ANALYZING THE DIGITAL TRACES OF POLITICAL MANIPULATION:
(THE 2016 RUSSIAN INTERFERENCE TWITTER CAMPAIGN)

Adam Badawy, Emilio Ferrara, Kristina Lerman

University of Southern California

BACKGROUND

- Studying large scale online political manipulation campaign
- US Presidential Elections
- Trolls

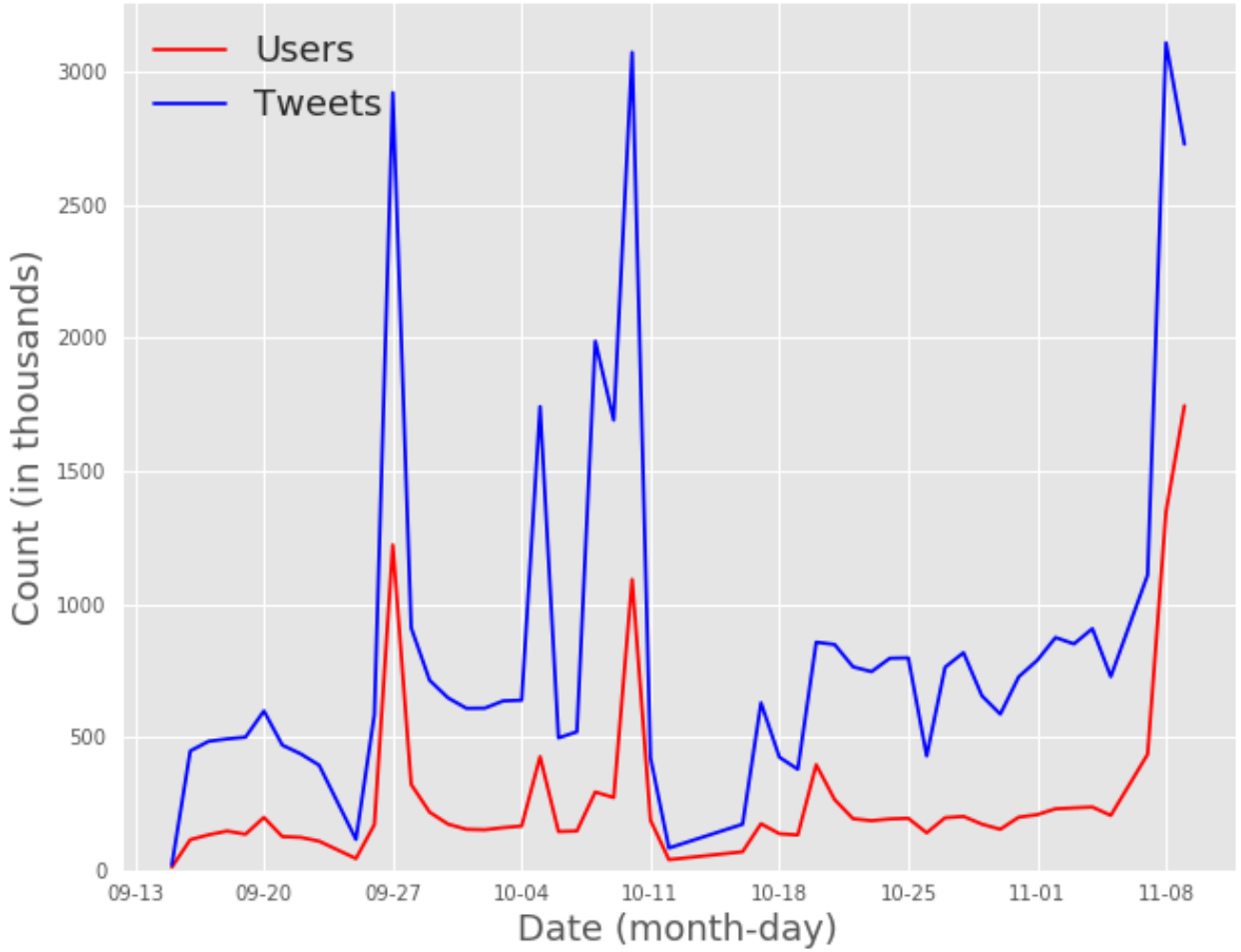
RESEARCH QUESTIONS

- What was the role of the users' political ideology?
- What was the role of social bots?
- Did trolls especially succeed in specific areas of the US?
- Can we predict which users will become susceptible to Russian trolls?
- What features distinguish users who spread trolls' messages?

DATA COLLECTION

- Twitter dataset: 43.7 M tweets posted by 5.7 M users from 15th of September to 9th of November 2016.
- Data collected using roughly equal number of hashtags and keywords (23 terms) associated with each major Presidential candidate.
- Over 31 M of the tweets are retweets and tweets/retweets with urls are over 22 M.

VOLUME OF TWEETS(BLUE) & USERS(RED)



RUSSIAN TROLLS

- Russian trolls were retweeted ~83K times, but most of the retwets came from 3 troll accounts:
  1. 'TEN GOP': 49,286
  2. 'Pamela Moore13': 16,532
  3. 'The-FoundingSon': 8,755; in total making over 89% of the retweets.

