RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

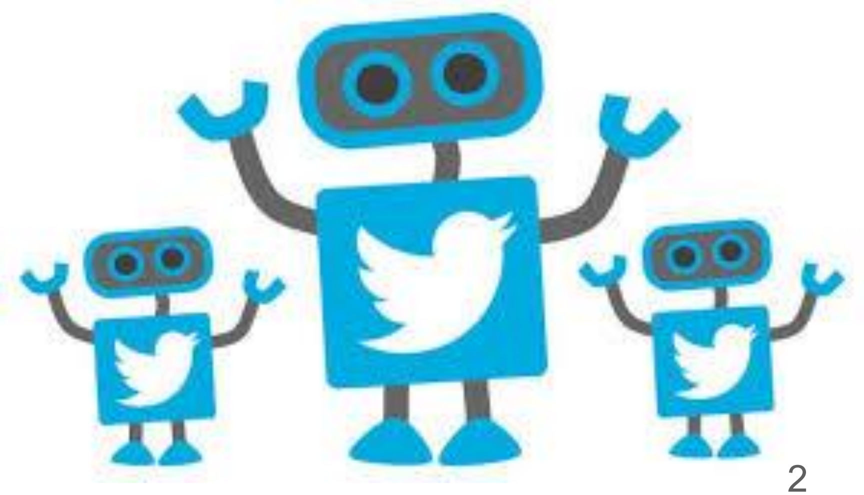
Survey of Political Bots on Twitter

A thesis submitted to the Graduate School - Camden
Rutgers, The State University of New Jersey

David Troupe
Candidate for Master of Science in Scientific Computing
December 2018

Outline

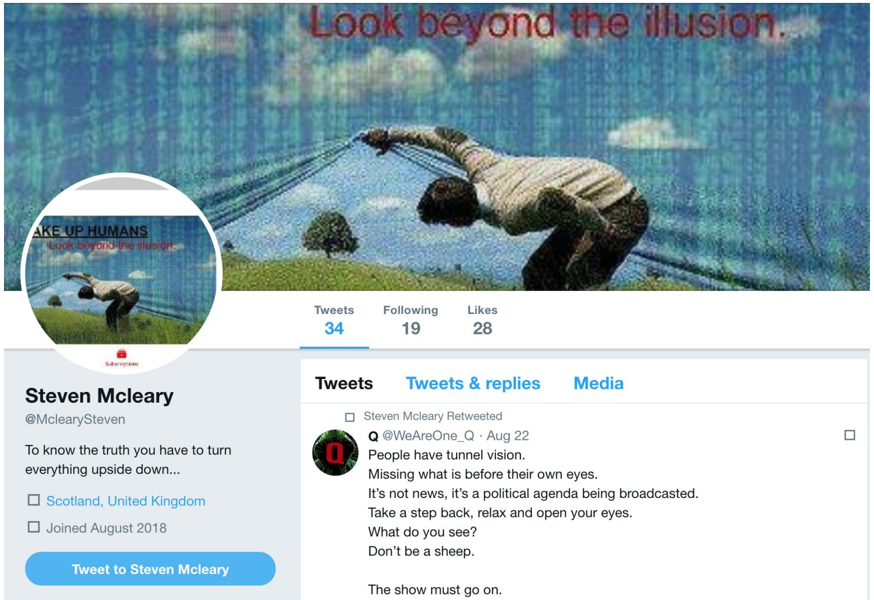
- Understanding Twitter bots
- Data collection
- Creation of training datasets
- Detection using machine learning
- Bot response to political events
- What is Twitter doing about political bots?
- Conclusions



What is a Twitter bot?

- Internet bots (web bots, or simply “bots”)
  - Automated tasks (scripts)
  - Simple, repetitive tasks at fast rate
- Cyborgs
  - Characteristics of both human & automated behavior
  - “Sophisticated bots”

focus: political bots



Tweets 34 | Following 19 | Likes 28

Steven Mcleary
@MclearySteven

To know the truth you have to turn everything upside down...

□ Scotland, United Kingdom
□ Joined August 2018

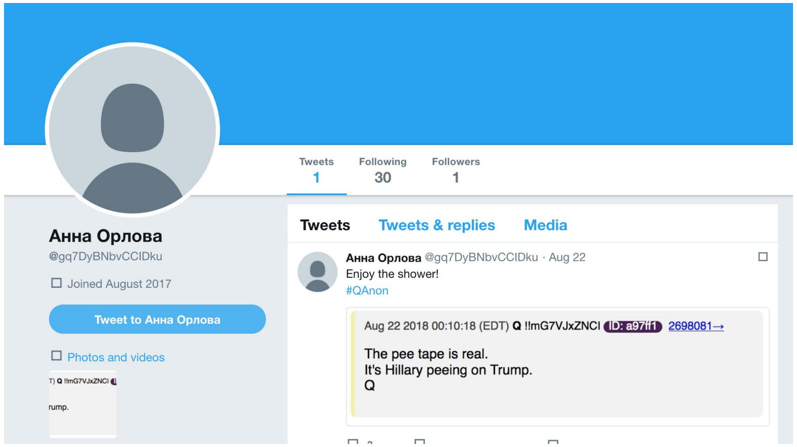
Tweet to Steven Mcleary

Tweets | Tweets & replies | Media

□ Steven Mcleary Retweeted

Q @WeAreOne_Q · Aug 22
People have tunnel vision.
Missing what is before their own eyes.
It’s not news, it’s a political agenda being broadcasted.
Take a step back, relax and open your eyes.
What do you see?
Don’t be a sheep.

