

CHARACTERIZING POLITICALLY ENGAGED USERS, THEIR MOOD VARIATION, AND HOMOPHILY DURING THE 2016 US PRESIDENTIAL CAMPAIGN

PROPPG/PPG – Graduate Program in Informatics

Josemar Alves Caetano

Advisor: Humberto Torres Marques-Neto

Social Networks and Political Campaigns



Reach of the Candidates on Twitter

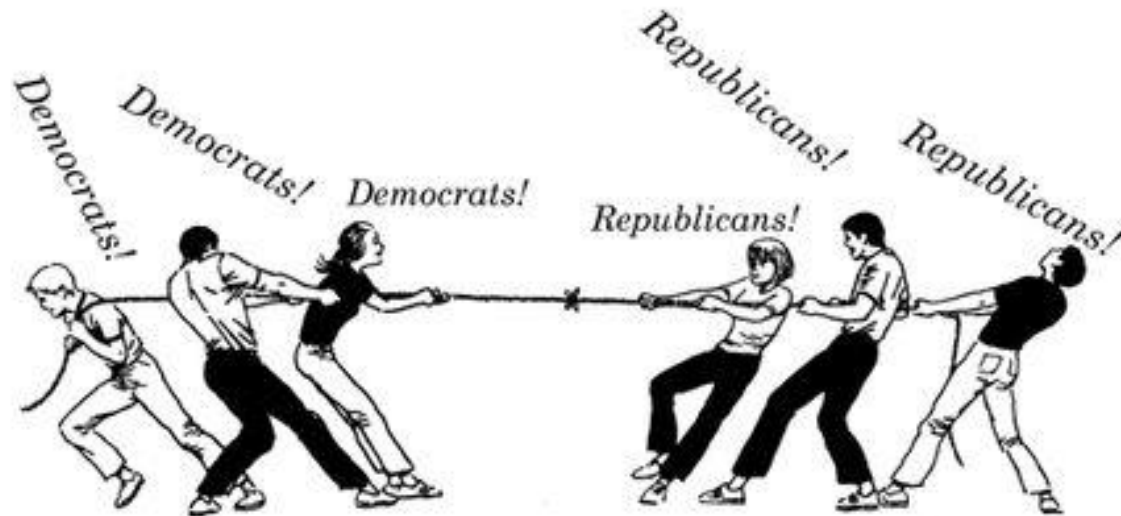


17 million followers
35 thousand published tweets



12 million followers
9 thousand published tweets

Political Biases on Social Networks



Candidates' Advocates



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



Negative

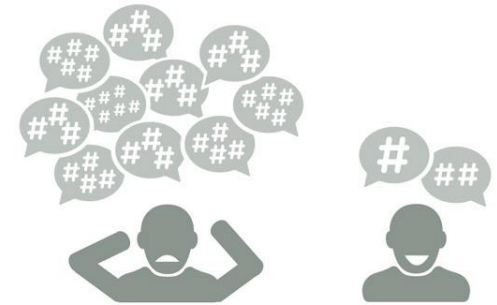


Neutral



Positive

Mean Sentiment
Analysis



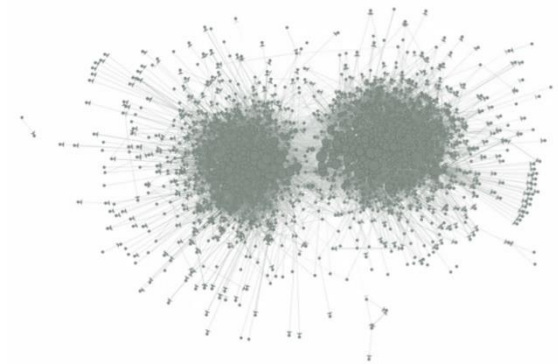
Language Patterns
Analysis



Popular users of
each group



Mood Variation
Analysis



Homophily
Analysis

Features Characterization



Which features
highlight each group



Negative

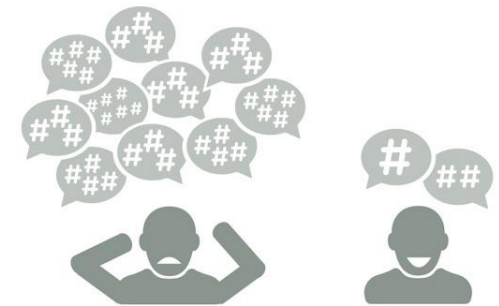


Neutral



Positive

Mean Sentiment
Analysis



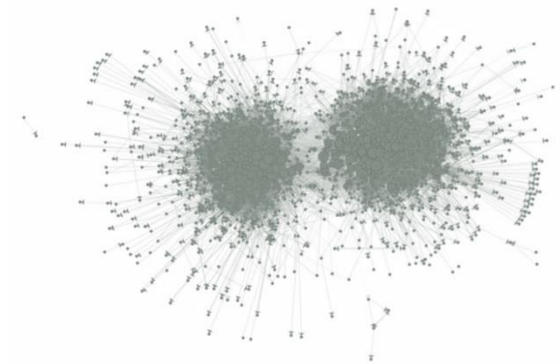
Language Patterns
Analysis



Popular users of
each group



Mood Variation
Analysis



Homophily
Analysis

Sentiment Characterization



Which features
highlight each group



Negative

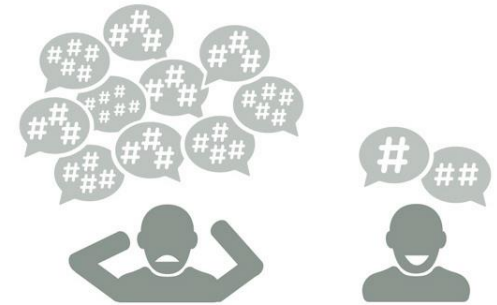


Neutral



Positive

Mean Sentiment
Analysis



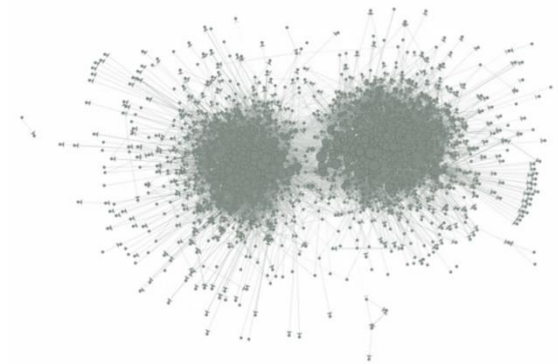
Language Patterns
Analysis



Popular users of
each group



Mood Variation
Analysis



Homophily
Analysis

Language Characterization



Which features
highlight each group



Negative



Neutral



Positive

Mean Sentiment
Analysis



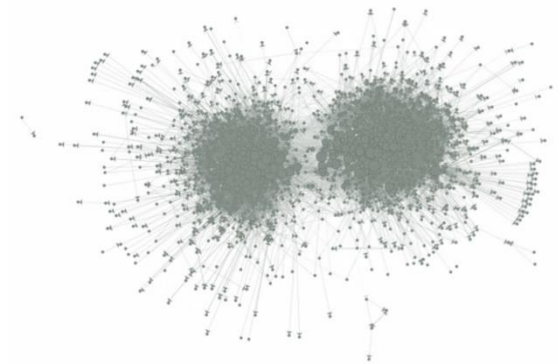
Language Patterns
Analysis



Popular users of
each group



Mood Variation
Analysis



Homophily
Analysis

Profile Characterization



Which features
highlight each group



Negative

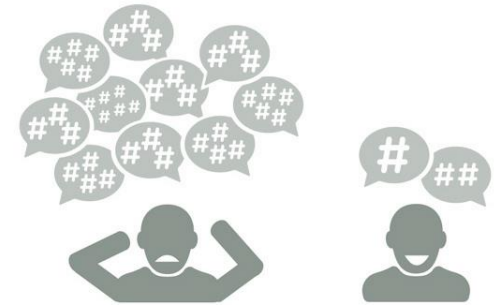


Neutral



Positive

Mean Sentiment
Analysis



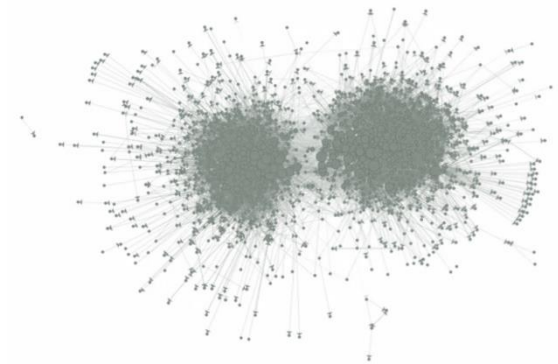
Language Patterns
Analysis



Popular users of
each group



Mood Variation
Analysis



Homophily
Analysis

Mood Characterization



Which features
highlight each group



Negative

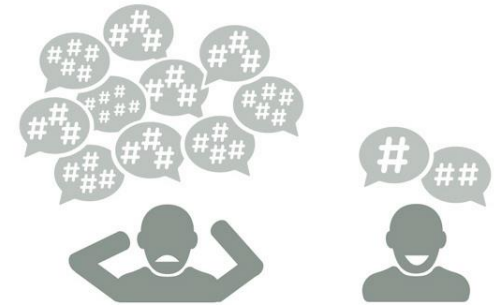


Neutral



Positive

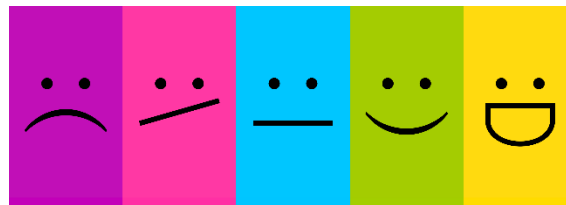
Mean Sentiment
Analysis



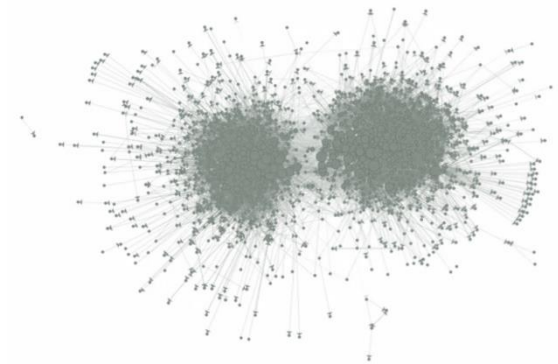
Language Patterns
Analysis



Popular users of
each group



Mood Variation
Analysis



Homophily
Analysis

Homophily Characterization



Which features
highlight each group



Negative

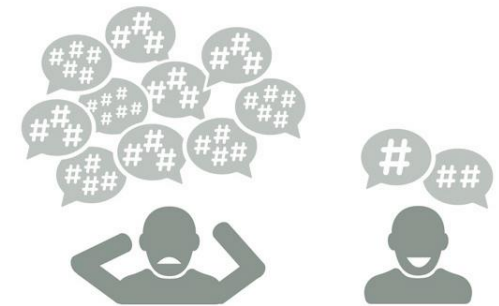


Neutral



Positive

Mean Sentiment
Analysis



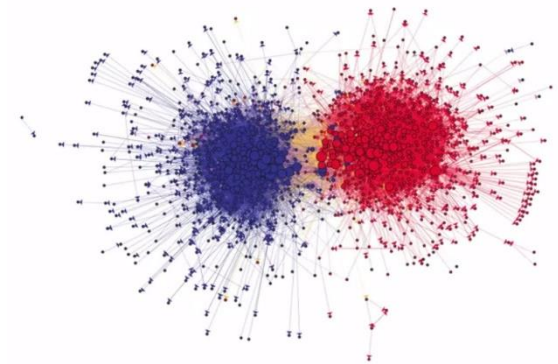
Language Patterns
Analysis



Popular users of
each group



Mood Variation
Analysis



Homophily
Analysis

To Perform these 6 Characterizations...



