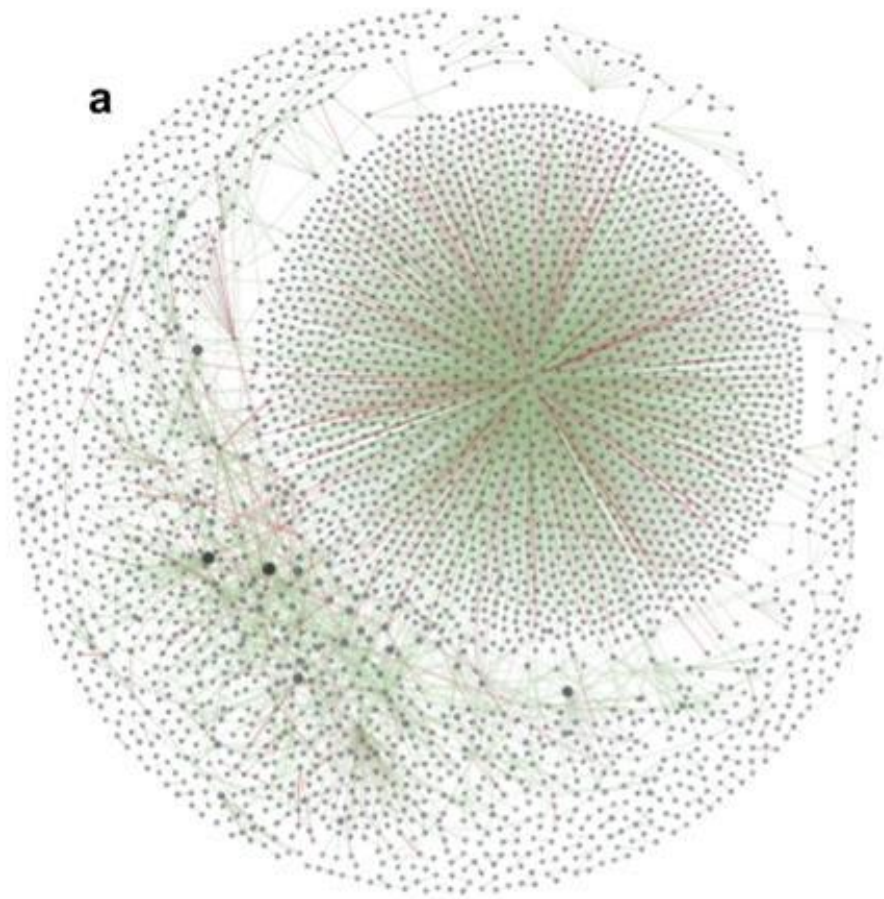
2. Change node size



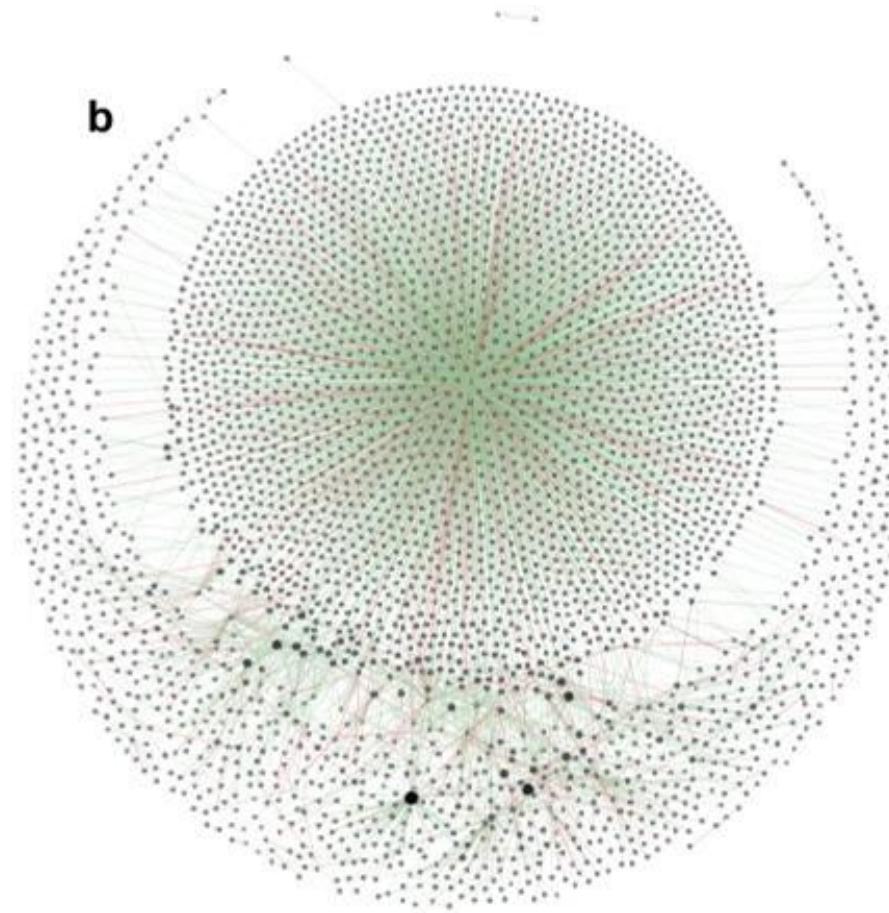
3. Change edge colour



Visualizing networks with Gephi



a) Replies to Donald Trump's tweet



b) Replies to Joe Biden's tweet

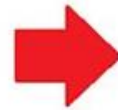
Nodes = Twitter accounts

Edges = Replies

Green edges = non-toxic replies

Red edges = likely **toxic** replies (toxicity scores > 0.8)

Identifying Active Users in the Network



2

Table view context menu options:

- Edit node
- Select on Overview
- Select neighbour nodes on table
- Select related edges
- Delete
- Move to...
- Copy to...



