



# CHARACTERIZING POLITICALLY ENGAGED USERS' BEHAVIOR DURING THE 2016 US PRESIDENTIAL CAMPAIGN

---

**Josemar Alves Caetano, Jussara Almeida,  
Humberto Torres Marques-Neto**

# Social networks and political campaigns



# Reach of the candidates on Twitter (election day)

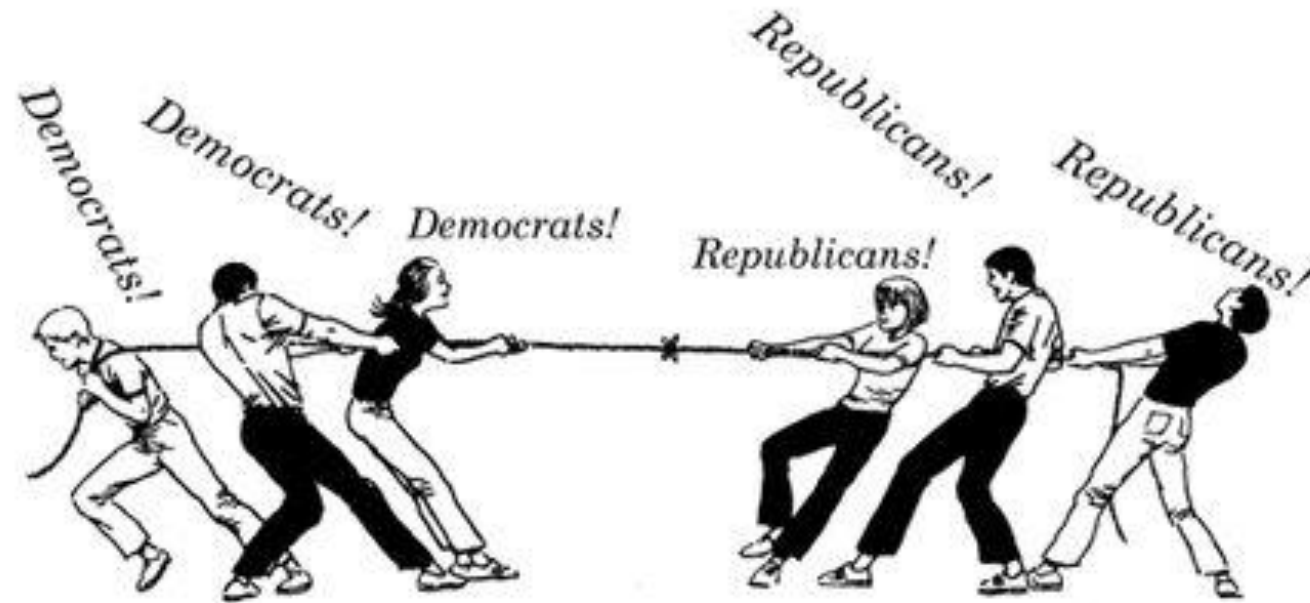


**17 million** followers  
**35 thousand** published tweets



**12 million** followers  
**9 thousand** published tweets

# Political biases on social networks



# Political biases on social networks



*Advocates*

# Other political groups



**Political Bots**



***Regular Users***

# Main objective

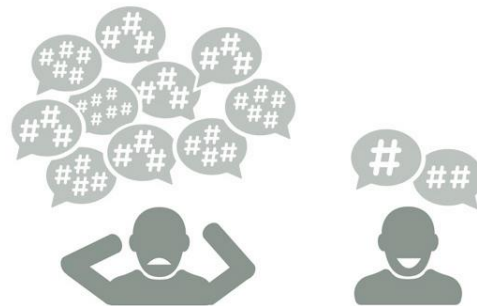
Characterize users in an online social network taking into account political biases and therefore different behaviors