Which features
highlight each group



Negative



Neutral



Positive

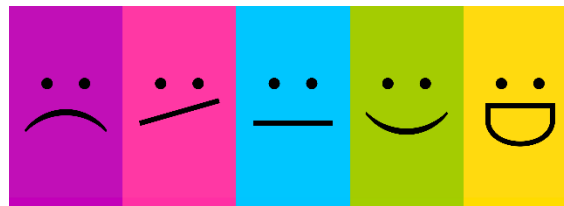
Mean Sentiment
Analysis



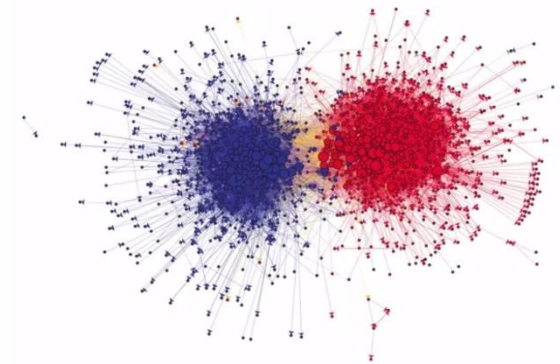
Language Patterns
Analysis



Popular users of
each group



Mood Variation
Analysis

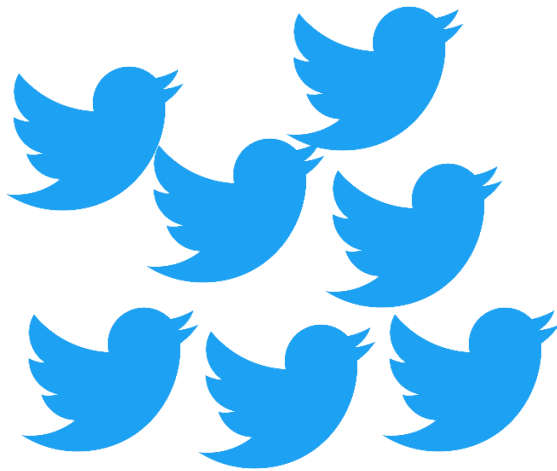


Homophily
Analysis

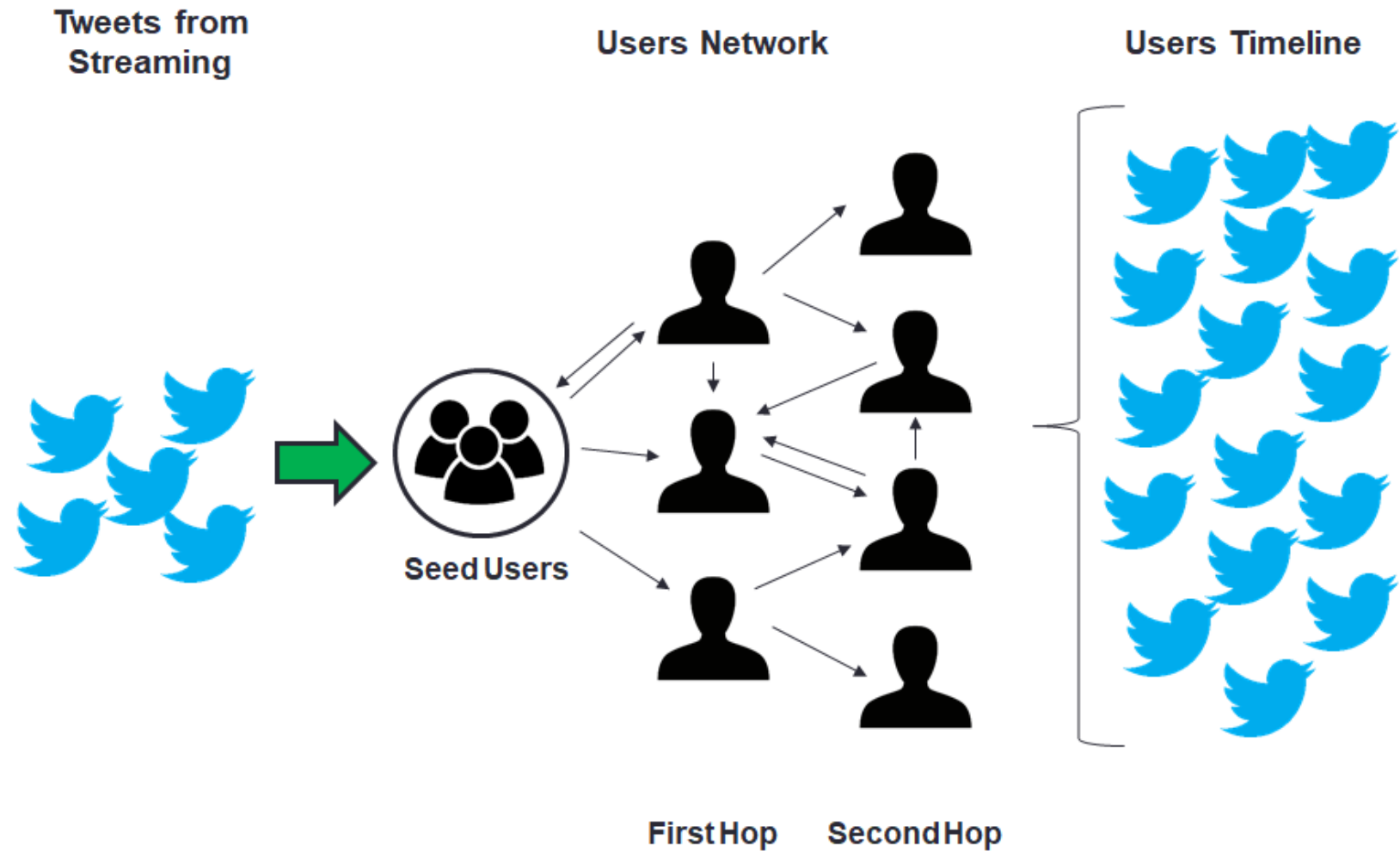
Methodology



Collecting Twitter Data

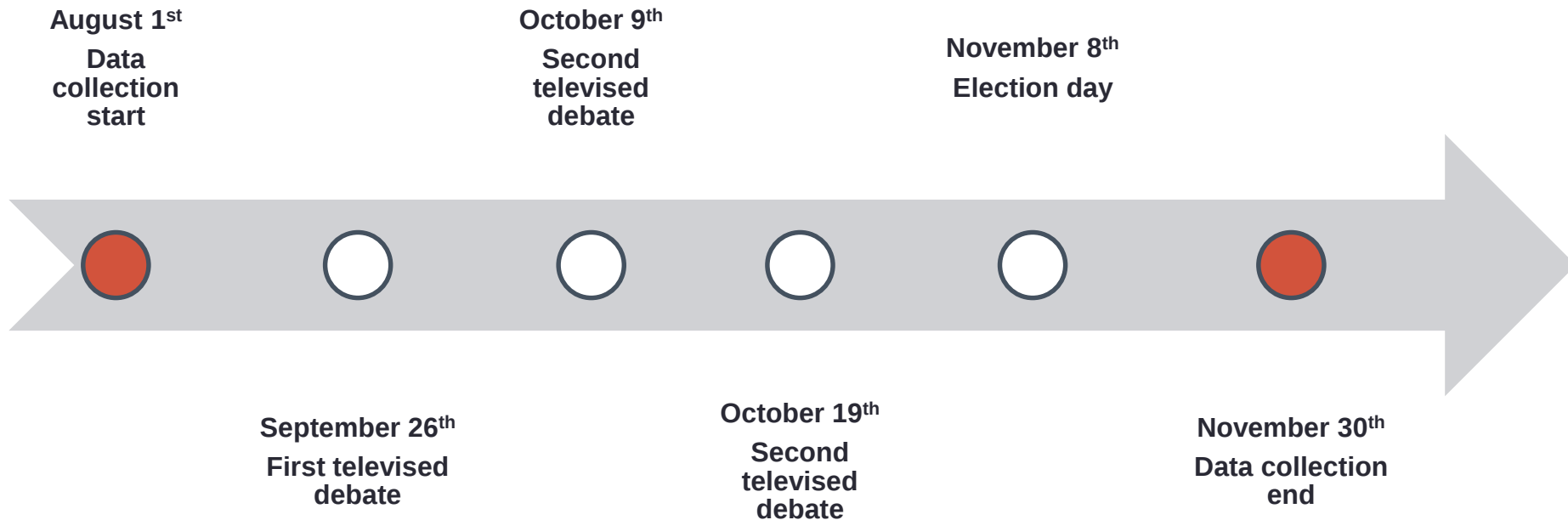


Data Collection Process



Data Collection Period

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



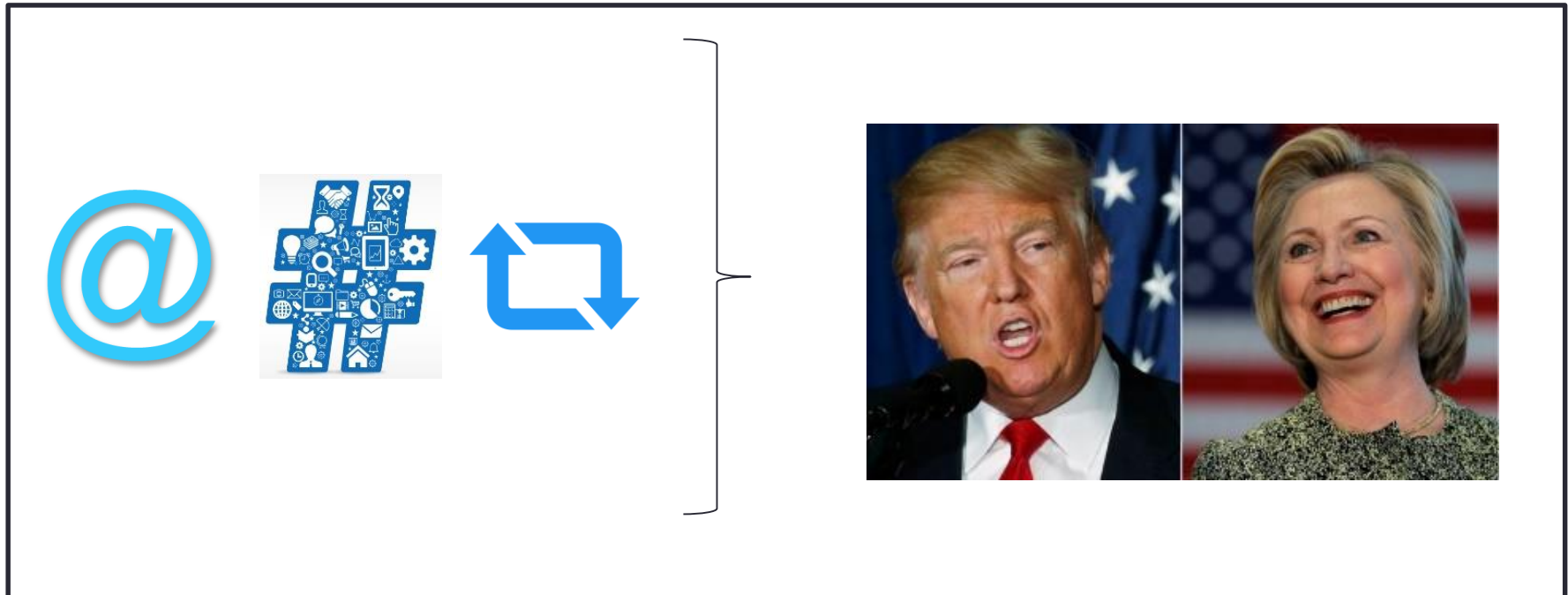
Dataset

# of tweets	23 mi
# of users	115 k
# of relationships	1,8 mi

Dataset Considerations

- Twitter API provides at most **the last 200** user's tweets
- Tweets published between January 1st and November 30th

Identifying Political Tweets



Political References Used in Tweets

@HillaryClinton @realDonaldTrump Hillary is a nice candidate.

Candidate Reference

*I endorse Trump and believe he wants to #MakeAmericaGreatAgain
#TrumpsArmy*

Political Hashtags

RT @HillaryClinton: It's time to fight back against Donald Trump...

Candidate Retweet

Candidates References Considered

Donald Trump	Hillary Clinton
@realDonaldTrump	@HillaryClinton
Trump	Hillary
DT	HC

Political Hashtags

	Donald Trump	Hillary Clinton
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

Political Tweets Conditions

-  The tweet is a retweet of a post by one of the two candidates
-  The tweet has a reference to at least one candidate
-  The tweet contains at least one political hashtag

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

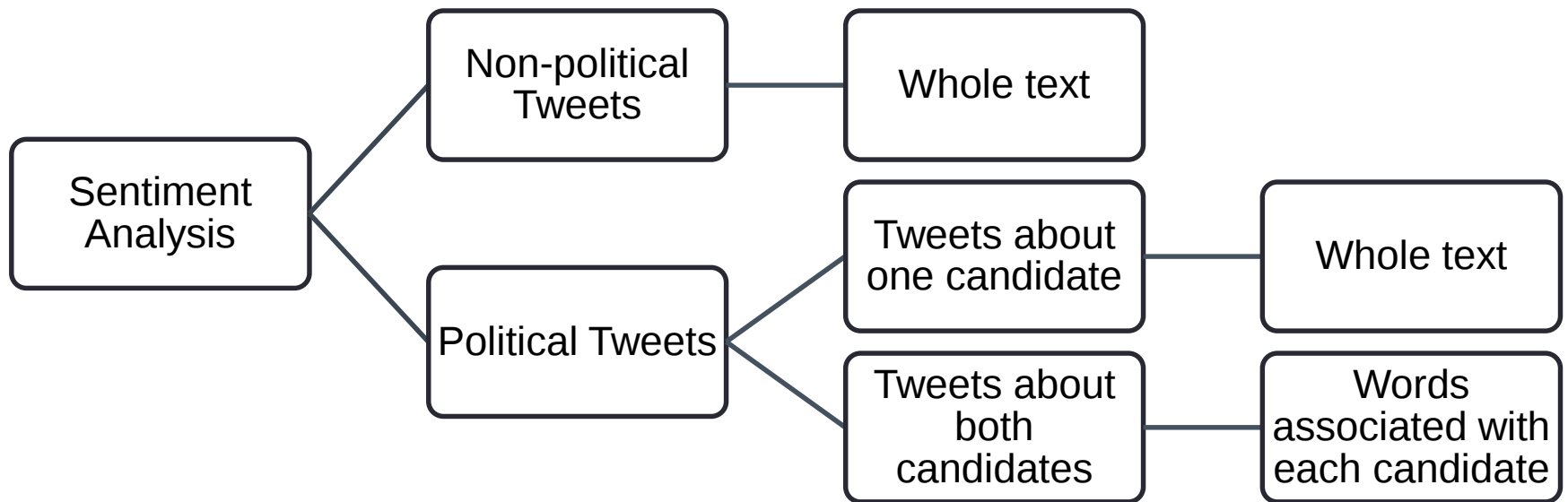


I **love** Hillary Clinton
but I **hate** Donald
Trump.

