





Networks of Hate Speech in COVID-19 Discourse

Joshua Uyheng
juyheng@cs.cmu.edu
CASOS Center, Institute for Software Research
Carnegie Mellon University

CASOS Summer Institute 2020




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



COVID-19 and Hate Speech

- **Hate speech** is:
 - Negative or abusive language
 - Targeting or discriminating against a disadvantaged group
- Distinct from merely **offensive language**
 - Offensive language may use profanities but not always be targeted toward some marginalized population
 - Hate speech may also include implicit negative cues without explicit use of abusive terms



June 2020

2



Carnegie Mellon
Institute for SOFTWARE RESEARCH

Definition/s of hate speech

- **Hate speech** is:
 - Negative or abusive language
 - Targeting or discriminating against a disadvantaged group
- Distinct from merely **offensive language**
 - Offensive language may use profanities but not always be targeted toward some marginalized population
 - Hate speech may also include implicit negative cues without explicit use of abusive terms

CASOS
June 2020

3

Carnegie Mellon
Institute for SOFTWARE RESEARCH

Hate speech as a *social phenomenon*

- Language does not exist in a vacuum
 - It is perpetuated *by* groups
 - It is committed *against* groups
- Over time, it is important to see how hate speech shapes social interaction
 - Formation of communities
 - Accrual of individual influence

CASOS
June 2020

4



Carnegie Mellon
Institute for
SOFTWARE
RESEARCH

Value of a dynamic network perspective

- Network science helps us:
 - Understand large-scale and complex patterns of relationships
 - See a social phenomenon at multiple scales
- Dynamic network methods are:
 - Interoperable with machine learning and other cutting-edge computational tools
 - Enable intuitive visualizations

CASOS
June 2020

5

Carnegie Mellon
Institute for
SOFTWARE
RESEARCH

Objectives of this case study

- In the context of the COVID-19 pandemic:
 - How can we empirically examine hate speech in its socially networked setting?
 - How can we characterize individuals and groups which do and do not engage in hate speech?

CASOS
June 2020

6



Carnegie Mellon
ISI Institute for SOFTWARE RESEARCH

A QUICK DETOUR

WHAT IS HATE SPEECH?

CASOS
June 2020

7

Carnegie Mellon
ISI Institute for SOFTWARE RESEARCH

Can we use a data-driven method to figure out what hate speech "is"?

