



How Political Bots influence Brexit – A new way of classifying bots

Who am I?

Zeshan Aziz



CENTER for
COMPUTATIONAL
SCIENCE



Audience questions

Familiar with Brexit?

Heard of Cambridge Analytica?

Heard of IRA (Internet Research Agency)?

Heard of Aggregate IQ?

Iranian disinfo campaigns?

Chinese disinfo campaigns?

“Around the world, government actors are using social media to manufacture consensus, automate suppression, and undermine trust in the liberal national order.”

Oxford, Computational Propaganda Research Project

In a nutshell...

- Brainspace, product of Cyxtera
- Novel way of approaching disinformation and social media
- Beyond basic stats: ML and NLP
- Brexit was trending and good topic

Breakdown / Outline

- Why old methods don't work anymore
- Goal of Study
- Dataset composition
- Technology Stack
- Data Exploration
- Examples
- Models
- Process
- Findings
- What does this all mean?
- Effects on Canada's election
- The bigger picture
- Lessons learned
- Other research

Why historical ways of detecting bots don't work anymore

- Posting 24/7, or at odd hours
- Extremely high amount of interactions
- Only retweets, no original posts
- Obvious reverse image searchable profile pics
- Botometer (formerly Bot-or-Not) and other analysis tools
- Strange handles ending with too many numbers
- Marketing programs used to post tweets
- Don't account for sock puppets (need trained analyst)

Goal of study?

- To find bots discussing highly ideological political discourse on Twitter
- Do it in a way that others have not.

Dataset composition



Target: Teresa May Target (PM of the UK at the time)

- 4.8 Million tweets
- Scraped 100k of most recent followers (77k active accounts, 33k dormant accounts with no tweets)
- Tweet attributes
- Scraped and pulled up to the last 200 tweets from each account (including retweets), and which accounts they talked to



Brainspace™

Python Twitter API

For reach account, collect up to 200 of their tweets and mentioned accounts

for each follower gather their timeline (including retweets)

Store as a "pickle"

Python script that inputs pickles

strips out unnecessary data

formats result as a CSV file

Brainspace Input as a DAT

Certain fields mapped as if they were email fields

Brainspace does its magic organizing the data (search, NLP, ML, translation, etc)

Note: Twitter API essentially hostile to this sort of data gathering, with many restrictions and undocumented rate limits.

Data Exploration

The screenshot displays the Brainspace interface, which is used for data exploration. The top navigation bar includes the Brainspace logo, a 'Brainspace Demo' dropdown, and two selection menus: 'Select a Focus' and 'Select CMMML Classifier'. On the right, there are tabs for 'Analytics' and 'Notebook'.

The main content area is titled 'Expanded Concept' and features a search bar with the text 'wine'. Below the search bar, there is a grid of sliders for various terms, each with a white dot indicating its current value. The terms and their approximate slider positions are as follows:

Term	Approximate Slider Position (0 to 100)
wine	75
oakville	25
take-out	50
syrah	50
tasting	55
raspberries	55
grenache	70
pinot noir	70
domaine	85
gary vaynerchuk	85

