



# Case Study: Social Media Analytics for Stance Mining

With Examples From COVID-19 Twitter Analysis  
Sumeet Kumar

sumeetku@andrew.cmu.edu



Carnegie Mellon

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



## Let's Define the Terms

- **Stance** is defined as a **mental or emotional position** adopted with respect to a proposition, a person, an idea, etc. [1].
- Users' Stance is categorized as:
  - **Pro** (Favor)
  - **Con** (Anti)
  - **Neutral** (or unknown)

1. <https://www.thefreedictionary.com/stance>



7 June 2020


Sumeet Kumar

2



**Carnegie Mellon**  
**IST** Institute for Software Research

## How to Learn Users' Stance (Pro/Anti)? Prior research on stance mining has appeared in two flavors



1. Language (Text) based Approach [1]

2. Network based Approach [2]

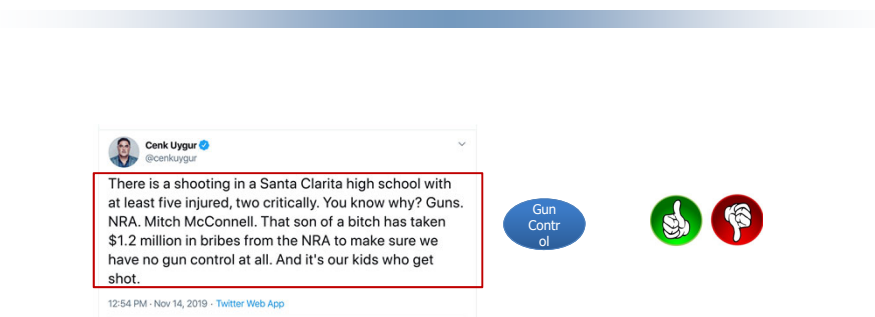
1. [SemEval-2016 Task 6: Detecting Stance in Tweets](#), Mohammad et al., 2016

2. 2011, Conover, Michael, Jacob Ratkiewicz, Matthew R. Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. ["Political polarization on twitter." ICWSM 133 \(2011\): 89-96](#)

**CASOS**  
7 June 2020 Sumeet Kumar 3

**Carnegie Mellon**  
**IST** Institute for Software Research

## Prior work on Language Based Stance Learning is Mostly Supervised which Requires Labeled data. Labeling data is Expensive.



[SemEval-2016 Task 6: Detecting Stance in Tweets](#), Mohammad et al., 2016

**CASOS**  
7 June 2020 Sumeet Kumar 4



**Carnegie Mellon**  
IST Institute for SOFTWARE RESEARCH

## Stance could also be learned from other multi-modal interactions (Networks)

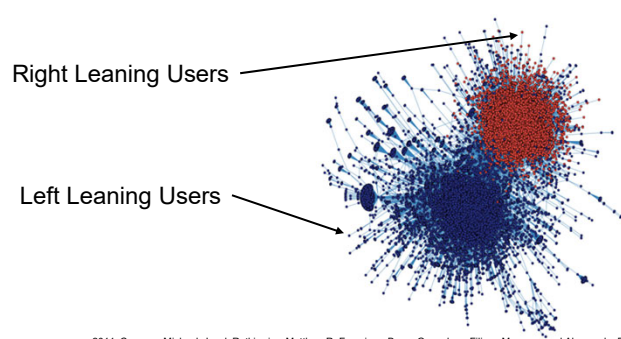


**CASOS**

7 June 2020 Sumeet Kumar 5

**Carnegie Mellon**  
IST Institute for SOFTWARE RESEARCH

## Network Based Stance Learning Methods are often Semi-Supervised, so Require Less Labeled Data. However, they can't handle isolates



Right Leaning Users

Left Leaning Users

2011, Conover, Michael, Jacob Ratkiewicz, Matthew R. Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini.  
["Political polarization on Twitter." ICWSM 133 \(2011\): 89-98](#)

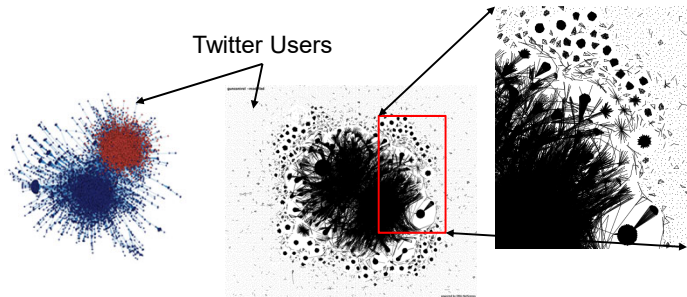
**CASOS**

7 June 2020 Sumeet Kumar 6



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## In a Real (un-processed) Network, the Isolates in the Network form a Good Fraction of the Dataset



Twitter Users

A retweets-based Network after removing the isolates  
Conover et al. [Political polarization on Twitter](#) ICWSM 133 (2011), 89-96

Unprocessed gun-control conversations on Twitter Collected by searching gun-control related terms. Links are based on Retweets.

CASOS

7 June 2020 Sumeet Kumar 7

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Three Main Challenges in Existing Approaches to Stance Mining

1. Most language-based stance mining models use supervised machine learning **which is expensive**
2. Network based semi-supervised approaches require less labeled data but **cannot handle isolates**
3. Topics change fast and new topics emerge **which make the problem more challenging**

CASOS

7 June 2020 Sumeet Kumar 8



**Carnegie Mellon**  
**IST** Institute for SOFTWARE RESEARCH

## Goal of this New Methodology: Can we Combine the Strengths of Text based Methods and Networks based Methods?

Network based Stance Learner

Text based Stance Learner

Predict the Stance of All Users in a Realistic Network

**CASOS**

7 June 2020 Sumeet Kumar 9

**Carnegie Mellon**  
**IST** Institute for SOFTWARE RESEARCH

## Co-Training on Social Networks: A Joint Network Label Propagation and Text Classification Approach for Stance Mining [2]

Extract Data

Step 1

Step 2

Input

#GunControlNow: Pro  
#2ndAmendment: Anti

Model Training

Step 3

Gun-control users' Network. Links represent retweets-based interactions.