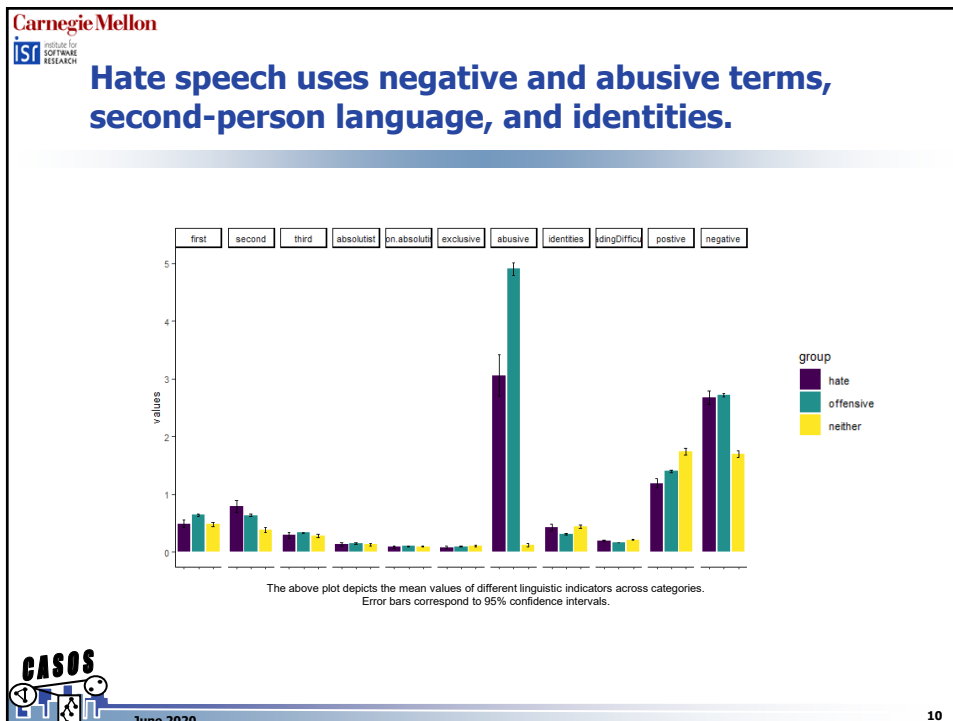
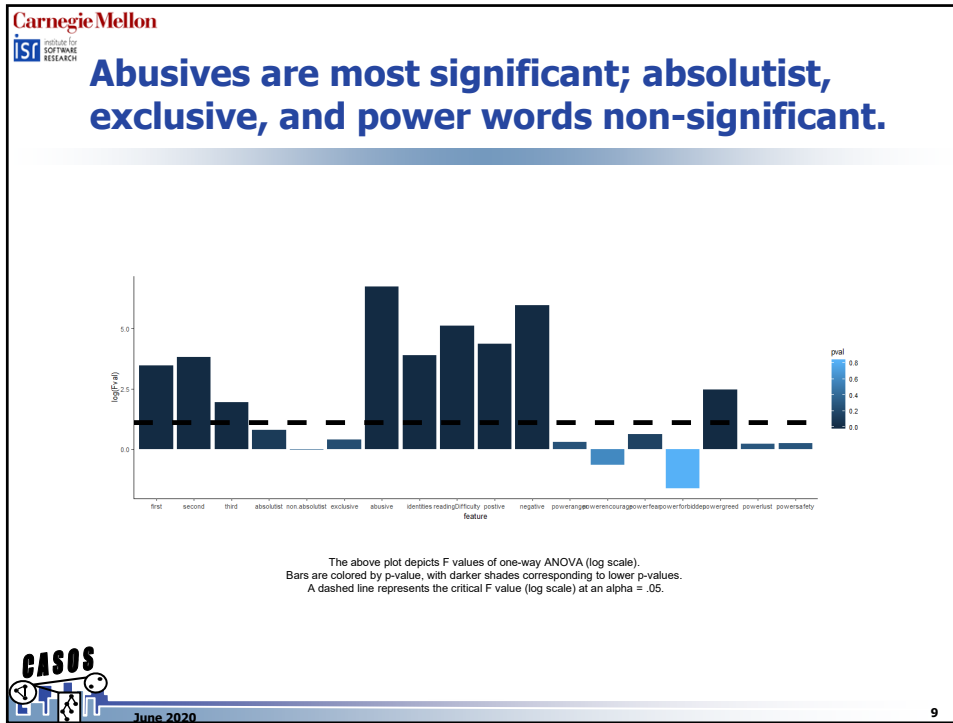
- 24K tweets labeled as hate speech, offensive language, or neither
 - 1430 hate speech (5.77%)
 - 191909 offensive language (77.43%)
 - 4163 neither (16.80%)
- Measured linguistic cues using Netmapper
 - Ran ANOVA tests to see statistically significant differences