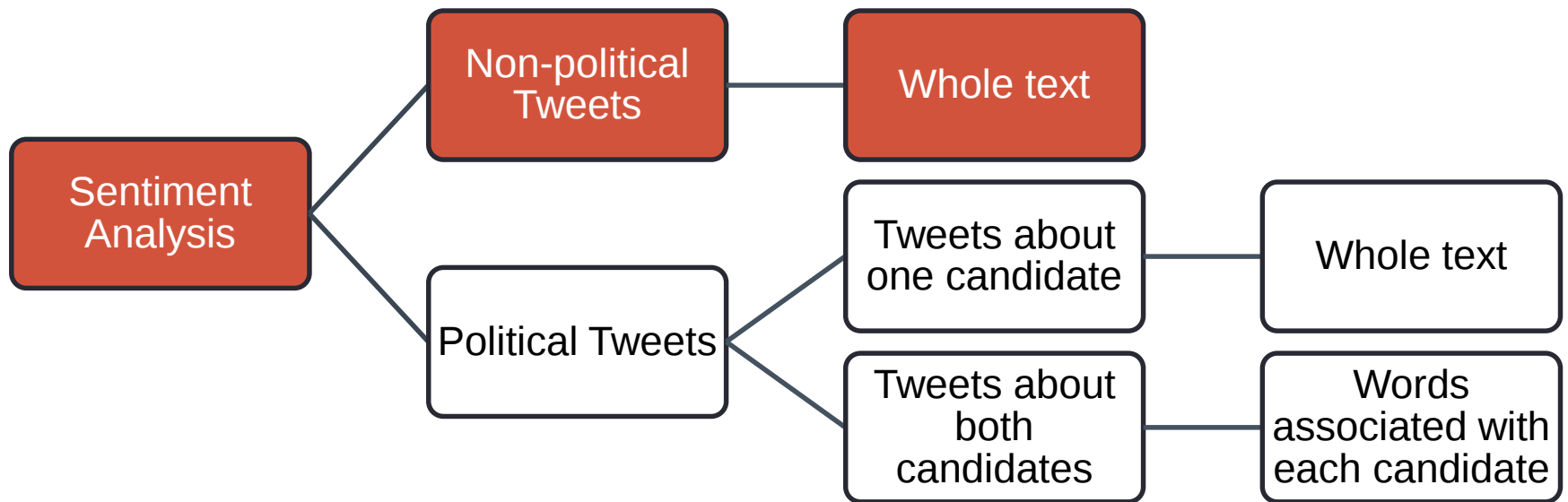
I **hate** Hillary Clinton
but I **love** Donald
Trump.



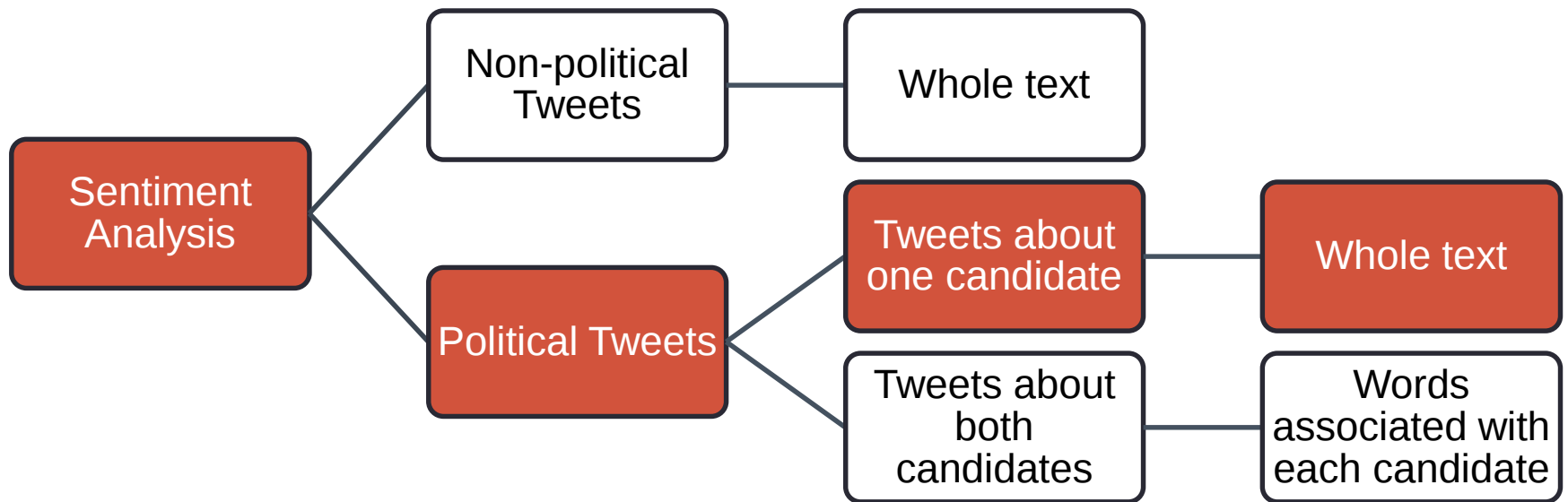
Sentiment Analysis Approaches



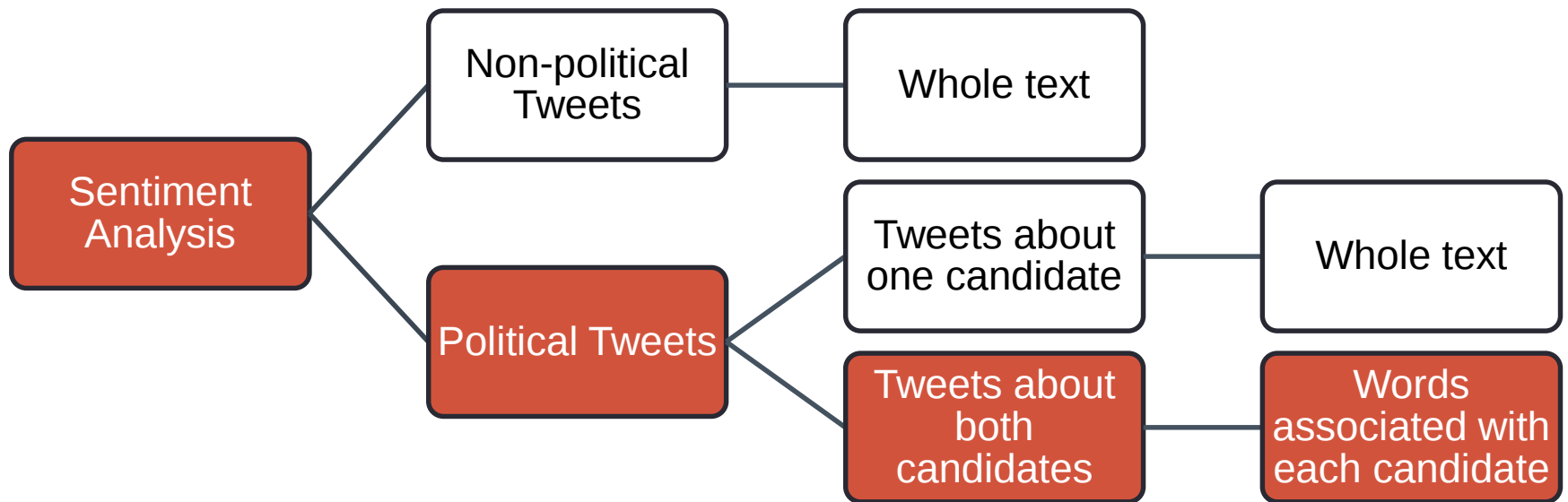
Non-political Tweets



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**

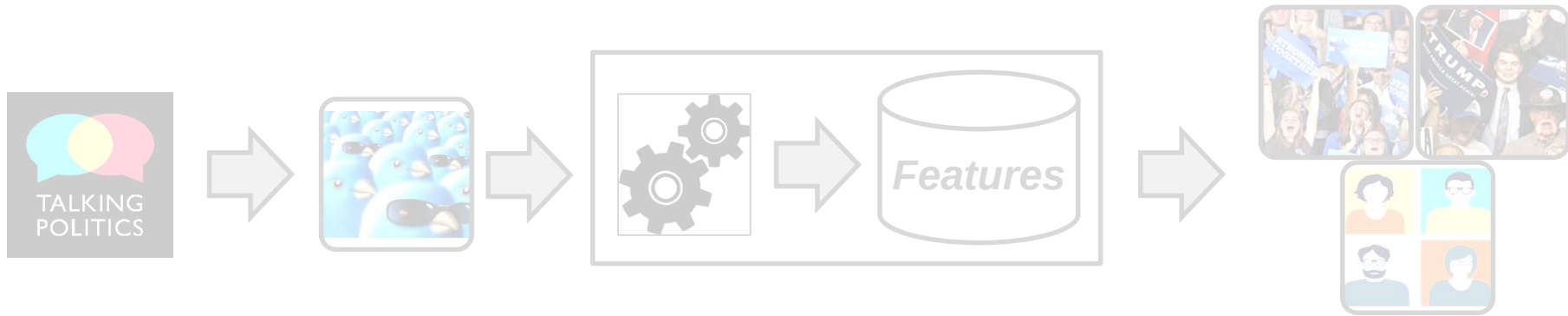


**Political
Bots**

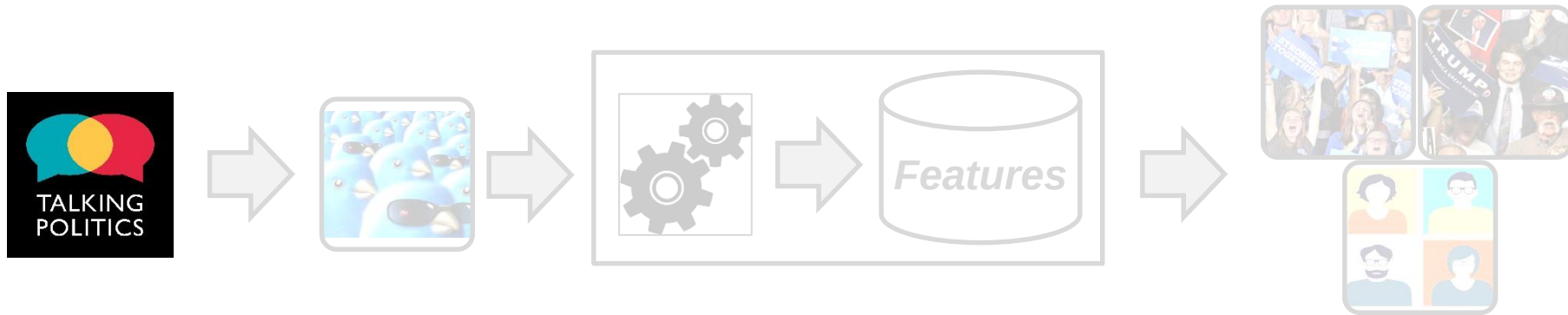


**Regular
Users**

Data Mining Process

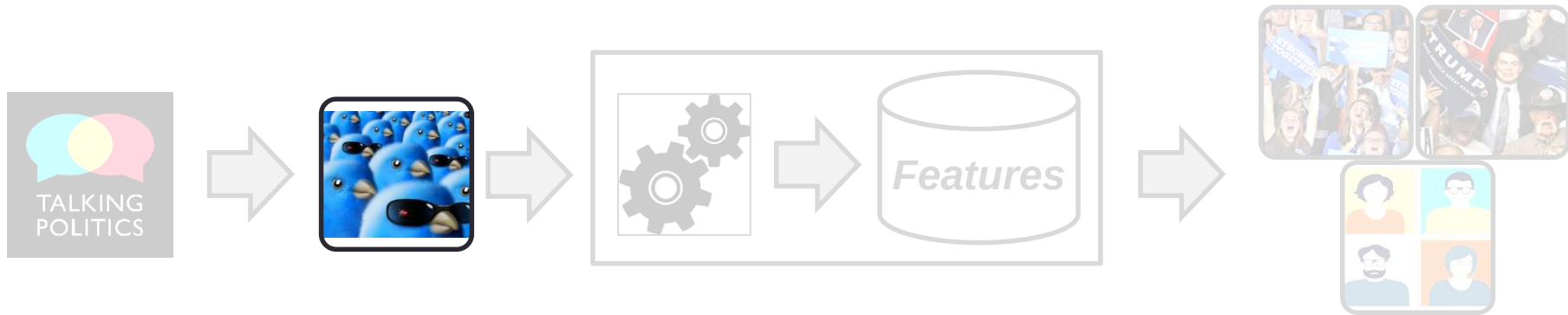


Removing Outliers



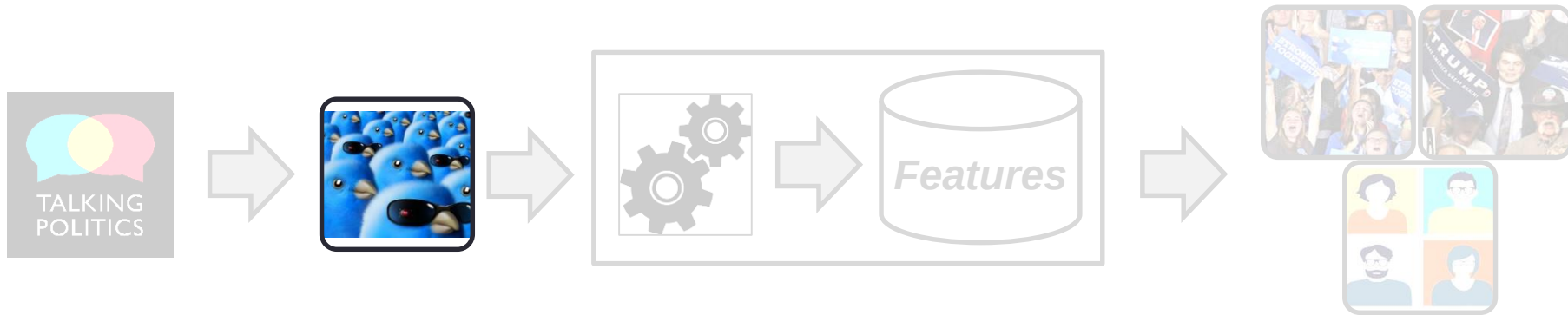
- Eliminated users with **Political Engagement** = 0
- Total eliminated: **1,318** users of the initial 115,664
- Total users to group: **114,346**