# Characterizations



Which features highlight each group



Language Patterns Analysis



Popular users of each group



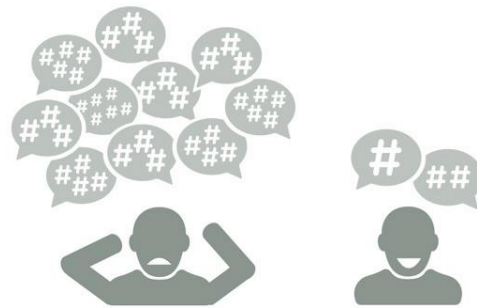
Mood Variation Analysis



# Feature characterization



Which features highlight each group



Language Patterns Analysis



Popular users of each group

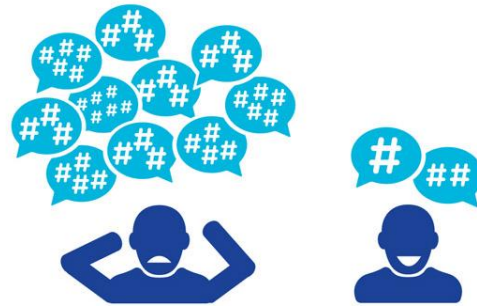


Mood Variation Analysis

# Language characterization



Which features  
highlight each group



Language Patterns  
Analysis



Popular users  
of each group

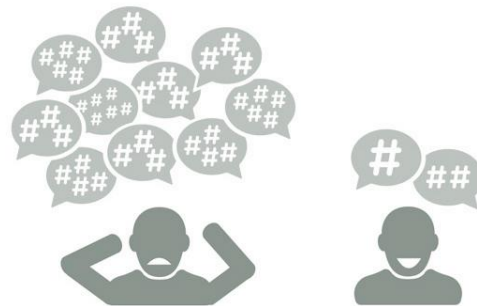


Mood Variation  
Analysis

# Profile characterization



Which features highlight each group



Language Patterns Analysis



Popular users of each group

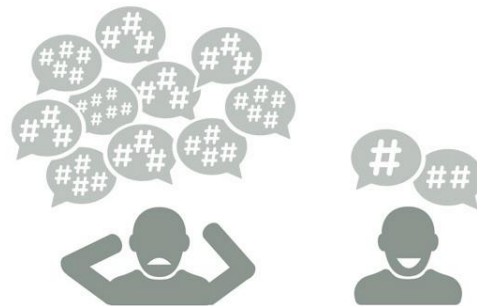


Mood Variation Analysis

# Mood characterization



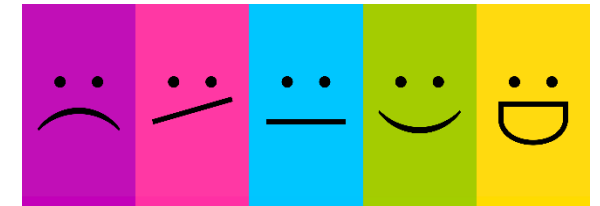
Which features highlight each group



Language Patterns Analysis



Popular users of each group



Mood Variation Analysis

# To perform these characterizations...



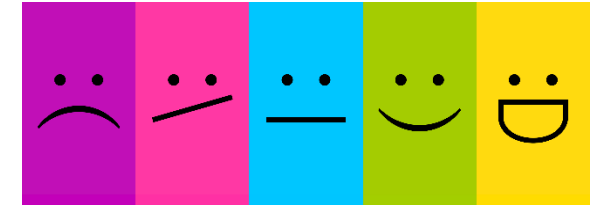
Which features  
highlight each group



Language Patterns  
Analysis

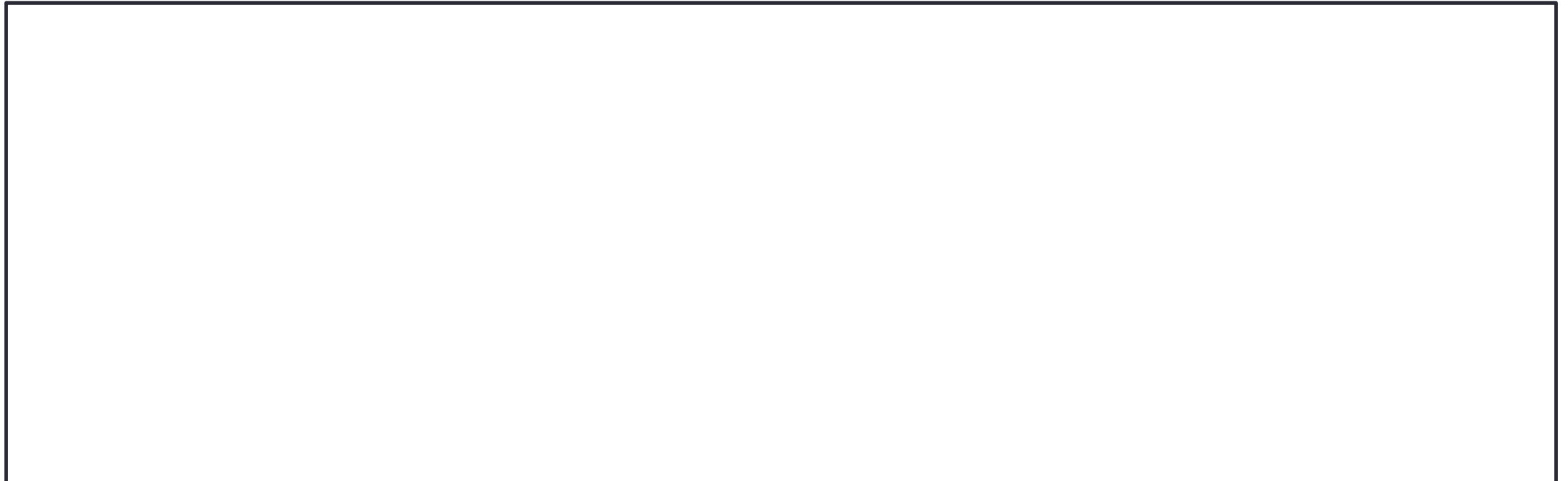


Popular users  
of each group

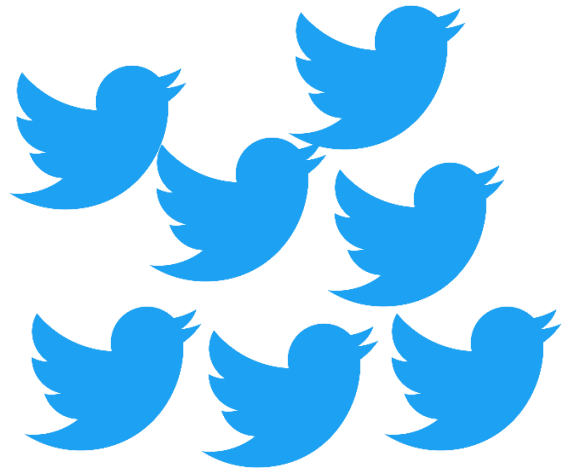


Mood Variation  
Analysis

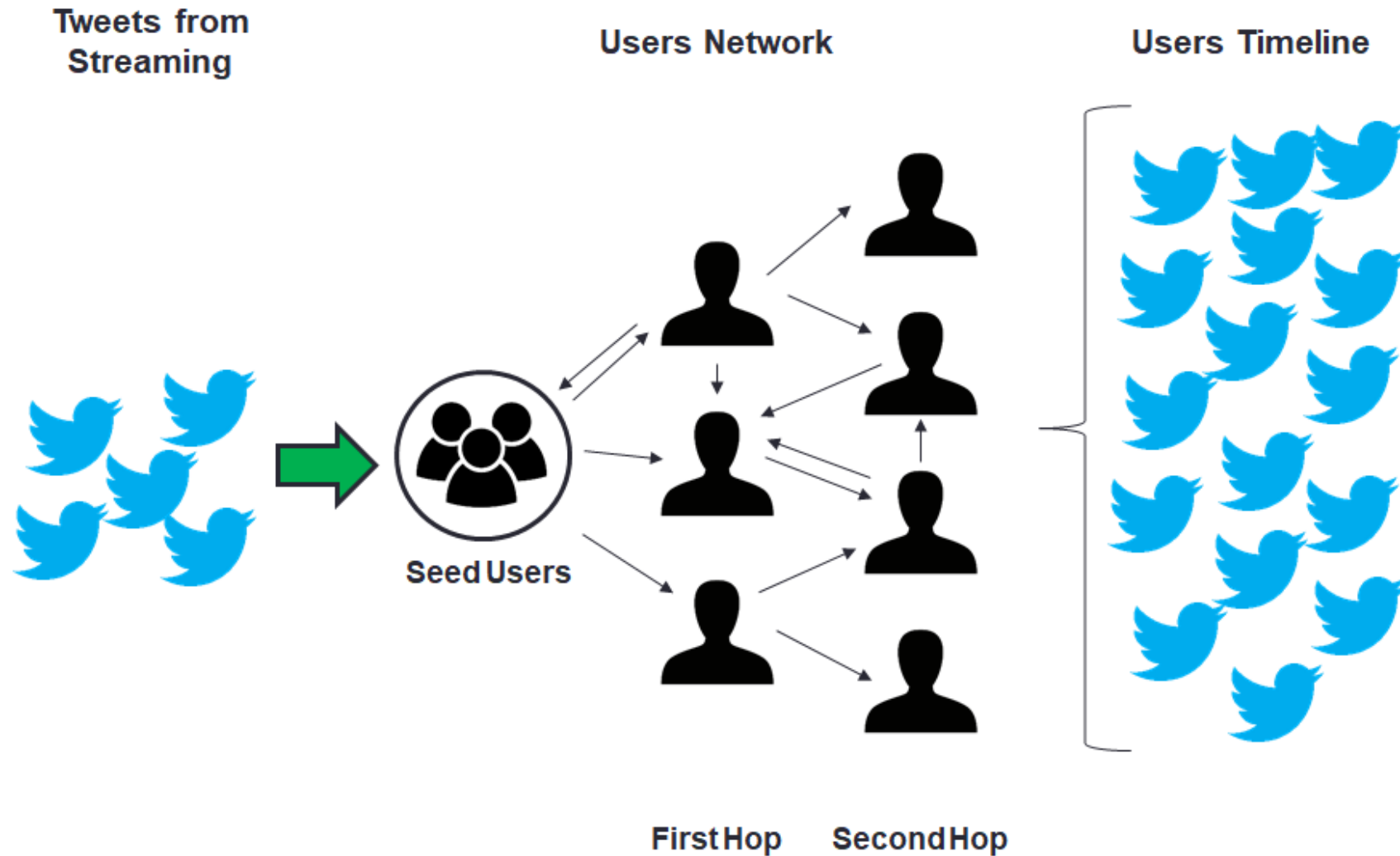
# Methodology



# Collecting Twitter data



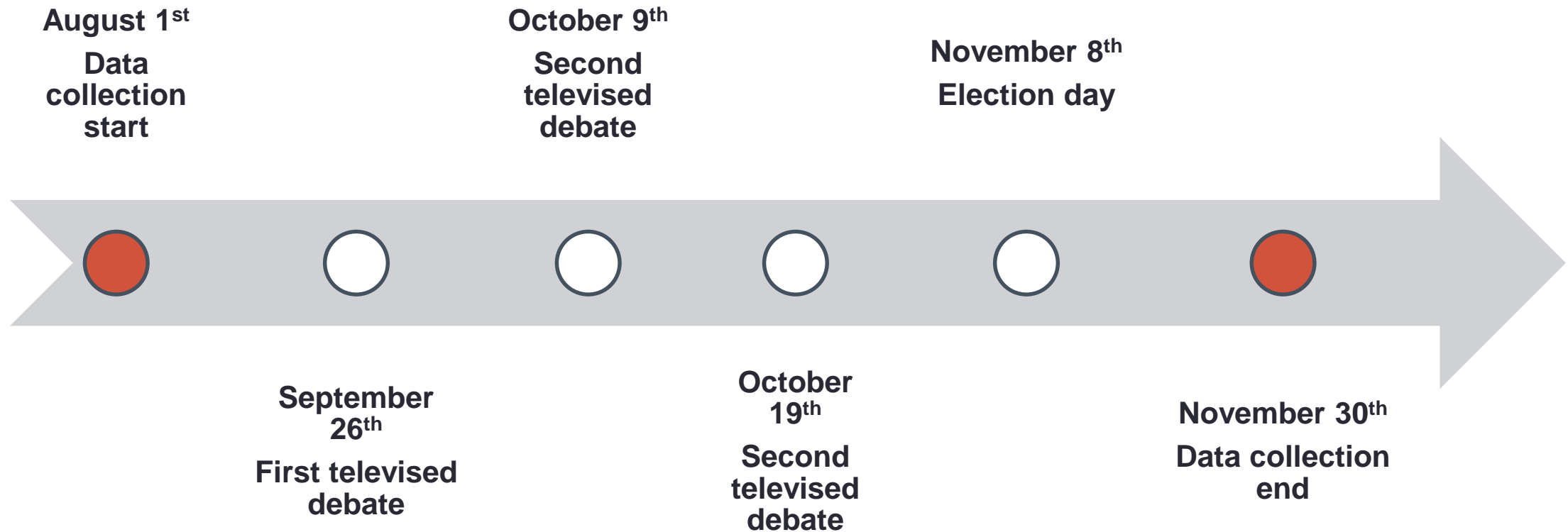
# Data collection process





# Data collection period

- Data collected over 122 days (August 1<sup>st</sup> to November 30<sup>th</sup> 2016)



# Dataset

<b># of tweets</b>	23 mi
<b># of users</b>	115 k
<b># of relationships</b>	1.8 mi

# Identifying political tweets



# Candidates references considered

<b>Donald Trump</b>	<b>Hillary Clinton</b>
@realDonaldTrump	@HillaryClinton
Trump	Hillary
DT	HC

# Political hashtags

	<b>Donald Trump</b>	<b>Hillary Clinton</b>
1	#Trump	#ImWithHer
2	#MAGA	#NeverTrump
3	#TrumpTrain	#Hillary
4	#TrumpPence16	#HillaryClinton
5	#DrainTheSwamp	#Hillary2016
6	#tcot	#UniteBlue
7	#Trump2016	#VoteBlue
8	#GOP	#HillaryBecause
9	#PJNET	#OHHillYes
10	#cco	#HillYes

# Tweet sentiment analysis



Negative



Neutral



Positive

# How sentiment analysis works?

- **SentiStrength** tool
  - Dictionary containing emotional words



# Political sentiment analysis



I **love** Hillary Clinton  
and her ideas.