Red nodes are 'Pro' and Green nodes are 'Anti' Users

**CASOS**

2. Sumeet Kumar, Tom Mitchell, Kathleen M. Carley, Co-Training on Social Networks, Currently under review

7 June 2020 Sumeet Kumar 10



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

### Proposed Idea: A Three Step Process

**Step 1**  
Extract Text Features and Users Networks

**Step 2**  
Label 2 to 4 hashtags  
#GunControlNow: Pro  
#2ndAmendment: Anti

**Step 3**  
Derive stance of other users from seed users

Seed Labeled Users  
Label Propagation to Unlabeled Nodes  
New Label  
Network with Text features  
Text Classifiers  
Predictions of Unlabeled Nodes  
Add new 'Confident' Node Labels  
New Node Labels  
Updates for the Next Iteration

CASOS

7 June 2020 Sumeet Kumar 11

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

### Step 1: Extract users' text features and users' networks from data

Interactions

Extracted text-data and Networks

Users - Text  
Users - Hashtags Graph  
Users - Retweets Graph

CASOS

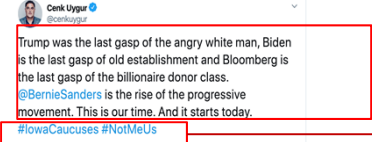
7 June 2020 Sumeet Kumar 12



**Carnegie Mellon**  
IST Institute for SOFTWARE RESEARCH

## Step 1: Extract text features and users' networks from data

1. Extract users text data
2. Extract networks



Users' Text

User	Tag	Weight
cenkuygur	#IowaCaucuses	1
cenkuygur	#NotMeUS	1

Users-Hashtags (Networks)

User	Retweet	Weight
sphursby	cenkuygur	1

Users-Retweets (Networks)

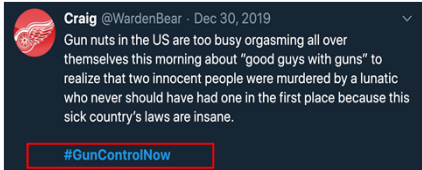
**CASOS**  
7 June 2020 Sumeet Kumar 13

**Carnegie Mellon**  
IST Institute for SOFTWARE RESEARCH

## Step 2: Label 2 to 4 popular hashtags with clear stance

**Steps:**

1. Use hashtags that appear at the end of tweets
2. Sort hashtags by their popularity
3. Label a few popular hashtags that have clear stance e.g. #GunControlNow



**CASOS**  
7 June 2020 Sumeet Kumar 14



Carnegie Mellon Institute for Software Research

### Step 3: A Semi-supervised Approach (Co-Training + Label Propagation)

- Semi-supervised approaches of machine learning is suitable for partially labeled data

Training Data

Supervised Learning: All Labeled Data → Model

Semi-Supervised Learning: Some Labeled Data + Lots of Unlabeled Data → Model

Unsupervised Learning: All Unlabeled Data → Model

- We use a **co-training setting**

CASOS 7 June 2020 Sumeet Kumar 15

Carnegie Mellon Institute for Software Research

### What is Co-Training?

- Co-training requires **two independent views to train two separate classifiers** (weak learners) iteratively [1]
- In the training process, **more confident predictions are used as new training data** [1]

Co-Training

Iteration: i

C1: A Classifier trained on view 1

C2: A Classifier trained on view 2

Allow C1 to label Some instances

Allow C2 to label Some instances

Add self-labeled instances to the pool of training data

Iteration: i+1

New labeled example

Image Source: <https://www.slideshare.net/butes/semisupervised-learning>

1: Blum, Avrim, and Tom Mitchell. "Combining labeled and unlabeled data with co-training." *Proceedings of the eleventh annual conference on Computational learning theory*. ACM, 1998.

CASOS 7 June 2020 Sumeet Kumar 16

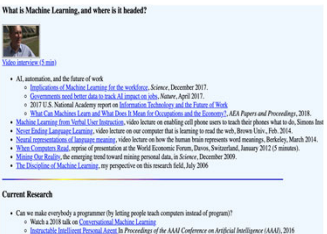




**Carnegie Mellon**  
IST Institute for SOFTWARE RESEARCH

## What is Co-Training? Applied to Website Classification

**View 1 (website)**



**View 2 (Text on the Links to the website)**

My advisor is Tom Mitchell and I work on.....  
Prof. Mitchell's work on never ending learning ...  
Prof. Mitchell, an expert in machine learning, mentioned ...

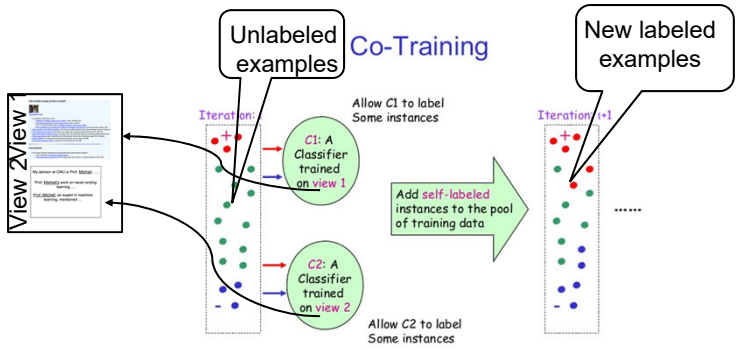
Academic / Non- Academic Webpage Classification

Blum, Avrim, and Tom Mitchell. "Combining labeled and unlabeled data with co-training." *Proceedings of the eleventh annual conference on Computational learning theory*. ACM, 1998.

**CASOS**  
7 June 2020 Sumeet Kumar 17

**Carnegie Mellon**  
IST Institute for SOFTWARE RESEARCH

## Co-Training could be useful if each data point has two (or more) views



**Unlabeled examples**

**Co-Training**

**New labeled examples**

Iteration 0

Iteration +1

Allow C1 to label Some instances

Allow C2 to label Some instances

Add self-labeled instances to the pool of training data

C1: A Classifier trained on view 1

C2: A Classifier trained on view 2

View 2 view

Blum, Avrim, and Tom Mitchell. "Combining labeled and unlabeled data with co-training." *Proceedings of the eleventh annual conference on Computational learning theory*. ACM, 1998.

**CASOS**  
7 June 2020 Sumeet Kumar 18



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

### Co-Training on Social-Networks.. What could be the multiple views?

The diagram illustrates a co-training process on social network data. It starts with 'Social Networks Data' which is split into two views: 'View 1' (a network graph) and 'View 2' (a text snippet). In 'Iteration 1', two classifiers are trained: 'C1: A Classifier trained on view 1' and 'C2: A Classifier trained on view 2'. 'C1' is used to label some instances, and 'C2' is used to label some instances. These 'self-labeled instances' are added to the training pool. The process repeats in 'Iteration i+1', resulting in 'New labeled examples'. The CASOS logo is in the bottom left, and the date '7 June 2020' and name 'Sumeet Kumar' are at the bottom.

Unlabeled examples

Co-Training

New labeled examples

Iteration 1

Iteration i+1

Allow C1 to label Some instances

Allow C2 to label Some instances

Add self-labeled instances to the pool of training data

View 1

View 2

Social Networks Data

C1: A Classifier trained on view 1

C2: A Classifier trained on view 2

CASOS

7 June 2020

Sumeet Kumar

19

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

### Co-Training on Social Networks - Texts and Networks Could be Considered as Different Views

The diagram shows a co-training process between two views: 'View 1 - Network based' and 'View 2 - Text based'. 'View 1' starts with 'Seed Labeled Users' and 'Label Propagation to Unlabeled Nodes' to identify 'New Label' and 'Add new 'Confident' Node Labels'. 'View 2' uses 'Text Classifiers' to make 'Predictions of Unlabeled Nodes' and provides 'Updates for the Next Iteration'. The two views are connected by 'Label Mixing'. The final output is a network graph with a color distribution and a screenshot of a social media post. The CASOS logo is in the bottom left, and the date '7 June 2020' and name 'Sumeet Kumar' are at the bottom.