Identifying Political Bots



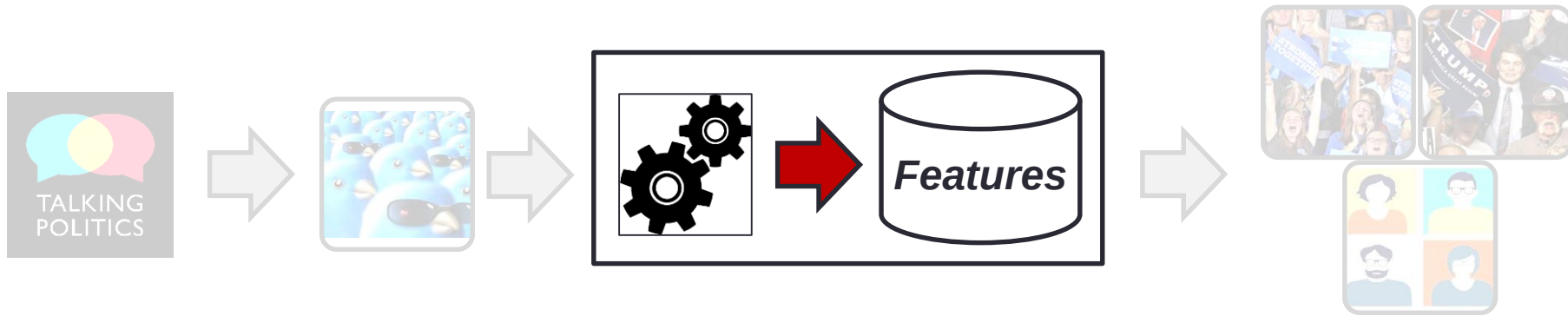
BotOrNot

Identifying Political Bots

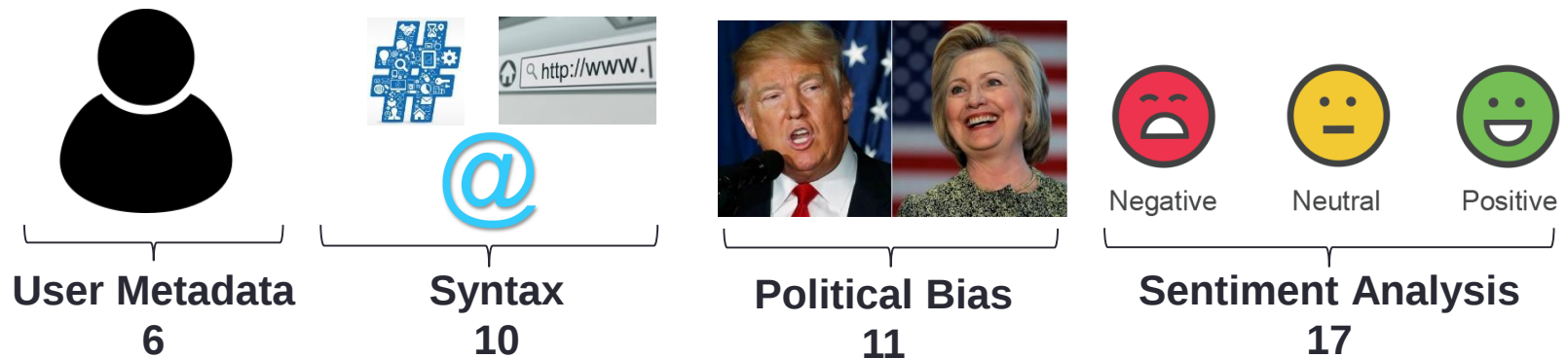


- Users with **BotOrNot score ≥ 0.75**
- Total selected: **4,053** users from 114,346
- Total remaining users: **110,293**

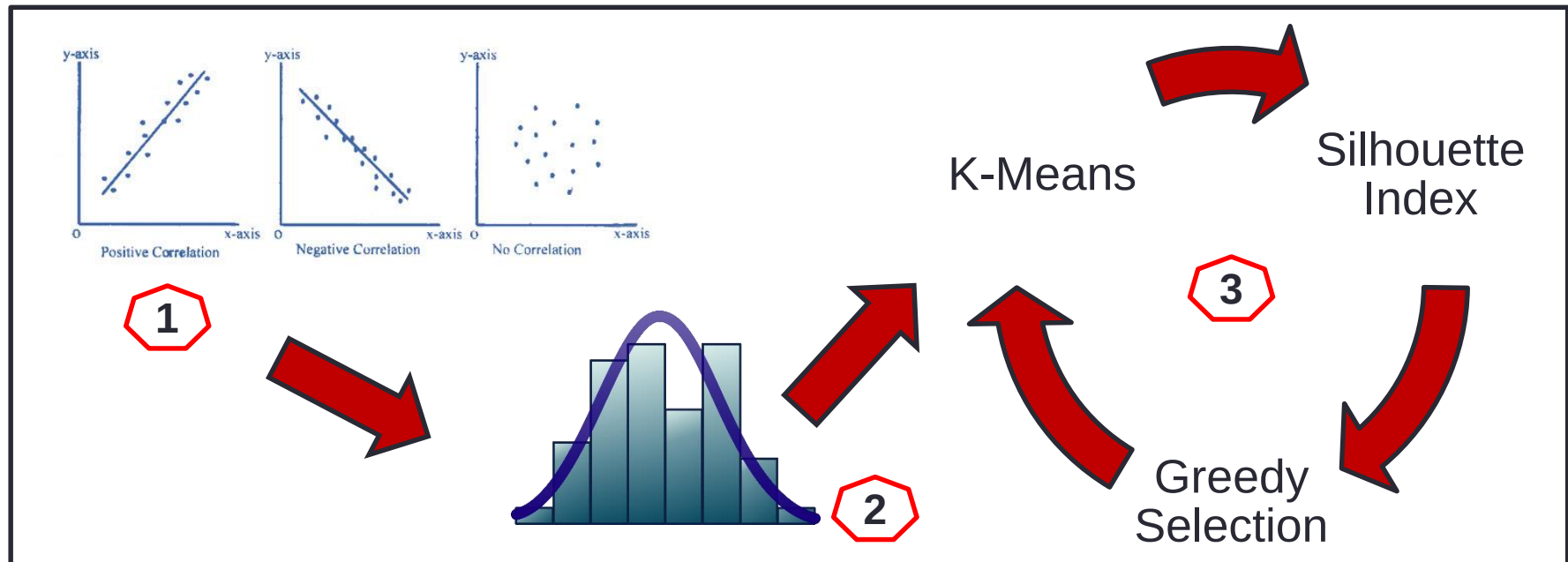
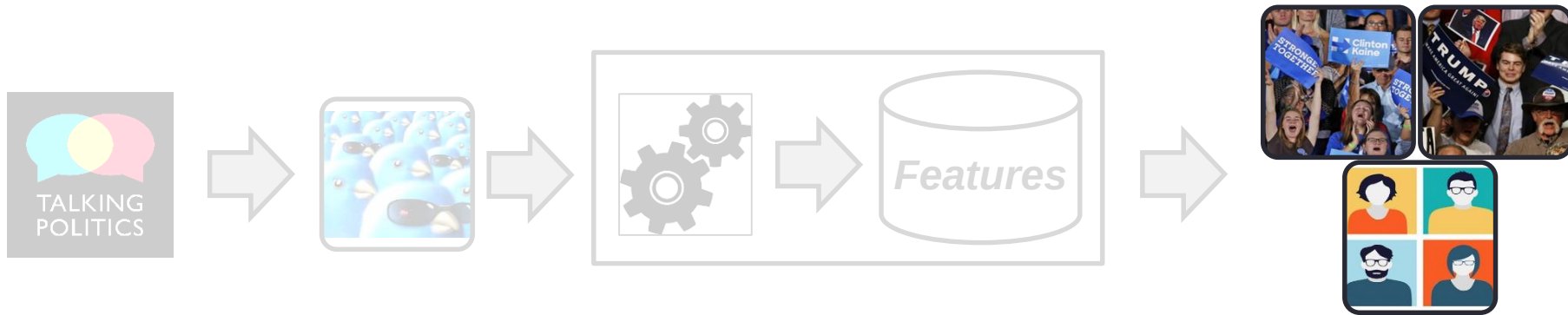
Feature Set Engineering