I **hate** Hillary Clinton  
and her ideas.





# Political sentiment analysis problem

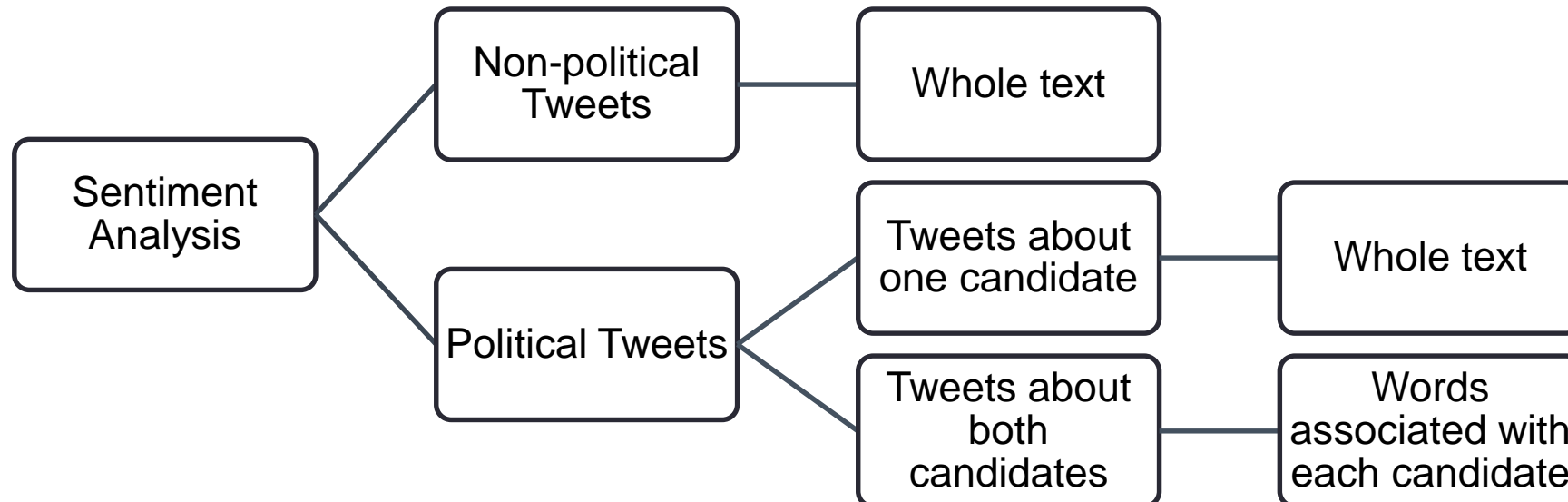


I **love** Hillary Clinton  
but I **hate** Donald  
Trump.

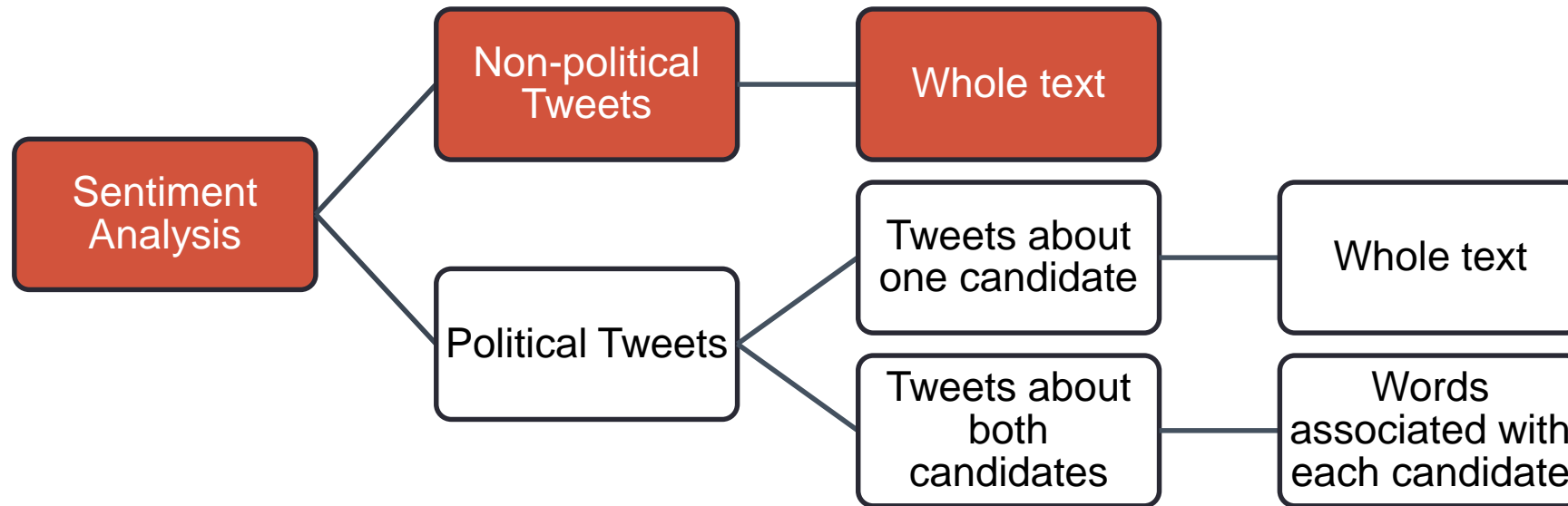
I **hate** Hillary Clinton  
but I **love** Donald  
Trump.