View 1 - Network based

View 2 - Text based

Label Mixing

Seed Labeled Users

Label Propagation to Unlabeled Nodes

New Label

Add new 'Confident' Node Labels

Network with Text features

Text Classifiers' Predictions of Unlabeled Nodes

Updates for the Next Iteration

CASOS

7 June 2020

Sumeet Kumar

20



Carnegie Mellon  
 ist Institute for SOFTWARE RESEARCH

## Co-Training on Social Networks. Texts and Networks form Different Views

---

**Algorithm 3** Co-Training: Joint training of hashtag based label-propagation model and text classifier

**Require:**  $T$  is the set of tweets collected

- 1: **function** CO-TRAIN SOCIAL NETWORKS( $T$ )
- 2: Extract User-Text  $D$
- 3: Extract User-Hashtag Network  $H$
- 4: Label Seed hashtags
- 5: Get Seed users ( $UL$ )
- 6: Get Unlabeled users ( $UU$ )
- 7: **while** until convergence **do**
- 8:     **STEP 1: Label Propagation**
- 9:     Label hashtags ( $H$ ) using  $UL$
- 10:     Predict the Stance  $S^U$  of  $UU$  using  $H$
- 11:     Estimate stance confidence ( $C^U$ )
- 12:     **STEP 2: Text based classification**
- 13:     Train a text classifier  $f_{text}(\theta)$  using  $UL$
- 14:     Use  $f_{text}(\theta)$  to Predict  $UU$  stance  $S^T$
- 15:     Estimate stance confidences ( $C^T$ )
- 16:     **STEP 3: Update  $UL$**
- 17:      $L = \text{LabelMixing}(S^U, C^U, S^T, C^T)$
- 18:      $UL = UL + L$
- 19: **endwhile**
- 20: **return**  $UL$
- 21: **end function**

> e.g. #Prochoice +1

**1 - Network based**

Label Propagation to Unlabeled Nodes

Network with Text features

Seed Labeled Users

Unlabeled Nodes

UL Seed Users

UU Unlabeled Users

New Label

**2 - Text based**

Text Classifier

Predictions of Unlabeled Nodes

Add new 'Confident' Node Labels

Updates for the Next Iteration

Proposed Algorithm

7 June 2020

Sumeet Kumar

21

Carnegie Mellon  
 ist Institute for SOFTWARE RESEARCH

## Classifier 1: Network Classifier – A Label Propagation Model

---

Initialize     $\dashrightarrow$     Step 1     $\dashrightarrow$     Step 2

7 June 2020

Sumeet Kumar

22



Carnegie Mellon  
IST Institute for Software Research

### Classifier 1: Label propagation on user-user networks has shortcomings

- Many Social-Media Networks are bi-partite i.e. users relate to other entities
- Often entities on Social Media follow power law distribution
- **Converting user-posts network to user-user network explodes the size**
  - For example. 100,000 users and 200 hashtags get converted to 100,000 x 100,000 size user-user network

CASOS

7 June 2020 Sumeet Kumar 23

Carnegie Mellon  
IST Institute for Software Research

### Label Propagation Model on Bipartite Networks

Stance = -1  
Stance = +1

Hashtag Stance

Users Stance

- New users are labeled by propagating hashtag stance to users

CASOS

7 June 2020 Sumeet Kumar 24



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Label Propagation Model on Bipartite Networks With Influence Functions

Stance = -1  
Stance = +1

Hashtag Stance

Users Stance

$\sigma$ : Influence Function to Filter Less Confident Examples

If  $(W'_{43} - W'_{23}) > K$

- Influence functions are used to filter less confident predictions
- In a **Linear Threshold** function, if a user gets higher than a certain level of influence from the influencers, the user gets influenced
- New users are labeled by propagating hashtag stance to users

CASOS

7 June 2020 Sumeet Kumar 25

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Classifier 1: Label Propagation Model on Bipartite Networks Better Suits our Needs

S1: ?  
S2: +1  
S3: ?  
S4: -1  
S5: ?  
S Users Stance

$\sigma'$  ( $S1 * W_{11}$ )  
 $\sigma'$  ( $S1 * W_{22}$ )  
 $\sigma'$  ( $S5 * W_{53}$ )  
 $\sigma'$  ( $S3 * W_{34}$ )

#T1  
#T2  
#T3  
#T4  
Hashtags Stance

$\sigma$  ( $S1 * W'_{11}$ )  
 $\sigma$  ( $S2 * W'_{22}$ )  
 $\sigma$  ( $S4 * W'_{43}$ )  
 $\sigma$  ( $S1 * W'_{14}$ )  
 $\sigma$  ( $S4 * W'_{44}$ )  
 $\sigma$  ( $S3 * W'_{35}$ )

Updated S

Iterate till convergence

New labeled Users

- Influence functions are used to filter less confident predictions
- Influence functions  $\sigma'$  and  $\sigma$  are threshold functions and used to filter out not confident hashtags and users respectively

CASOS

7 June 2020 Sumeet Kumar 26



Carnegie Mellon Institute for Software Research

### Classifier 2: Learn Stance from Text in Users' Tweets

The diagram illustrates a network-based classifier for learning stance from tweets. It starts with a network of users and their tweets. Seed labeled users are identified, and their labels are propagated to unlabeled nodes. Text classifiers are used to predict labels for unlabeled nodes, and these predictions are used to update the network for the next iteration. The process involves adding new 'confident' node labels and mixing labels.

Seed Labeled Users

Label Propagation to Unlabeled Nodes

Text Classifiers' Predictions of Unlabeled Nodes

Updates for the Next Iteration

Label Mixing

Add new 'Confident' Node Labels

New Label

Network with Text features

7 June 2020 Sumeet Kumar 27

Carnegie Mellon Institute for Software Research

### Classifier 2: A Typical Text Based Classifier

- A simple text classifier (e.g. Support Vector Machine) uses labeled data to train a model
- The trained model is used to predict labels of unlabeled data

The flowchart shows the process of a typical text-based classifier. Labeled data (S2: +1, S4: -1) is used to initialize the model parameters ( $\theta, C^1$ ). Unlabeled data (S1: ?, S3: ?, S5: ?) is then processed by a text classifier to produce classifier predictions.

Labeled

Unlabeled

S2: +1

S4: -1

S1: ?

S3: ?

S5: ?