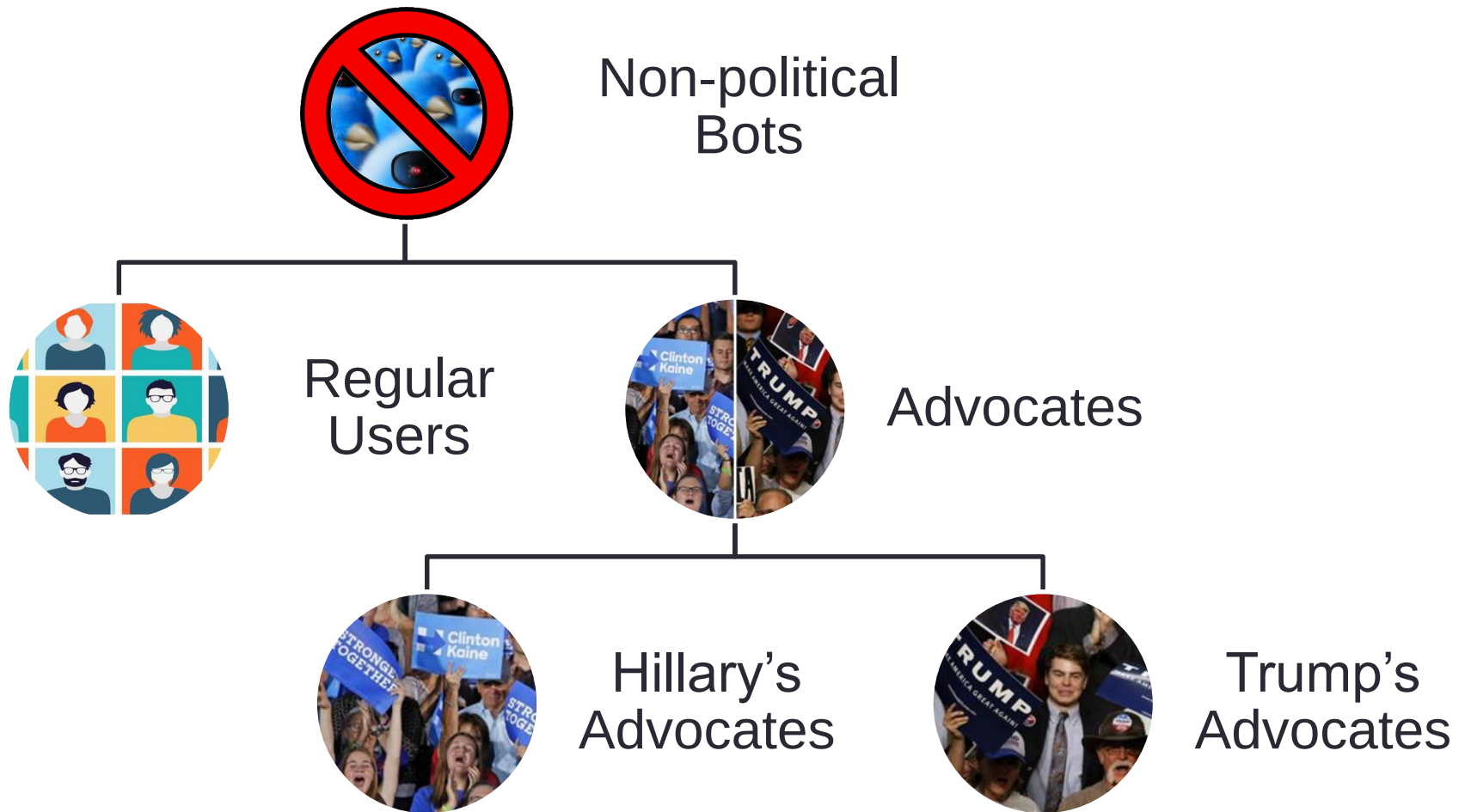
44 Features



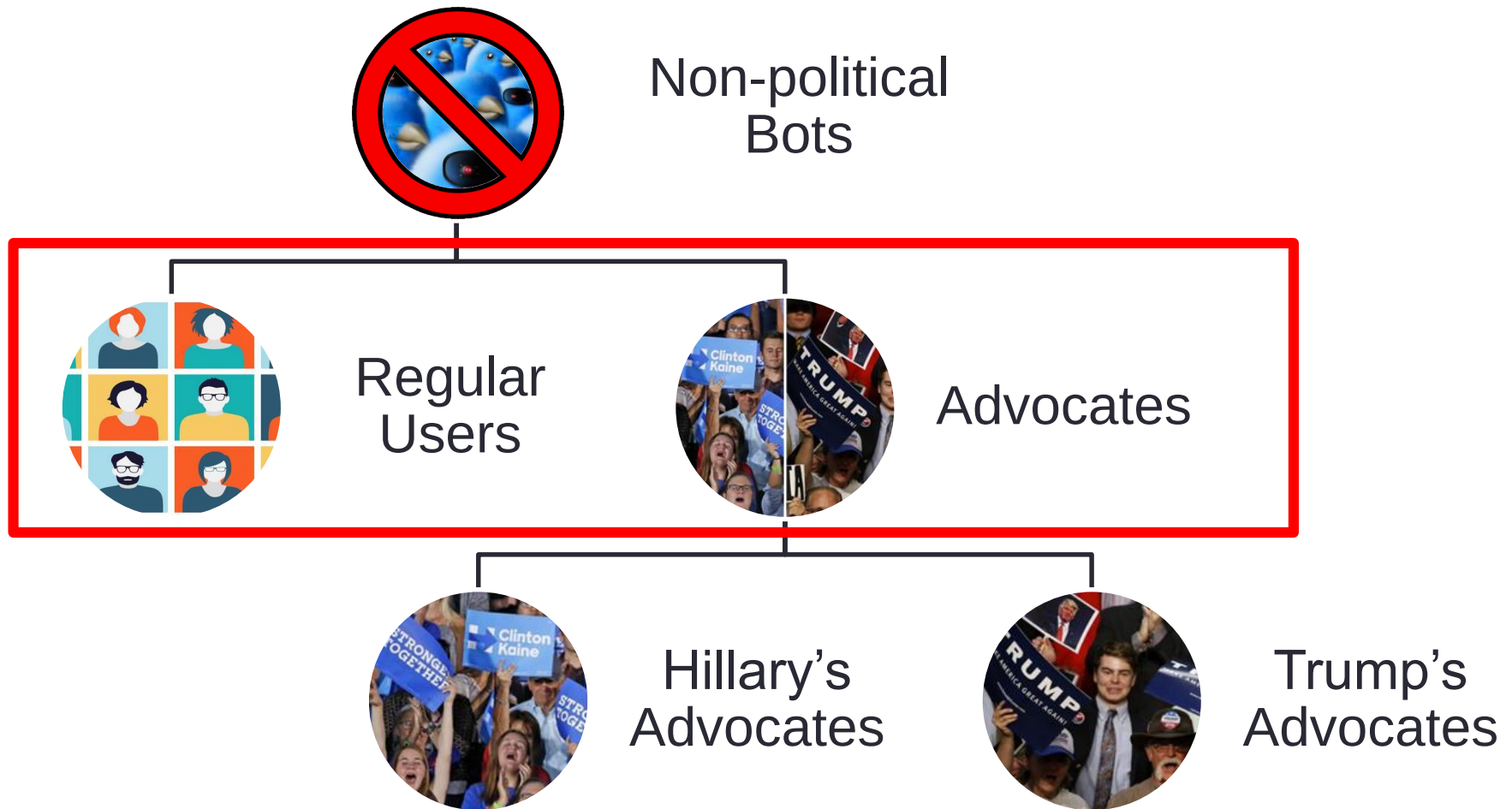
Identifying Regular Users, Hillary's Advocates, and Trump's Advocates



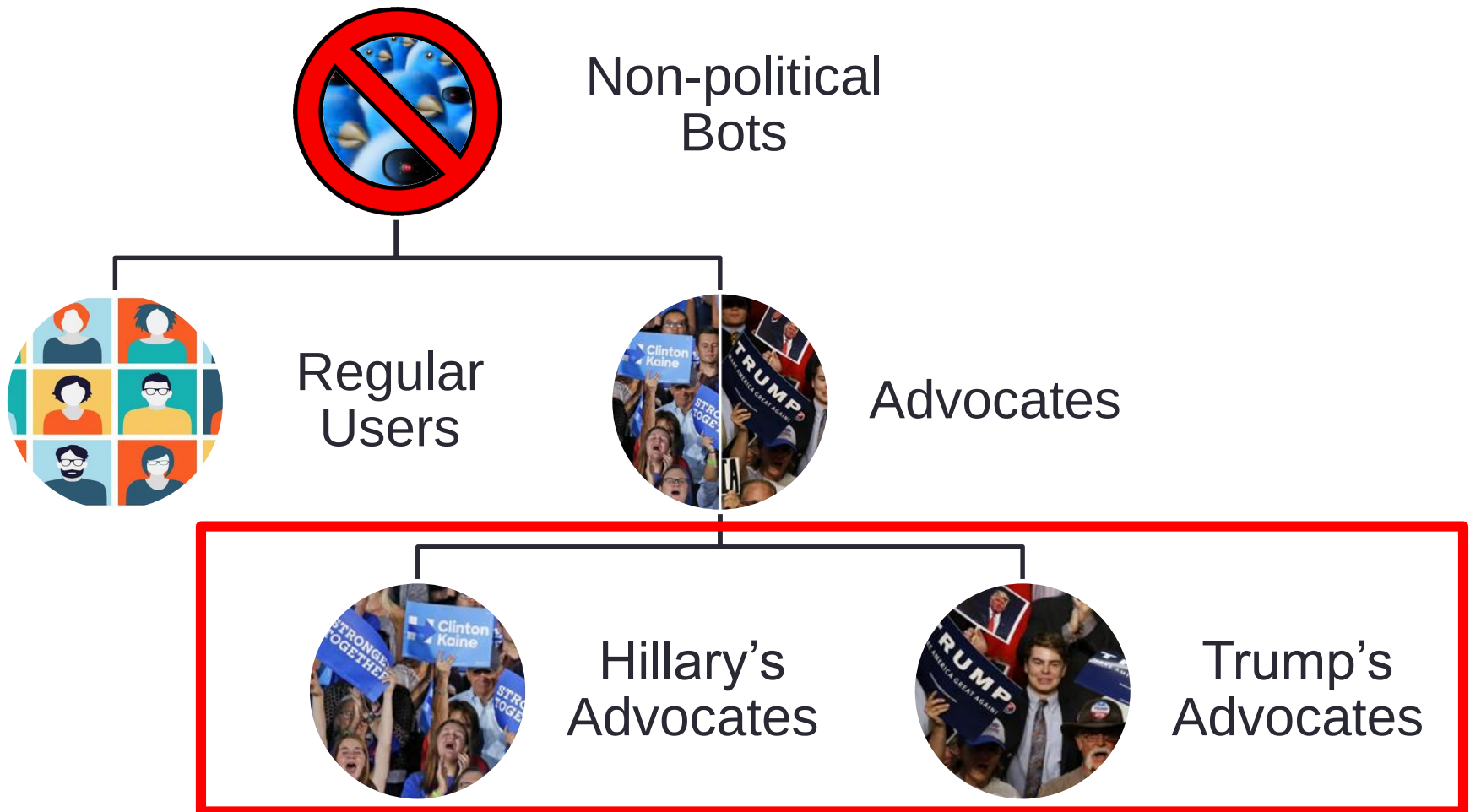
Hierarchical Clustering



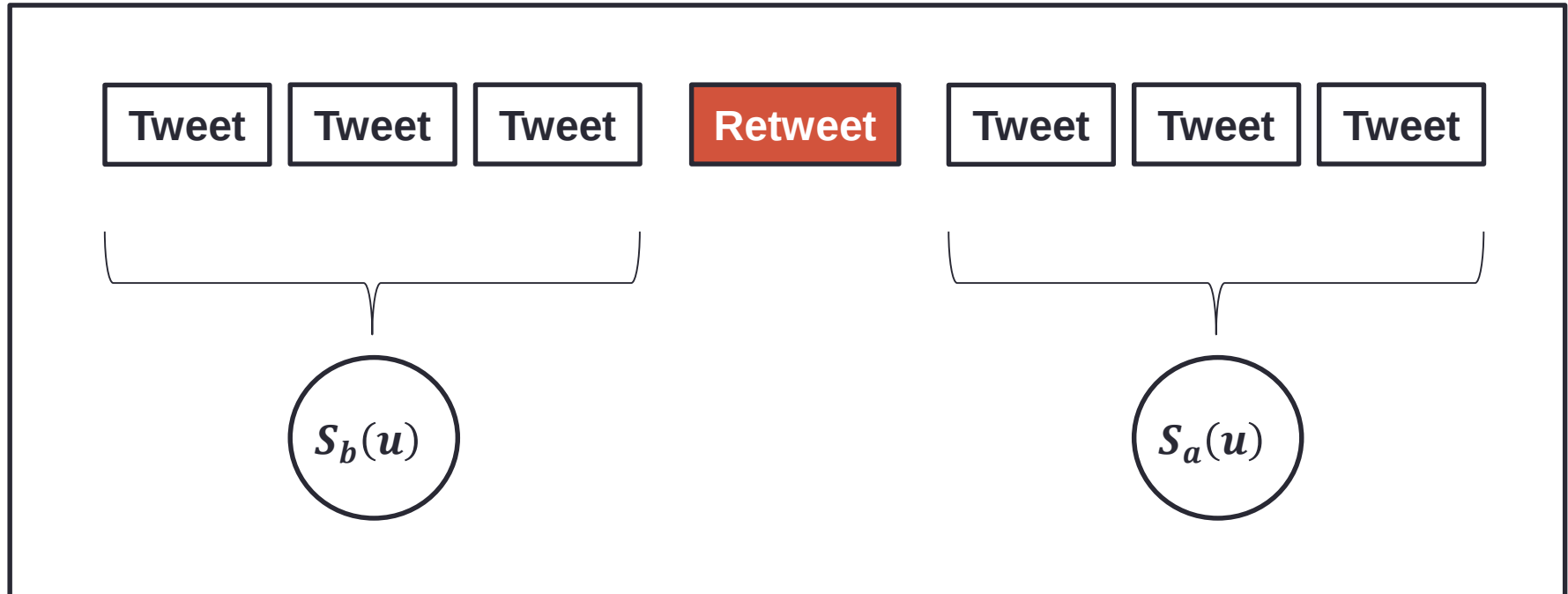
Identifying Regular Users and Advocates



Identifying Trump's Advocates and Hillary's Advocates



Mood Variation Analysis



Subjective Well-Being Definition

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

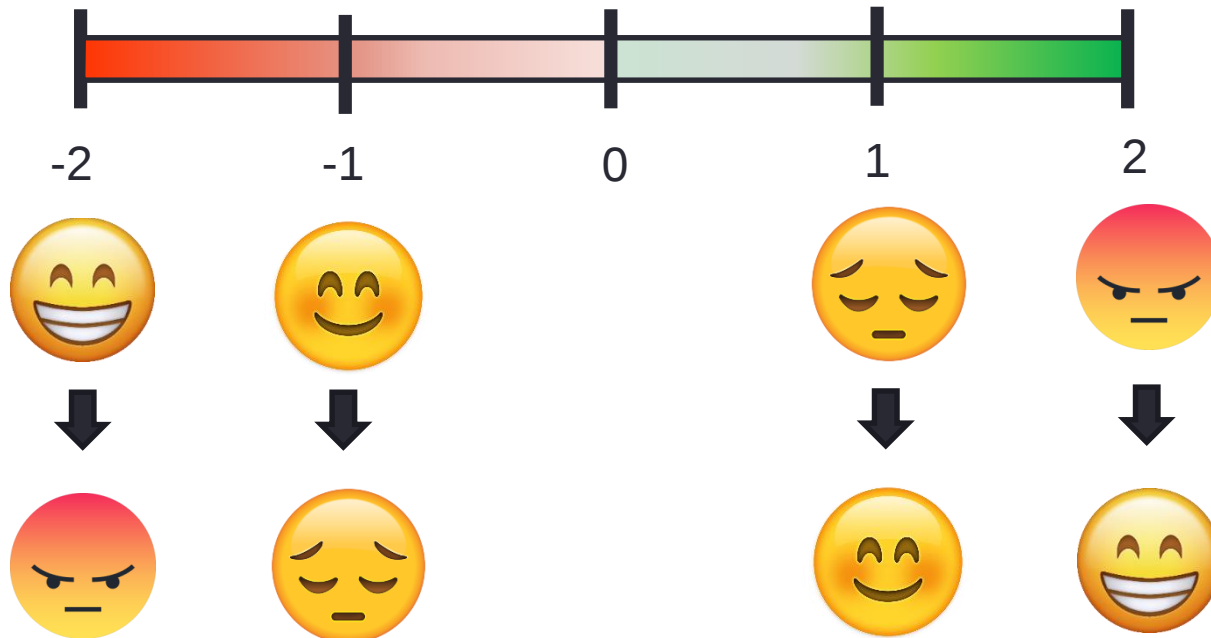
- $N_{p_u}(t_1, t_2)$: positive tweets total
- $N_{n_u}(t_1, t_2)$: negative tweets total
- $S_u(t_1, t_2)$: $-1 \leq S_u \leq 1$

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
- Δ_{S_u} values: $-2 \leq \Delta_{S_u} \leq 2$

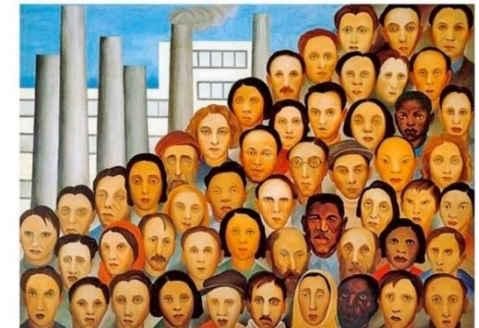
What Does It Means?



But What If...