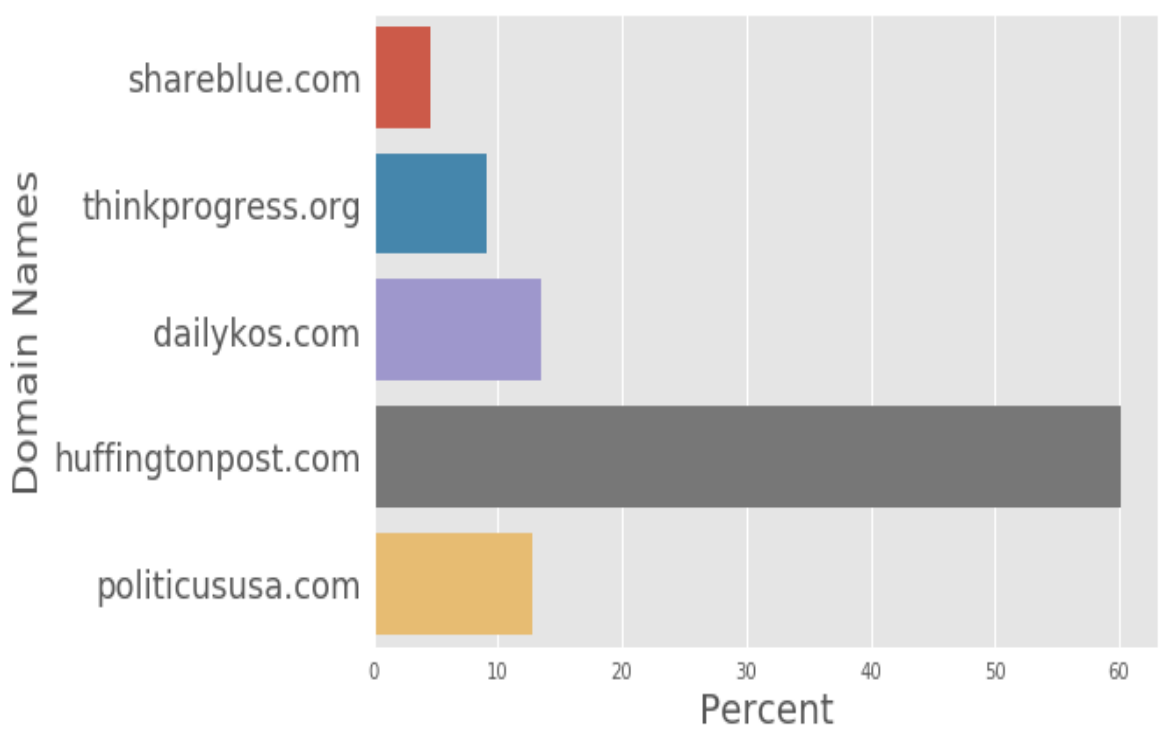
	Value
# of Russian trolls	2,735
# of trolls in our data	221
# of trolls wrote original tweets	85
# of original tweets	861

POLITICAL IDEOLOGY

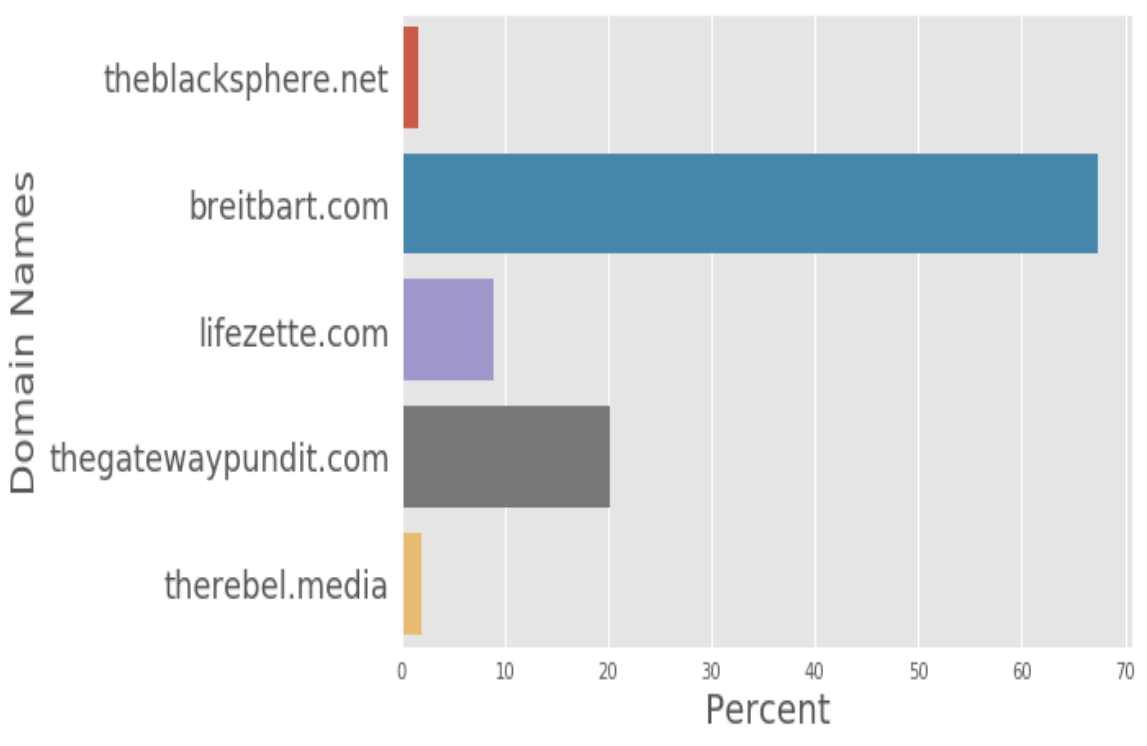
CLASSIFICATION OF MEDIA OUTLETS

- Classification of users based on the political leaning of the media outlets they share.
- Complied lists of partisan media outlets by AllSides and Media Bias/Fact Check.
- Picked five media outlets from each partisan category that appeared most frequently in the Twitter dataset and compiled a list of users who tweeted most from these outlets.
- For top-five liberal media outlets, we have ~161K tweets and ~10K users; for the conservative outlets: ~184K tweets and ~7K users.

