



Conversations Around Insider and Organizational Threat

Luke Osterritter

losterritter@cmu.edu



CarnegieMellon

Center for Computational Analysis of
Social and Organizational Systems
<http://www.casos.cs.cmu.edu/>



What is an "Insider Threat"?

- Malicious Insider
 - a current or former employee, contractor, or business partner who meets the following criteria:
 - has or had authorized access to an organization's network, system, or data
 - has intentionally exceeded or intentionally used that access in a manner that negatively affected the **confidentiality, integrity, or availability** of the organization's information or information systems
- Can also be inadvertent (non-malicious)



Source: The CERT Insider Threat Center

11 June 2020

Osterritter

2



Carnegie Mellon
IST Institute for Software Research

Conversations around Insider Threat

- Why look at public conversation? Unlikely to find any insider threats...
- ...but, there may be actors trying to shape the conversation to their own ends – corporations, nation states, etc.
- Understanding the conversation will lead to informed research
- **Research question:** Can dynamic network analysis be used to discover the nature of public conversations around insider threat and related organizational threats?

CASOS
11 June 2020 Osterritter 3

Carnegie Mellon
IST Institute for Software Research

Table 1. Set of hashtags used for tweet collection by conversation category

Hashtag Collection

Category	Hashtags			
General	#insiderthreat	#insiderattack	#cyberespionage	#dataloss
Corporate	#industrialespionage	#tradesecrets	#embezzlement	#embezzling
Nation-state	#militarysecrets	#spy	#spying	#spies

CASOS
11 June 2020 Osterritter 4



Carnegie Mellon
IST Institute for Software Research

Collection Method

- Use Python package twarc to retrieve tweets from Twitter Search API based on hashtag query
- Tweets collected between March 27th and April 15th 2020 (data has some gaps)
- Import Twitter JSON data into ORA – ORA handles creating derived networks and basic stats.
- Use ORA for reporting and visualization

CASOS
11 June 2020 Osterritter 5

Carnegie Mellon
IST Institute for Software Research

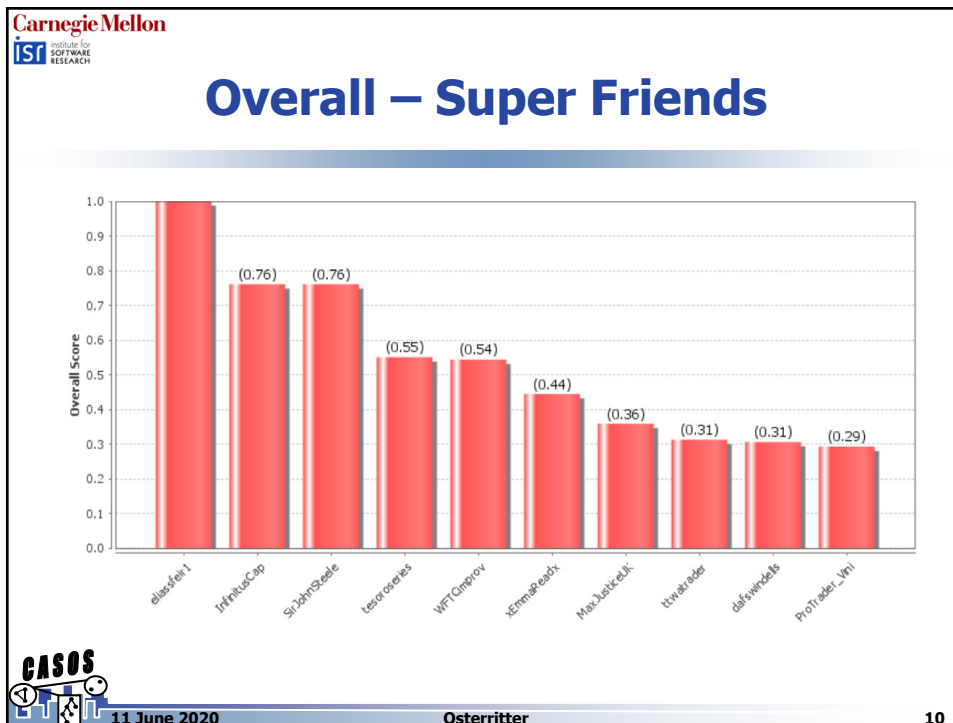
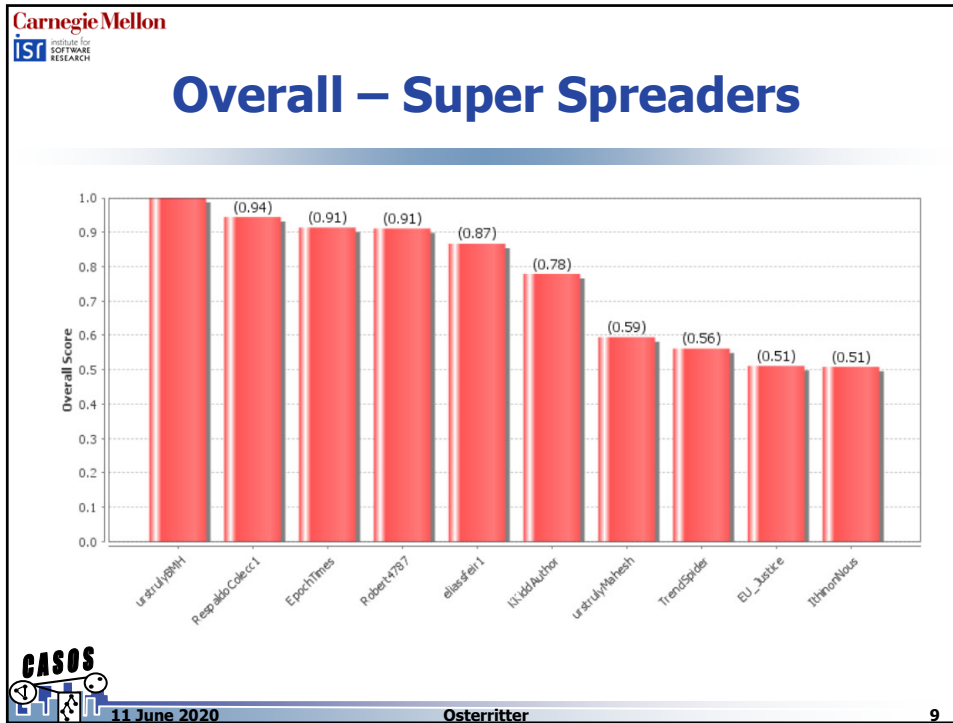
Data Description