The show must go on.



Tweets 1 | Following 30 | Followers 1

Анна Орлова
@gq7DyBNbvCCIDku

□ Joined August 2017

Tweet to Анна Орлова

□ Photos and videos

Tweets | Tweets & replies | Media

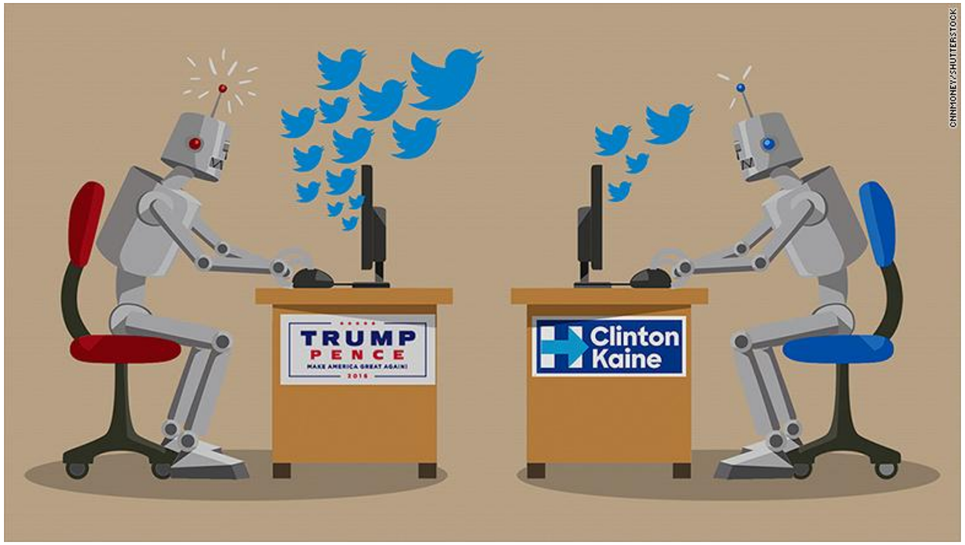
Анна Орлова @gq7DyBNbvCCIDku · Aug 22
Enjoy the shower!
#QAnon

Aug 22 2018 00:10:18 (EDT) **Q** !!mG7VJxZNCI ID: a97if1 2698081→

The pee tape is real.
It's Hillary peeing on Trump.
Q

Motivation for focus on political bots

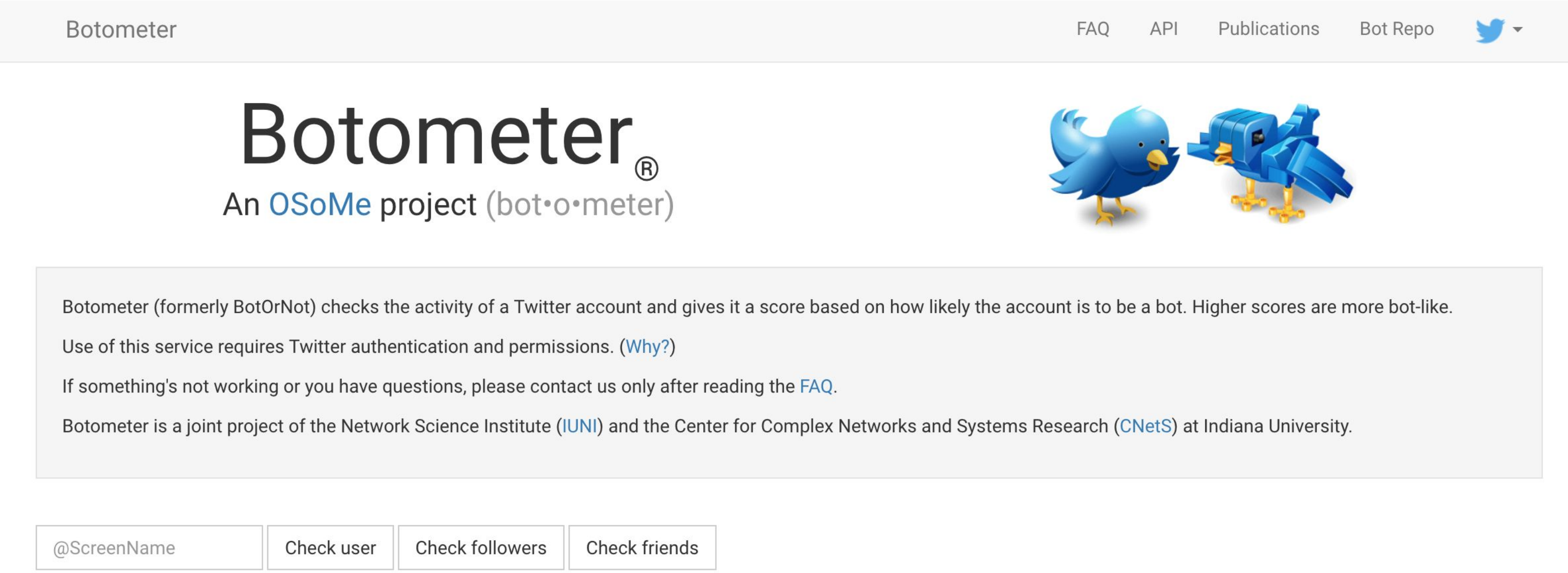
- Fueling political hysteria
- Used to influence elections around the world
- Large-scale spread of misinformation
- Problem can be solved