Initialize ( $\theta, C^1$ )

Text Classifier

E-Step

Classifier Predictions

7 June 2020 Sumeet Kumar 28



Carnegie Mellon  
 IST Institute for SOFTWARE RESEARCH

## Classifier 2: A Text Classifier with Self-Training

- When plenty of unlabeled data is available, models' predictions could be used to train a better model... also called self-training [1]
- Self-training exploits unlabeled data**
- In self-training, in every iteration, **new 'confident' predictions are used as new training examples**

1. Nigam, Kamal, and Rayid Ghani. "Analyzing the effectiveness and applicability of co-training." In *Proceedings of the ninth international conference on Information and knowledge management*, pp. 86-93, 2000.

CASOS

7 June 2020 Sumeet Kumar 29

Carnegie Mellon  
 IST Institute for SOFTWARE RESEARCH

## Classifier 2: An SVM Text Classifier with Confidence Estimate and a Decreasing Threshold Function

$$s_j^T = \begin{cases} 1, & \text{if } \frac{\sum_{k=1}^m 1_{s_k > 0}}{\sum_{k=1}^m 1} > f_t(\theta^T) \\ -1, & \text{elif } \frac{\sum_{k=1}^m 1_{s_k < 0}}{\sum_{k=1}^m 1} > f_t(\theta^T) \\ 0, & \text{otherwise} \end{cases}$$

$$c_j^T = \begin{cases} \frac{\sum_{k=1}^m 1_{s_k > 0}}{\sum_{k=1}^m 1}, & \text{if } s_j^T > 0 \\ \frac{\sum_{k=1}^m 1_{s_k < 0}}{\sum_{k=1}^m 1}, & \text{elif } s_j^T < 0 \\ 0, & \text{otherwise} \end{cases}$$

$s_j^T$  = stance of  $j^{\text{th}}$  user  
 $s_k$  = stance of  $k^{\text{th}}$  text message of user  $j$   
 $f_t$  = Uniformly decreasing function  
 $\theta^T$  = text threshold  
 $c_j^T$  = user text-based confidence estimate

Stance and confidence estimate of user  $j$  based on his/her tweets' text

CASOS

7 June 2020 Sumeet Kumar 30



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## In 'Label Mixing', add the top 5% confident predictions as new training examples in the next iteration

- In **co-training**, more confident predictions (of both classifiers) are added as new training data in each iteration
- In each iteration, we use the **top 5% predictions of both classifiers** as new training examples. In case of a conflict among classifiers, we use the the more confident prediction

CASOS  
7 June 2020 Sumeet Kumar 31

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Joint Model: It Combines the Predictions of Both the Text and the Network Classifier

$$s_j = \begin{cases} 0, & \text{if } C_j^T = C_j^I = 0 \\ s_j^T, & \text{elif } C_j^T \geq C_j^I \\ s_j^I, & \text{Otherwise} \end{cases}$$

$s_j$  = stance of  $j^{\text{th}}$  user (joint model)  
 $s_j^T$  = stance of  $j^{\text{th}}$  user based on text  
 $s_j^I$  = stance of  $j^{\text{th}}$  user based on interaction  
 $c_j^T$  = user text based confidence estimate  
 $c_j^I$  = user interaction based confidence

The joint model uses the predictions of the more confident of the two classifiers (text and network) to predict the final stance

CASOS  
7 June 2020 Sumeet Kumar 32





**Carnegie Mellon**  
**IST** Institute for SOFTWARE RESEARCH

## Summary - Three Steps to Train Two Stance Classifiers

**Step 1**  
 Extract Text Features and Users Networks

**Step 2**  
 Label 2 to 4 hashtags  
 #GunControlNow: Pro  
 #2ndAmendment: Anti

**Step 3**  
 Derive stance of other users from seed users

**CASOS**  
 7 June 2020 Sumeet Kumar 33

**Carnegie Mellon**  
**IST** Institute for SOFTWARE RESEARCH

## Experiment: Users' Stance Dataset on Three Controversial Topics

Events	Users	Tweets	RT Users	Hashtags	Endtags
Gun-control	70387	117679	15635	8587	5505
Obamacare	67937	123320	14807	11559	7376
Abortion	111463	173236	26818	15646	9784

**Dataset**

Events	Neutral	Pro	Anti	Total
Gun-control	60	156	288	504
Obamacare	33	108	363	504
Abortion	55	169	280	504

**Labeled Users in the Dataset**

Haokai Lu, James Caverlee, and Wei Niu. 2015. BiasWatch: A Lightweight System for Discovering and Tracking Topic-Sensitive Opinion Bias in Social Media. In Proceedings of the 24th ACM International on Conference on Information and Knowledge Management (CIKM '15). ACM, New York, NY, USA, 213-222. DOI: <https://doi.org/10.1145/2806416.2806573>

**CASOS**  
 7 June 2020 Sumeet Kumar 34



Carnegie Mellon  
IST Institute for Software Research

## Experiment: Manually Labeled Four Hashtags in Each Dataset

Dataset	Seed Hashtags	No. of Seed Users
Gun-control	#guncontrolnow: Pro, #endgunviolence: Pro, #2ndamendment: Anti, #secondamendment: Anti	Pro:782, Anti:321
Obamacare	#uniteblue: Pro, #ilikeobamacare: Pro, #defundobamacare: Anti, #dontfundit: Anti	Pro:1342, Anti:3883
Abortion	#prochoice: Pro, #reprorights: Pro, #prolife: Anti, #stand4life: Anti	Pro:499, Anti:2183