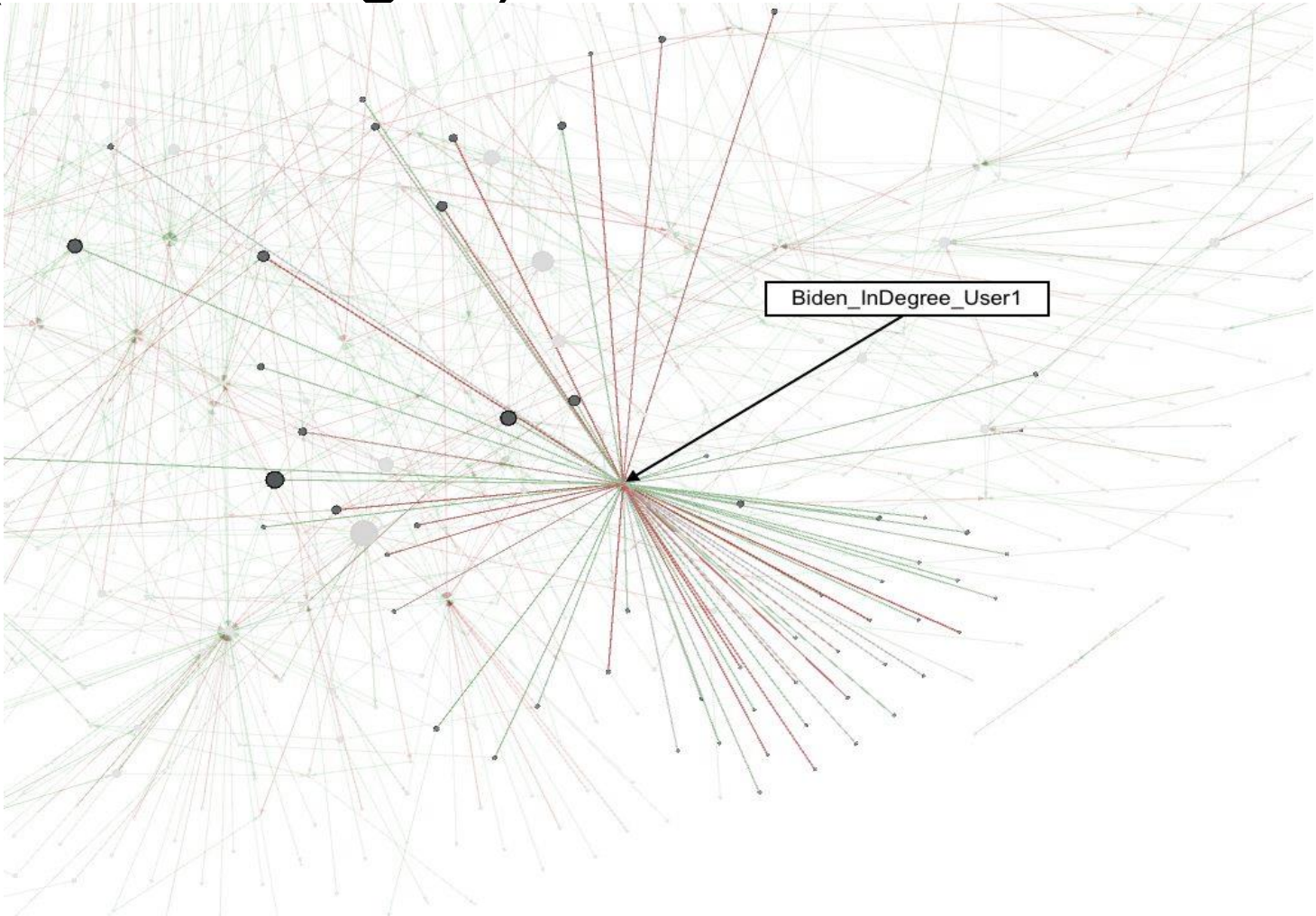