Homophily Analysis



Homophily Definition

$$H_i = \frac{s_i}{s_i + d_i}$$

- s_i : represents homogeneous connections
- d_i : represents heterogeneous connections

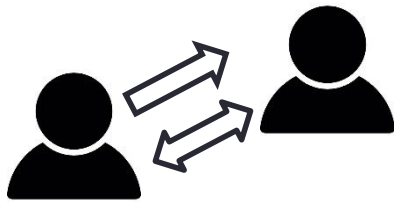
Inbreeding Homophily

$$IH_i = \frac{H_i - w_i}{1 - w_i}$$

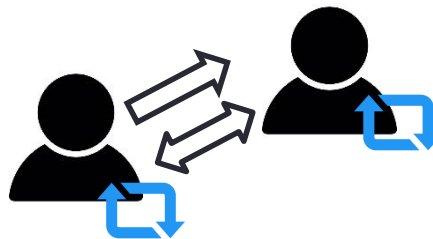
- w_i : represents the probability of occurrence of individuals of group i in the population

$$\begin{cases} IH_i > 0 \rightarrow \text{homophily} \\ IH_i < 0 \rightarrow \text{heterophily} \end{cases}$$

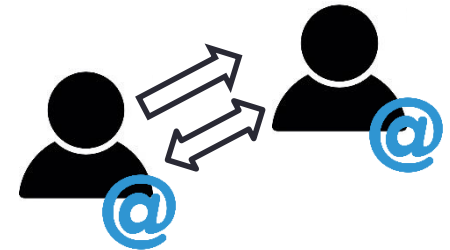
Connections Types



**Follow
Connections**



**Retweet
Connections**



**Mention
Connections**

Results



Which features
highlight each group



Negative



Neutral



Positive

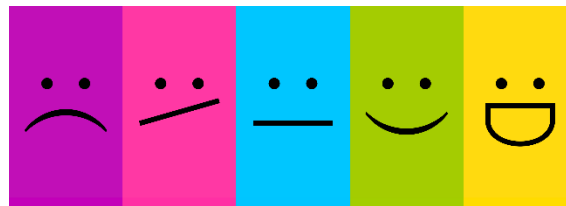
Mean Sentiment
Analysis



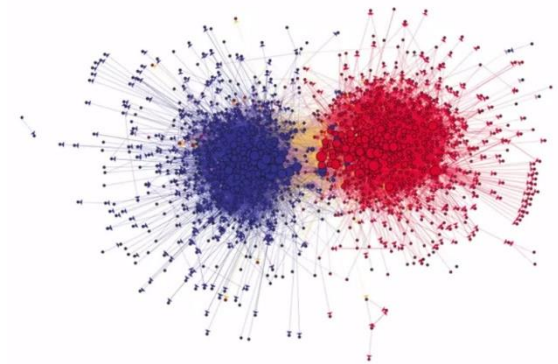
Language Patterns
Analysis



Popular users of
each group



Mood Variation
Analysis



Homophily
Analysis

Clustering Regular Users and Advocates

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

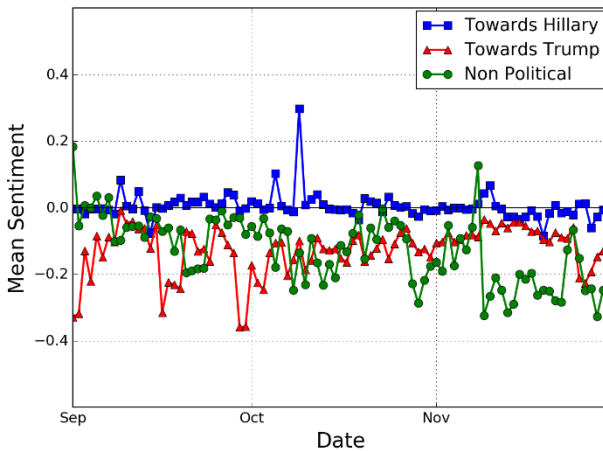
Silhouette Index: 0.81

Clustering Hillary's Advocates and Trump's Advocates

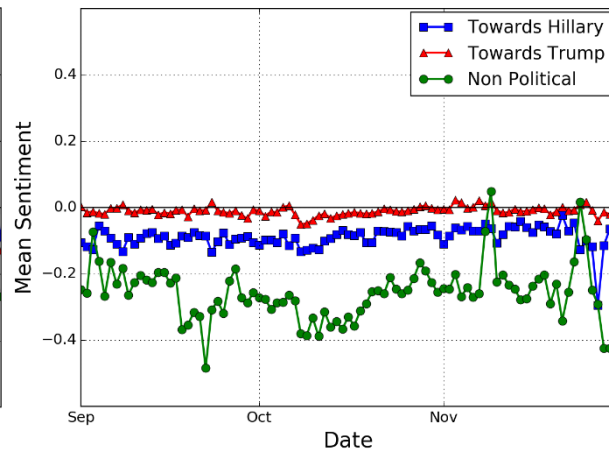
	Hillary's Advocates (26,230)		Trump's Advocates (13,733)	
	μ	σ	μ	σ
# hashtags in user's description	0.4030	0.1494	0.3516	0.1886
avg number of words per tweet	0.2787	0.1934	0.3429	0.1961
% tweets with some reference to Trump	0.5578	0.2349	0.7702	0.2532
% tweets with some reference to Hillary	0.8355	0.1864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