CNNMONEY/SHUTTERSTOCK

Data Collection: Related work

- *Online Human-Bot Interactions: Detection, Estimation, and Characterization* [1]
  - Measured bot and human behavioral dynamics
  - CAP score
- Utilized for Bot vs. Human status for training data



Botometer (formerly BotOrNot) checks the activity of a Twitter account and gives it a score based on how likely the account is to be a bot. Higher scores are more bot-like.

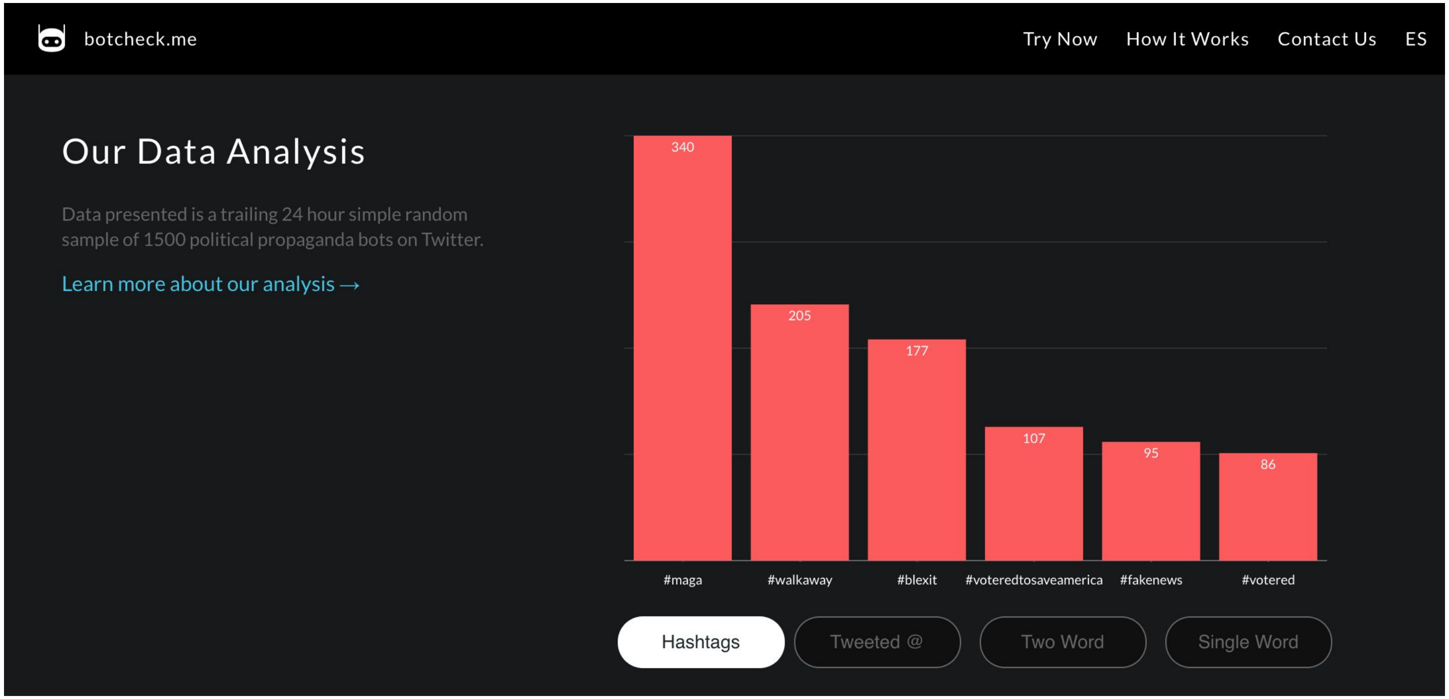
Use of this service requires Twitter authentication and permissions. (Why?)

If something's not working or you have questions, please contact us only after reading the FAQ.

Botometer is a joint project of the Network Science Institute (IUNI) and the Center for Complex Networks and Systems Research (CNetS) at Indiana University.