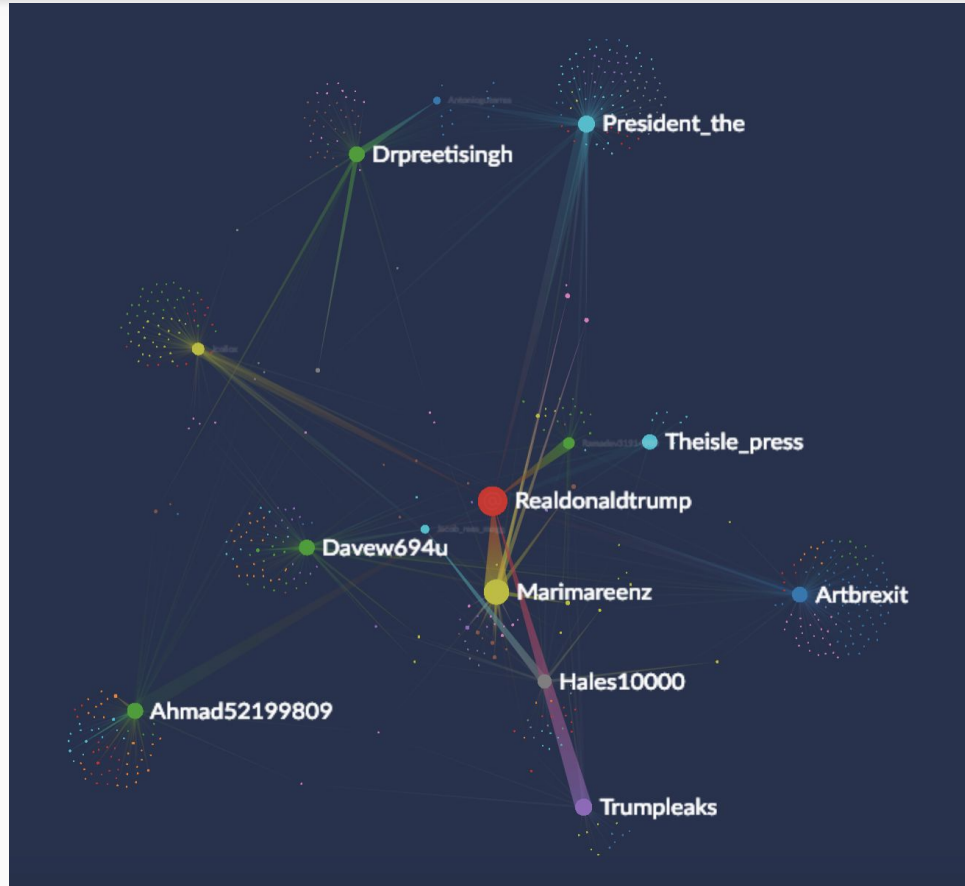
Below the sliders, there is a 'Reset Concept' button on the left and a 'Load More Terms' button on the right. At the bottom of the interface, there is a 'Collapse' button with up and down arrows.

Analysis based purely on content and behavior

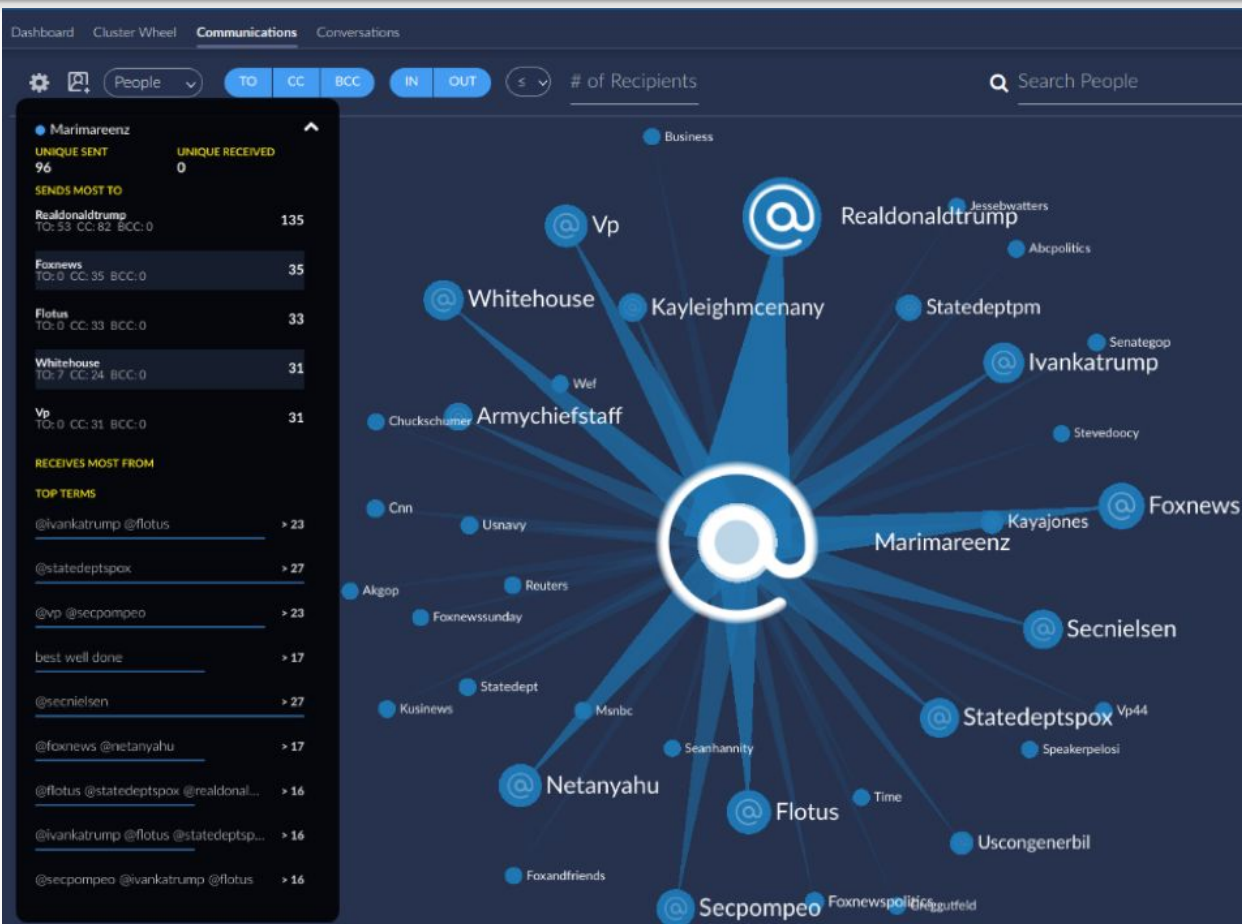
- Bots must push out tweets in order to stay relevant
- Bots must behave in a manner similar to marketing bots