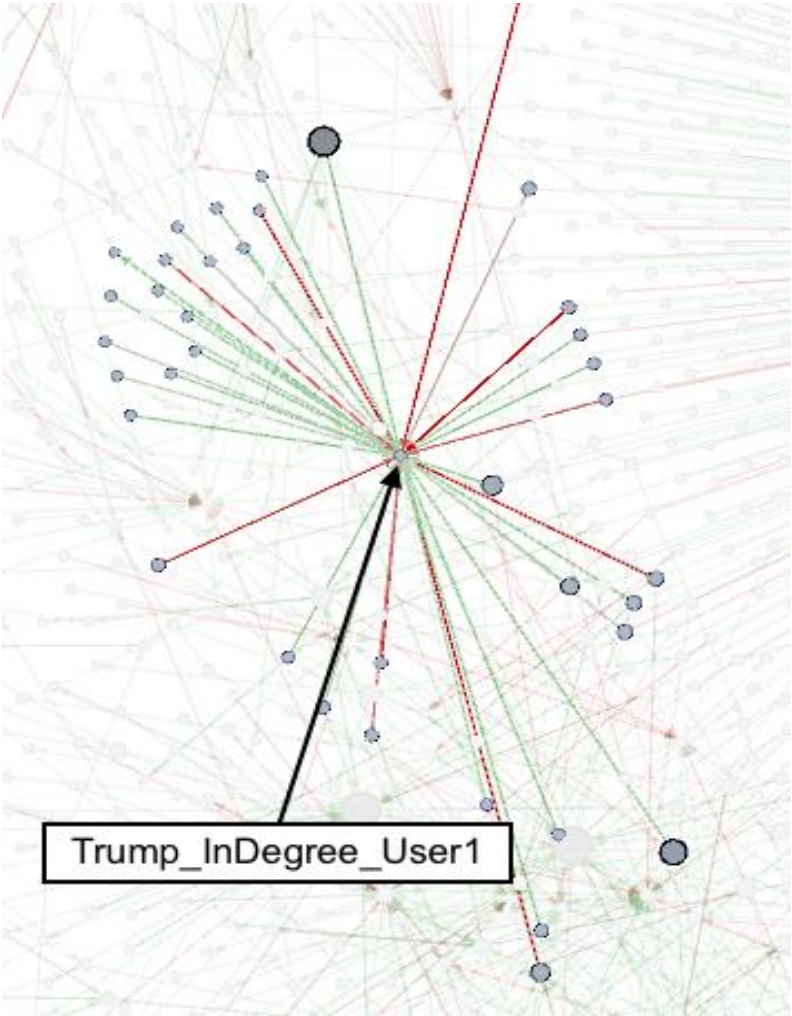
3