- 5 nodesets: Agent, Hashtag, Location, Tweet, URL

Network	Twitter JSON All Hashtags
First tweet date	2013-01-15 07:06:07-05
Last tweet date	2020-04-15 08:45:03-04
Number of tweets	13640
Number of tweets with geotag	9
Number of tweets with URL	4939
Number of retweets	5826
Number of tweeters	6260
Number of verified tweeters	145
Number of news agency tweeters	17
Number of mentions	4233
Number of distinct hashtags	6795
Number of distinct hashtags used more than once	3212
Number of distinct words	0
Number of distinct words used more than once	0
Number of distinct locations	9

CASOS
11 June 2020 Osterritter 6






Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Overall Takeaways

- Difficult to find anything of note in the whole collection
- “Spy” hashtag has a lot of out-of-scope discourse
 - Movie and TV
 - Video games (Team Fortress 2)
 - Novels, books, stories, etc.
 - ES Futures vs SPY (refuse to look deeper into this)



CASOS

11 June 2020 Osterritter 11

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

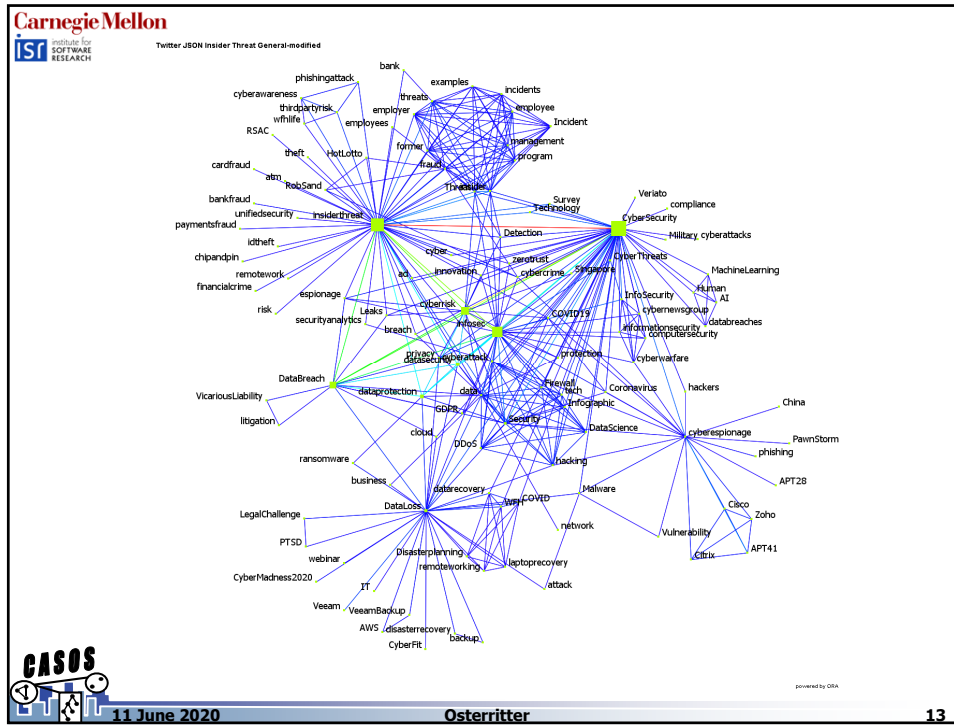
Inside Threat Tweets

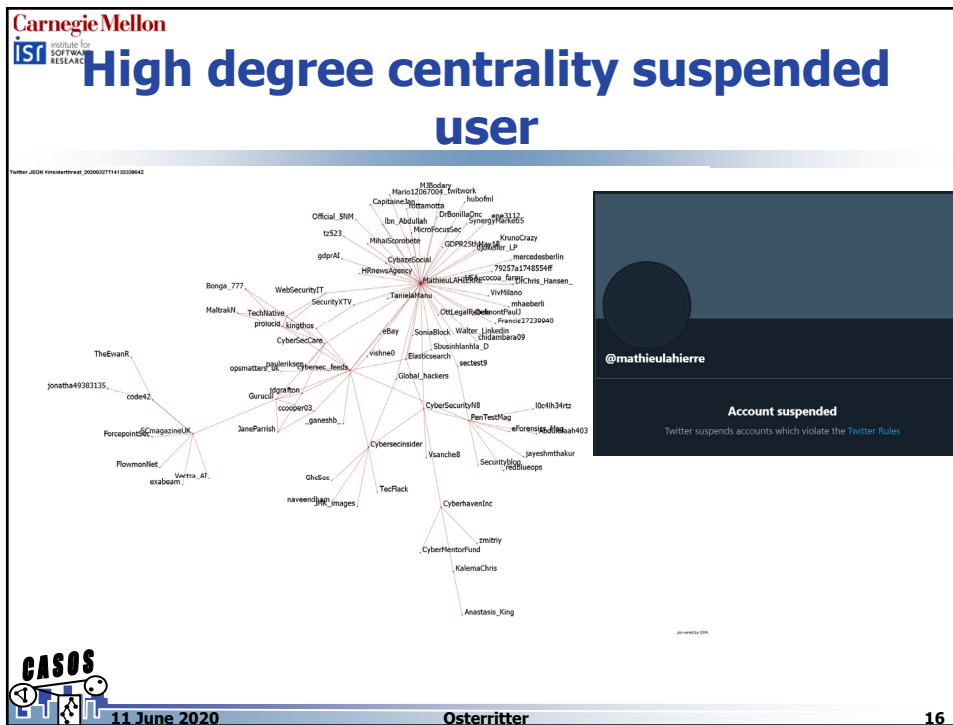
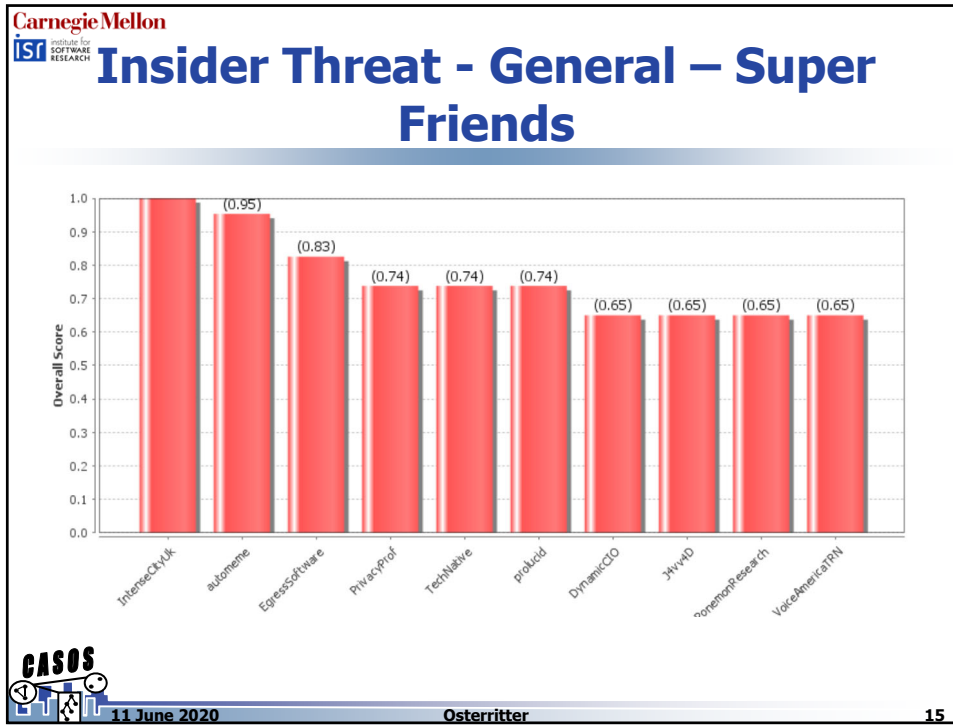
“GENERAL” GROUPING

CASOS

11 June 2020 Osterritter 12








Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

Bot or not?




Twinybots

Artificial Intelligence made to help with social media

Be active any time
Twinybots keeps your presence on social networks at every second

Smart content (AI)