Pro-Trump bots

accounts
tweeting the
most at Trump



Bot example



 **GEN Mariy Mareenz**   
@MariMareenz

My name is Mariy Mareenz, Im a woman, Im forty five years old, Im From America. Im a Judge of Anchorage America now. I Love You Sisster Ivanka Real Turmp

United States Joined September 2018

2,279 Following 331 Followers

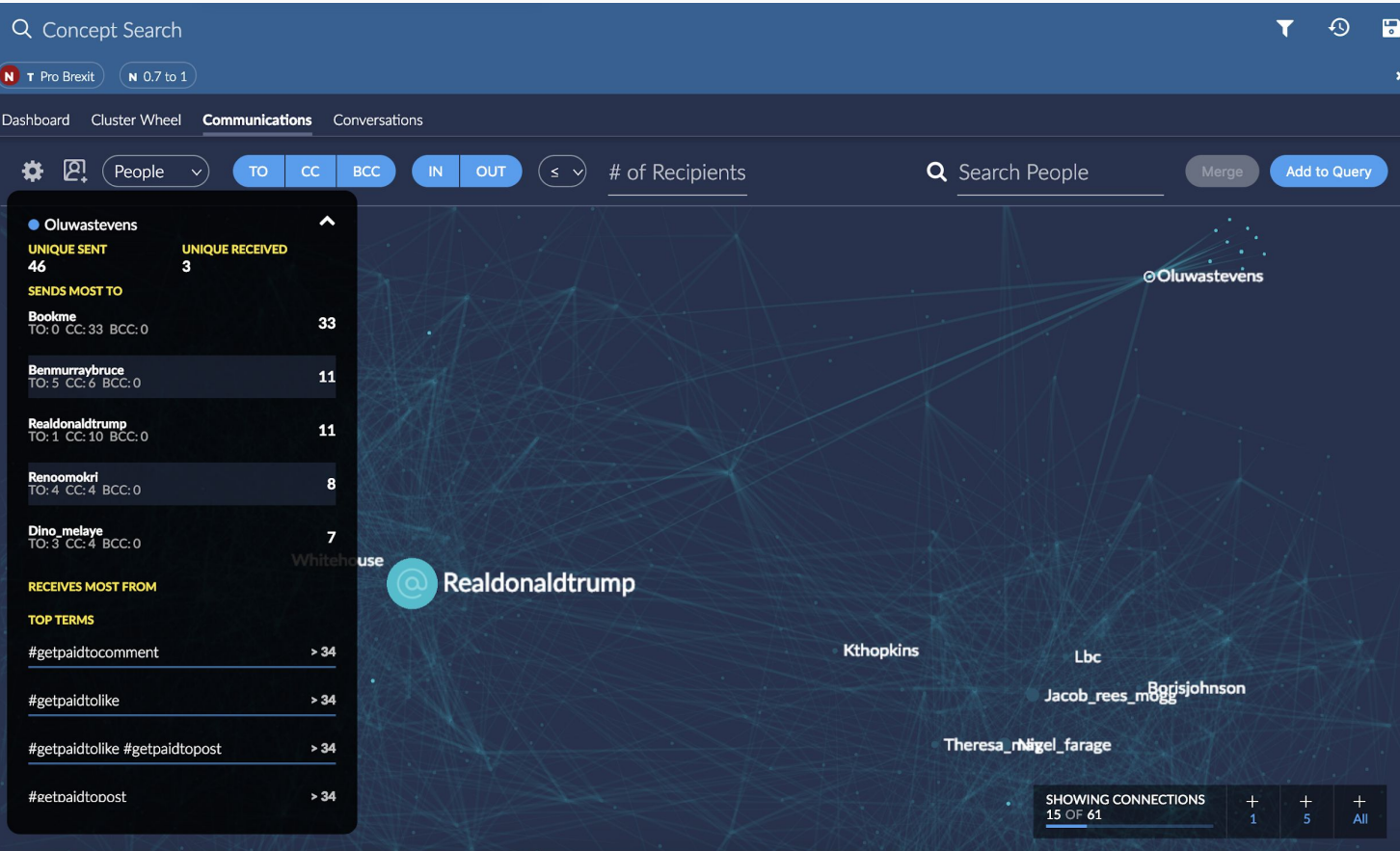
 **Sarhang GEN Malcolem**  
@PrinceSarhang