Data Collection: Related work

- *BotCheck.me* [2]
  - Displays most common hashtags by political bots
- Used to find Twitter accounts of interest



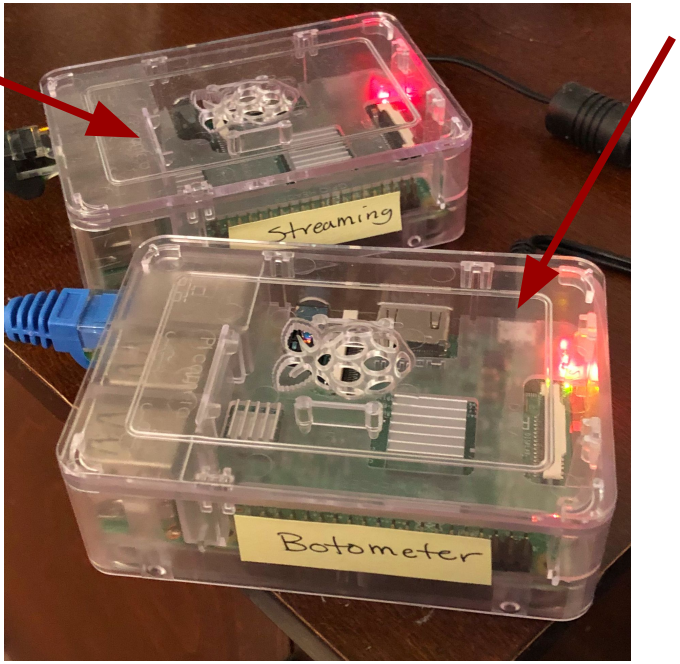
Our Data Analysis

Data presented is a trailing 24 hour simple random sample of 1500 political propaganda bots on Twitter.

Learn more about our analysis →

Data Collection: Technical implementation

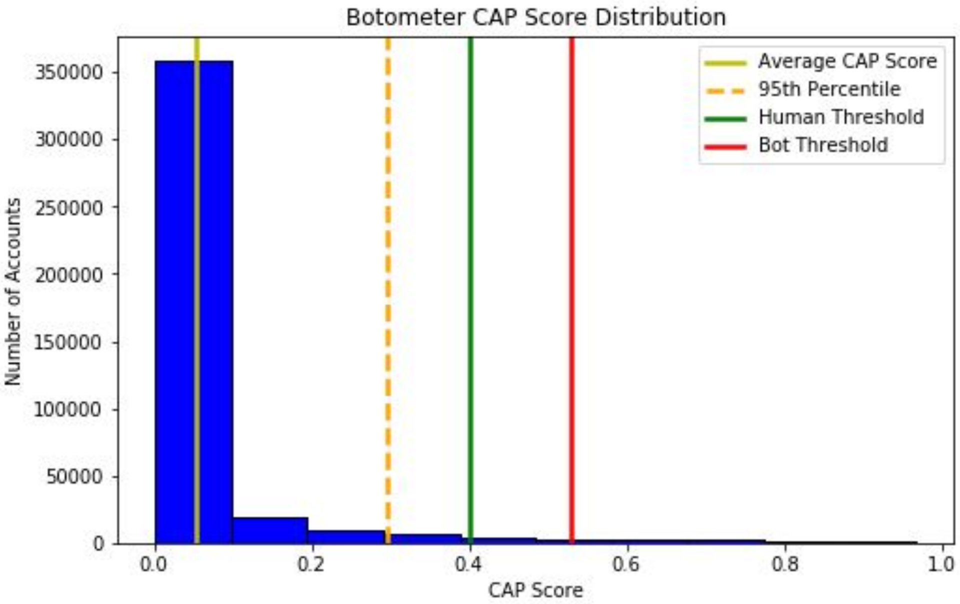
- Twitter stream
- Data collected (CSV file):
  - Public profile information: username, screen name, description, tweet count, friend count, account creation date
  - Tweet information: tweet text, tweet creation timestamp, retweet count, like count, additional information



- Botometer CAP score assignment
  - Last 200 tweets
  - Complete user profile
  - All recent “mentions”

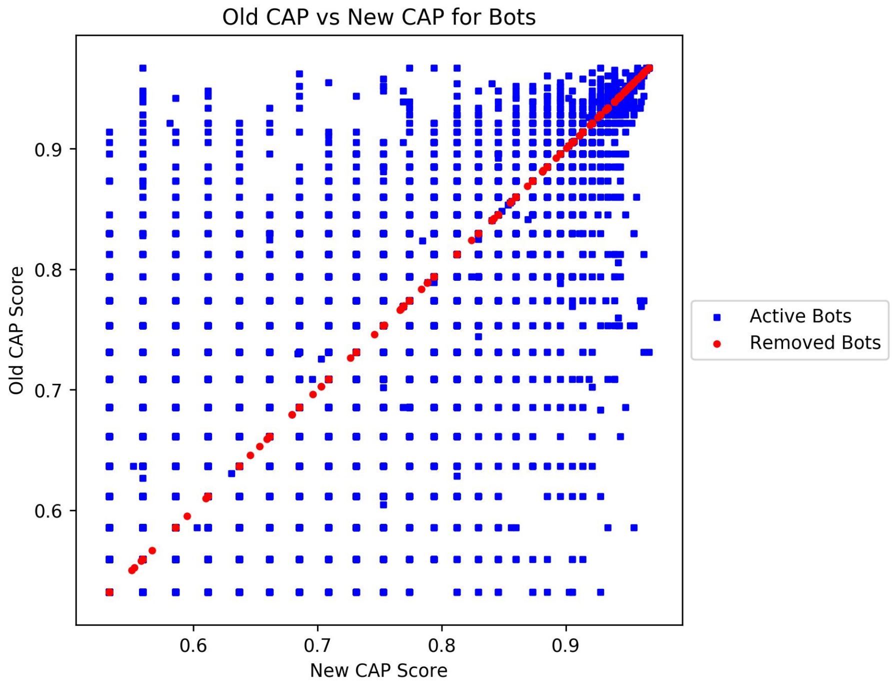
Creation of Training Datasets: Using Botometer