LIBERAL



CONSERVATIVE



RETWEET NETWORK

	Count
# of nodes	~4.6M
# of edges	~19M
# of nodes in the weak giant component	~4.4M

LABEL PROPAGATION

- Semi-supervised network-based algorithm:
  1. Each node is assigned a label, which is updated iteratively based on the labels of node's neighbors.
  2. Each node takes the most frequent label of its neighbors as its own new label.
  3. The algorithm proceeds updating labels iteratively and stops when the labels no longer change.
- The algorithm takes as parameters:
  1. weights (in-degree)
  2. seeds (the list of labeled nodes).
- The seeds' labels are fixed so they do not change in the process, since this seed list serves as the ground truth

VALIDATION

- ~ 3.4 M labeled as Liberals and ~ 1M as Conservatives.
- Applied a stratified cross 5-fold validation to the set of ~29K seeds.
- The precision and recall are around 0.91.
- Same technique with a hyper-partisan list of users, precision and recall are about 0.93.

TROLLS BY IDEOLOGY

	Liberal	Conservative	Ratio
# of trolls	107	108	1
# of trolls w/ original tweets	15	64	4.3
# of original tweets	44	844	19

TOP 20 STEMMED WORDS

Liberal	count	Conservative	count
trump	14	trumpforpresid	486
debat	10	trump	241
nevertrump	6	trumppence16	227
like	5	hillaryforprison2016	168
2016electionin3word	5	vote	127
elections2016	4	maga	113
imwithh	4	neverhillari	106
obama	3	election2016	102
need	3	hillari	100
betteralternativetodeb	3	hillaryclinton	85
women	3	trump2016	80
would	3	draintheswamp	50
vote	3	trumptrain	48
mondaymotiv	2	debat	48
last	2	realdonaldtrump	45
oh	2	electionday	43
thing	2	clinton	41
damn	2	makeamericagreatagain	34
see	2	votetrump	32
defeat	2	america	31

SPREADERS

	Value
# of spreaders	40,224
# of times retweeted trolls	83,719
# of spreaders with original tweets	28,274
# of original tweets	>1.5 Million
# of original tweets and retweets	>12 Million

	Liberal	Conservative	Ratio
# of spreaders	892	27,382	31
# of tweets	>42,000	>1.5 Million	36