My name is Sarhang, im a son, Im twenty eight years old, im from America. I Love You Real D.Trump

Washington, DC Born September 25, 1990 Joined September 2018

1,579 Following 390 Followers

Screenshot of Brainspace



Model methodology for Content-based sentiment analysis

Pro-Brexit	Anti-Brexit
Pro-Trump	Anti-Trump

Trained 55~90 tweets per ideology into each model.

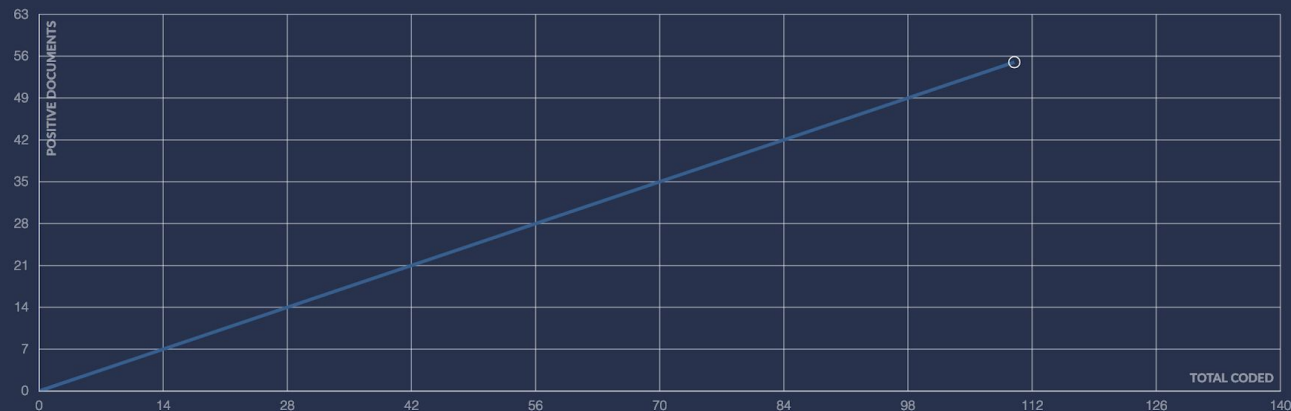
The screenshot displays the CMML (Content-based Machine Learning) interface for two trained models. The top model is 'Pro Brexit' (ID: 33), which has completed 1 training round with 55 positive and 55 negative documents. The bottom model is 'Pro Trump' (ID: 36), which has completed 3 training rounds with 88 positive and 92 negative documents. Both models are in a 'Ready' status, indicated by a green dot. The interface includes a 'CMML' label, training details, a 'POSITIVE' button with the model's name, and a status bar at the bottom with icons for a bar chart, settings, and a trash can.