- Threshold determination based on established research [3]:
  - Human threshold: 0.4
  - Bot threshold: 0.53
- **21,418 accounts classified as bots** (avg CAP score 0.71)



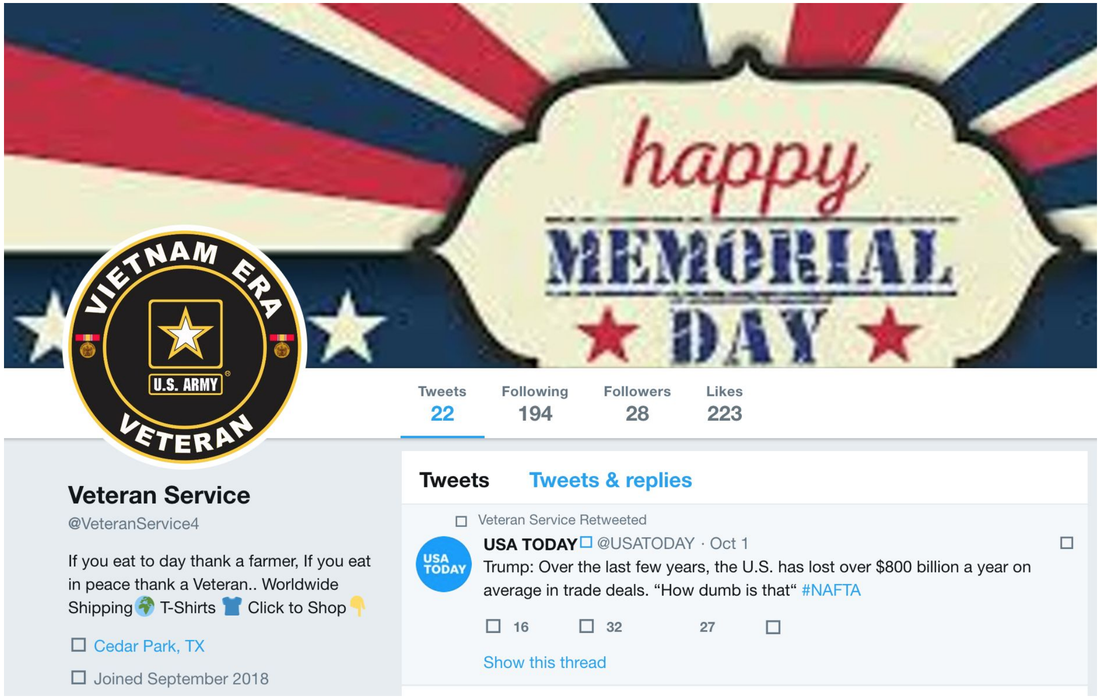
Creation of Training Datasets: Using Botometer

- Decided to *recheck* with Botometer:
  - **9,126 accounts (42.6%) *removed*** (avg CAP score 0.74)
  - **7,496 accounts remained classified as bots**
  - (3,378 accounts reclassified as humans)
  - (1,417 accounts fell between thresholds)
- **Training datasets:**
  - Once-classified bots (21,418)
  - Removed bots (9,126)
  - Twice-classified bots (7,496)



Detection Using Machine Learning

- Selection of account features for training: *limited to user profile information:*
  - ~~user id~~
  - favorites count
  - statuses count
  - description
  - ~~location~~
  - account creation date
  - verification status
  - ~~urls (account page, profile/ bkgd images)~~
  - listed count
  - followers count
  - default profile image
  - friends count
  - default profile
  - name
  - screen name
  - language
  - geo-enabled status



Tweets	Following	Followers	Likes
22	194	28	223

Veteran Service
@VeteranService4

If you eat to day thank a farmer, If you eat in peace thank a Veteran.. Worldwide Shipping 🌍 T-Shirts 👕 Click to Shop 👇

□ Cedar Park, TX

□ Joined September 2018

Tweets Tweets & replies

□ Veteran Service Retweeted

USA TODAY □ @USATODAY · Oct 1
Trump: Over the last few years, the U.S. has lost over $800 billion a year on average in trade deals. "How dumb is that" #NAFTA