SOCIAL BOTS

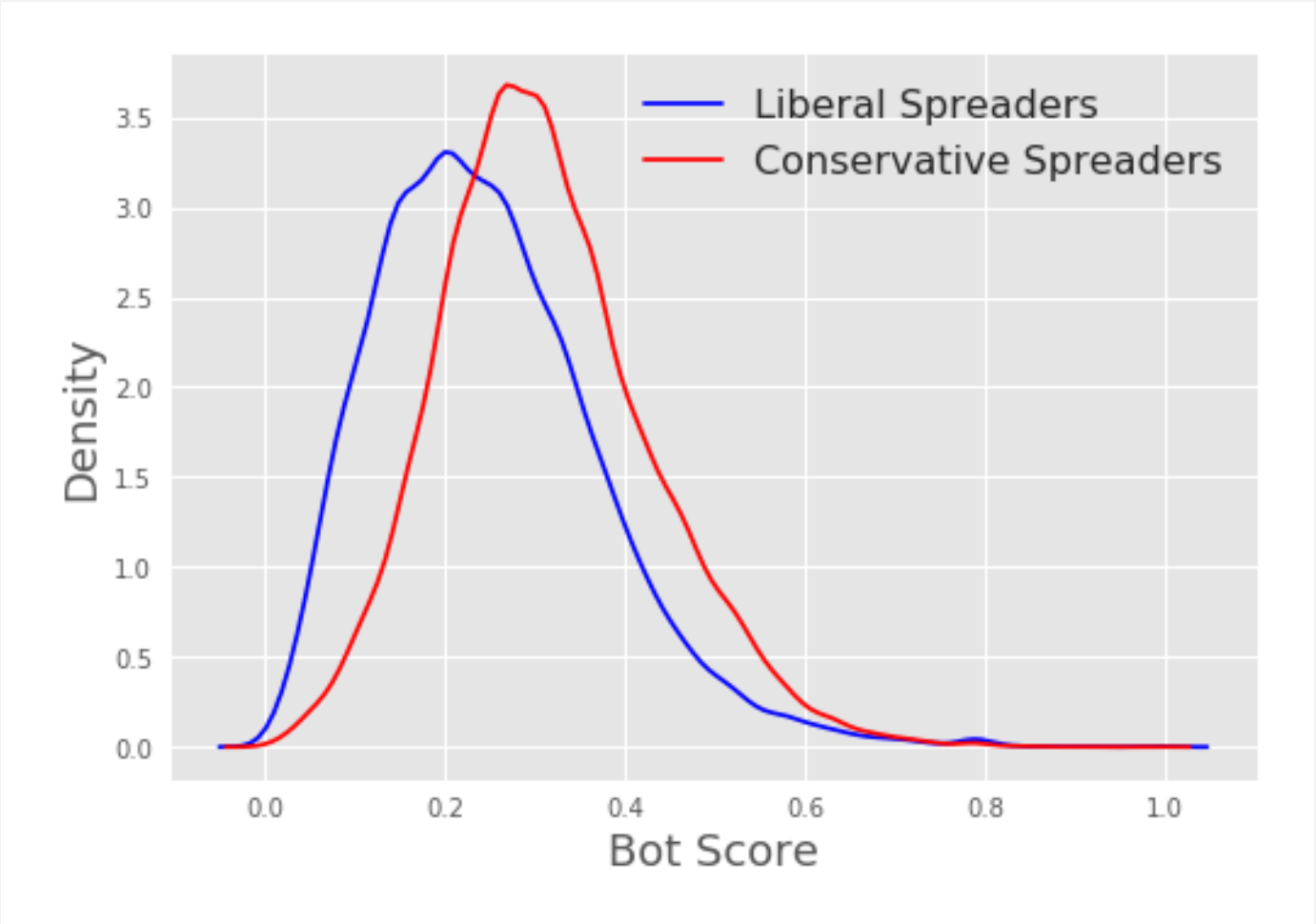
BOT DETECTION

- Botometer (BotOrNot)
- It extracts and analyses a set of over one 1,000 features spanning:
  1. content
  2. network structure
  3. temporal activity
  4. user profile data
  5. sentiment analysis
- Produces a score for the likelihood that the inspected account is a social bot, [0,1], above 0.5 is considered a bot.

SPREADER BOT ANALYSIS

- Liberal bots:
  1. 4.9% of total liberal users
  2. 8.3% of total tweets by liberal uses
- Conservative bots:
  1. 6.2% of total conservative users
  2. 8% of total tweets by conservative users

	Liberal	Conservative	Ratio
# of spreaders	1,506	32,513	22
# of tweets	224,943	11,928,886	53
# of bots	75	2,018	27
# of tweets by bots	18,749	955,583	51



PROBABILITY DENSITY DISTRIBUTION (LIBERALS VS. CONSERVATIVES)

- Liberal Bot Scores' Mean: 0.24
- Conservative Bot Scores' Mean: 0.3
- p-value < 0.0 (two-sided t-test)

GEOSPATIAL ANALYSIS

GEOLOCATION

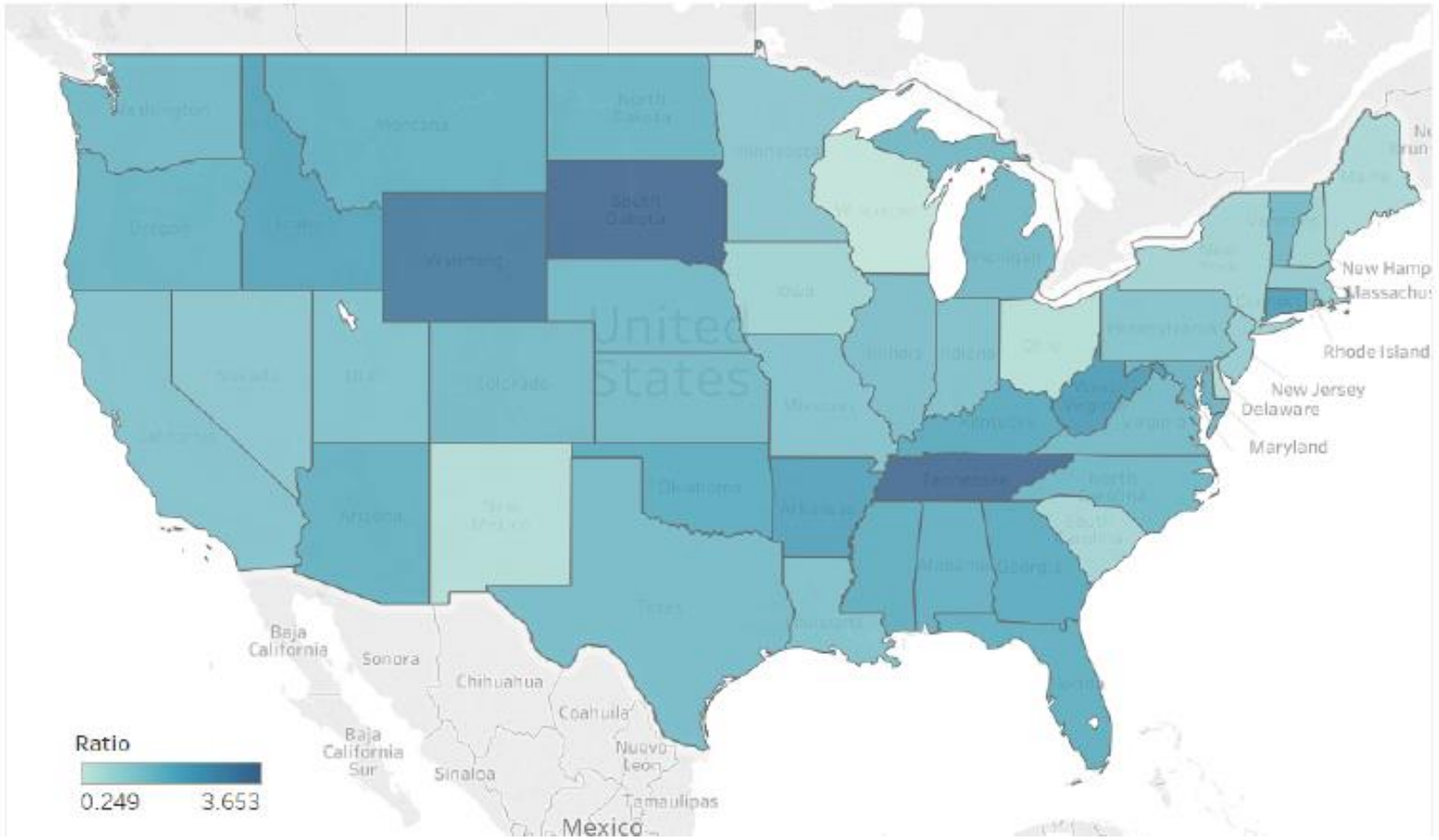
- Two ways to get users' locations:
  1. Tweets' geo-location
  2. Self-reported account location

- Only about ~36K tweets are geo-located (~0.001% of the total volume of tweets). Concentrated mainly in the South with Kentucky being the state with the most geo-located tweets.