Model Name	ID	Training Rounds	Positive Documents	Negative Documents	Status
Pro Brexit	33	1	55	55	Ready
Pro Trump	36	3	88	92	Ready

CMML (Continuous Multimodal Learning)

Classifiers » Pro Brexit (ID:33) » Pro Brexit : yes [HELP](#)

Continuous Multimodal Learning Progress



CODED
110

POSITIVE DOCS
55

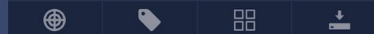
NEGATIVE DOCS
55

Training Rounds

TRAINING		Create Validation Set		New Training Round		CODED	POSITIVE	NEGATIVE	CONSISTENCY	
ROUND 1	• Net 110 documents tagged manually.					110	55	55	100%	

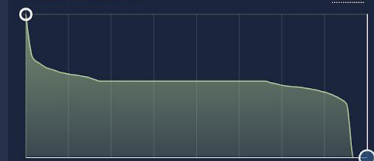
4,809,559

RESULTS



PREDICTIVE RANK

TRIM



Show only Untagged Documents ☒

ORIGINAL

[1081489154234638336](#)

January 5, 2019 - @AnEnglishRevolt Boot most of em out. Any that stay need electronic tags or microchipping, private i ...

ORIGINAL

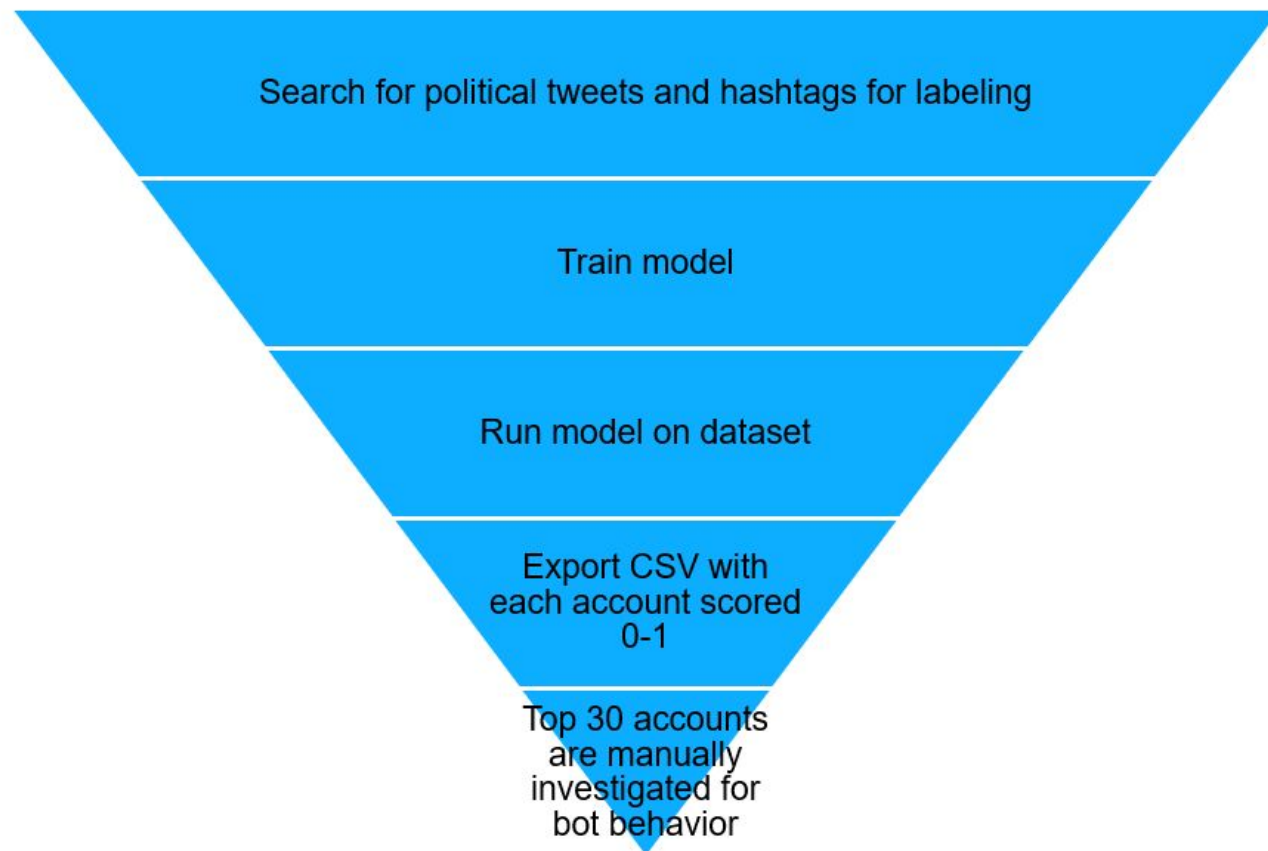
[1085131408404897793](#)