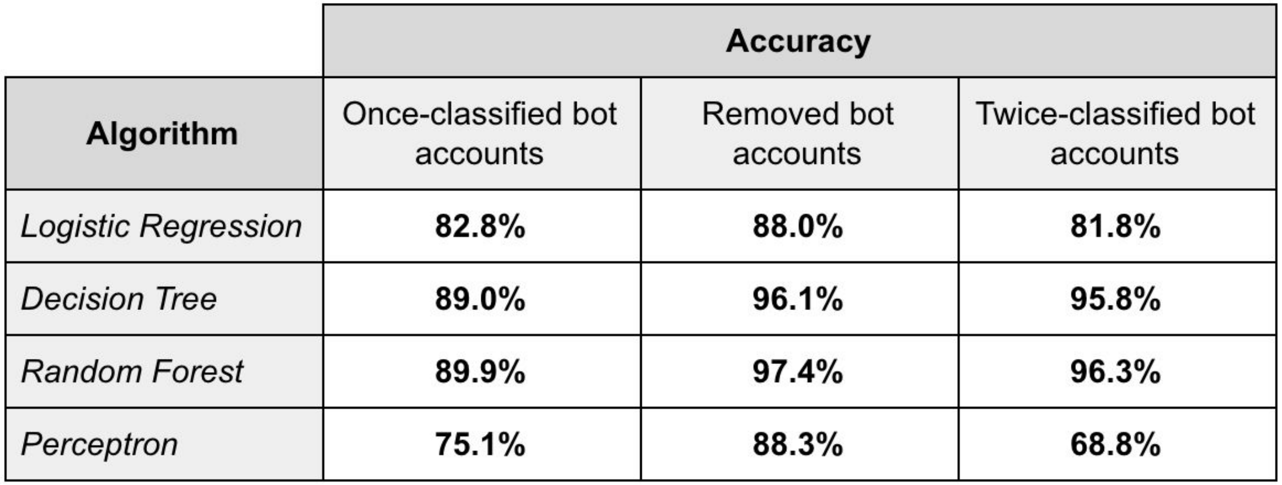
□ 16 □ 32 27 □

Show this thread

Detection Using Machine Learning: Results

- Machine learning algorithms:
  - Logistic Regression
  - Perceptron
  - Decision Tree
  - Random Forest

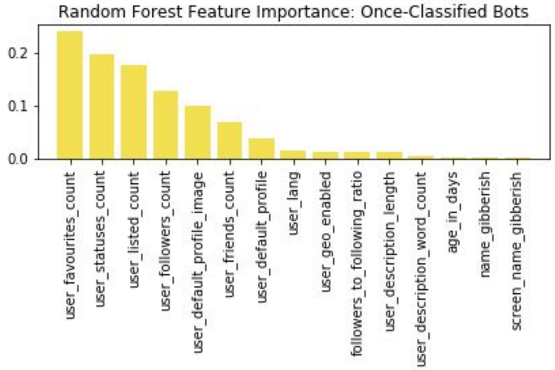
- Trained on:
  - Once-classified bots (n = 21,418)
  - Removed bots (n = 9,126)
  - Twice-classified bots (n = 7,496)

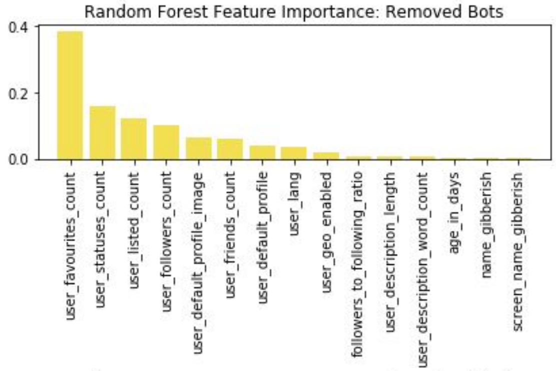


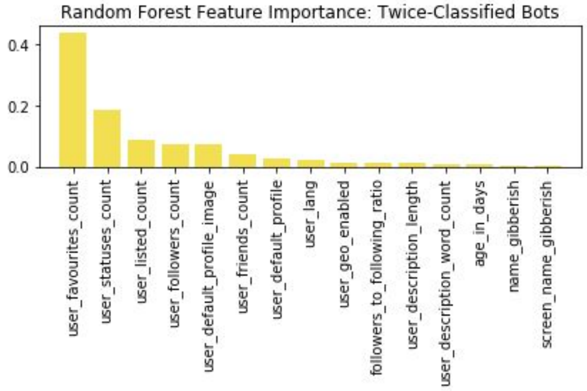
	Accuracy		
Algorithm	Once-classified bot accounts	Removed bot accounts	Twice-classified bot accounts
Logistic Regression	**82.8%**	**88.0%**	**81.8%**
Decision Tree	**89.0%**	**96.1%**	**95.8%**
Random Forest	**89.9%**	**97.4%**	**96.3%**
Perceptron	**75.1%**	**88.3%**	**68.8%**

Detection Using Machine Learning: Results

- Examined the bot predictiveness of each user profile feature
  - Specifically when using Random Forest model
- Most predictive: **number of tweets a user has favorited (liked)****
  - Human user avg: > 15,000
  - Bot avg: 3,100
- Other predictive features: **number of tweets, number of public lists, number of accounts following the user**