Source	Target	flirtation	sexually_explicit	threat	profanity	insult	identity_attack	severe_toxicity	toxicity	like_count	reply_count	retweet_count	quote_count	lan...	text	date
1092725112...	51213244	...	0	...	0.287639	0.059538	0.152658	0.155984	0.216419	0.129527	0.100359	0.198815	0	0	0	0	en	@DianeNoMoreLes @srale...	2020-12-17T12:32:31.000Z
51213244	2973969367	...	1	...	0.333201	0.038577	0.114487	0.09265	0.306189	0.055418	0.066937	0.241932	1	1	0	0	en	@fanofequality @sralex6 ...	2020-12-14T14:24:11.000Z
51213244	51213244	...	2	...	0.37718	0.051429	0.067251	0.060751	0.143291	0.038314	0.038633	0.111387	1	2	0	0	en	@fanofequality @sralex6 ...	2020-12-14T14:29:54.000Z
51213244	36711022	...	3	...	0.219366	0.009202	0.023362	0.011361	0.128474	0.034705	0.016183	0.098784	1	0	0	0	en	@DanRather The vaccine i...	2020-12-14T07:07:44.000Z
51213244	2812144567	...	4	...	0.250825	0.033409	0.030972	0.029352	0.194106	0.046589	0.02665	0.13122	13	11	1	1	en	@sralex6 @DanRather Th...	2020-12-14T07:04:54.000Z
9563245230...	8751062236...	...	5	...	0.266388	0.068176	0.052847	0.075507	0.03165	0.021926	0.025043	0.062953	23	1	0	0	en	@SusanWLinden1 @DanR...	2020-12-14T07:25:44.000Z
9563245230...	2804379390	...	6	...	0.200353	0.019463	0.024878	0.029595	0.02728	0.017175	0.013471	0.05203	0	0	0	0	en	@dpwisey @SusanWLinde...	2020-12-17T10:06:29.000Z

Identify users that are **sending** toxic tweets (larger nodes)



Identify users that are **receiving** toxic tweets (lots of edges)



Examining the Most “Toxic” Accounts in the Data Laboratory feature in Gephi

1. Apply a filter to the network

Step 1: Click on the 'Filters' tab in the top left panel.

Step 2: Click on the 'Statistics' sub-tab in the top left panel.

Step 3: Select the 'Range' filter from the 'Library' list.

Step 4: Click on the 'Range (toxicity)' filter in the 'Queries' section.

Step 5: Adjust the 'Range (toxicity) Settings' slider to 0.994484.

Step 6: Click the 'Filter' button at the bottom of the settings panel.

On the right side of the interface, 'Step 1' points to the 'Filters' tab, 'Step 2' points to the 'Run' button for 'Average Degree', and 'Step 3' points to the 'Run' button for 'PageRank'.

2. Further investigate in the Data Laboratory tab

The 'Data Laboratory' tab is selected in the top navigation bar. Below it, the 'Data Table' is displayed with columns for node attributes and network metrics.

Id	Label	Interval	user_tweet...	user_listed...	user_followin...	user_follower...	user_des...	user_crea...	user_lo...	user_pro...	user_v...	In-Degree	Out-De...	Degree
36711022	DanRather		18069	7270	683	1759341	Journalist, ...	2009-04-30...	New Yor...	False	False	2536	0	2536
29128240	KareemF...		41515	18	712	800	Husband, ...	2009-04-06...	Pennsylv...	False	False	71	15	86
8751062...	SusanWL...		8506	1	166	60	Love being ...	2017-06-14...	St Alban...	False	False	53	4	57
69293662	KittyKahlo		13762	36	1006	1208	Dr-Mrs. mo...	2009-08-27...		False	False	43	1	44
989797831	KeithGra...		19095	2	1160	656	I'm a Celtic...	2012-12-04...	McDonou...	False	False	36	20	56
2812144...	sralex6		30742	3	1012	716	****Let's ...	2014-09-15...		False	False	30	3	33
1264454...	Meidas_...		16633	23	21751	20551	RN_CCRN...	2020-05-24...	Honolulu...	False	False	21	1	22
412494921	Masterof...		6171	11	510	371	Master Des...	2011-11-14...	Worldwide	False	False	20	1	21
32643406	emkinger		467	0	764	61	Emily Ro...	2009-04-17...	Bremerto...	False	False	16	2	18
51213244	DianeNo...		20706	6	1228	900	Catholic - Y...	2009-06-26...	Boston MA	False	False	13	4	17
266204577	tabby808		36861	4	4574	496		2011-03-14...	Hawaii	False	False	13	3	16
86770661	WallyE...		10670	10	266	270	#Trump?	2017-10-05	Central T...	False	False	11	7	18



Session 5: Getting Started with Communal analytic Social Network Analysis of Signed Networks

July 8, 2021

SocialMediaLab.ca