CASOS

11 June 2020 **Osterritter** 17

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

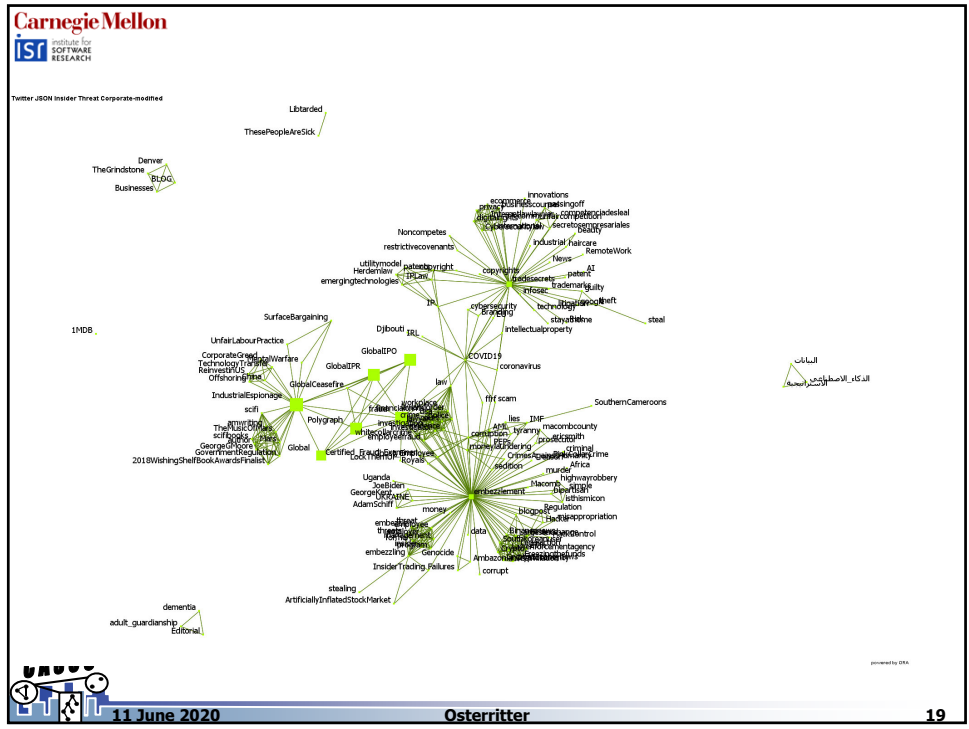
Inside Threat Tweets

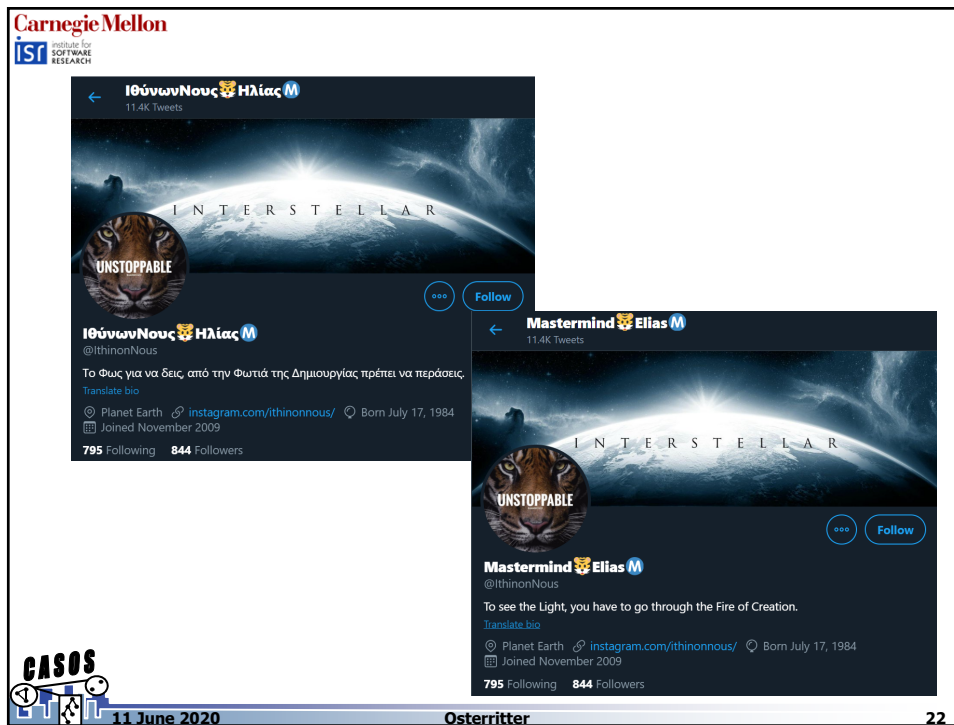
"CORPORATE" GROUPING

CASOS


11 June 2020 **Osterritter** 18







Carnegie Mellon
IST Institute for SOFTWARE RESEARCH



EU Justice @EU_Justice
9,969 Tweets

Building a European area of Justice. Updates from the European Commission's Justice and Consumers DG. RT, quotes from 3rd parties and links are not endorsements

Brussels, Europe ec.europa.eu/justice Joined January 2012
1,160 Following 55.1K Followers

Mastermind Elias @lthinonNous · Apr 28
Communities involved and insist to this fraud and industrial espionage, will be eventually banned. @EU_Justice
#Polygraph#IndustrialEspionage#Fraud#GlobalIPR#GlobalIPO#Global

Mastermind Elias @lthinonNous · Apr 28
You before me ain't gonna happen. It's fraud and industrial espionage. @EU_Justice
#Polygraph#IndustrialEspionage#Fraud#GlobalIPR#GlobalIPO#Global

Mastermind Elias @lthinonNous · Apr 28
Without "fast track" justice, you will not sell "Hellenism"! Do you hear Europe!? @EU_Justice
#Polygraph#IndustrialEspionage#Fraud#GlobalIPR#GlobalIPO#Global

Mastermind Elias @lthinonNous · Apr 28
Albania and albanians to be fully charged and banned. @EU_Justice
#Polygraph#IndustrialEspionage#Fraud#GlobalIPR#GlobalIPO#Global