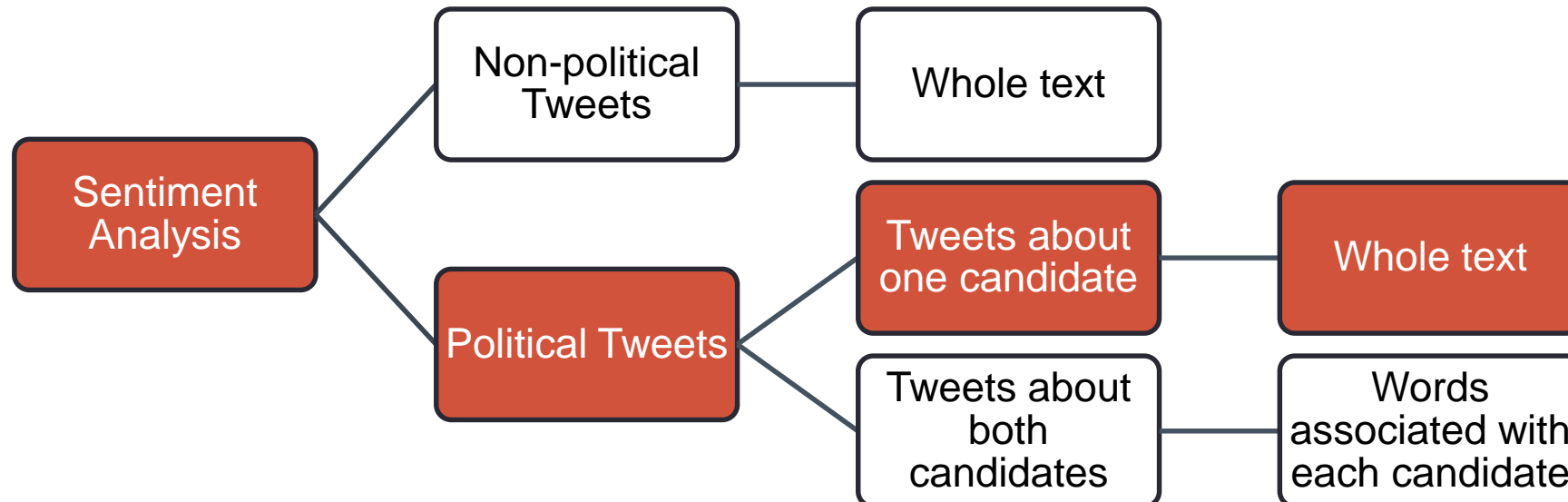
# Sentiment analysis approaches



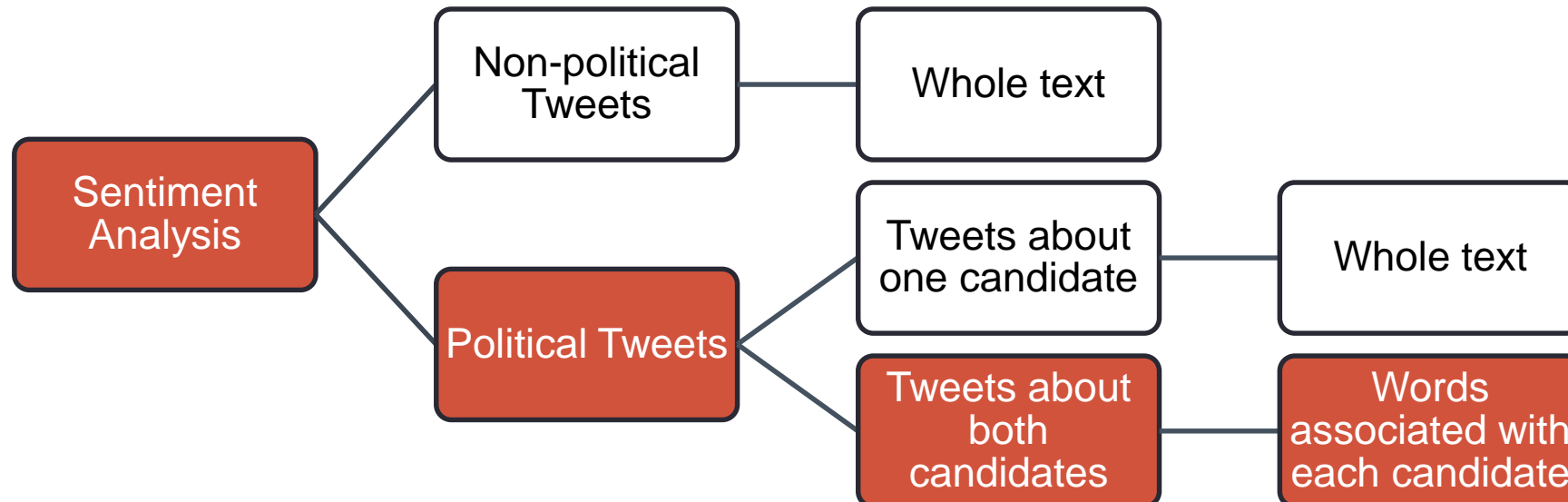
# Non-political tweets



# Political tweets about one candidate



# Political tweets about both candidates



# Identifying words related to candidates

- **Stanford Parser** tool
  - Natural language processor



# Identifying politically engaged user groups



**Hillary's  
Advocates**



**Trump's  
Advocates**

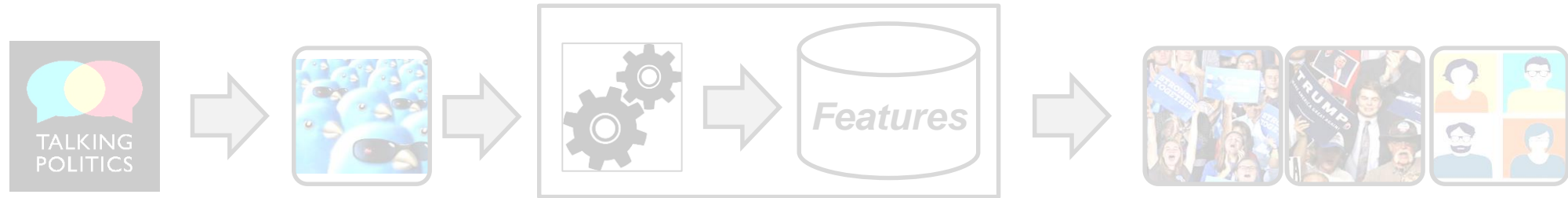


**Political  
Bots**



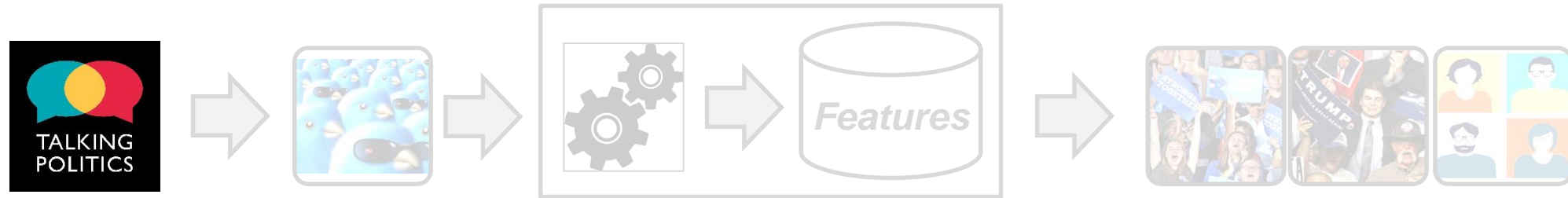
**Regular  
Users**

# Data mining process



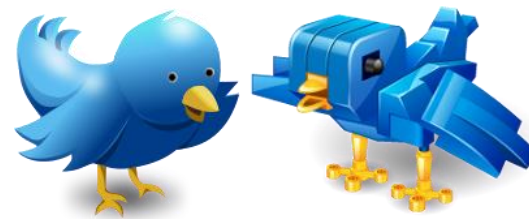
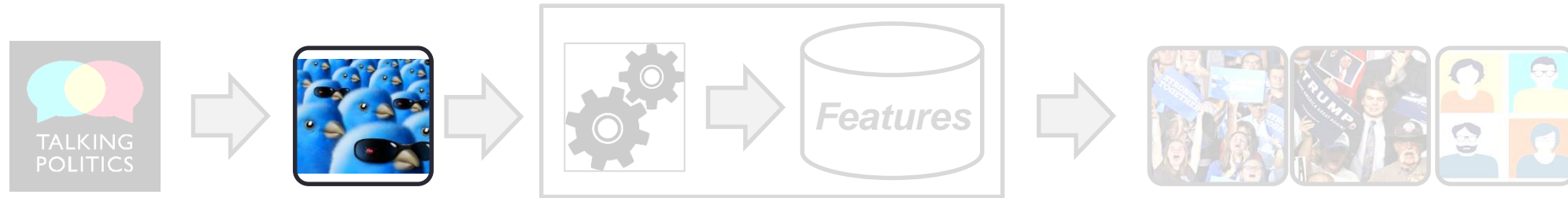