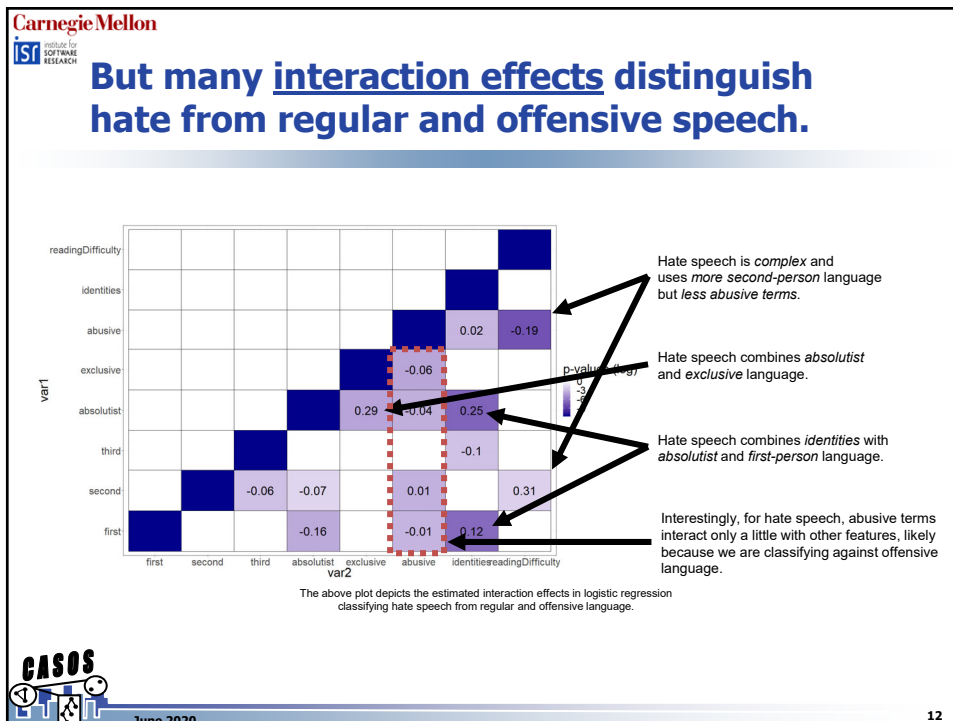
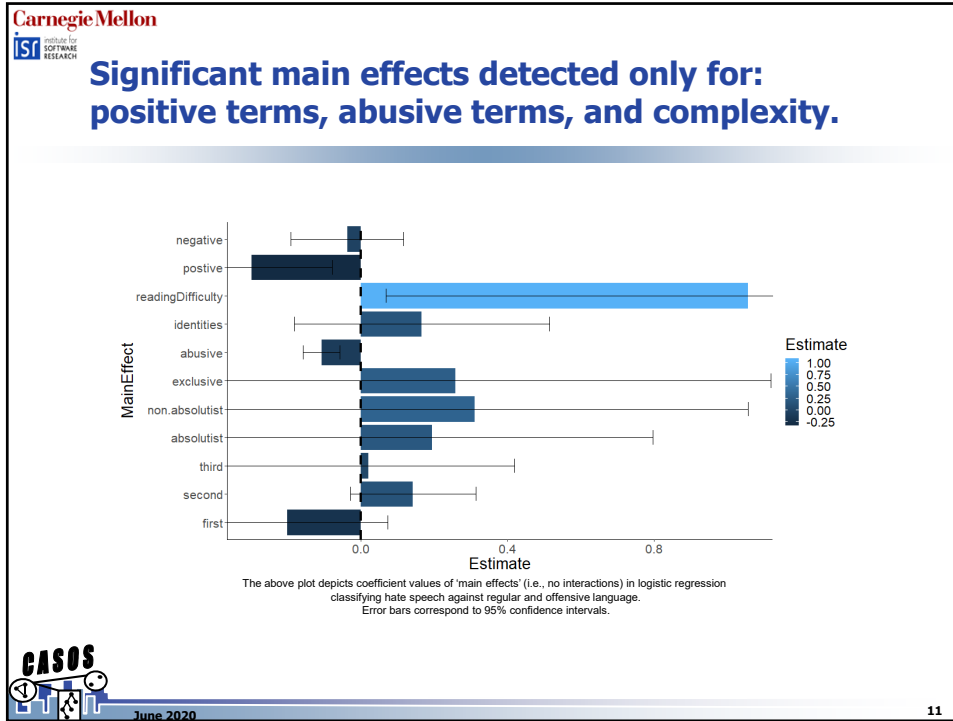
Davidson, T., Warmley, D., Macy, M., & Weber, I. (2017, May). Automated hate speech detection and the problem of offensive language. In *Proc. ICWSM*.