CASOS

11 June 2020 Osterritter 23

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

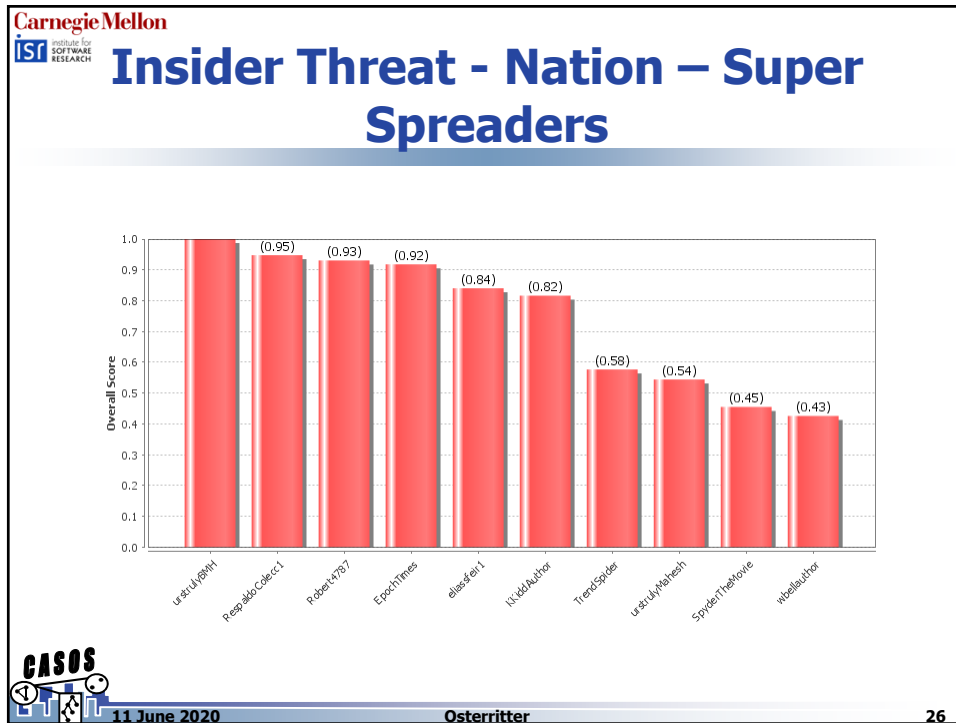
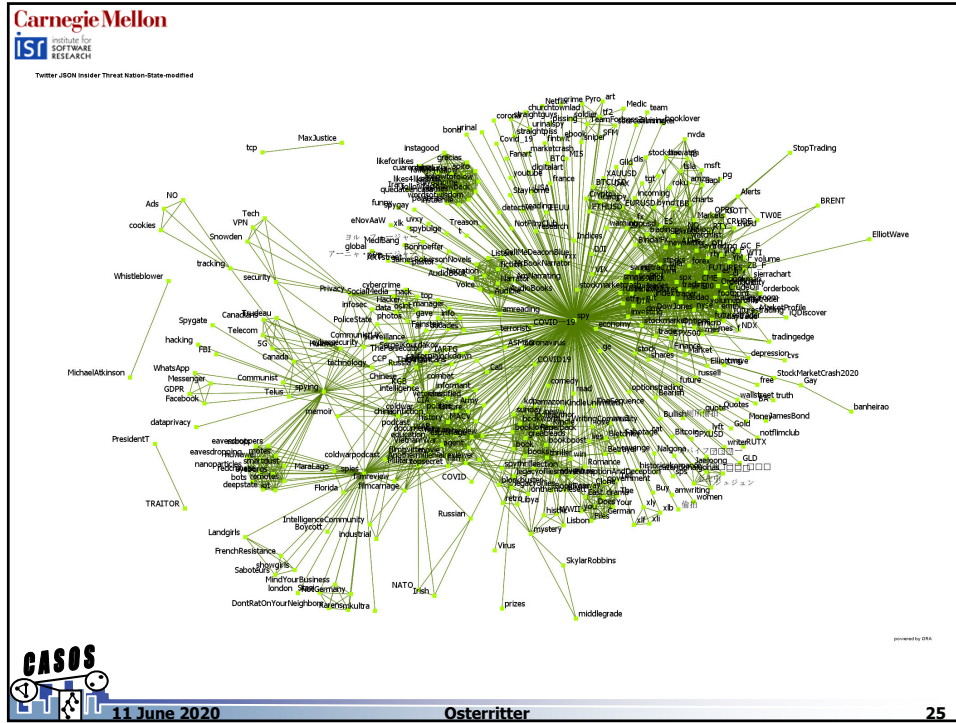
Inside Threat Tweets