# Removing outliers



Eliminated users that did not have political tweets

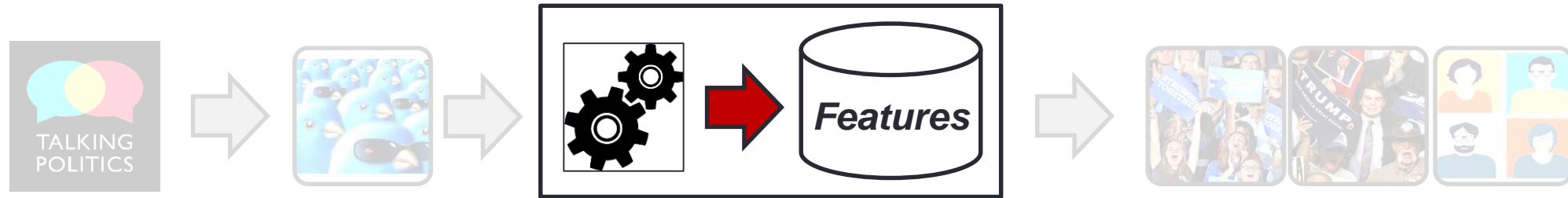
# Identifying political bots



**BotOrNot**

Users with **BotOrNot score  $\geq 0.75$**

# Feature set engineering



## 44 Features

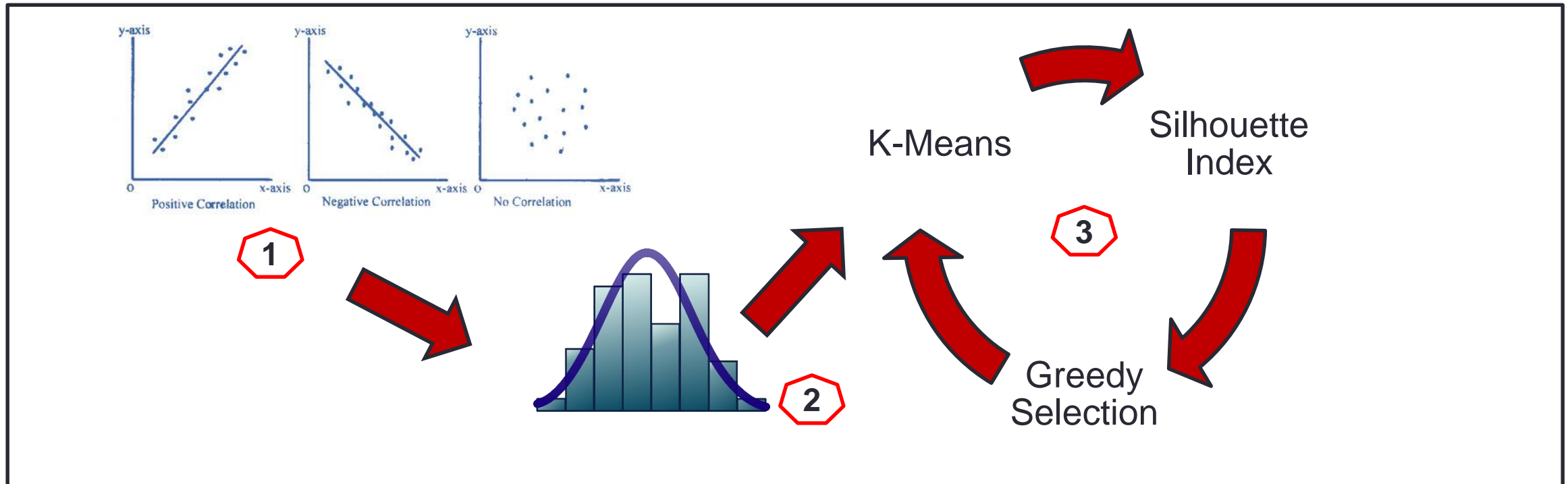
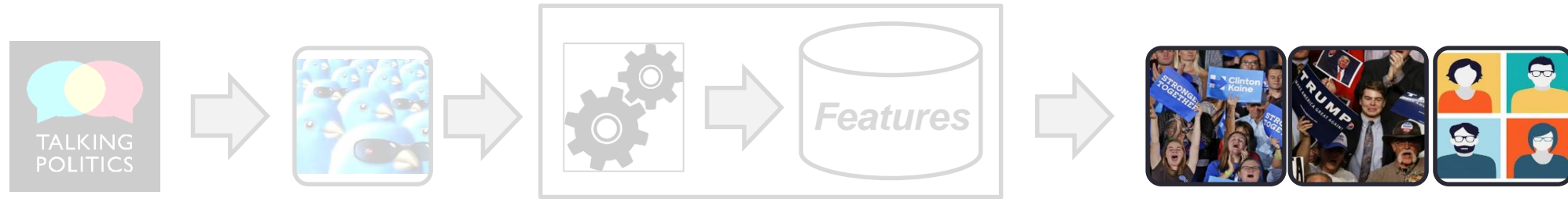
**User Metadata**  
6

**Syntax**  
10

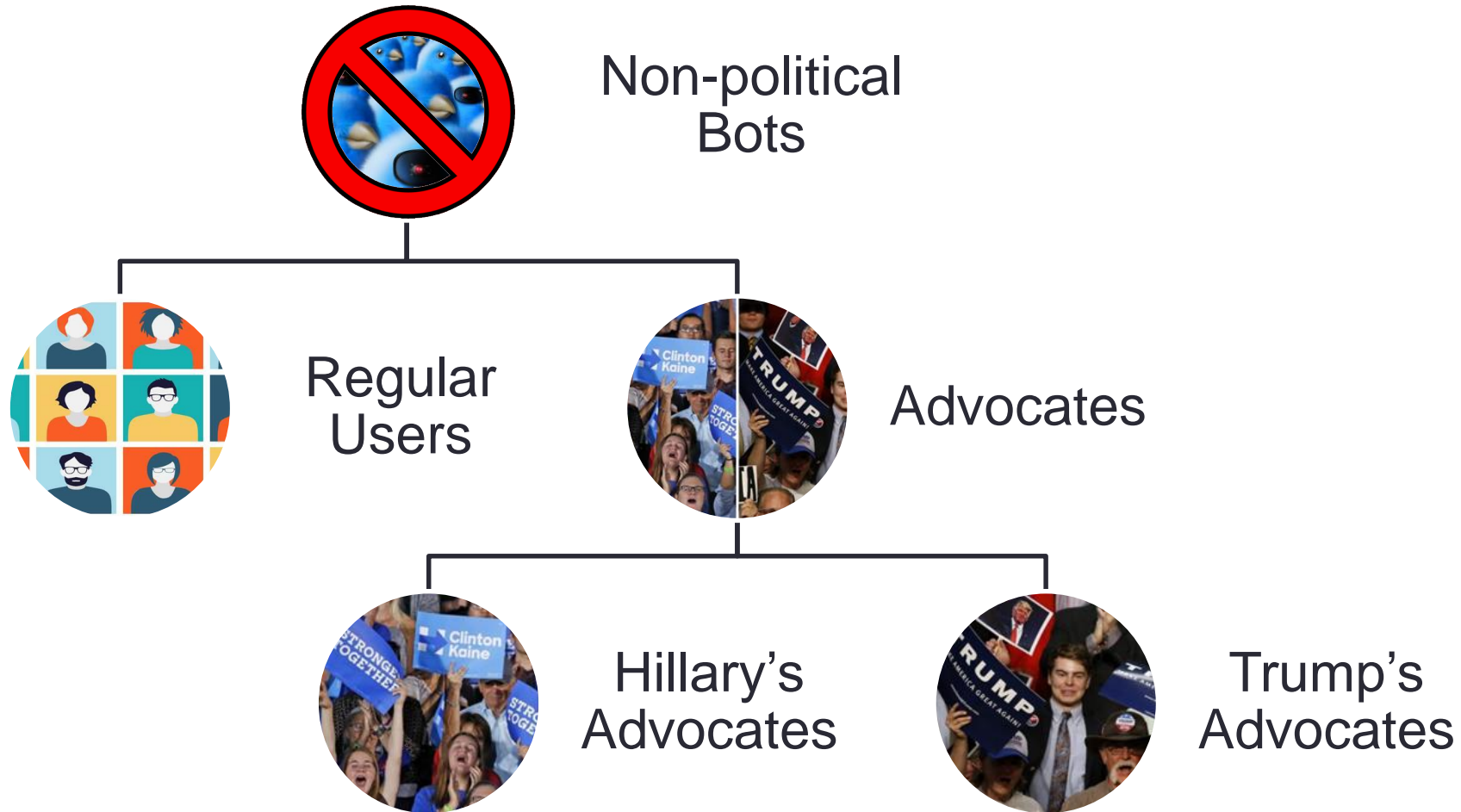
**Political Bias**  
11

**Sentiment Analysis**  
17

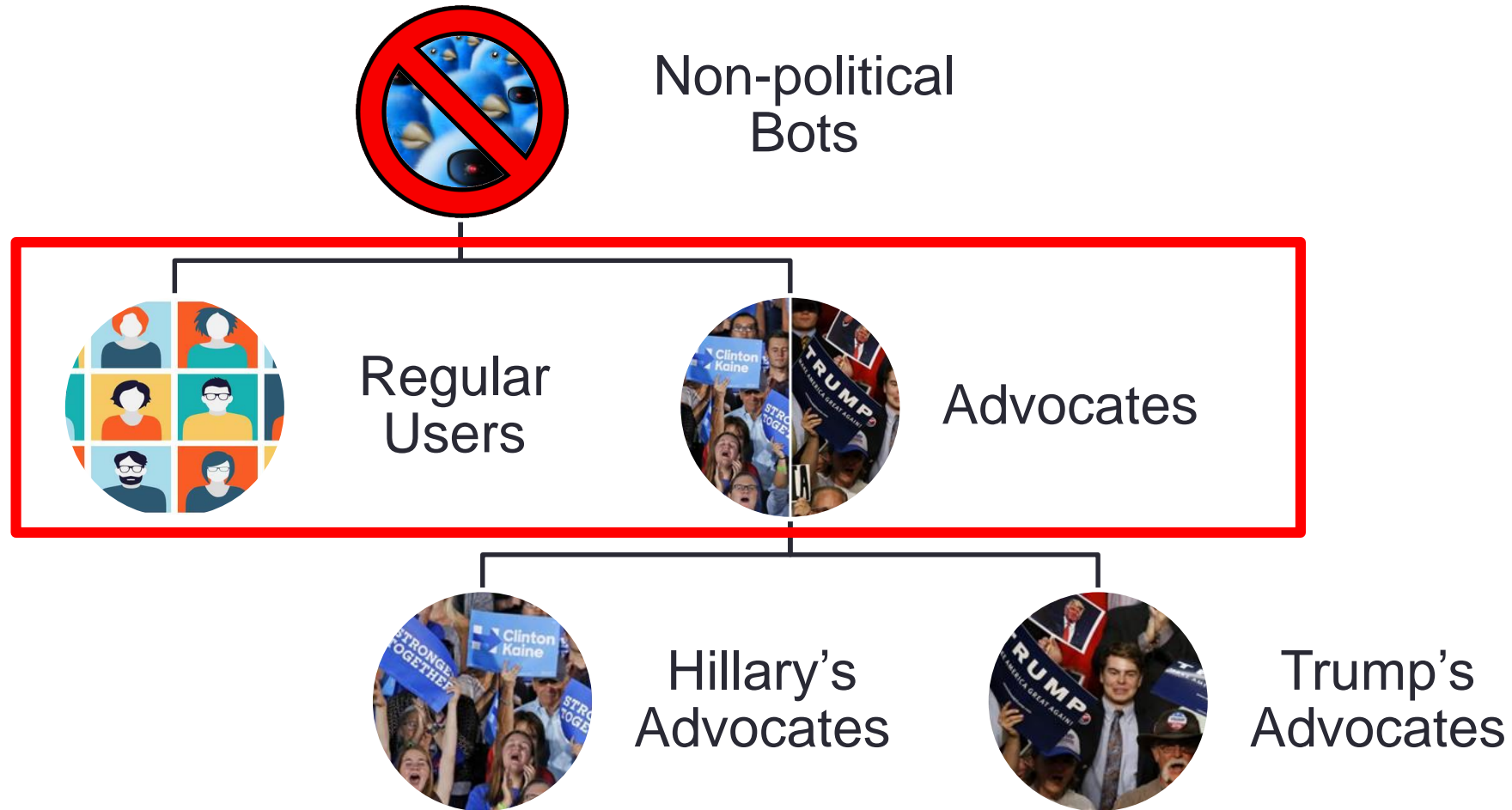
# Identifying Regular Users, Trump's Advocates, and Hillary's Advocates