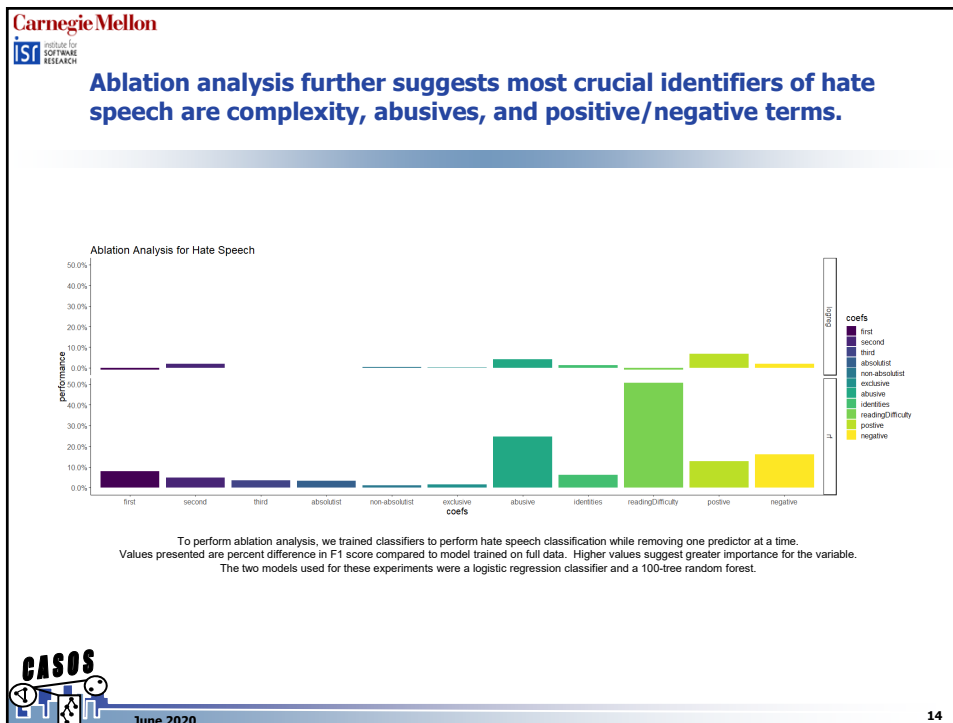
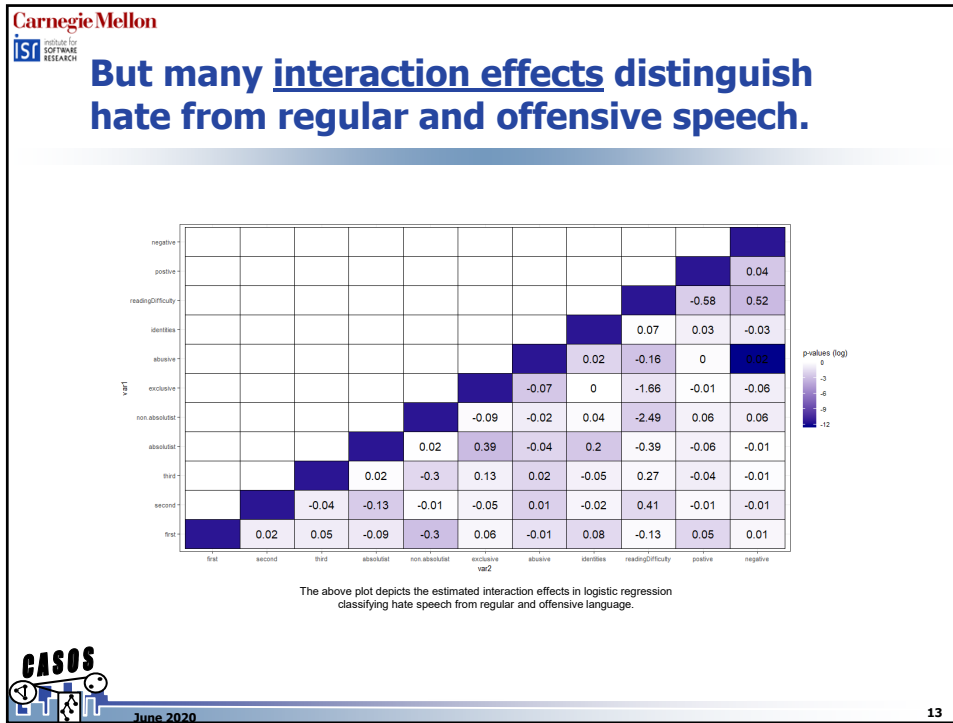
CASOS
June 2020

8









Carnegie Mellon
ISI Institute for SOFTWARE RESEARCH

Machine Learning Classifier

- Training Procedure
 - Oversampling during training to have equal proportions across categories
 - 70-20-10 train-validate-test split
- Evaluation
 - Measure accuracy, F1 ('weighted') scores
 - Compare against random baseline
 - Choose classifier with best validation performance
 - Final evaluation on test set

CASOS June 2020 15

Carnegie Mellon
ISI Institute for SOFTWARE RESEARCH

Random forest with 50 trees gives best validation performance with decent improvement over baseline.

Performance Measures

log(ntrees)	train_acc	train_f1	valid_acc	valid_f1	dumb_acc	dumb_f1
0	93.0%	93.0%	72.0%	70.0%	63.0%	63.0%
1	94.0%	94.0%	74.0%	73.0%	63.0%	63.0%
2	94.5%	94.5%	75.0%	74.5%	63.0%	63.0%
3	94.5%	94.5%	75.5%	75.0%	63.0%	63.0%
4	94.5%	94.5%	76.0%	75.5%	63.0%	63.0%

Improvement Measures

log(ntrees)	diff_acc	diff_f1
0	13.0%	15.0%
1	11.0%	16.0%
2	14.0%	17.0%
3	16.0%	18.0%
4	17.0%	18.5%

Test accuracy is 76.40% ||| Test F1 score is 76.74%
Accuracy improvement is 22.51% ||| F1 improvement is 21.85%

CASOS June 2020 16



Carnegie Mellon
ISI Institute for SOFTWARE RESEARCH

RESULTS

CASOS
June 2020 17

Carnegie Mellon
ISI Institute for SOFTWARE RESEARCH

Data (Preliminary – to be expanded)

- Twitter data
 - Collected using REST API
 - Terms: #COVID19US
 - At some point official hashtag used for pandemic discourse specific to the United States
 - Dates: March 5 – 25 (21 days)
 - > data available already up to May still processing

CASOS
June 2020 18



Carnegie Mellon
ISI Institute for SOFTWARE RESEARCH

Exploratory questions

- How much hate speech and offensive language do we detect in online discussion of the #COVID19US hashtag?
- How much bot activity do we detect in online discussion of the #COVID19US hashtag?
- Are the two quantities related?