- For the self-reported location, we used Google map api for the top used locations.

ANALYSIS

- $\rho = (T_S / P_S) \times 100$:
  1. T_s is the total number of retweets of trolls from a given state S
  2. P_s is the total number of tweets per each State
- After calculating the deviations by using a two-tailed t-test on the z-scores of each deviation calculated on the distribution of ratios, we see that some states exhibit high proportions of retweets per total number of tweets for conservatives:
  1. South Dakota (ρ=3.65, p-value < 0.001)
  2. Tennessee (ρ=3.61, p-value < 0.001)
  3. Wyoming (ρ=3.20, p-value = 0.019)


Rhode Island
New Jersey
Delaware
Maryland
Baja California
Sonora
Chihuahua
Coahuila
Baja California Sur
Sinaloa
Nuevo Leon
Tamaulipas
México
Ratio
0.249
3.653

FEATURES FOR PREDICTION OF SPREADERS

FEATURES

Metadata	LIWC	Engagement	Activity	Other
# of followers	Word Count	Retweet variables	# of characters	Political Ideology
# of favourites	Postive Emotion	Mention variables	# of hashtags	Bot Score
# of friends	Negative Emotion	Reply variables	# of mentions	Tweet Count
Status count	Anxiety	Quote variables	# of urls	
Listed count	Anger			
Default Profile	Sadness			
Geo-enabled	Analytic			
Background-image	Clout			
Verified	Affection			
Account Age	Tone			

LINGUISTIC INQUIRY AND WORD COUNT (LIWC) I

- Psychological Processes:
  1. Positive emotion: love, nice, & sweet
  2. Negative emotion: hurt, ugly, & nasty
  3. Anxiety: worried, & fearful
  4. Anger: hate, kill, & annoyed
  5. Sadness: crying, grief, & sad

LINGUISTIC INQUIRY AND WORD COUNT (LIWC) II

- Summary Language Variables:
  1. Analytical thinking : formal, logical, and hierarchical thinking
  2. Clout: speaking from the perspective of high expertise and confidence
  3. Authentic: honest, personal, and disclosing text
  4. Emotional tone: positive and upbeat style text

ENGAGEMENT

- User engagement in four activities:
  - Retweets
  - Mentions
  - Replies
  - Quotes
- Engagement of a user is measured through three components: the quantity, longevity, and stability in each activity

ENGAGEMENT VARIABLES

- For a set of N users, we calculate 15 engagement scores for user $i \in N$ by calculating the following:
  1. number of retweets, replies, mentions, and quotes by $N - i$ users for user i
  2. time difference between the last and the first quote, reply, and retweet per tweet
  3. consistency of mentioning, replying, retweeting, and quoting by $N - i$ users for user i across time (per day)
  4. number of unique users who retweeted, commented, mentioned, and quoted user i

ENGAGEMENT (H-INDEX)

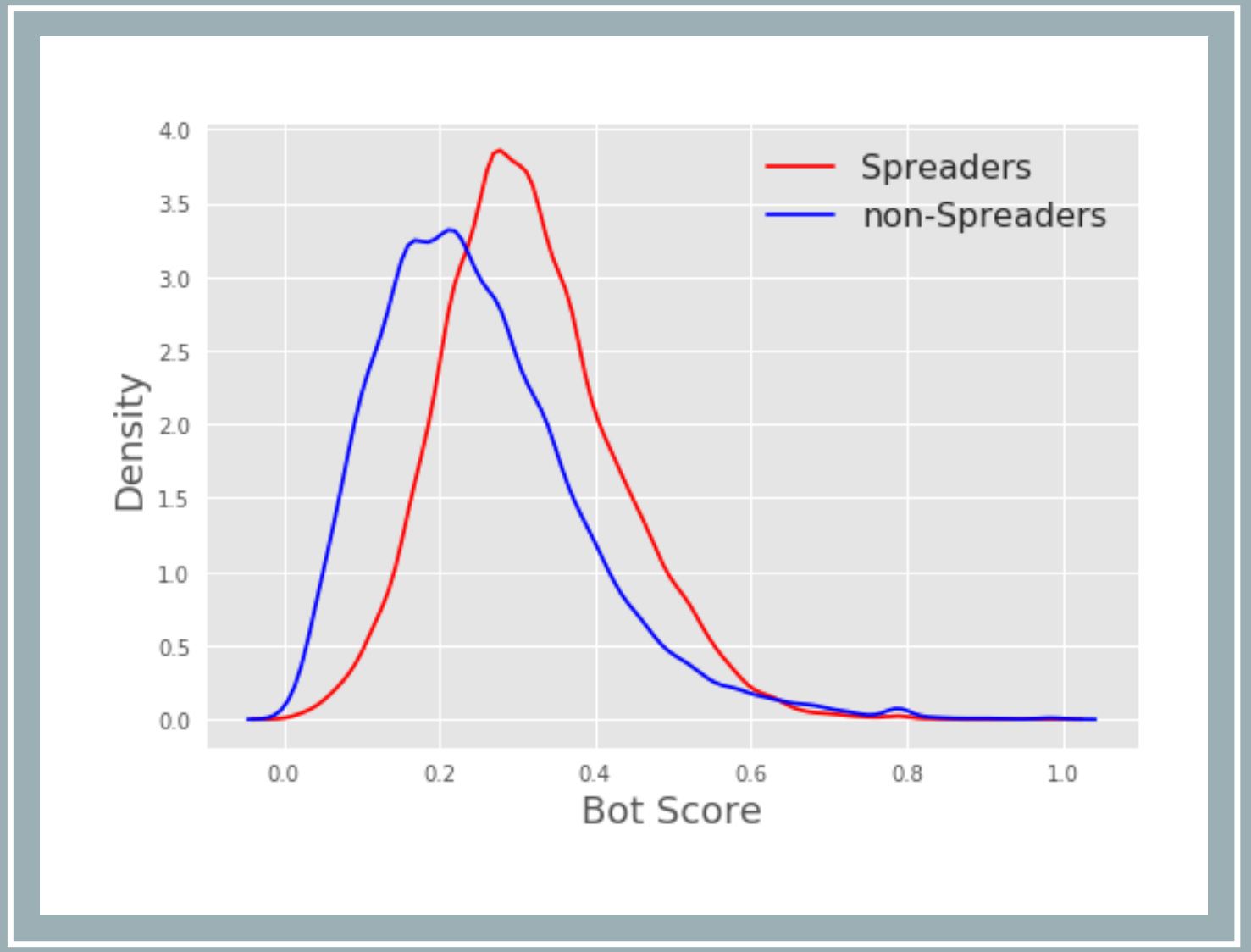
- The measure captures two notions: how highly referenced and how continuously highly referenced by the rest of the network