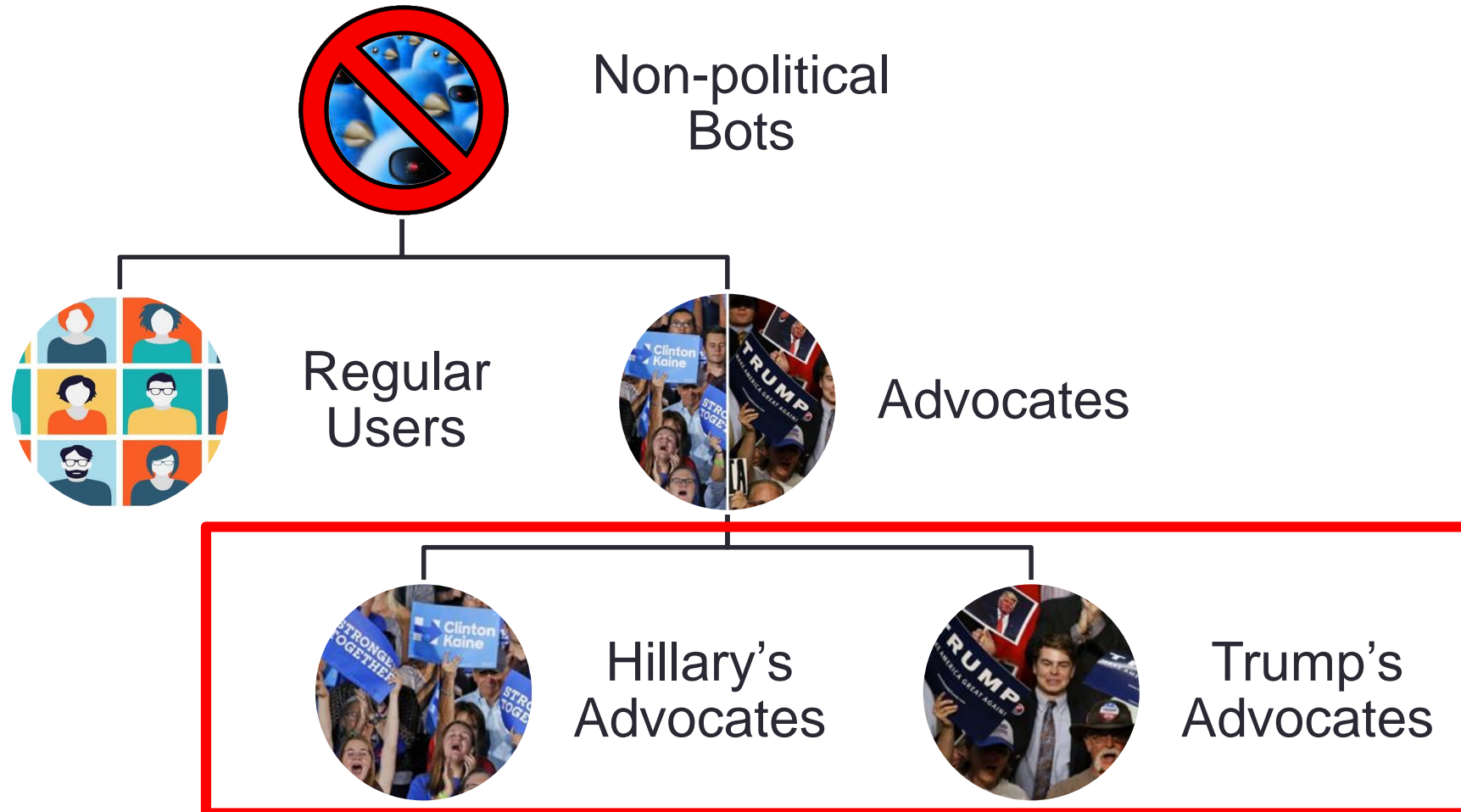
# Two steps clustering



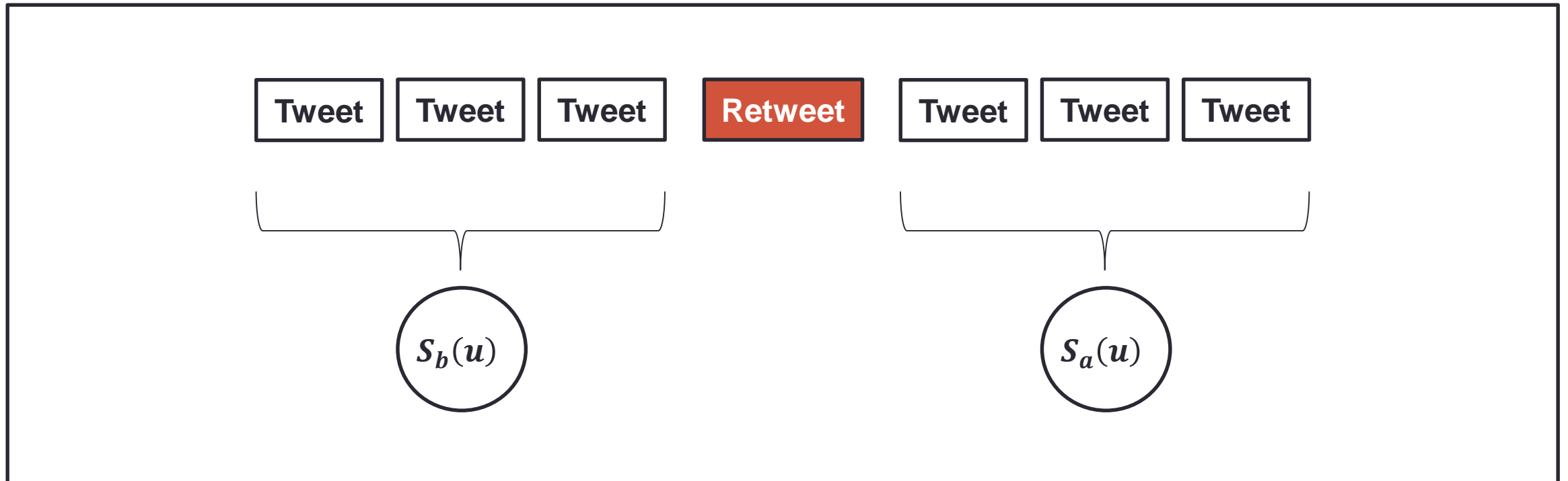
# Identifying regular users and advocates



# Identifying Trump's Advocates and Hillary's Advocates



# Mood variation analysis





# Subjective Well-Being definition

$$S_u(t_1, t_2) = \frac{N_{p_u}(t_1, t_2) - N_{n_u}(t_1, t_2)}{N_{p_u}(t_1, t_2) + N_{n_u}(t_1, t_2)}$$

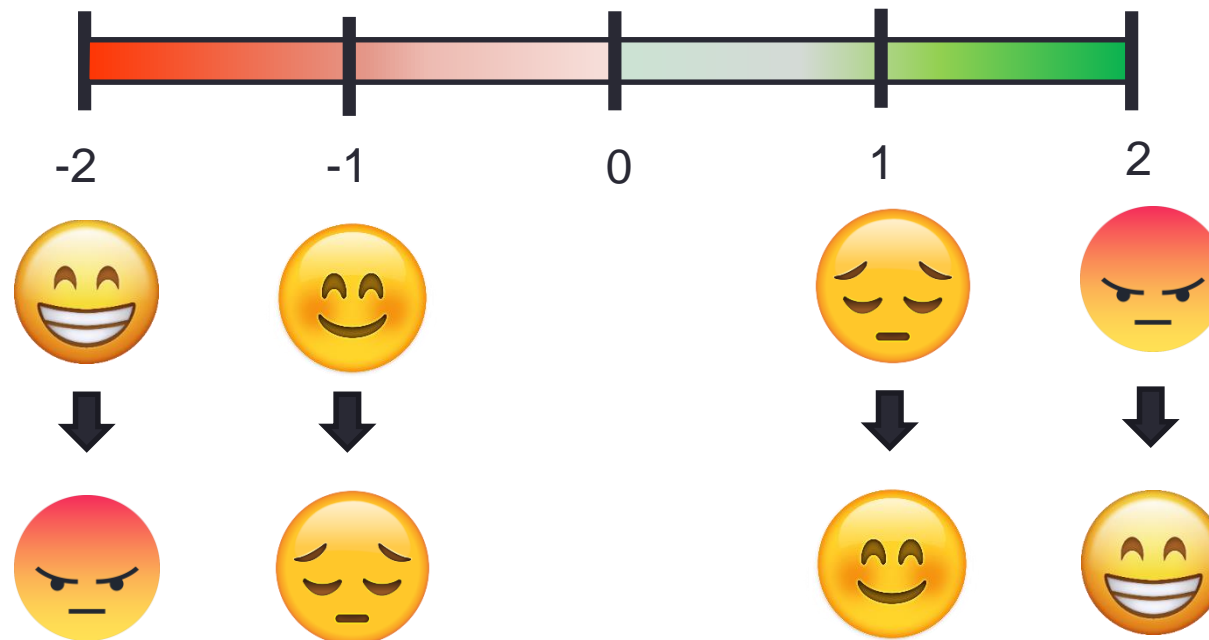
- $N_{p_u}(t_1, t_2)$ : positive tweets total
- $N_{n_u}(t_1, t_2)$ : negative tweets total
- $S_u(t_1, t_2)$ :  $-1 \leq S_u \leq 1$