Labeled two pro and two anti hashtags in each dataset

CASOS

7 June 2020 Sumeet Kumar 35

Carnegie Mellon  
IST Institute for Software Research

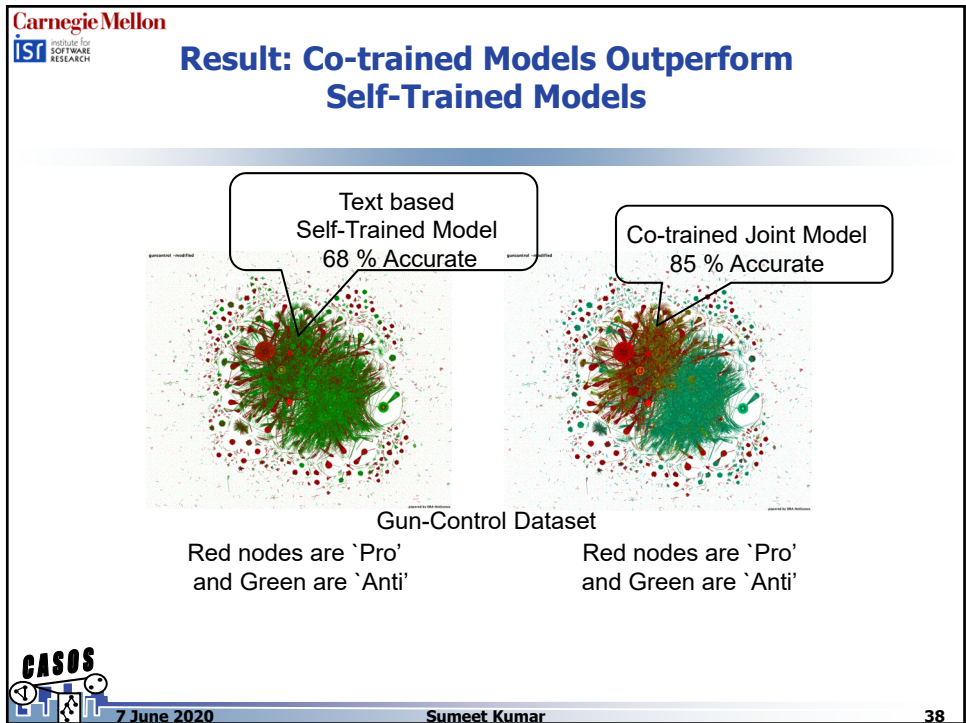
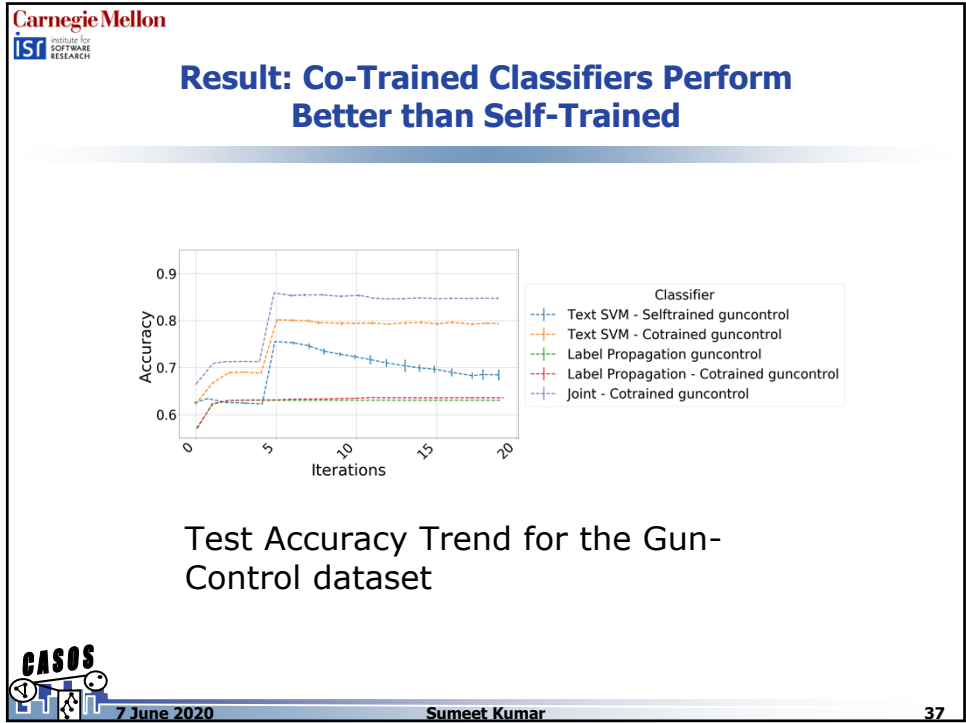
## Experiment Details

- 3 hop label propagation is used by the network classifier
- SVM classifier is used as the text classifier
  - TF-IDF features
  - Unigrams and bigrams are used
- Hyper-parameters were determined by evaluating them on the gun-control dataset
  - Top 250 hashtags are used
  - Top 5000 retweets are used

CASOS

7 June 2020 Sumeet Kumar 36





**Carnegie Mellon**  
**IST** Institute for SOFTWARE RESEARCH

## Result: Co-Trained Classifiers Perform Better than Self-Trained on All Dataset

- Text classifier improves by more than 17% on all three datasets
- LP in the figure implies bi-partite label propagation

**CASOS**  
 7 June 2020 Sumeet Kumar 39

**Carnegie Mellon**  
**IST** Institute for SOFTWARE RESEARCH

## Results: Comparing Different Seed Hashtags

Seed Hashtags	Classifier Type	Metric→ Dataset↓	Accuracy	Precision	Recall	F1-Score	Labeled Users (Fraction)
#stand4life:+1, #prochoice:-1	Hashtags + Retweets + Text Joint	abortion	0.38	0.38	0.98	0.54	0.98
#prolife:+1, #prochoice:-1	Hashtags + Retweets + Text Joint	abortion	<b>0.83</b>	0.77	0.78	0.78	0.94
#reprorights:+1, #stand4life:-1	Hashtags + Retweets + Text Joint	abortion	<b>0.83</b>	0.76	0.80	0.78	0.94
#endgunviolence:+1, #secondamendment:-1	Hashtags + Retweets + Text Joint	guncontrol	0.68	1.00	0.09	0.16	0.96
#2ndamendment:+1, #guncontrolnow:-1	Hashtags + Retweets + Text Joint	guncontrol	<b>0.81</b>	0.87	0.55	0.67	0.92
#2ndamendment:+1, #endgunviolence:-1	Hashtags + Retweets + Text Joint	guncontrol	0.68	1.00	0.10	0.18	0.96
#likeobamacare:+1, #defundobamacare:-1	Hashtags + Retweets + Text Joint	obamacare	0.80	0.81	0.16	0.26	0.95
#uniteblue:+1, #defundobamacare:-1	Hashtags + Retweets + Text Joint	obamacare	0.90	0.87	0.67	0.75	0.92
#uniteblue:+1, #dontfundit:-1	Hashtags + Retweets + Text Joint	obamacare	<b>0.91</b>	0.84	0.75	0.79	0.90

Comparison of Seed Hashtags: Some Seed Hashtags May lead to Poor Models

**CASOS**  
 7 June 2020 Sumeet Kumar 40



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH


## Any Questions So Far?


**CASOS**  
7 June 2020 Sumeet Kumar 41

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Case Study- Twitter COVID-19 Data Analyze Users Stance on `Fire Dr. Fauci`

### Topic -- Fire Dr. Fauci

 **Donald J. Trump** @realDonaldTrump  
Sorry Fake News, it's all on tape. I banned China long before people spoke up. Thank you @OANN

 **DeAnna Lorraine** @DeAnna4Congress · 1h  
Fauci is now saying that had Trump listened to the medical experts earlier he could've saved more lives.  
Fauci was telling people on February 29th that there was nothing to worry about and it posed no threat to the US public at large.  
Time to #FireFauci...