January 15, 2019 - @supercooper216 @markcartline44 a lot of people in Surrey and England (probably most leavers) would I ...

ORIGINAL

[1082302485094690816](#)

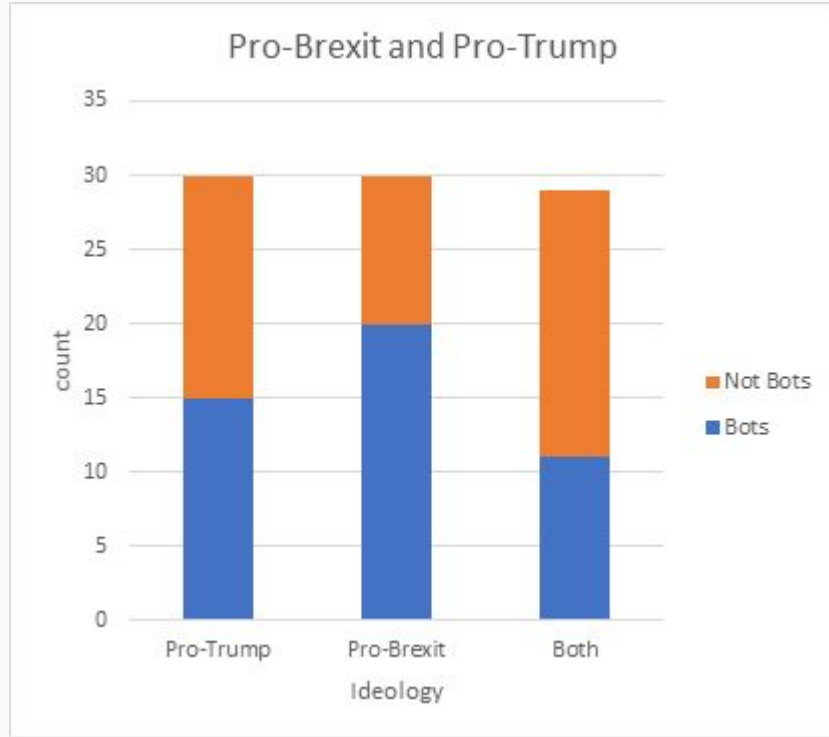
January 7, 2019 - @ThomasPilchard fake news. EUSSR nationals need to be deported on 30th March then reapply for leave ...