CASOS
June 2020

19

Carnegie Mellon
ISI Institute for SOFTWARE RESEARCH

Method

- Hate speech detection
 - Features: Linguistic cues associated with psychological states (see Pennebaker)
 - Model: Random forest with 40 estimators
 - Trained on open dataset of hate speech, offensive language, normal language
 - Achieved ~97% training accuracy and F1; ~75% testing accuracy and F1
- Network analysis with ORA
 - Visualization of agent x agent networks
 - Visualization of lexical networks for hate speech

CASOS
June 2020

20



Carnegie Mellon
Institute for SOFTWARE RESEARCH

Relative levels of hate appear to fluctuate over time.

- #COVID19US discourse is dominated by language that is neither offensive nor hate speech
- However, noticeable proportions of the latter persist
 - Between 8-17% hate speech
 - Between 7-30% offensive

CASOS
June 2020

21

Carnegie Mellon
Institute for SOFTWARE RESEARCH

Are bots driving hate speech and offensive language? Results suggest they do not.

- Bot activity over time is negatively correlated to both offensive language and hate speech
- Bot activity instead positively correlated with normal speech

	hate	offensive	normal	bots
hate	1.0000000	0.1269184	-0.4097136	-0.3371000
offensive	0.1169184	1.0000000	-0.9568375	-0.8094684
normal	-0.4097136	-0.9568375	1.0000000	0.8431966
bots	-0.3371000	-0.8094684	0.8431966	1.0000000

CASOS
June 2020

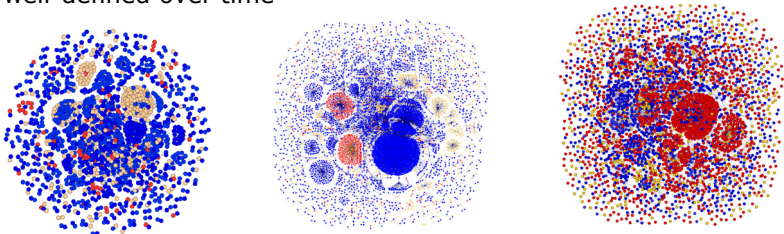
22



Carnegie Mellon
ISI Institute for Software Research

What is striking, however, is the apparent formation of *hate communities*.

- Networks of users deploying hate speech appear to grow more well-defined over time



March 5 March 14 March 25

Figures depict agent x agent networks (replies + retweets + mentions).
Agents colored based on use of hate speech (red), offensive language (orange), and neither (blue).

CASOS
June 2020 23

Carnegie Mellon
ISI Institute for Software Research

Quantifying community formation: Hate entropy as a measure of randomness

- Entropy measures level of disorder or randomness in a system
- Computation
 - Suppose there are N possible labels for a system of nodes
 - Then for label k in {1, 2, ... N}, we define:
$$p_k = \frac{\#nodes\ with\ label\ k}{\#total\ nodes}$$
 - Entropy = $-\sum_{k=1}^N p_k \log p_k$
- Higher-entropy system: Less homophily
 - ● ● ● $p_1 = 0.5, p_2 = 0.5$
● ● ● ● Entropy = 0.6931472
- Lower-entropy system: More homophily
 - ● ● ● $p_1 = 0.875, p_2 = 0.125$
● ● ● ● Entropy = 0.3767702
- As hate speech grows more clustered, we expect *hate entropy to go down*

CASOS
June 2020 24



Carnegie Mellon
ISI Institute for SOFTWARE RESEARCH

Hate entropy metric shows that distribution of hate speech is *less random*, more clustered.

- Procedure for calculation:
 - Produce Louvain clusters over Agent x Agent network (All Communication)
 - Take only subset of Louvain clusters with size ≥ 10
 - Compute entropy of hate class labels per cluster
 - Take mean over time

Interestingly, still not correlated to bot activity – is the hate speech organic?