**CASOS**  
7 June 2020 Sumeet Kumar 42



Carnegie Mellon  
IST Institute for Software Research

### Case Study- Twitter COVID-19 Data Analyze Users Stance on `Fire Dr. Fauci`

**Input:**

1. Twitter data as Json file
2. Labeled Hashtags

**Output:**

1. Users Stance Labels
2. Other Hashtags Stance Labels
3. URL Stance labels

CASOS

7 June 2020 Sumeet Kumar 43

Carnegie Mellon  
IST Institute for Software Research

### Model for Propagating Stance from Users to Other Entities E.g. From Users' Stance to Stance Given by Hashtags

$S1: ?$   $S2: +1$   $S3: ?$   $S4: -1$   $S5: ?$

$W_{11}$   $W_{22}$   $W_{41}$   $W_{34}$   $W_{44}$   $W_{53}$

$\sigma' \left( \begin{matrix} S1 * W_{11} \\ S4 * W_{41} \end{matrix} \right) \rightarrow \#T1$

$\sigma' (S1 * W_{22}) \rightarrow \#T2$

$\sigma' (S5 * W_{53}) \rightarrow \#T3$

$\sigma' (S3 * W_{34}) \rightarrow \#T4$

S Users Stance       $\tilde{S}$  Hash'

Users Stance  $\rightarrow$  Hashtags Stance  
Users Stance  $\rightarrow$  URLs (Websites) Stance  
Users Stance  $\rightarrow$  Media URLs (Pictures) Stance

CASOS

7 June 2020 Sumeet Kumar 44



Carnegie Mellon  
IST Institute for Software Research

## Stance Mining Applied to 'Fire Dr. Fauci' in Covid Data

- Fire Dr. Fauci (vs Save Dr. Fauci)  
Tags used for data filtering: 'fauci', 'firing fauci', '#firefauci', '#firetrump', '#savefauci',
- Labeled seed hashtags for stance analysis  
firefauci:1, firedrfauci:1, faucithefraud:1, savefauci:-1, fauciisahero: -1, keepfauci: -1, firetrumpkeepfauci: -1

CASOS

7 June 2020 Sumeet Kumar 45

Carnegie Mellon  
IST Institute for Software Research

## Agenda - I Try to Answer Two Questions in This part of the Talk

1. How to identify the users that are pro (or anti) a given topic?
2. How the users differ in their usage of hashtags?

CASOS

7 June 2020 Sumeet Kumar 46



**Carnegie Mellon**  
**IST** Institute for SOFTWARE RESEARCH

## Start ORA

- Start ORA and Import Data

**CASOS** 7 June 2020 Sumeet Kumar 47

**Carnegie Mellon**  
**IST** Institute for SOFTWARE RESEARCH

## Import Twitter Data

- Pick Twitter Data

**CASOS** 7 June 2020 Sumeet Kumar 48





Carnegie Mellon  
IST Institute for Software Research

## Import Twitter Data

- Pick Twitter Data
- Select Import Options

Import Data into ORA

Select the twitter data format:  
Twitter JSON

Select one or more data files:  
/Users/sumeetku/Downloads/Fauci/output/fauci\_tweets\_apr24.1590010540003.anonymized.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 3 at 00:00:00  
 Ignore tweets after: 2020 June 3 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:

Cancel < Back Next > Finish

CASOS

7 June 2020 Sumeet Kumar 49

Carnegie Mellon  
IST Institute for Software Research

## Import Twitter Data

- Pick Twitter Data
- Select Import Options

Import Data into ORA

Select the twitter data format:  
Twitter JSON

Select one or more data files:  
/Users/sumeetku/Downloads/Fauci/output/fauci\_tweets\_apr24.1590010540003.anonymized.json Browse

Create a separate dynamic meta-network per file

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

Tweet message options:  
 Store the tweet message as a tweet node attribute  
 Parse the tweet message into words with options:  
Characters to remove before applying thesaurus and delete list: !,?\*,<>&@&#%&'><|

Use the universal thesaurus  
 Use domain thesaurus Browse  
 Use the universal delete list  
 Use domain delete list Browse

Use built-in delete lists:  stopwords  numbers  ordinal numbers  
 Create usage measures tweet node attributes

Date aggregation options:  
 Aggregate by 6 Hours  
 Create uniform time periods  
 Create retweets in creation time period

Geospatial options:  
 Automatically locate country and add as an attribute  
 Automatically locate US state and add as an attribute

Cancel < Back Next > Finish

CASOS

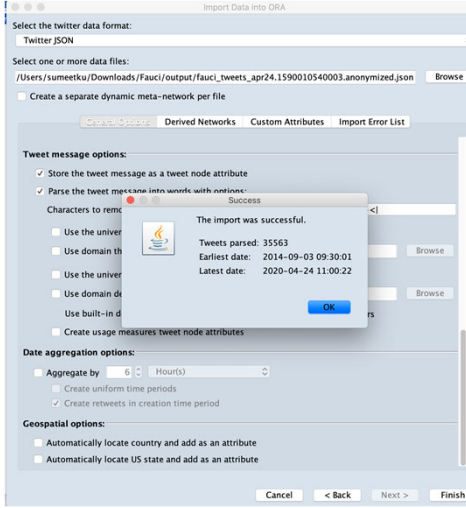
7 June 2020 Sumeet Kumar 50



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Import Twitter Data

- Pick Twitter Data
- Select Import Options



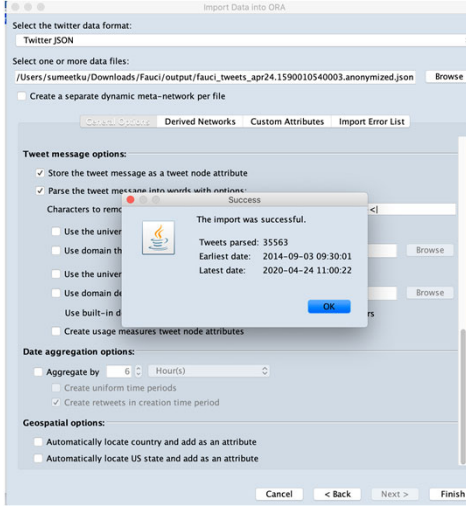
CASOS

7 June 2020 Sumeet Kumar 51

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Import Twitter Data

- Pick Twitter Data
- Select Import Options
- Import Data



CASOS

7 June 2020 Sumeet Kumar 52



**Carnegie Mellon**  
**IST** Institute for SOFTWARE RESEARCH

## Import Twitter Data

The screenshot shows the ORA 3.0.9.9.107 interface. The 'Meta-Network Manager' window on the left lists various network statistics such as Agent size (36184), Hashtag size (2666), Location size (12), Tweet size (6881), and URI size (14271). The 'Meta-Network' window on the right displays detailed statistics for the dataset 'Twitter JSON faul\_tweets\_apr24\_15901054003\_anonymized'. It includes general statistics like source count (0), node count (123167), and network count (21). It also provides link statistics, all link values (Min: 1, Max: 5335, Mean: 1.857628, StdDev: 22.721647, Sum: 2601614), non self-loops (1399001), self-loops (1502), and component statistics including isolates (0), dyads (0), triads (5), and larger sizes (38).