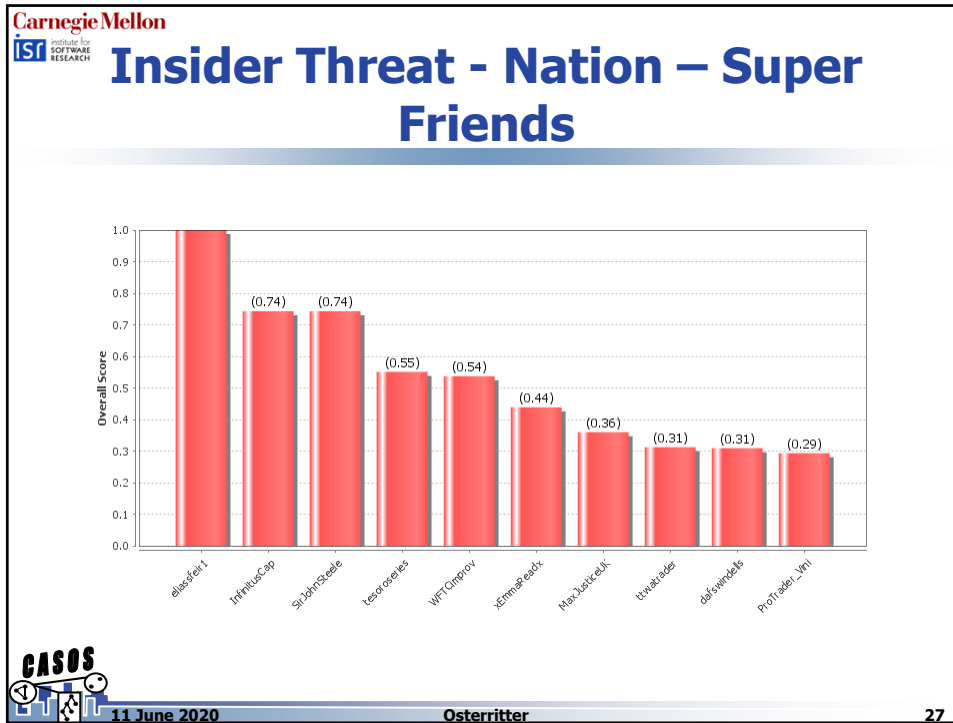
"NATION-STATE" GROUPING

CASOS

11 June 2020 Osterritter 24







- Carnegie Mellon
IST Institute for Software Research
- ## Findings
- Much of the conversation around insider threat are news aggregators and companies marketing services
 - ...but, there is more to do!
- CASOS
- 11 June 2020 Osterritter 28



Carnegie Mellon
IST Institute for Software Research

Next Steps

- Bot analysis
- NetMapper
- Network comparison (corporate vs nation-state vs general)
- Get list of disabled users in data collected

Future Work

- Explore other hashtags (APT28, APT29, APT41, etc.)
- Possibly cross-reference with other social media (Facebook, YouTube) Maltego?

CASOS

11 June 2020 Osterritter 29

Carnegie Mellon
IST Institute for Software Research

Questions for future thought

- What other insights would be useful to show?
 - Other analyses from ORA Twitter report?
 - Other network visualizations?
- What would we want to know about this conversation?
 - Possibly: Geographic or group attribution of conversation drivers – how to divine this?
 - What companies are present here?
- Best practices for analyzing a conversation?
 - Overall methods to go from large set of Twitter data to meaningful insights

CASOS

11 June 2020 Osterritter 30



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

ORA Walkthrough

ORA 3.0.9.107
File Edit Preferences Data Management Generate Networks Analysis Simulations Visualizations System Help
Open Meta-Network... Ctrl+O
Large-Scale Load...
Data Import Wizard... Ctrl+W
Data Export...
Manage CASOS database...
Open Workspace...
PileSort...
Text to Network Wizard...
Save Meta-Network Ctrl+S
Save Meta-Network As...
Save Workspace
Exit

Meta-Network: Scale Free
Meta-Network Time Click to create... Date Period
Filename Load...
Generate Reports... Visualize Measure Charts...

General statistics:
Source count: 0
Nodeset count: 1
Node count: 100
Network count: 1
Total density: 0.049495

Link statistics:
All links: 490
All link values: Binary
Non self-loops: 490
Non self-loop values: Binary
Self-loops: 0
Self-loop values: Binary

Component statistics:
Isolates: 12
Trwarks: n

CASOS
11 June 2020 Osterritter 31

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

ORA Walkthrough

Import Data into ORA

What would you like to do? Description
Import one or more twitter files and create one new dynamic meta network per file.

- Design a meta-network
- Import Excel or text delimited files
- Import from another analysis tool
- Import IRL network data
- Import other data formats
 - JSON data
 - Amazon data
 - Blogtrackers data
 - GitHub data
 - YouTube data
 - Talkwalker data
 - Pulse data
 - NewsIntelligence data
 - VC data
 - Survey Monkey data
 - Shazam data
 - Bibliography & Citations data
 - TAT data
 - TRIM data
 - Reddit data
- Import Email
- Import from a database

Cancel < Back Next > Finish

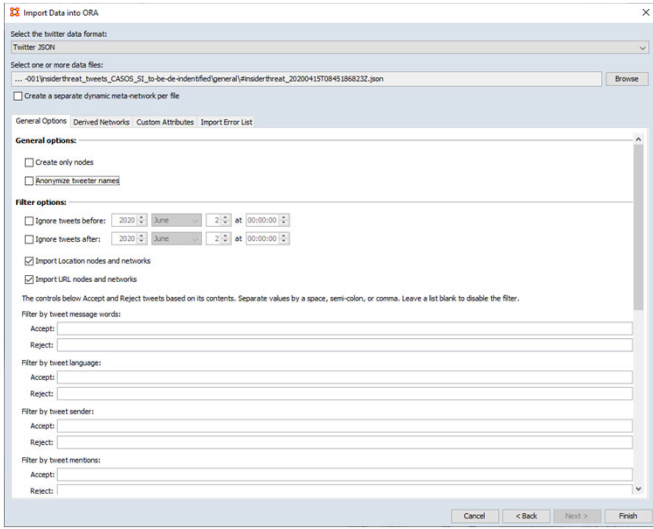
CASOS
11 June 2020 Osterritter 32



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

ORA Walkthrough

* Can choose to anonymize tweeter names if needed for real data



Import Data into ORA

Select the twitter data format:
Twitter JSON

Select one or more data files:
...-001insidethreat_tweets_CASOS_SI_to-be-de-identified/general/insidethreat_202004131708451868232.json [Browse]

Create a separate dynamic meta-network per file

General Options | Derived Networks | Custom Attributes | Import Error List

General options:

Create only nodes
 Anonymize tweeter names

Filter options:

Ignore tweets before: 2020 June 2 at 00:00:00
 Ignore tweets after: 2020 June 2 at 00:00:00

Import Location nodes and networks
 Import URL nodes and networks

The controls below Accept and Reject tweets based on its contents. Separate values by a space, semi-colon, or comma. Leave a list blank to disable the filter.