CASOS
June 2020 25

Carnegie Mellon
ISI Institute for SOFTWARE RESEARCH

DISCUSSION

CASOS
June 2020 26



Carnegie Mellon
ISI Institute for SOFTWARE RESEARCH

Some Takeaways

- Hate speech is an important yet challenging problem to examine in the context of a global pandemic
- It is important to see hate speech as both a linguistic and socially networked phenomenon
- Interoperable pipelines of network science and machine learning tools can help us approach the problem empirically
- Policies designed to respond to hate speech and other social cyber-security issues must be grounded in multidisciplinary and multi-methodological perspective

CASOS
June 2020 27

Carnegie Mellon
ISI Institute for SOFTWARE RESEARCH

METHODOLOGY

CASOS
June 2020 28



Carnegie Mellon
Institute for SOFTWARE RESEARCH

Tools

1. Netmapper

- To measure use of abusive terms
- To measure use of identity terms

2. ORA

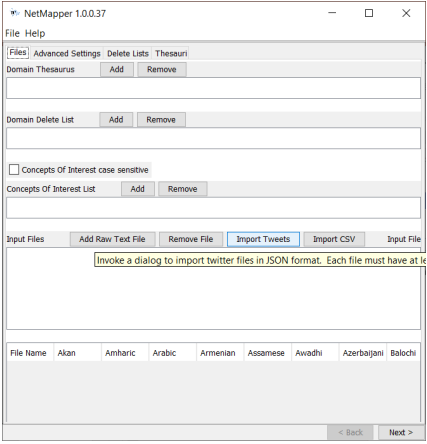
- To visualize social interactions
- To measure important network metrics

CASOS
June 2020 29

Carnegie Mellon
Institute for SOFTWARE RESEARCH

Instructions for Netmapper: Loading data

- Load files into Netmapper using the Import Tweets button
- We want the following files:
 - covidhate_20200309.json
 - covidhate_20200314.json
 - covidhate_20200319.json



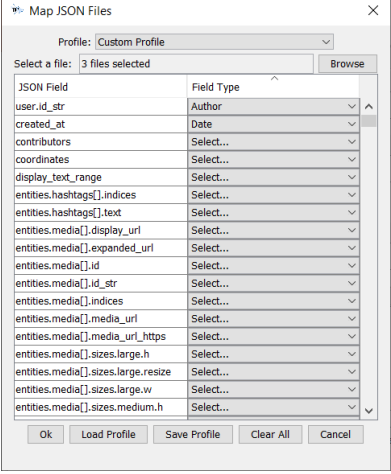
CASOS
June 2020 30



Carnegie Mellon Institute for SOFTWARE RESEARCH

Instructions for Netmapper: Analysis

- Make sure relevant Netmapper fields match their corresponding JSON fields
 - Author: user.id_str
 - Date: created_at
 - Tweet ID: id_str
 - Text: full_text
- Run and save Netmapper files
 - Make sure we are getting "usage measures"

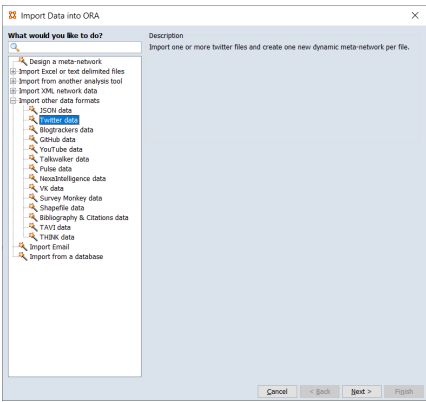


CASOS June 2020 31

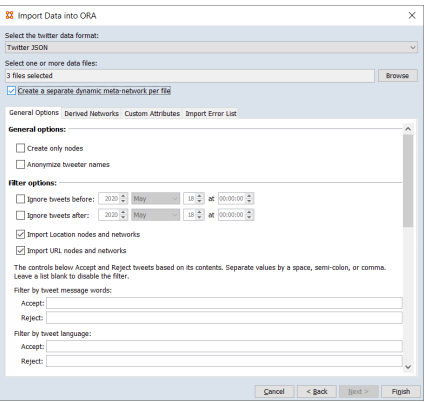
Carnegie Mellon Institute for SOFTWARE RESEARCH

Instructions for ORA: Loading data

Import Twitter data



Create a separate dynamic meta-network per file



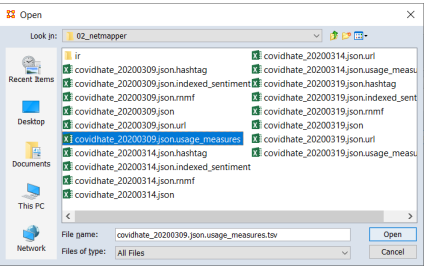
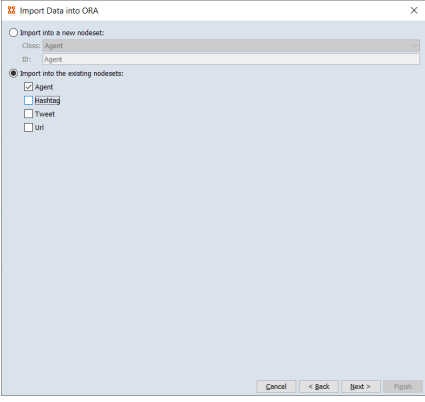
CASOS June 2020 32