Silhouette Index: 0.72

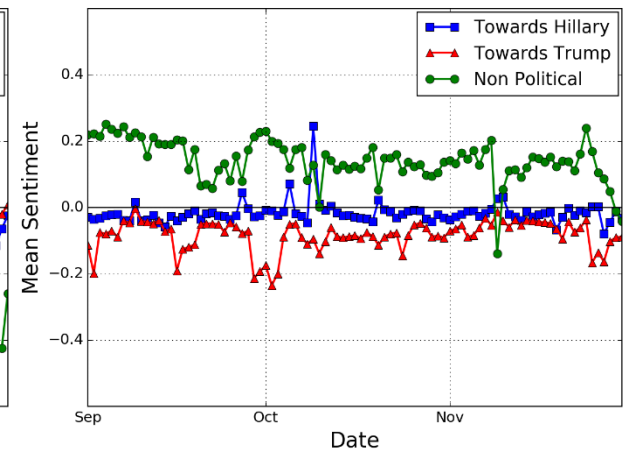
Daily Mean Sentiment



Hillary's Advocates

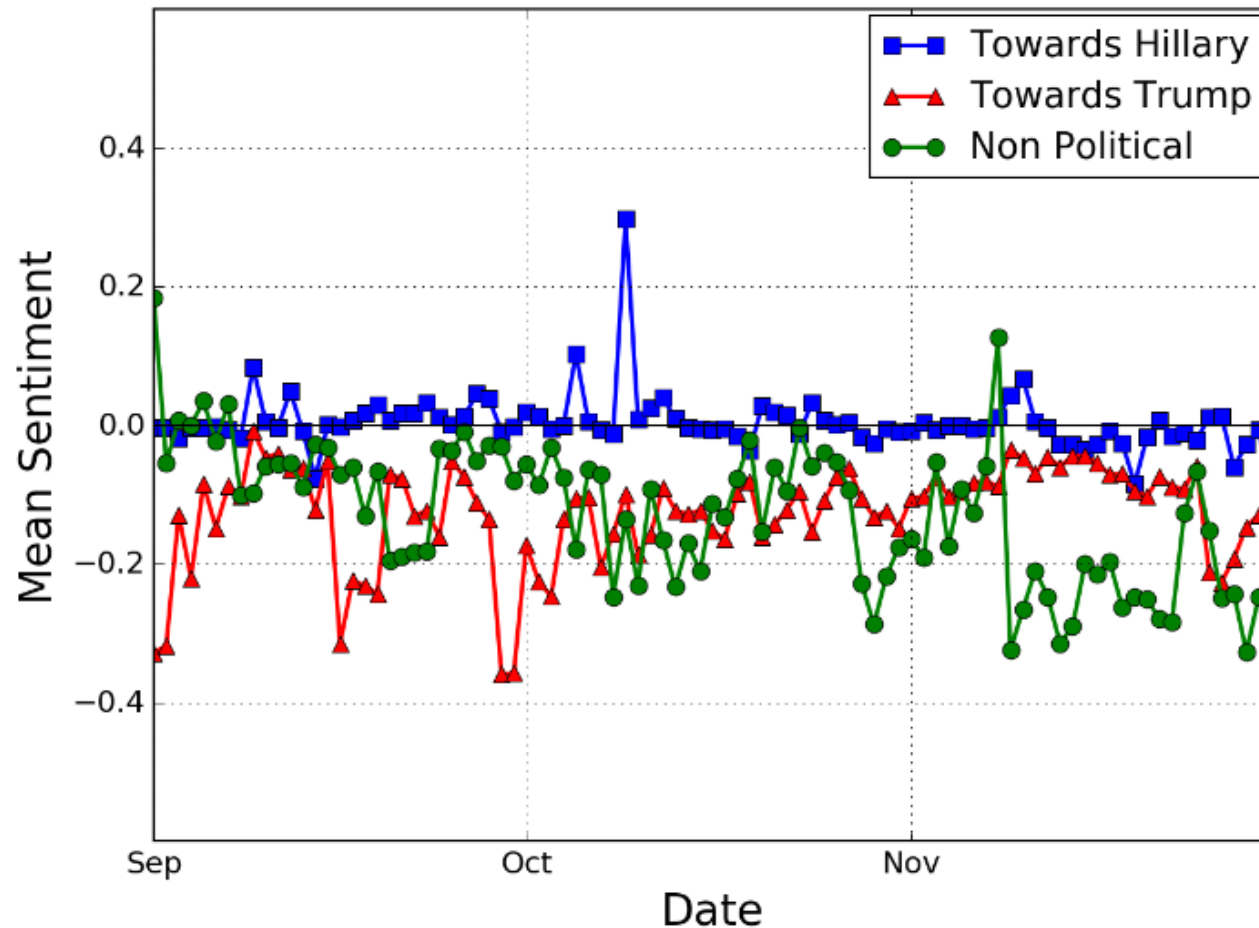


Trump's Advocates

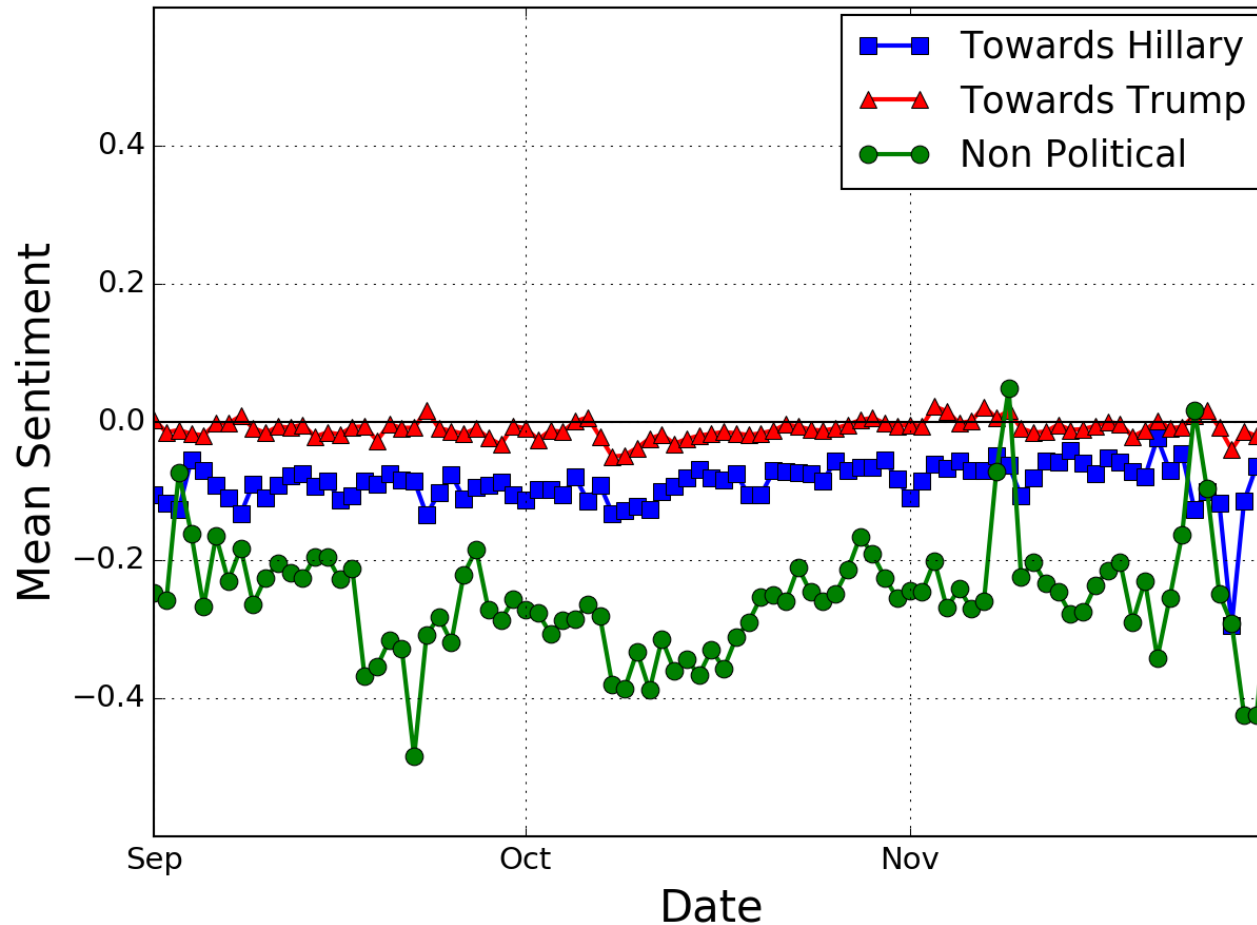


Regular Users

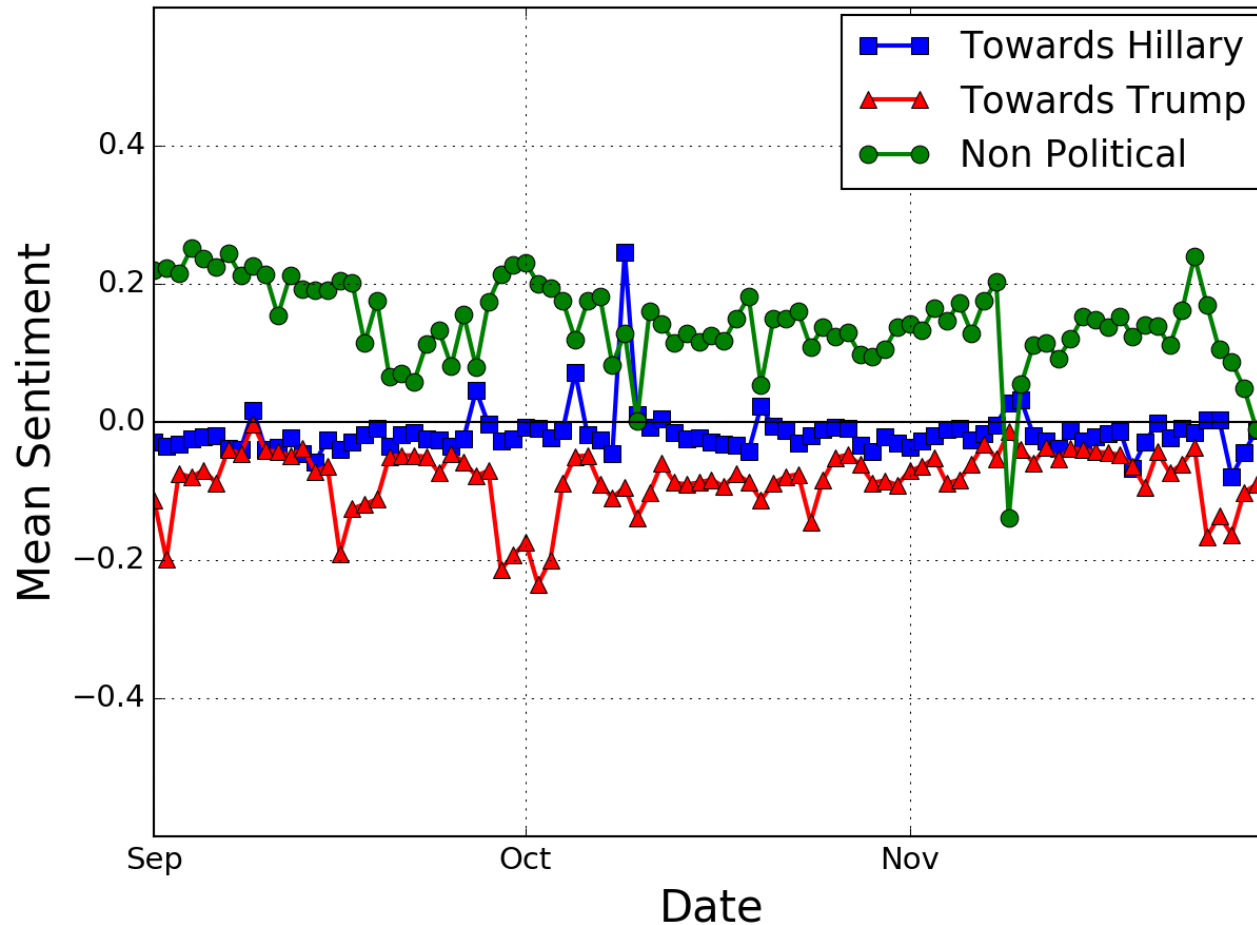
Daily Mean Sentiment – Hillary's Advocates



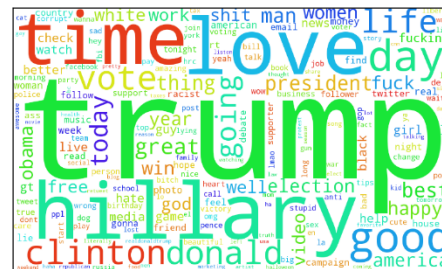
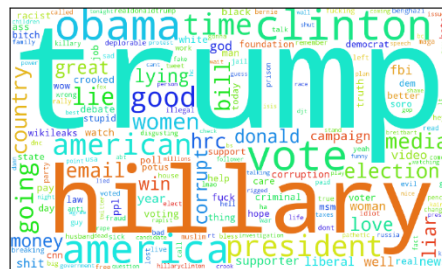
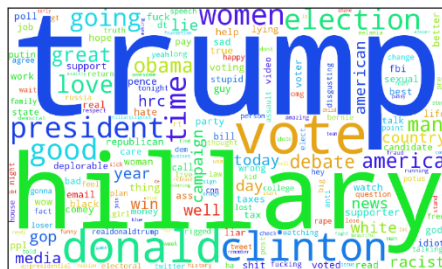
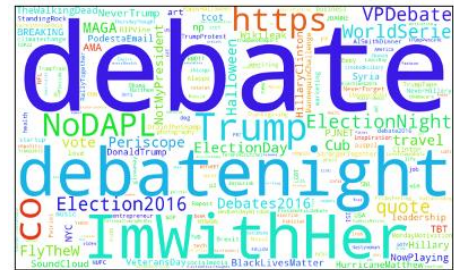
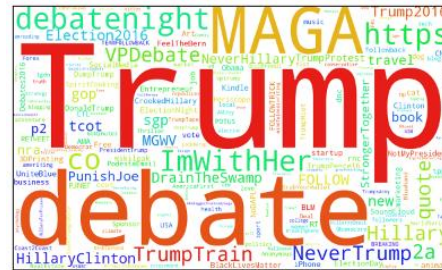
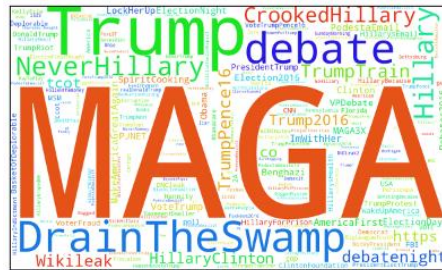
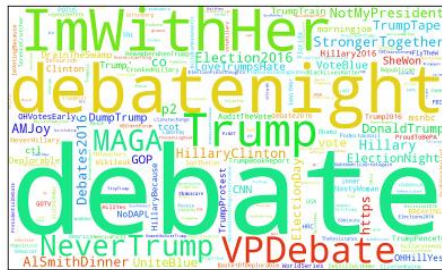
Daily Mean Sentiment – Trump's Advocates



Daily Mean Sentiment – Regular Users



Language Patterns – Hashtags and Words



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/Xgo5kjt8c>

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/YapuoqX0HS>.

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

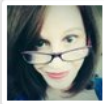
2



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

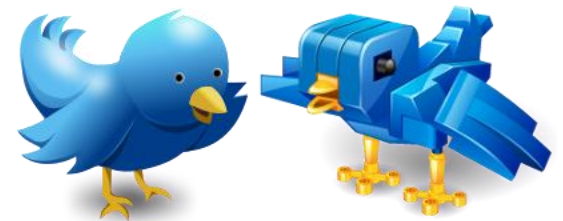


Lou Dobbs @LouDobbs

Lou Dobbs Tonight, Fox Business Network, 7 & 11 pm IG: <https://t.co/Mqnx3lgtA>

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

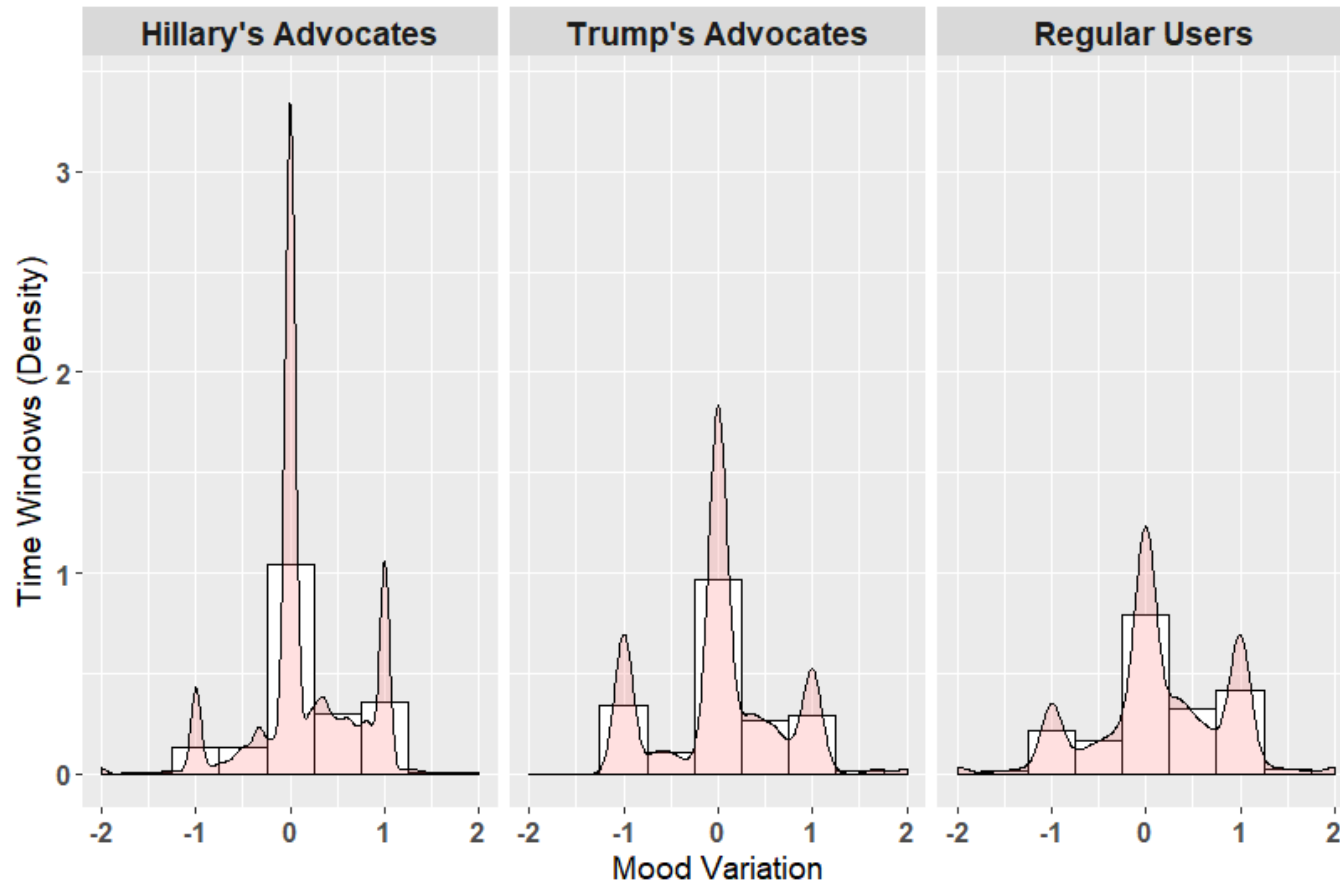
Sen. Bernie Sanders is the longest serving independent in congressional history. Tweets ending in -B are from Bernie, and all others are from a staffer.