- Example: number of retweets of a user in 7 days: {27, 4, 2, 40, 100, 50, 60}.
- Reorder from highest to lowest: {100, 60, 50, 40, 27, 4, 2}
- This user h-index is 5

POLITICAL IDEOLOGY

	Liberal	Conservative
# of users	>3.4 M	>1 M
# of trolls	107	108
# of spreaders	1,991	38,233

PROBABILITY DENSITY DISTRIBUTION (SPREADERS VS. NON-SPREADERS)

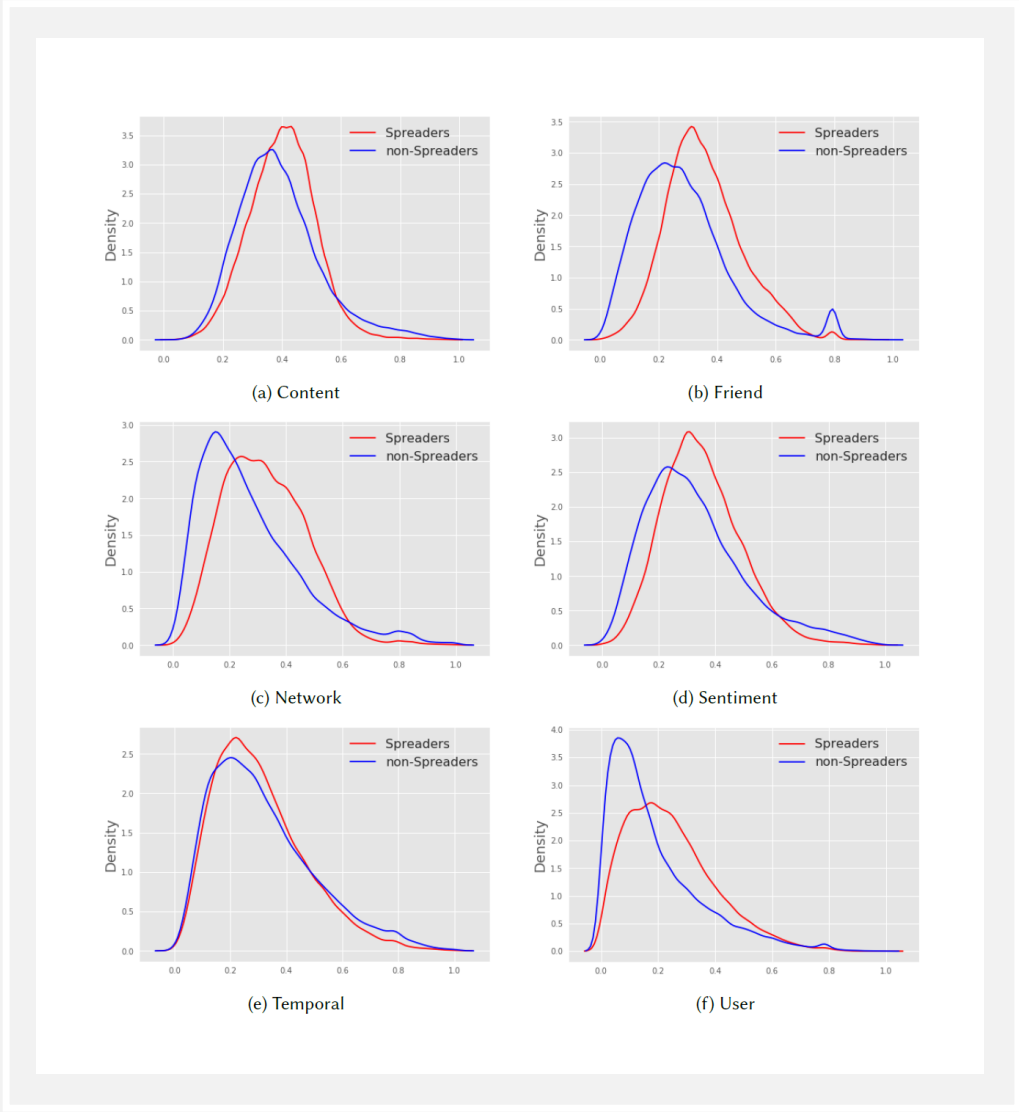


BOTOMETER SUB-CLASS FEATURES

Spreaders are different on almost all the Botometer subclass scores, except for the temporal features

Characteristics (metadata), friends, and network distributions, are the most different respectively