Carnegie Mellon
ISI Institute for SOFTWARE RESEARCH

Instructions for ORA: Loading attributes

Load attributes only for Agents **Use the appropriate "usage_measures" files**



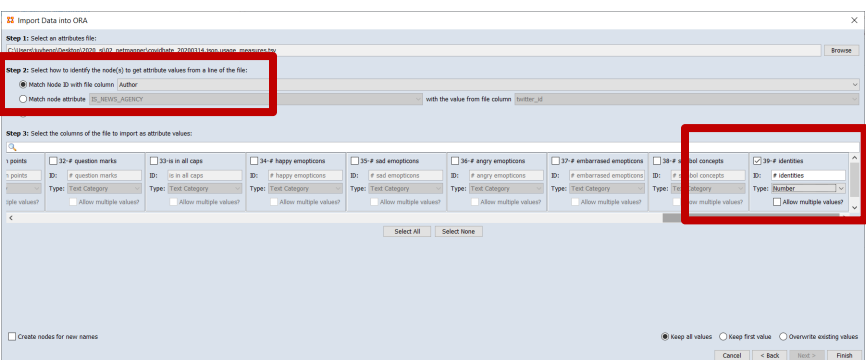
CASOS
June 2020

33

Carnegie Mellon
ISI Institute for SOFTWARE RESEARCH

Instructions for ORA: Loading attributes

Match NODE ID with file column Author **Make sure to click only abusives and #identities**



CASOS
June 2020

34



Carnegie Mellon
ISI Institute for SOFTWARE RESEARCH

Instructions for ORA: Visualize!

Visualize All Communication

- Twitter JSON covidhate_20200309
 - Agent : size 10492
 - Hashtag : size 1297
 - Tweet : size 11131
 - Url : size 861
 - Agent x Agent - All Communication**
 - 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

Remove components smaller than 3 nodes

Display Options for Large Networks

The network you are displaying has 10492 nodes and 10779 links. For networks with more than 5000 entities and/or more than 10000 links, display might be very slow. Consider these options to parse down the meta-network before loading to improve performance.

- Remove links with weight less than or equal to
- Select networks to load
 - Agent x Agent - All Communication
- Remove components of size less than or equal to
- Hide labels
- Hide links
- Save as new meta-network in ORA

Cancel Continue

CASOS
June 2020

35

Carnegie Mellon
ISI Institute for SOFTWARE RESEARCH

Instructions for ORA: Visualize!

Size by identities invoked

Node Size Selector

File

Use this window to manipulate nodes by measure value or attribute.

Select an attribute:
identities

Select a measure:
<select>

Invert links for geodesic measures

Geodesic measure radius:

Expand to control individual values.

Apply Changes Close

identities Values: [1,0,5,0]

Color by use of abusive terms

Node Color Selector

File

Use this window to manipulate nodes by measure value or attribute.

Select an attribute:
abusive

Select a measure:
<select>

Invert links for geodesic measu...

Geodesic measure radius:

Lacks an Attribute value 1.0

Apply Changes Close

abusive Values: [1,0,1,0]

CASOS
June 2020

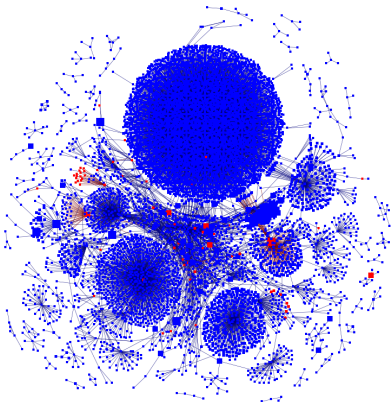
36



Carnegie Mellon
ISI Institute for SOFTWARE RESEARCH

Sample ORA network visualization

Twitter JSON covariate_20200309-modified



powered by ORA

CASOS

June 2020 37

Carnegie Mellon
ISI Institute for SOFTWARE RESEARCH

Instructions for ORA: Run Reports

Select Key Entities Ranking

Generate Reports - Key Entities Ranking

Select Report: Filter Data, Measures, Negative Links, Union by Thirds, Transform Data, Remove Nodes

Reports: select a report to run from the list or by category. Key Entities Ranking, Twitter, Key Entities Ranking, Locate Groups, All Measures by Category, Change in Key Entities, Network Comparison, Critical Sets, Immediate Impact, Topic Analysis, Communicative Power, Triad Census, Simmelian Ties Analysis

Categories: position in the report.