5

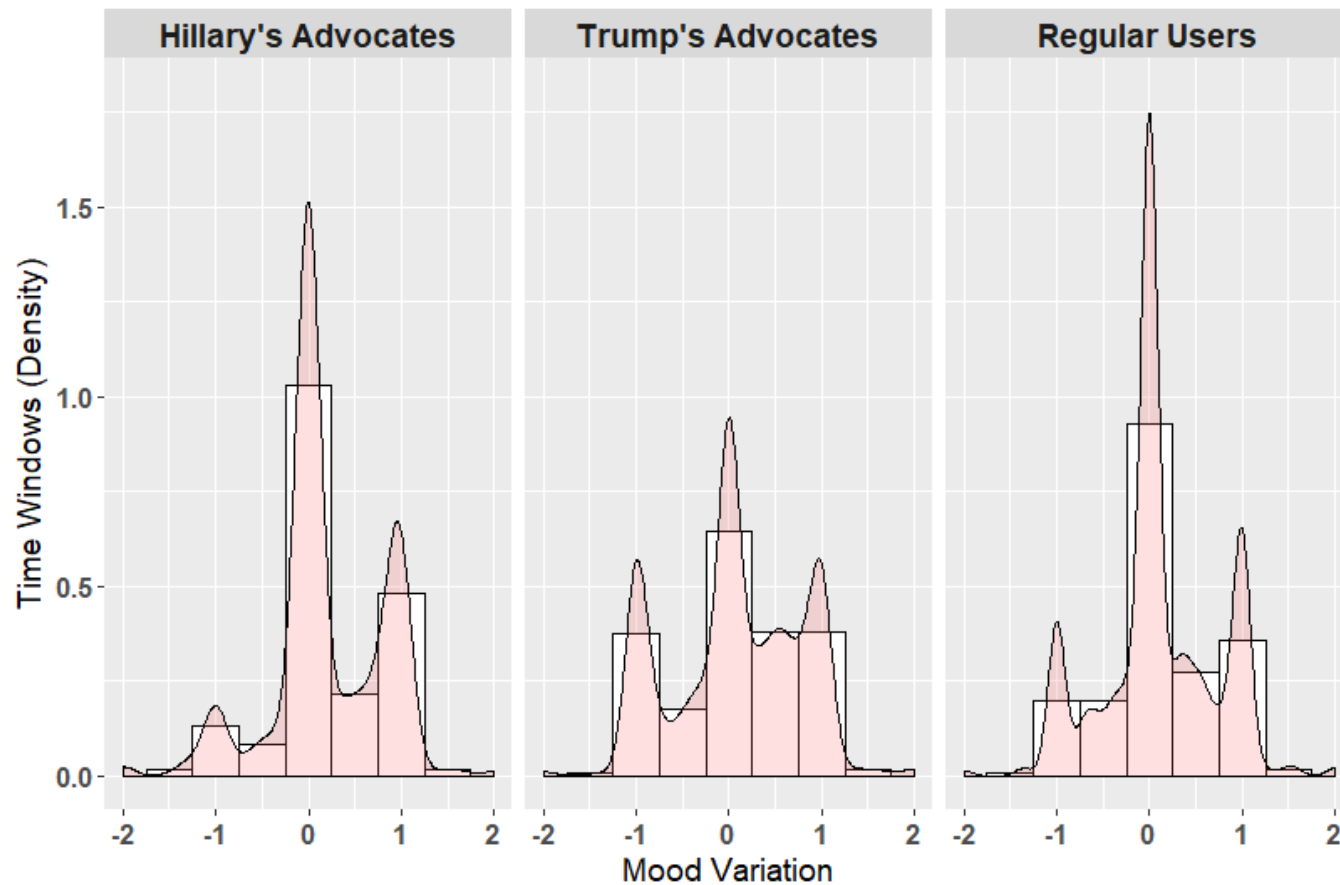
**Chris Hayes** @chrishayes

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

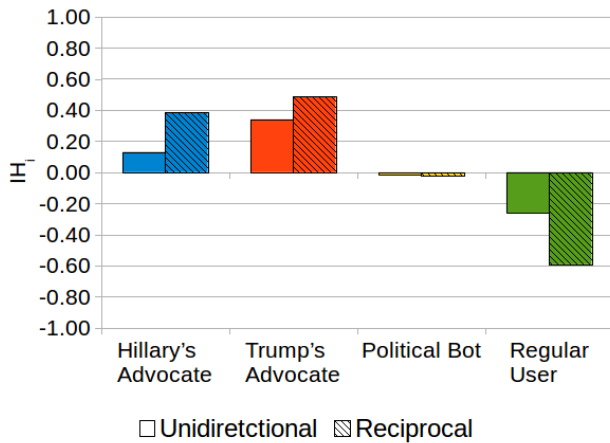


Mood Variation – Trump's Tweets

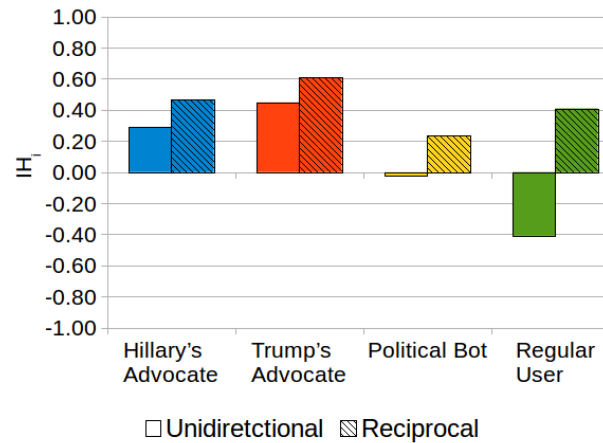


Homophily Analysis

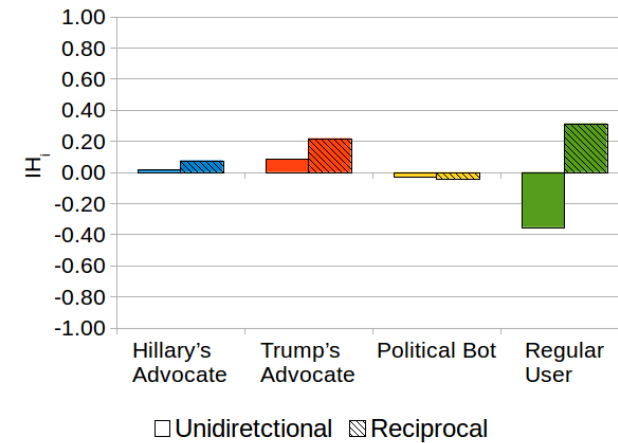
Homophily - Follow Connections



Homophily - Retweet Connections



Homophily - Mention Connections



Almost There...



Performing Sentiment Analysis on Tweets

SentiBench - a benchmark comparison of state-of-the-practice sentiment analysis methods

Filipe N Ribeiro, Matheus Araújo, Pollyanna Gonçalves, Marcos André Gonçalves,
Fabrício Benevenuto
EPJ Data Science, 2016

A Survey and Comparative Study of Tweet Sentiment Analysis via Semi-Supervised Learning

Nadia Felix F. Da Silva, Luiz F. S. Coletta, Eduardo R. Hruschka
ACM Computing Surveys, 2016

Identifying Political Bots

The rise of social bots

Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, Alessandro Flammini
Communications of the ACM, 2016

Online Human-Bot Interactions: Detection, Estimation, and Characterization

Onur Varol, Emilio Ferrara, Clayton Davis, Filippo Menczer, Alessandro Flammini
ICWSM, 2017

Identifying Advocates

Understanding and Identifying Advocates for Political Campaigns on Social Media

Suhas Ranganath, Xia Hu, Jiliang Tang, Huan Liu

International Conference on Web Search and Data Mining, ACM, 2016

Understanding Anti-Vaccination Attitudes in Social Media

Tanushree Mitra, Scott Counts, James W. Pennebaker

ICWSM 2016

Mood Variation Analysis

Happiness is assortative in online social networks

Johan Bollen, Bruno Gonçalves, Guangchen Ruan, Huina Mao
Artificial Life, Springer, 2011

Emotions in Social Networks: Distributions, Patterns, and Models

Jin Shengmin, Reza Zafarani
Conference on Information and Knowledge Management, 2017

Homophily Analysis

Birds of a feather: Homophily in social networks

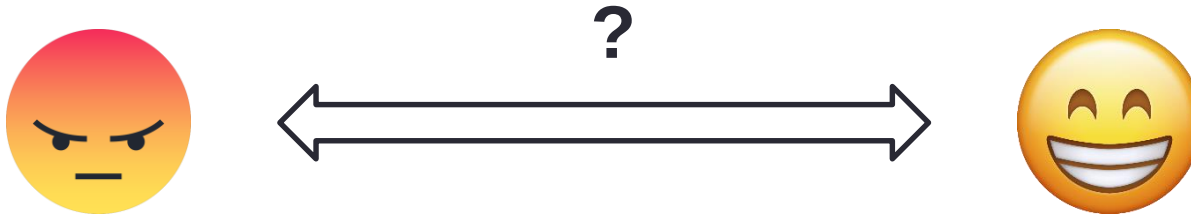
McPherson, Miller and Smith-Lovin, Lynn and Cook, James M
Annual review of sociology, 2001

Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data

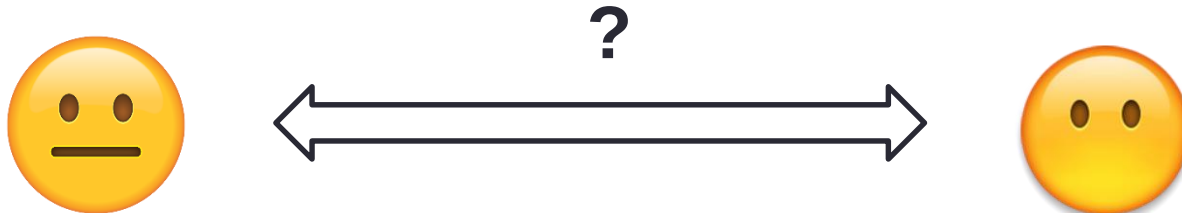
Colleoni, Elanor, Rozza, Alessandro, Arvidsson, Adam
Journal of Communication, 2014

Threats to Validity

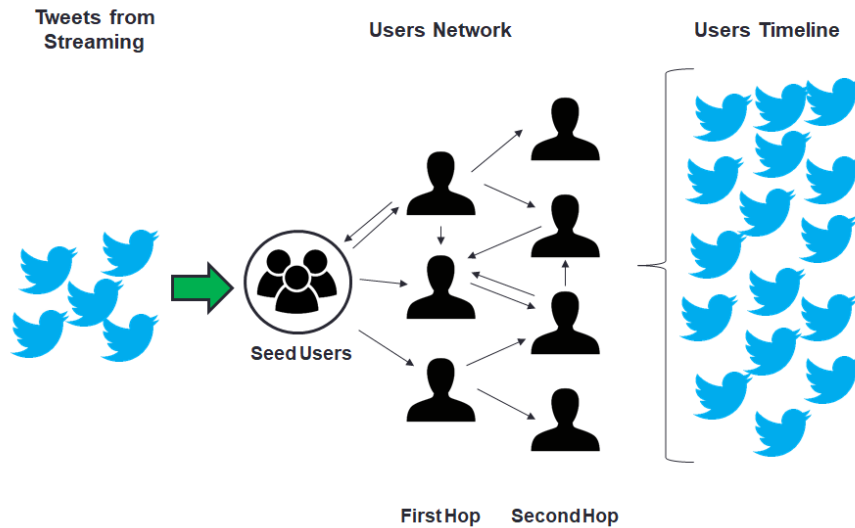
- Sarcasm in tweets



- Twitter API limitation (max 200 last tweets)



Concluding Remarks



Concluding Remarks



Which features
highlight each group



Negative



Neutral



Positive

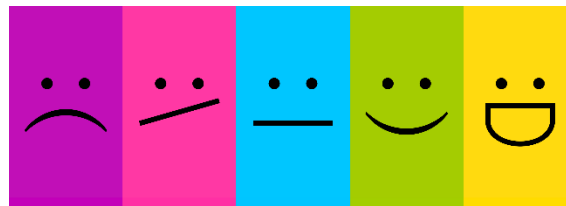
Mean Sentiment
Analysis



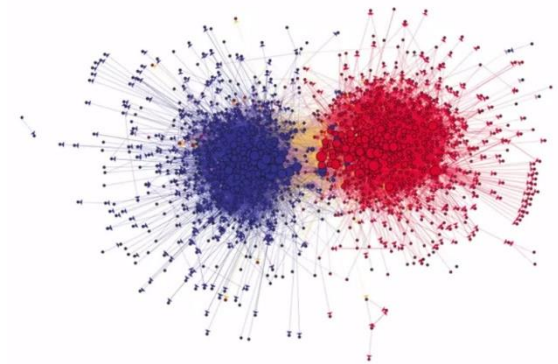
Language Patterns
Analysis



Popular users of
each group



Mood Variation
Analysis

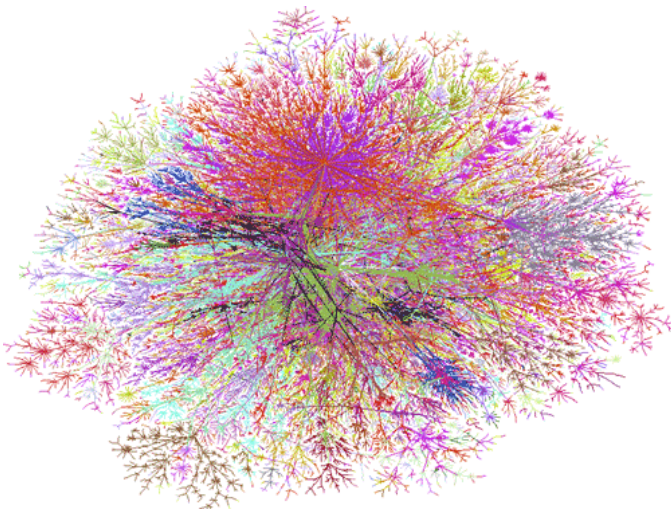


Homophily
Analysis

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

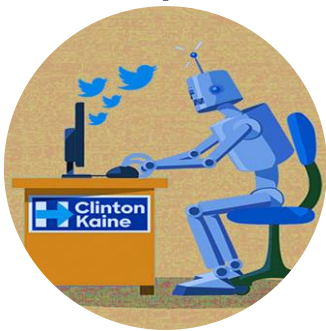
Future Work – Characterizations



Future Work – Political Groups



Political Bots

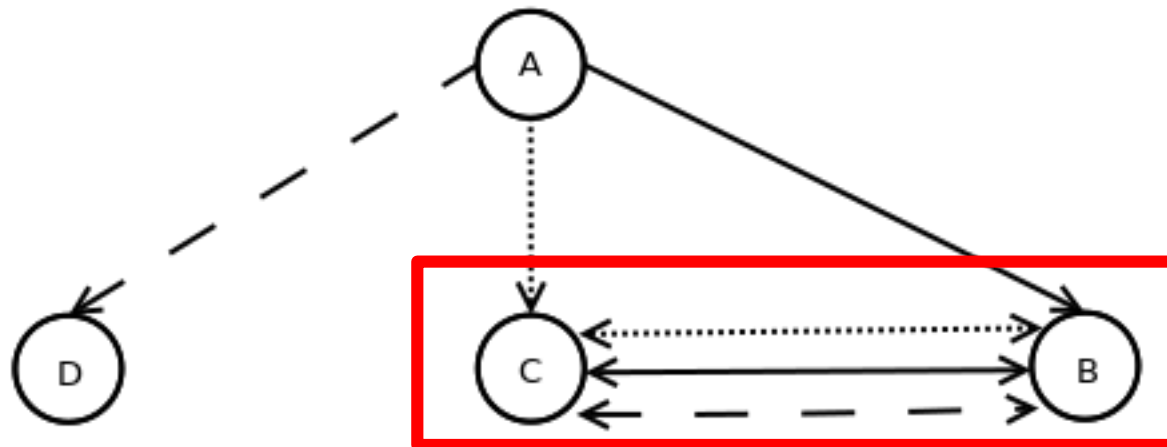


Hillary's Bots



Trump's Bots

Future Work - Homophily

**Legend:**

- Follow connection
- Mention connection
- - - Retweet connection

Publications

Utilizando Análise de Sentimentos para Definição da Homofilia Política dos Usuários do Twitter durante a Eleição Presidencial Americana de 2016

BraSNAM 2017 (best paper)

Using the Sentiment Analysis to Define Twitter Political Users' Classes and Their Homophily during the 2016 American Presidential Election

JISA 2017 (under review)

Characterizing Politically Engaged Users and their Mood Variation during the 2016 US Presidential Campaign

ICWSM 2018 (under review)

Thank You!