**CASOS**

7 June 2020 **Sumeet Kumar** 53

**Carnegie Mellon**  
**IST** Institute for SOFTWARE RESEARCH

## Start Stance Detection Analysis

- Pick the option shown below

The screenshot shows the ORA 3.0.9.9.107 interface. The 'Meta-Network Manager' window on the left is the same as in the previous slide. The 'Analyze' menu is open, showing options like 'Knowledge Networks & Network Text Analysis', 'Statistical Procedures and Diagnostics', 'Social Media', 'Geospatial', 'Classify Groups and Networks', 'BND & Community Assessment...', 'Core Networks...', 'Ego Net...', 'Locate Groups...', 'Stance Detection', and 'Triad Census...'. The 'Stance Detection' option is highlighted.

**CASOS**

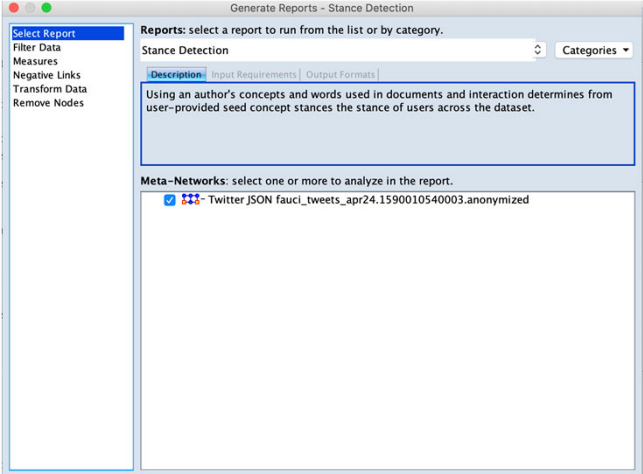
7 June 2020 **Sumeet Kumar** 54



Carnegie Mellon Institute for Software Research

## Start Stance Detection Analysis

- Pick the option shown below

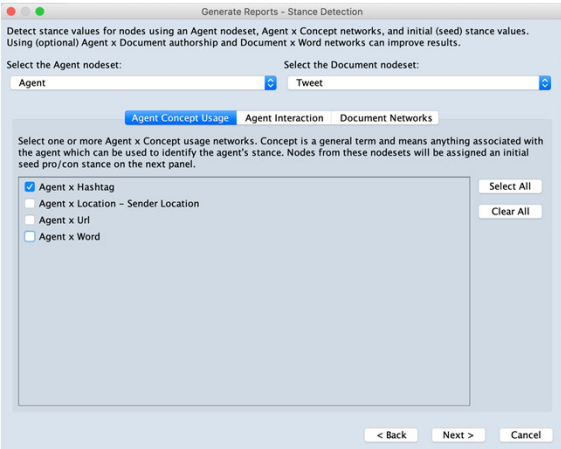


7 June 2020 55

Carnegie Mellon Institute for Software Research

## Stance Detection Analysis

- Pick the option shown below



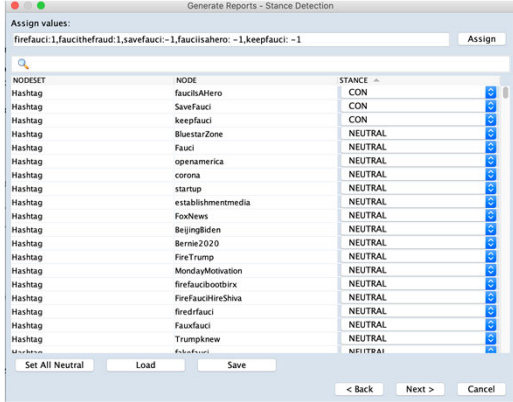
7 June 2020 Sumeet Kumar 56



**Carnegie Mellon**  
IST Institute for SOFTWARE RESEARCH

## Stance Detection Analysis

- Assign stance values to a selected set of hashtags
- You can copy paste the values from the slide (or enter it manually)



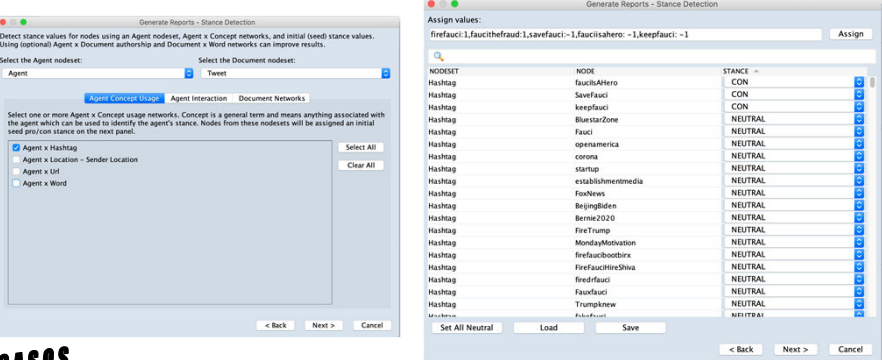
NODESET	NODE	STANCE
Hashtag	faucisAHero	CON
Hashtag	SaveFauci	CON
Hashtag	keepfauci	CON
Hashtag	BluestarZone	NEUTRAL
Hashtag	Fauci	NEUTRAL
Hashtag	openamerica	NEUTRAL
Hashtag	corona	NEUTRAL
Hashtag	startup	NEUTRAL
Hashtag	establishmentmedia	NEUTRAL
Hashtag	FoxNews	NEUTRAL
Hashtag	BeijingBiden	NEUTRAL
Hashtag	Bernie2020	NEUTRAL
Hashtag	FireTrump	NEUTRAL
Hashtag	MondayMotivation	NEUTRAL
Hashtag	firefauciboothrx	NEUTRAL
Hashtag	FireFauciHireShiva	NEUTRAL
Hashtag	firedfauci	NEUTRAL
Hashtag	Fauxfauci	NEUTRAL
Hashtag	Trumpknew	NEUTRAL

7 June 2020 Sumeet Kumar 57

**Carnegie Mellon**  
IST Institute for SOFTWARE RESEARCH

## Stance Detection Analysis

- Assign stance values to a selected set of hashtags
- You can copy paste the values from the slide (or enter it manually)