Mean of spreaders is higher in all the subclass features



PEARSON CORRELATION HEATMAP

Engagement variables: "rich get richer“ effect

Word Count and Tweet Count, LIWC Positive Emotion and Affection, Anxiety and Anger, all these pairs show very high correlation

PREDICTION

Model	Features
1	Metadata
2	Metadata + LIWC
3	Metadata + LIWC + Activity
4	Metadata + LIWC +Activity + Engagement
5	Metadata + LIWC +Activity + Engagement + Other

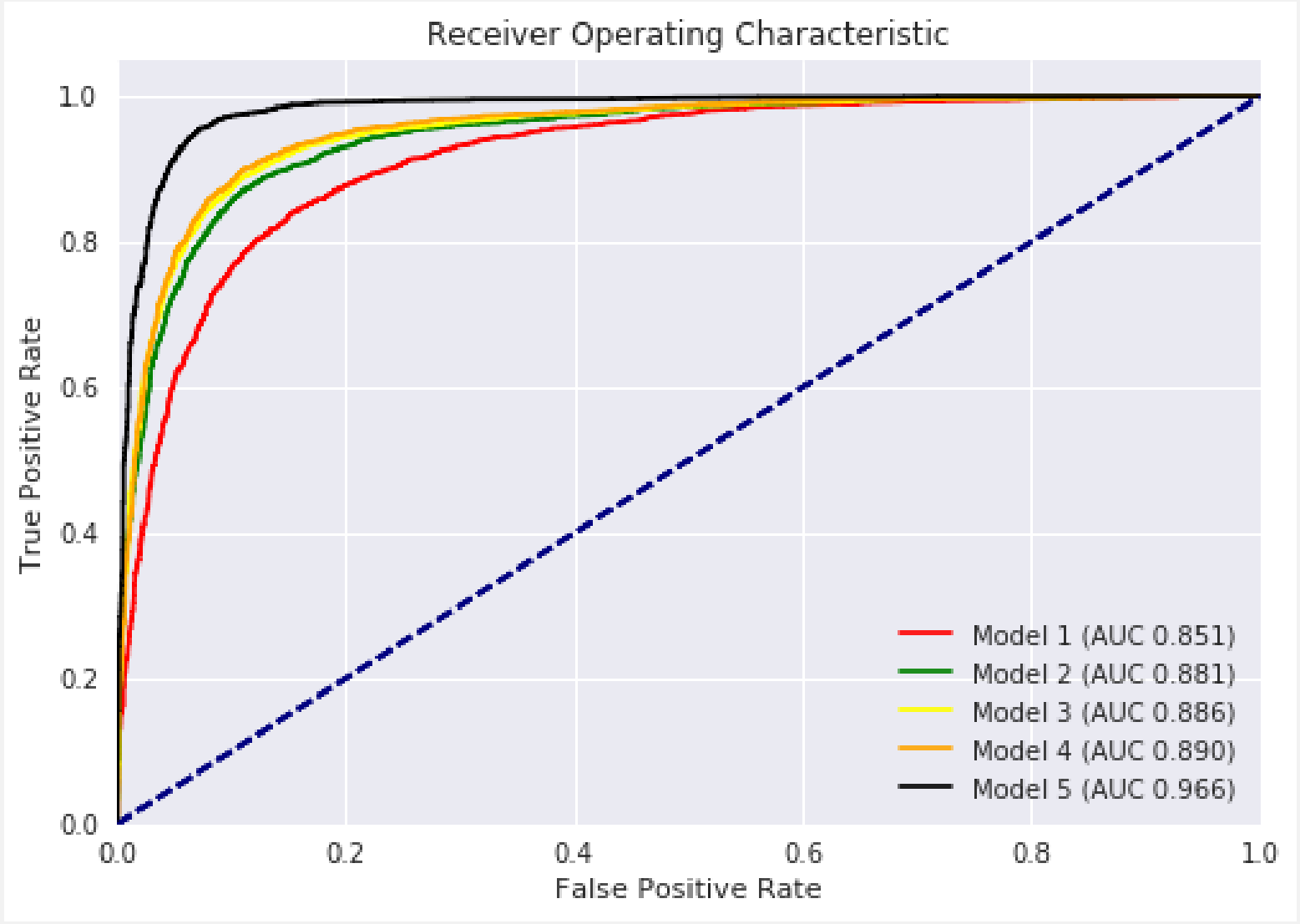
CLASSIFIERS & PREPROCESSING

- Four off-the-shelf machine learning algorithms:
  1. Extra Trees
  2. Random Forest
  3. Adaptive Boosting
  4. Gradient Boosting
- Stratified 10-fold cross-validation:
  1. replace categorical missing values with the most frequent value
  2. replace continuous missing values with the mean of the variable

GRADIENT BOOSTING (BALANCED DATASET)

- Balanced dataset has about 72K users,:
  1. 34K spreaders
  2. 38K non-spreaders