Filter by tweet message words:
Accept:
Reject:

Filter by tweet language:
Accept:
Reject:

Filter by tweet sender:
Accept:
Reject:

Filter by tweet mentions:
Accept:
Reject:

Cancel < Back Next > Finish

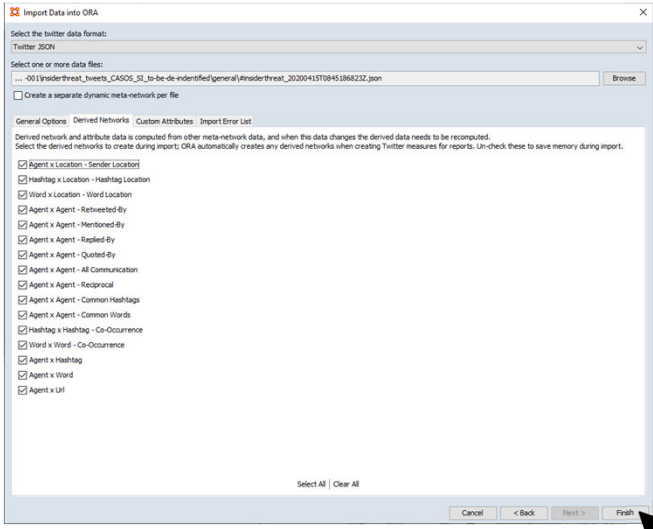
CASOS 11 June 2020 Osterritter 33

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

ORA Walkthrough

'Derived Networks' tab - you can choose non-default networks if desired.

At this point, click 'Finish' to import your data



Import Data into ORA

Select the twitter data format:
Twitter JSON

Select one or more data files:
...-001insidethreat_tweets_CASOS_SI_to-be-de-identified/general/insidethreat_202004131708451868232.json [Browse]

Create a separate dynamic meta-network per file

General Options | **Derived Networks** | Custom Attributes | Import Error List

Derived network and attribute data is computed from other meta-network data, and when this data changes the derived data needs to be recomputed. Select the derived networks to create during import; ORA automatically creates any derived networks when creating Twitter measures for reports. Un-check these to save memory during import.

Agent x Location - Sender Location
 Hashtag x Location - Hashtag Location
 Word x Location - Word Location
 Agent x Agent - Retweeted-By
 Agent x Agent - Mentioned-By
 Agent x Agent - Replied-By
 Agent x Agent - Quoted-By
 Agent x Agent - All Communication
 Agent x Agent - Reciprocal
 Agent x Agent - Common Hashtags
 Agent x Agent - Common Words
 Hashtag x Hashtag - Co-Occurrence
 Word x Word - Co-Occurrence
 Agent x Hashtag
 Agent x Word
 Agent x Url

Select All | Clear All

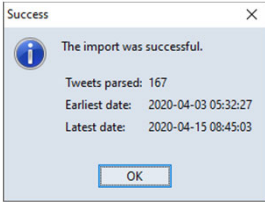
Cancel < Back Next > Finish

CASOS 11 June 2020 Osterritter 34



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

ORA Walkthrough



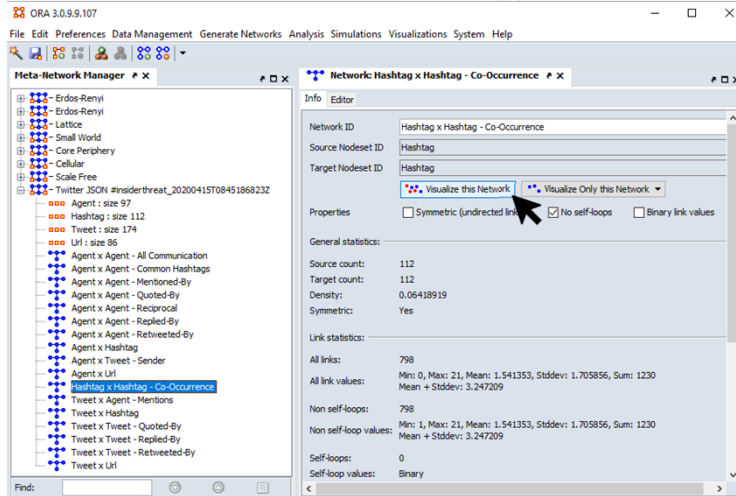
Success
The import was successful.
Tweets parsed: 167
Earliest date: 2020-04-03 05:32:27
Latest date: 2020-04-15 08:45:03
OK

CASOS
11 June 2020 Osterritter 35

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

ORA Walkthrough

Select 'Hashtag x Hashtag - Co-occurrence' network, then choose 'Visualize this Network'



ORA 3.0.9.9.107
File Edit Preferences Data Management Generate Networks Analysis Simulations Visualizations System Help