Random Forest Feature Importance: Once-Classified Bots


Random Forest Feature Importance: Removed Bots

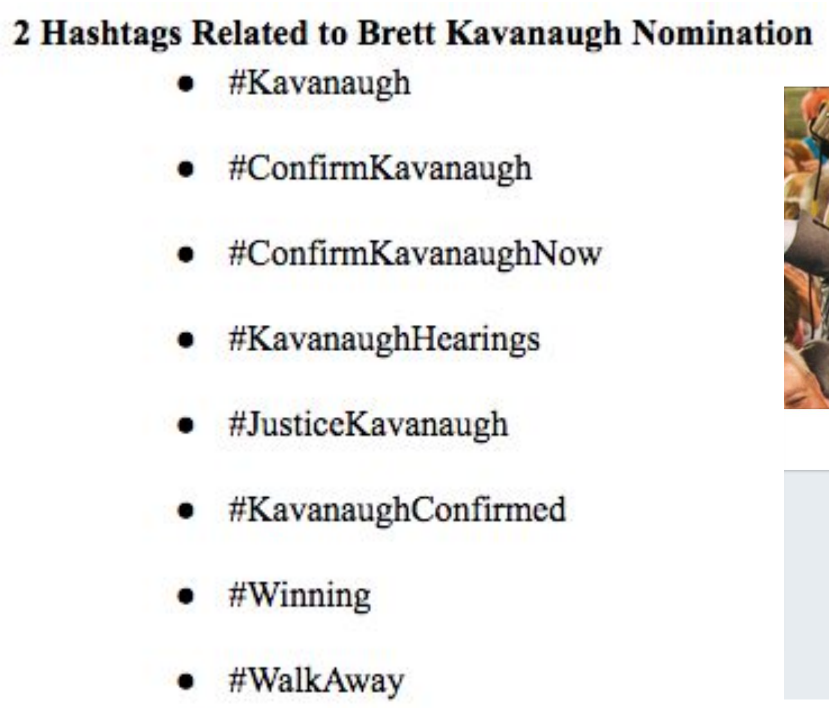

Random Forest Feature Importance: Twice-Classified Bots

Detection Using Machine Learning: Limitations

- Reliance on Botometer
- Gray zone between bot and human thresholds
- *Truth* not known
- Focus on political Twitter bots
- Research not conducted during a major election
- New data needs to be collected to keep our model up-to-date

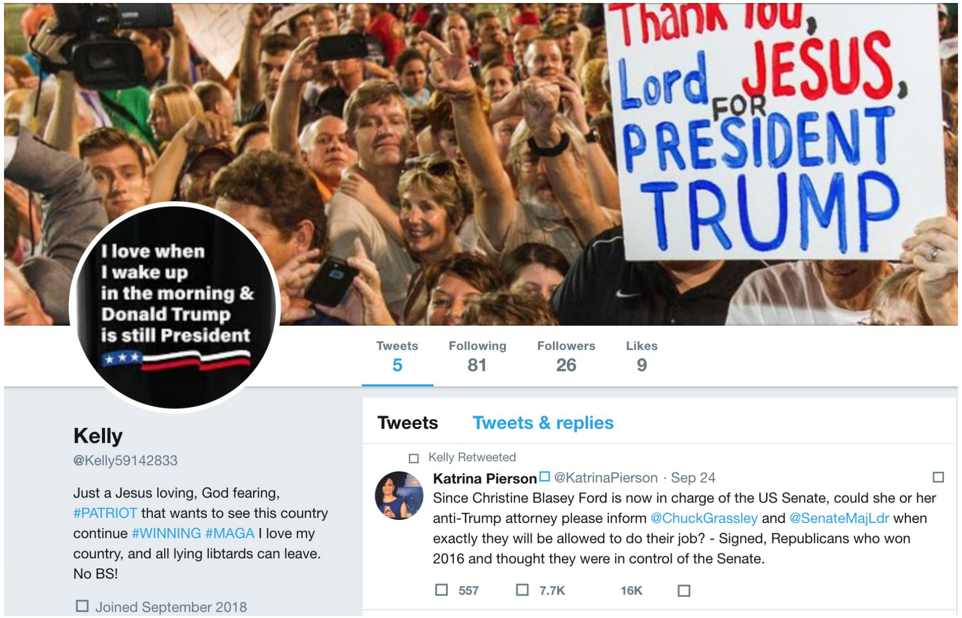
Bot Response to Political Events

- Political bots respond quickly to real-world political events
  - Shift in most popular political bot hashtags (*BotCheck.me* [2])



2 Hashtags Related to Brett Kavanaugh Nomination

- #Kavanaugh
- #ConfirmKavanaugh
- #ConfirmKavanaughNow
- #KavanaughHearings
- #JusticeKavanaugh
- #KavanaughConfirmed
- #Winning
- #WalkAway



Thank You, Lord JESUS, FOR PRESIDENT TRUMP
I love when I wake up in the morning & Donald Trump is still President

Tweets	Following	Followers	Likes
5	81	26	9

Kelly
@Kelly59142833

Just a Jesus loving, God fearing, #PATRIOT that wants to see this country continue #WINNING #MAGA I love my country, and all lying libtards can leave. No BS!