# Mood variation definition

$$\Delta S_u = S_u(t, t + \delta) - S_u(t, t - \delta)$$

- $S_u(t, t + \delta)$ : SWB after retweet
- $S_u(t, t - \delta)$ : SWB before retweet
- $\Delta S_u$  values:  $-2 \leq \Delta S_u \leq 2$

# What does it mean?



# Results



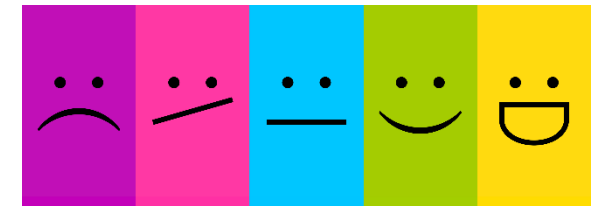
Which features highlight each group



Language Patterns Analysis



Popular users of each group



Mood Variation Analysis

# Clustering Regular Users and Advocates

	Regular Users (70,290)		Advocates (40,003)	
	$\mu$	$\sigma$	$\mu$	$\sigma$
political discourse	0.0871	0.4083	0.4614	1.5802
avg number of political hashtags related to <b>Trump</b> per tweet	-0.0005	0.0088	-0.0080	0.0297
avg number of political hashtags related to <b>Hillary</b> per tweet	-0.0066	0.0141	-0.0318	0.0385
positive/negative bias towards <b>Trump</b>	0.0759	0.0617	0.3431	0.1050
positive/negative bias towards <b>Hillary</b>	0.0833	0.4276	0.6592	2.1534

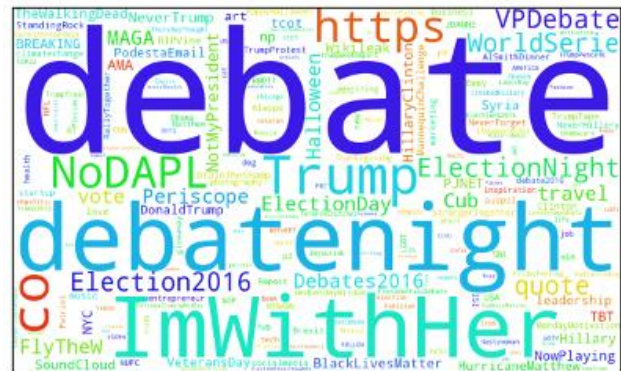
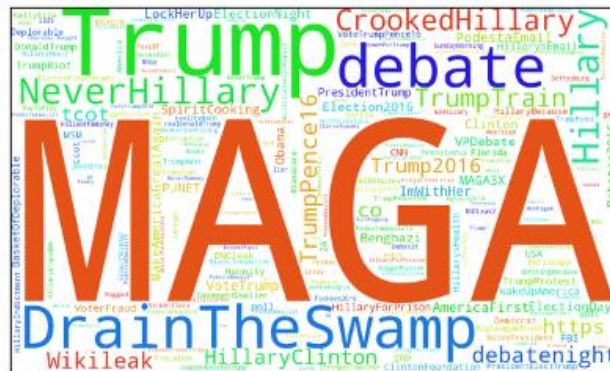
**Silhouette index: 0.81**

# Clustering Hillary's Advocates and Trump's Advocates

	Hillary's Advocates (26,230)		Trump's Advocates (13,733)	
	$\mu$	$\sigma$	$\mu$	$\sigma$
# hashtags in user's description	0.4030	0.1494	0.3516	0.1886
avg number of words per tweet	0.2787	0.1934	0.3429	0.1961
% tweets with some reference to <b>Trump</b>	0.5578	0.2349	0.7702	0.2532
% tweets with some reference to <b>Hillary</b>	0.8355	0.1864	0.6504	0.2624
std of the sentiment score of tweets with some reference to <b>Trump</b>	3.7241	4.8296	7.8192	5.2273
std of the sentiment score of tweets with some reference to <b>Hillary</b>	0.4692	1.4341	0.7009	1.7545

**Silhouette index: 0.72**

# Language patterns



Hillary's Advocates

Trump's Advocates


Political Bots

Regular Users

# Top 5 Hillary's Advocates

1

**CNN** @CNN

It's our job to #GoThere & tell the most difficult stories. Join us! For more breaking news updates follow @CNNBRK & Download our app   
<https://t.co/Xqo5kilt8c>

2

**Senator Tim Kaine** @timkaine

U.S. Senator from Virginia. Husband and father of 3. Avid reader and outdoorsman. Bluegrass and harmonica enthusiast.

3

**The New York Times** @nytimes

Where the conversation begins. Follow for breaking news, special reports, RTs of our journalists and more from <https://t.co/YapuqX0HS>.

4

**The Hill** @thehill

The Hill is the premier source for policy and political news. Follow for tweets on what's happening in Washington, breaking news and retweets of our reporters.

5

**ABC News** @ABC

See the whole picture with @ABC News. Facebook: <https://t.co/ewMNZ54axm> Instagram: <https://t.co/pPIGmNHztz>



# Top 5 Trump's Advocates

1



**Bill Mitchell** @mitchellvii

Host of YourVoice™ America at <https://t.co/B7i6W1n0cB>, Mon-Fri 7pm ET! Support the show: <https://t.co/0cVNqg8Pts> #TrustTrump

2



**Tennessee** @TEN\_GOP

Unofficial Twitter of Tennessee Republicans. Covering breaking news, national politics, foreign policy and more. #MAGA #2A

3



**Linda Suhler, Ph.D.** @LindaSuhler

I support PRESIDENT Donald Trump AMERICA FIRST Christian supports Family~Constitution~Capitalism~1A~2A~10A~NRA~Military~Police~Israel #PresidentTrump #MAGA 🇺🇸

4



**BRIAN FRASER** @bfraser747

PROUD Supporter of #PresidentTrump fighting one tweet at a time #MAGA !! Retweeted by @realDonaldTrump #AmericaFirst 🇺🇸

5

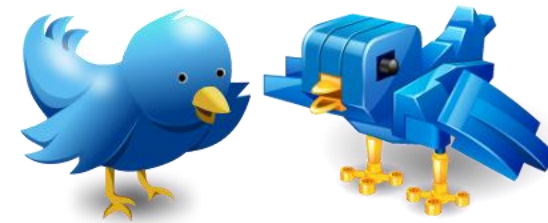


**Lou Dobbs** @LouDobbs

Lou Dobbs Tonight, Fox Business Network, 7 & 11 pm IG: <https://t.co/Mqnxd3lgtA>

# Top 5 Political Bots

Twitter suspended the **top 10** Political Bots accounts



# Top 5 Regular Users

1



**Joy Reid** @JoyAnnReid

"Ignorance, allied with power, is the most ferocious enemy justice can have." - James Baldwin #AMJoy #reiders

2



**Kurt Eichenwald** @kurteichenwald

Contributing editor, Vanity Fair; MSNBC Contributor, New York Times bestselling author.

3



**WikiLeaks** @wikileaks

We open governments // Contact: <https://t.co/676V6mG02v> // PGP: A04C 5E09 ED02 B328 03EB 6116 93ED 732E // Editor: @JulianAssange // Artwork: @WLArtForce

4



**Bernie Sanders** @SenSanders

Sen. Bernie Sanders is the longest serving independent in congressional history. Tweets ending in -B are from Bernie, and all others are from a staffer.