Meta-Network Manager
Erdos-Renyi
Erdos-Renyi
Lattice
Small World
Core Periphery
Cellular
Scale Free
Twitter JSON #insiderthreat_20200415T0845186823Z
Agent : size 97
Hashtag : size 112
Tweet : size 174
URL : size 96
Agent x Agent - All Communication
Agent x Agent - Common Hashtags
Agent x Agent - Mentioned-By
Agent x Agent - Quoted-By
Agent x Agent - Reciprocal
Agent x Agent - Replied-By
Agent x Agent - Retweeted-By
Agent x Hashtag
Agent x Tweet - Sender
Agent x URL
Hashtag x Hashtag - Co-Occurrence
Tweet x Agent - Mentions
Tweet x Hashtag
Tweet x Tweet - Quoted-By
Tweet x Tweet - Replied-By
Tweet x Tweet - Retweeted-By
Tweet x URL

Network: Hashtag x Hashtag - Co-Occurrence
Info Editor
Network ID: Hashtag x Hashtag - Co-Occurrence
Source Nodeset ID: Hashtag
Target Nodeset ID: Hashtag
Visualize this Network
Visualize Only this Network
Properties
 Symmetric (undirected links) No self-loops Binary link values
General statistics:
Source count: 112
Target count: 112
Density: 0.06418919
Symmetric: Yes
Link statistics:
All links: 798
All link values: Min: 0, Max: 21, Mean: 1.541353, Stddev: 1.705856, Sum: 1230
Mean + Stddev: 3.247209
Non self-loops: 798
Non self-loop values: Min: 1, Max: 21, Mean: 1.541353, Stddev: 1.705856, Sum: 1230
Mean + Stddev: 3.247209
Self-loops: 0
Self-loop values: Binary

CASOS
11 June 2020 Osterritter 36



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

ORA Walkthrough

Twitter JSON #insiderthreat_20200415T0845186823Z - ORA Network Visualizer

112 Nodes, 798 Links

11 June 2020 Osterritter 37

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

ORA Walkthrough

ORA 3.0.8.9.107

Meta-Network Manager

Meta-Network: Twitter JSON #insiderthreat_20200415T0845186823Z

General statistics:

- Source count: 0
- Nodeset count: 4
- Node count: 469
- Network count: 17
- Total density: 0.01674

Link statistics:

- All links: 6455
- All link values: Mean: 1.220604, Stddev: 0.960075, Sum: 7879
- Non self-loops: 6371
- Non self-loop values: Mean: 1.191807, Stddev: 0.874086, Sum: 7593
- Self-loops: 84
- Self-loop values: Mean: 3.404762, Stddev: 2.837332, Sum: 286

11 June 2020 Osterritter 38



Carnegie Mellon
IST Institute for Software Research

ORA Walkthrough

Categories: threat_20200415T0845186823Z
Date Period
Load...
Knowledge Networks & Network Text Analysis
Statistical Procedures and Diagnostics
Social Media
Geospatial
Characterize Groups and Networks
Dynamics
Locate Key Entities
Locate Key Relations
All Measures by Category
Measure Charts...
Blog
Twitter
You identifies key tweeters and tweets and in a meta-network derived from twitter data.

1.220604, Stride: 0.960075, Sum: 7879

CASOS
11 June 2020 Osterritter 39

Carnegie Mellon
IST Institute for Software Research

ORA Walkthrough

* Leave defaults for initial exploration

Generate Reports - Twitter
Parameters
Ranked Tables
Geospatial Options
 Create geospatial visualizations
Use a shapefile attribute to associate latitude/longitude coordinates with location names, such as cities or countries. This option might slow down report generation.
 Use the shapefile:
Attribute: Browse...
< Back Next > Cancel

CASOS
11 June 2020 Osterritter 40



Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

ORA Walkthrough

Generate Reports - Twitter

Save Options
Preferences

Reports can present their results in different formats. Each format produces one or more files that are saved to a specified location. When multiple files are created, each filename will be an extension of the one you give.

Select the report formats to create:

- Text
- HTML
- CSV
- JSON
- PowerPoint All slides
- PDF

Enter a directory in which to save the report:
C:\Reports

Enter a filename without extension:
Twitter

Run report once per meta-network

< Back Finish

CASOS

11 June 2020 Osterritter 41

Carnegie Mellon
IST Institute for SOFTWARE RESEARCH

ORA Walkthrough

TWITTER REPORT

Input data: Twitter JSON #insiderthreat_20200415T0845186823Z
Start time: Wed Jun 3 23:36:38 2020
[Data Description](#)

Import Summary

This is an overview of the import process that was used to create the dataset.

Twitter file(s) imported	C:\insiderthreat_tweets_CASOS_SI_to-be-de-identified-20200601T140824Z-001\insiderthreat_tweets_CASOS_SI_to-be-de-identified_general\insiderthreat_20200415T0845186823Z.json
Twitter file format	Twitter JSON
Dynamic meta-network?	No, all tweets are in one meta-network

Import Data Statistics

This is an overview of the tweet activity in the dataset. The dataset contains only one meta-network and all tweets are analyzed.

Network	Twitter JSON #insiderthreat_20200415T0845186823Z
First tweet date	2020-04-03 05:32:27-04
Last tweet date	2020-04-15 08:45:03-04
Number of tweets	174
Number of tweets with geotag	0
Number of tweets with TDT	100

Report will save to local machine and open in default web browser

Explore Data Statistic, Super Friends report, and Super Spreaders report

CASOS

11 June 2020 Osterritter 42