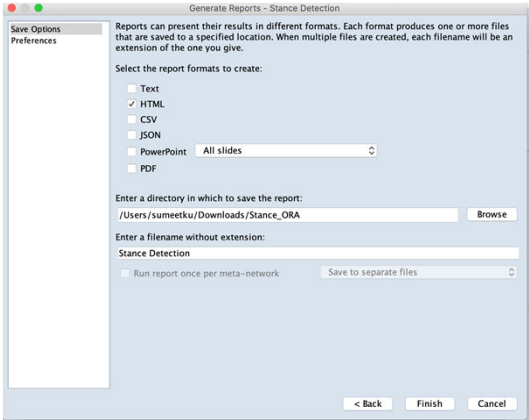
7 June 2020 Sumeet Kumar 58



Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Stance Detection Analysis

- Select save option
- Stance detection report will be generated



7 June 2020 Sumeet Kumar 59

Carnegie Mellon  
IST Institute for SOFTWARE RESEARCH

## Stance Detection Analysis

- Stance detection report – shows selected options

**STANCE DETECTION REPORT**

Input data: Twitter JSON files\_events\_apr24\_159000540003.amazonaws.com  
Start time: Wed Jun 3 10:03:29 2020  
[Data Description](#)

**Parameters**

Agent network	Agent
Concept network	Healing
Agent x Concept network	Agent x Healing
Agent Interaction network	Agent x Agent - Quoted By, Agent x Agent - Retweeted, Agent x Agent - Mentioned By, Agent x Agent - Retweeted By, Agent x Agent - Retweeted By, Agent x Agent - Retweeted By, Agent x Agent - AS Communication
Document network	Text
Document Authorship network	Agent x Tweet - Sender
Document Word network	Tweet x Healing

**Initial Node Stances (seeds)**

Network	Node	Stance
Healing	infocast	pro
Healing	FacebookFeed	pro
Healing	FacebookAlerts	contra
Healing	SwiftFacet	contra
Healing	Amplifacet	contra

Showing 1 to 5 of 5 entries

**Agent Stance Summary**

	Number of nodes	Mean node confidence
Pro Nodes	2427	0.688
Contra Nodes	493	0.793
Not Assigned	744	

7 June 2020 Sumeet Kumar 60



**Carnegie Mellon**  
IST Institute for SOFTWARE RESEARCH

## Stance Detection Analysis

- Stance detection report – shows Pro/Con Users

**Agent Pro Stance**

The pro-stance Agent nodes ranked by the confidence of the stance calculation.

If the node of interest has a higher than normal value (greater than 1 standard deviation(s) above the mean) the row is colored red. The row is green if the node is within 1 standard deviation of the mean. Finally, the row is colored blue if the node has a lower than normal value (less than one standard deviation(s) below the mean).

Show 10 entries

Rank	Agent	Value
1	00227030402005	1
2	00227030204006	1
3	00227030204008	1
4	00227030207008	1
5	00227030404007	1
6	00227030404007	1
7	00227030404007	1
8	00227030404007	1
9	00227030404007	1
10	00227030404007	1

Showing 1 to 10 of 100 entries

**Agent Contra Stance**

The contra-stance Agent nodes ranked by the confidence of the stance calculation.

If the node of interest has a higher than normal value (greater than 1 standard deviation(s) above the mean) the row is colored red. The row is green if the node is within 1 standard deviation of the mean. Finally, the row is colored blue if the node has a lower than normal value (less than one standard deviation(s) below the mean).

Show 10 entries

Rank	Agent	Value
1	00227030404007	1
2	00227030404007	1
3	00227030404007	1
4	00227030404007	1
5	00227030404007	1
6	00227030404007	1
7	00227030404007	1
8	00227030404007	1
9	00227030404007	1
10	00227030404007	1

Showing 1 to 10 of 100 entries

7 June 2020 Sumeet Kumar 61

**Carnegie Mellon**  
IST Institute for SOFTWARE RESEARCH

## Stance Detection Analysis

- Stance detection report – shows Pro/Con hashtags by confidence

**Hashtag Pro Stance**

The pro-stance Hashtag nodes ranked by the confidence of the stance calculation.

If the node of interest has a higher than normal value (greater than 1 standard deviation(s) above the mean) the row is colored red. The row is green if the node is within 1 standard deviation of the mean. Finally, the row is colored blue if the node has a lower than normal value (less than one standard deviation(s) below the mean).

Show 10 entries

Rank	Hashtag	Value
1	SD	1
2	BlackPan	1
3	SDConstance	1
4	TR	1
5	AN	1
6	ANSDSD	1
7	ANTHONYHAUCHE	1
8	aha	1
9	SDConstance	1
10	AnthonyHauchec	1

Showing 1 to 10 of 100 entries

**Hashtag Contra Stance**

The contra-stance Hashtag nodes ranked by the confidence of the stance calculation.

If the node of interest has a higher than normal value (greater than 1 standard deviation(s) above the mean) the row is colored red. The row is green if the node is within 1 standard deviation of the mean. Finally, the row is colored blue if the node has a lower than normal value (less than one standard deviation(s) below the mean).

Show 10 entries

Rank	Hashtag	Value
1	SDConstance	1
2	AnthonyHauchec	1
3	SDConstance	1
4	SDConstance	1
5	SDConstance	1
6	SDConstance	1
7	SDConstance	1
8	SDConstance	1
9	SDConstance	1
10	SDConstance	1

Showing 1 to 10 of 100 entries

CASOS 7 June 2020 Sumeet Kumar 62



**Carnegie Mellon**  
IST Institute for SOFTWARE RESEARCH

## Stance Detection Analysis

- Stance detection report – shows Pro/Con hashtags by usage

**Agent Pro Stance, Hashtag Usage**

The Hashtag nodes used the most by pro-stance agents, weighted higher when not used by contra-stance agents.  
If the node of interest has a higher than normal value (greater than 1 standard deviation(s) above the mean) the row is colored red. The row is green if the node is within 1 standard deviation of the mean. Finally, the row is colored blue if the node has a lower than normal value (less than one standard deviation(s) below the mean).

Rank	Hashtag	Value
1	#ObamaOTW	105
2	#Africa	89
3	#antiRac	65
4	#MuslimsAre	62
5	#Indivisibl	44
6	#open	34,091
7	#BillSanderForces	35
8	#todayRites	35
9	#MuslimsAre	28,333
10	#MuslimMood	28

Showing 1 to 10 of 100 entries

**Agent Contra Stance, Hashtag Usage**

The Hashtag nodes used the most by contra-stance agents, weighted higher when not used by pro-stance agents.  
If the node of interest has a higher than normal value (greater than 1 standard deviation(s) above the mean) the row is colored red. The row is green if the node is within 1 standard deviation of the mean. Finally, the row is colored blue if the node has a lower than normal value (less than one standard deviation(s) below the mean).

Rank	Hashtag	Value
1	#china	97,241
2	#ChinaAfrica	48
3	#Resistance	45,847
4	#Africa	35
5	#SelfHate	12,875
6	#open	35
7	#antiRac	27
8	#Resistance	20,500
9	#Free	16
10	#FreeBlackMinds	16

Showing 1 to 10 of 100 entries

**CASO**

7 June 2020 Sumeet Kumar 63

**Carnegie Mellon**  
IST Institute for SOFTWARE RESEARCH

## Thank You

Please feel free to ask/send your questions

**CASO**

7 June 2020 Sumeet Kumar 64