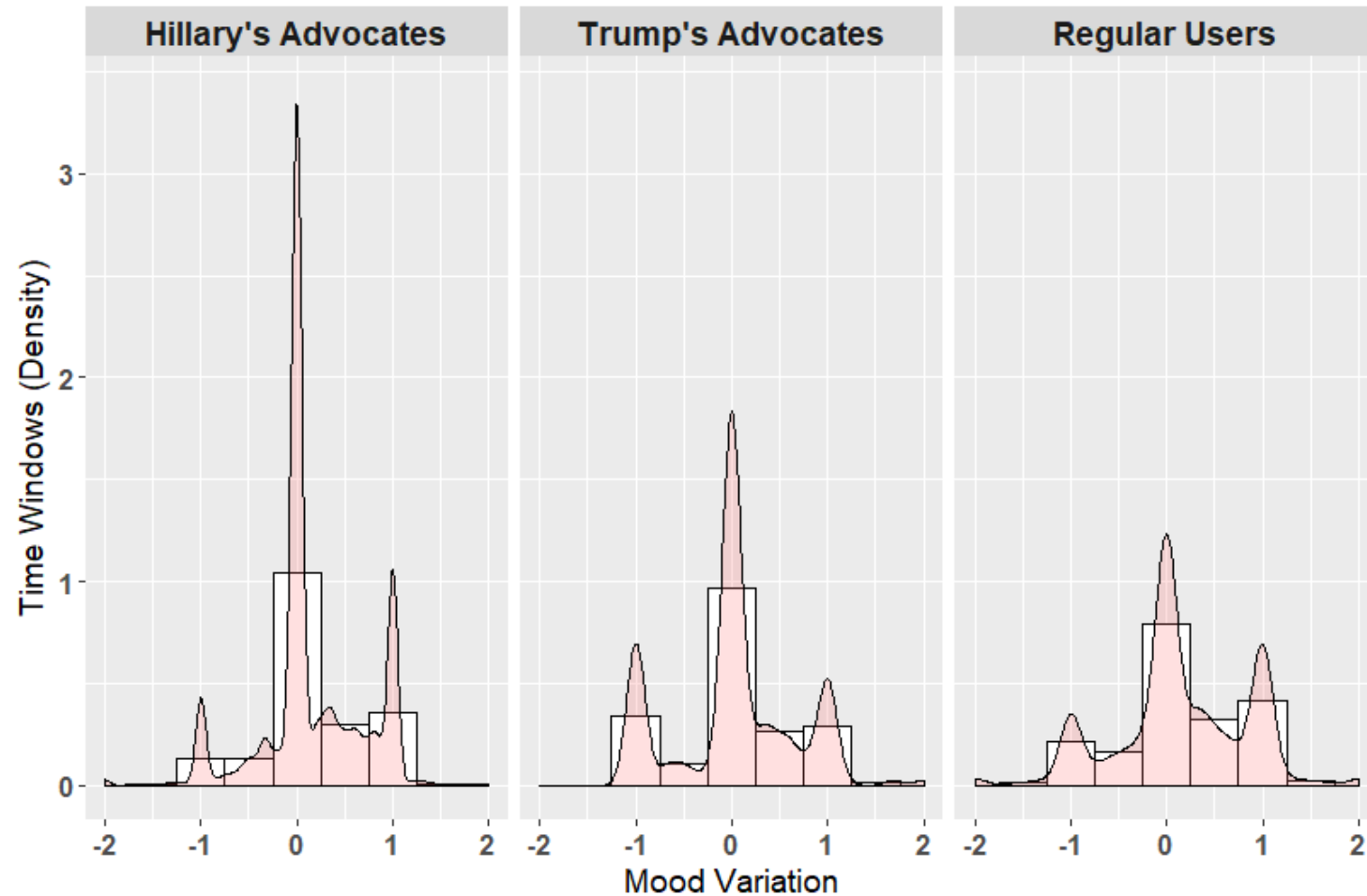
5



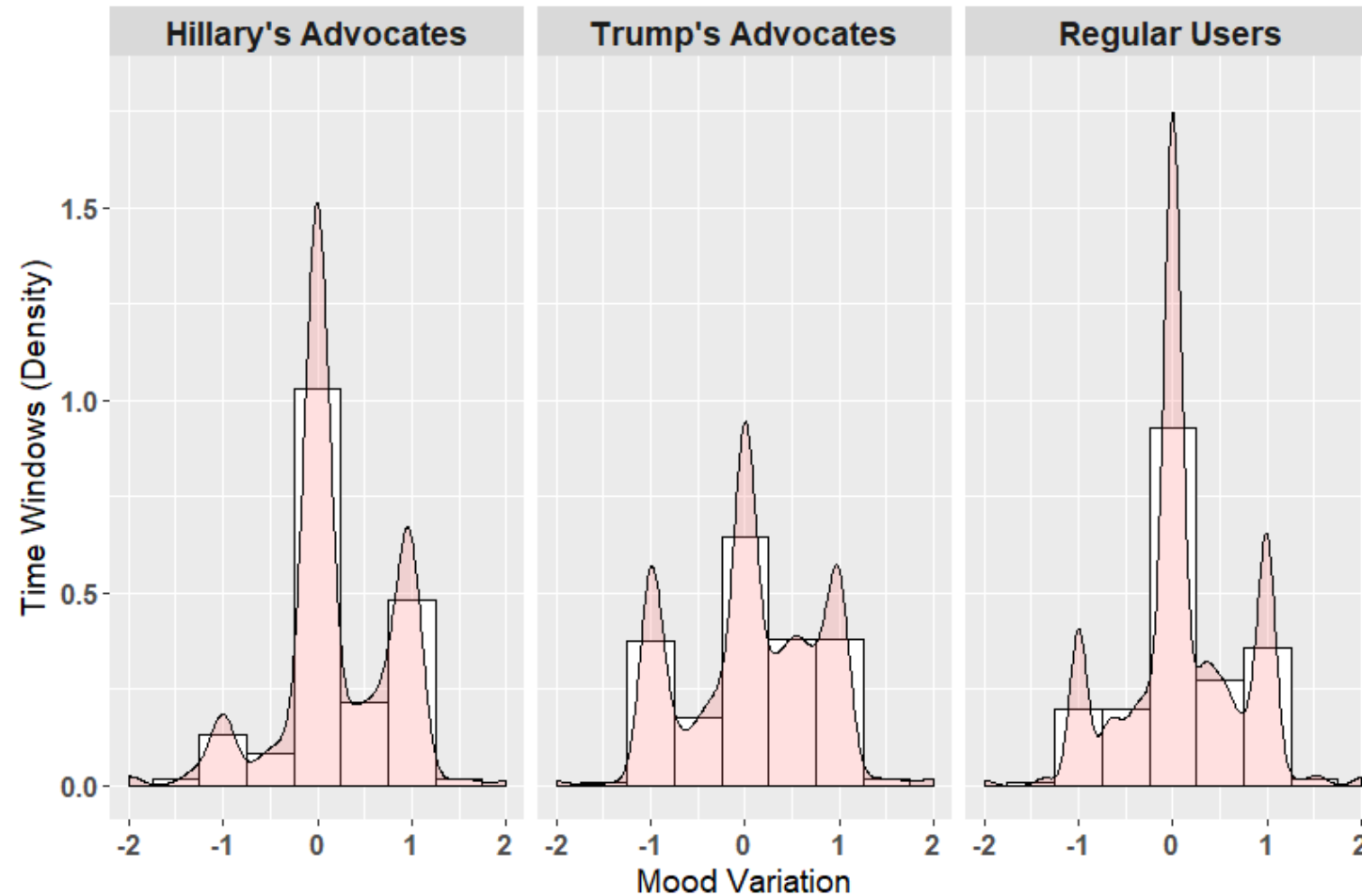
**Chris Hayes** @chrishayes

Host of All In with Chris Hayes on MSNBC, Weeknights at 8pm. Editor at Large at The Nation. Cubs fan. Instagram: chrishayes FB: <https://t.co/niNbW3BZcv>

# Mood variation – Hillary's tweets



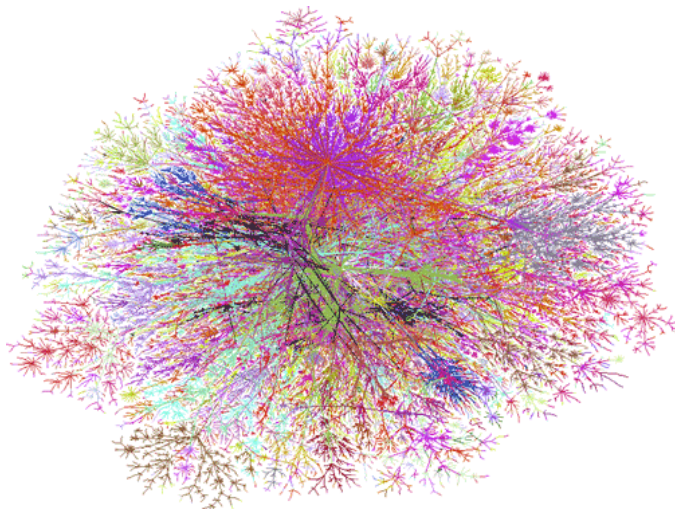
# Mood variation – Trump's tweets



# Main contributions

- Better understanding of the political engagement of users on online social networks
- How candidates may influence their voters using Twitter
- How users interact with each other

# Future work



# Questions?

[josemar.caetano@sga.pucminas.br](mailto:josemar.caetano@sga.pucminas.br)