< Back Next > Cancel

Choose Default Settings and Save HTML and CSV Output

Generate Reports - Key Entities Ranking

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:\Users\Jayheng\Desktop\2020_sl\03_ora Browse

Enter a filename without extension:
Key Entities Ranking

< Back Finish > Cancel

CASOS

June 2020 38



Carnegie Mellon
Institute for SOFTWARE RESEARCH

Instructions for ORA: Run Reports

Who Attribute Analysis is helpful for high-level view

CSVs provide raw metrics for downstream analysis

KEY ENTITIES RANKING REPORT

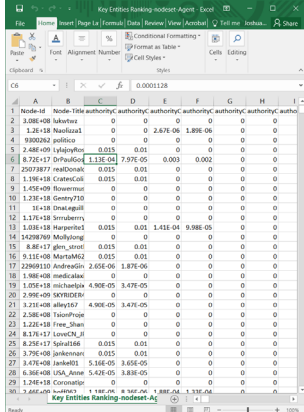
Input data: Twitter JSON covidhate_20200309
Start time: Mon May 18 12:40:57 2020

[Data Description](#)


Table of Contents

- [Key Asent - Who Analysis](#)
- [Key Asent - Who Attribute Analysis](#)
- [Key Hashtag - How Analysis](#)
- [Key Hashtag - How Attribute Analysis](#)
- [Key Uri - How Analysis](#)
- [Key Uri - How Attribute Analysis](#)
- [Performance Indicators - measures performance of the organization/1 as a whole](#)

Produced by ORA, a joint product of the CASOS center at Carnegie Mellon University and Netanomics



Node-Id	Node-Title	authority	authority	authority	authority	authority	authority	autho
1	3.08E+08	lakutera	0	0	0	0	0	0
2	1.2E+18	NasirAzad1	0	0	2.47E-06	1.89E-06	0	0
3	9.902E+08	judicop	0	0	0	0	0	0
4	2.48E+08	ylajayfms	0.015	0.01	0	0	0	0
5	8.72E+17	DPauloG4	1.13E-01	7.97E-05	0.003	0.002	0	0
6	2.0038E+17	evdionaw	0.025	0.01	0	0	0	0
7	1.19E+18	CratesC6	0.015	0.01	0	0	0	0
8	1.42E+08	Rouartm	0	0	0	0	0	0
9	1.23E+18	Genery750	0	0	0	0	0	0
10	1E+18	DnatAgall	0	0	0	0	0	0
11	1.13E+18	Srenaldery	0	0	0	0	0	0
12	1.03E+18	Hurperet1	0.015	0.01	1.41E-04	9.98E-05	0	0
13	1.429E+09	Mulyong	0	0	0	0	0	0
14	8.4E+17	gbes_ahed	0.015	0.01	0	0	0	0
15	9.11E+08	MartaM52	0.015	0.01	0	0	0	0
16	1.286E+19	Andresdin	2.85E-06	1.87E-06	0	0	0	0
17	1.98E+08	medicavc	0	0	0	0	0	0
18	1.05E+18	michaebo	4.90E-05	3.47E-05	0	0	0	0
19	2.9E+08	SPRICEDEV	0	0	0	0	0	0
20	1.21E+08	alway587	4.90E-05	3.47E-05	0	0	0	0
21	2.58E+08	ToushPou	0	0	0	0	0	0
22	1.22E+18	free_Shan	0	0	0	0	0	0
23	8.37E+17	loveCN_ji	0	0	0	0	0	0
24	8.25E+17	Spatiate	0.015	0.01	0	0	0	0
25	3.7E+08	jaksonant	0.015	0.01	0	0	0	0
26	3.47E+08	Jankah22	5.18E-06	3.85E-06	0	0	0	0
27	6.36E+08	USA_Ame	5.42E-05	3.83E-05	0	0	0	0
28	1.24E+18	Coronator	0	0	0	0	0	0
29	3.46E+08	huyuca	1.1E-06	8.1E-07	1.88E-04	1.33E-04	0	0
30		Key Entities Ranking - nodeest.Ag						




June 2020

39

Carnegie Mellon
Institute for SOFTWARE RESEARCH

DEMONSTRATION



June 2020

40



<Your Name>



Networks of Hate Speech in COVID-19 Discourse

Joshua Uyheng

juyheng@cs.cmu.edu

CASOS Center, Institute for Software Research
Carnegie Mellon University

CASOS Summer Institute 2020



Carnegie Mellon

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