□ Joined September 2018

Tweets Tweets & replies

□ Kelly Retweeted

Katrina Pierson□ @KatrinaPierson · Sep 24 □

Since Christine Blasey Ford is now in charge of the US Senate, could she or her anti-Trump attorney please inform @ChuckGrassley and @SenateMajLdr when exactly they will be allowed to do their job? - Signed, Republicans who won 2016 and thought they were in control of the Senate.

□ 557 □ 7.7K 16K □

Bot Response to Political Events

Before Kavanaugh controversy:

- Bots accounted for 3.17% of all tweets
- Bots accounted for 2.45% of all actively tweeting accounts
- Average account ag of bots = 698 days

During Kavanaugh controversy:

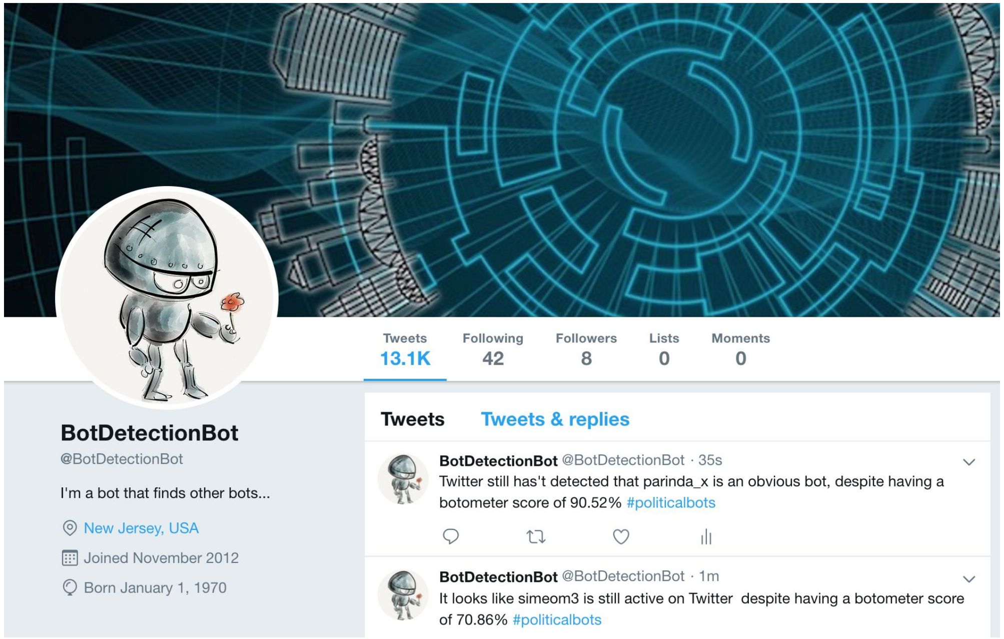
- Bots accounted for 4.02% of all tweets
- Bots accounted for 3.98% of all actively tweeting accounts
- Average account ag of bots = 613 days

What is Twitter doing about political bots?

- Twitter has become more active in its efforts to combat bot activity on its platform [4]
  - Also evidenced by our research
- Removed bot accounts had an average account age of 228 days - Twitter's response time?
  - Likely not
  - Authentic accounts can be hacked
  - Bot accounts may be dormant
- MORE can be done
  - Conflict of interest?

Creating our own Twitter bot

- *Informative bot*: shared the identity of political bots identified by our research
- Code to send automated tweets = simple
- Twitter's response:
  - Account fully suspended once
  - Account's ability to tweet revoked five times
  - QUICK response



Tweets 13.1K | Following 42 | Followers 8 | Lists 0 | Moments 0

BotDetectionBot
@BotDetectionBot

I'm a bot that finds other bots...

New Jersey, USA
Joined November 2012
Born January 1, 1970

Tweets | Tweets & replies

BotDetectionBot @BotDetectionBot · 35s
Twitter still has't detected that parinda_x is an obvious bot, despite having a botometer score of 90.52% #politicalbots

BotDetectionBot @BotDetectionBot · 1m
It looks like simeom3 is still active on Twitter despite having a botometer score of 70.86% #politicalbots

Conclusions

- Political bots comprised 3-8% of accounts tweeting on political hashtags from May-October 2018
- Designed a machine learning algorithm (with limited data) that achieves approximately 97.4% accuracy when classifying bots
- The number of tweets an account has favorited (liked) is a strong determinant of bot status
- Political bots respond in real-time to political events
- Twitter is making an effort to combat presence of political bots, but more can be done

https://techviral.net/write-a-twitter-bot-in-5-minutes/

References

1. Botometer by OSoMe. (n.d.). Retrieved September 24, 2018, from https://botometer.iuni.iu.edu/#!/faq#what-is-cap
2. Smiley, L. (2017, November 01). The College Kids Doing What Twitter Won't | Backchannel. Retrieved May 1, 2018, from https://www.wired.com/story/the-college-kids-doing-what-twitter-wont/
3. Pozzana, I., & Ferrara, E. (n.d.). Measuring bot and human behavioral dynamics. Retrieved June 1, 2018.
4. Russell, J. (2018, February 22). Twitter is (finally) cracking down on bots. Retrieved from https://techcrunch.com/2018/02/22/twitter-is-finally-cracking-down-on-bots/