- Average AUC scores for the 10 folds range from 85% to 96%

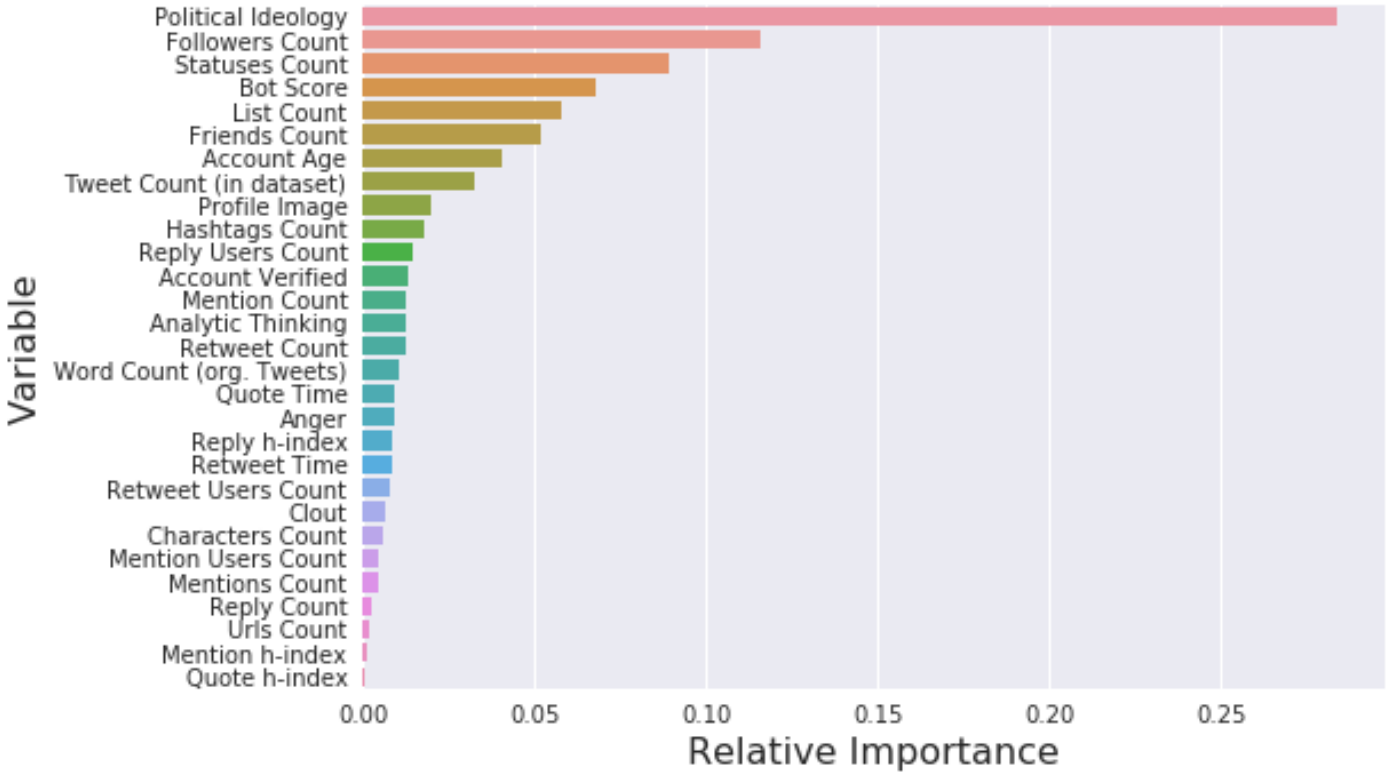

Receiver Operating Characteristic
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
Model 1 (AUC 0.851)
Model 2 (AUC 0.881)
Model 3 (AUC 0.886)
Model 4 (AUC 0.890)
Model 5 (AUC 0.966)

VALIDATION

- Two strategies:
  - Gradient Boosting (with the same preprocessing steps) on the whole dataset
  - Different models without imputations and with all missing observations deleted (using Gradient Boosting)
- First approach: average AUC scores (10-fold validation) ranged form 83% for the baseline model to 98% for the full model
- Second approach: 84% to 91%

FEATURE IMPORTANCE

VARIABLE IMPORTANCE



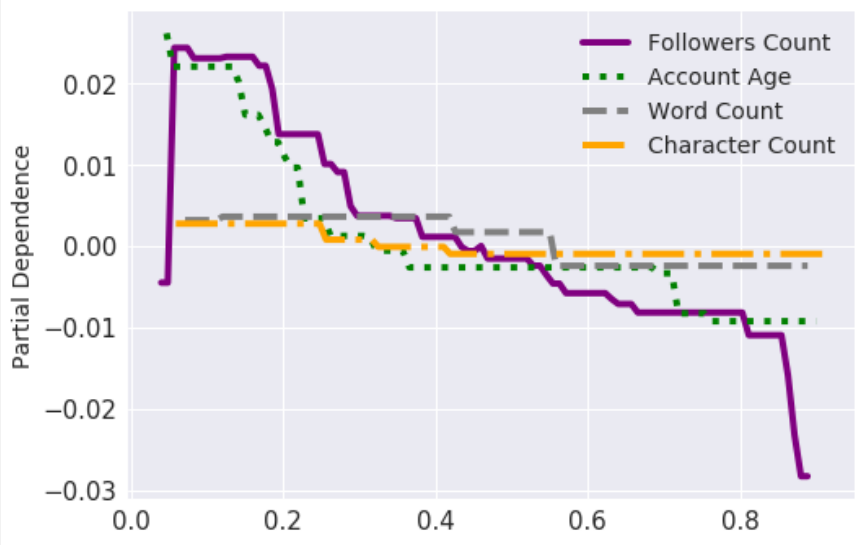
PARTIAL DEPENDENCE

- Feature importance plots reveal which features contribute most to classification performance, but they do not tell us the nature of the relationship between the outcome variable and the predictors
- Partial Dependence plots tell us a lot about the structure and direction of the relationship between the target and independent variables
- They show these relationships after the model is fitted, while marginalizing over the values of all other features

PARTIAL DEPENDENCE PLOTS

UPWARD TRENDS

DOWNWARD TRENDS



CONCLUSION

- Messages of Conservative trolls spread more than Liberal trolls.
- Conservative spreaders have a higher bot scores than Liberal spreaders.
- Some Southern states show anomalously high levels of retweeting of Conservative trolls.
- Predicting users who spread trolls' messages is feasible with high precision/recall.
- Political Ideology, metadata, and bot scores are predictive of users' susceptibility to share trolls' content