Tools used for manual analysis

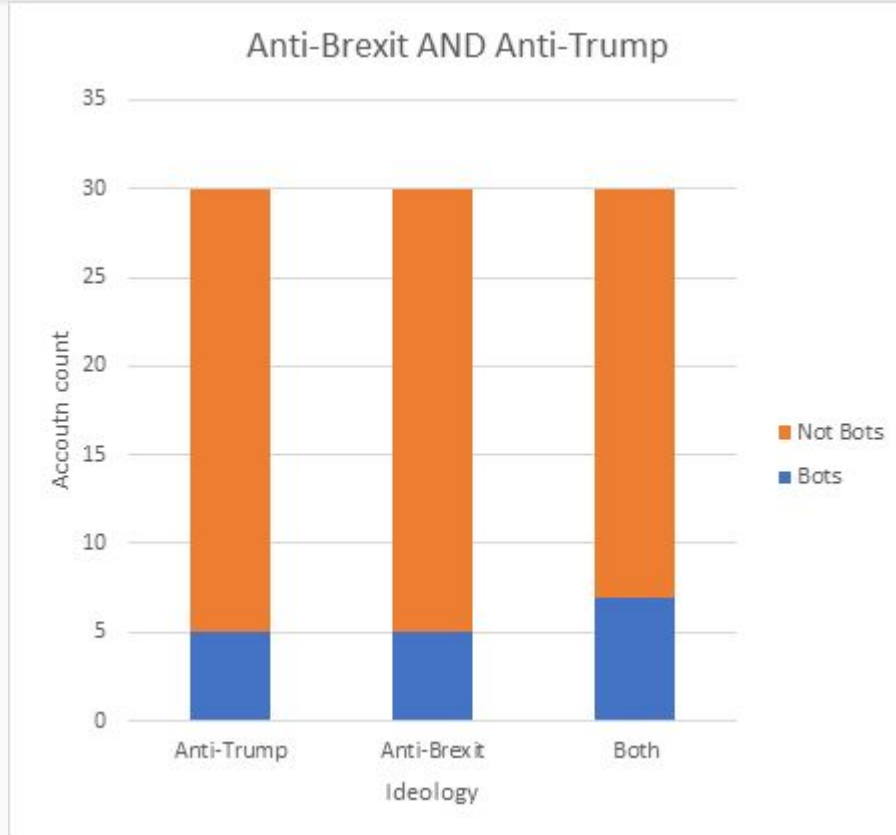
- Yandex / Google Reverse Image Search
- iseek.com/iseek (ideal for human targets with unique names)
- Botometer (formally bot-or-not)
- Smell test

Result: Pro-Brexit + Pro-Trump



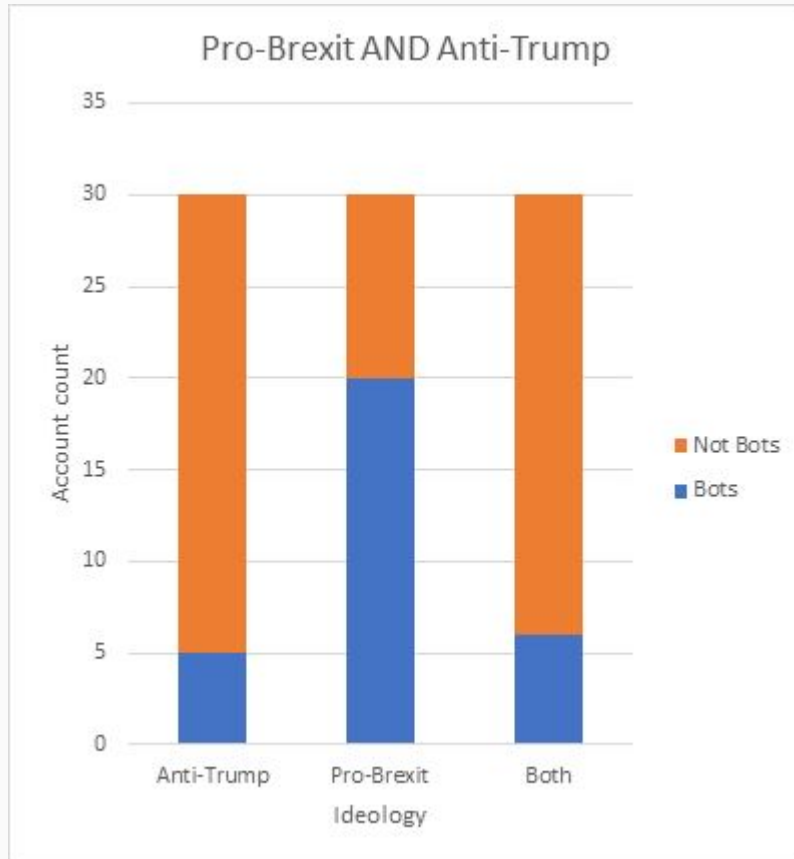
- Both: 38% of accounts are bots
- Pro-Trump: 50% are bots.
- Pro-Brexit: 66% are bots.

Result: Anti-Brexit + Anti-Trump



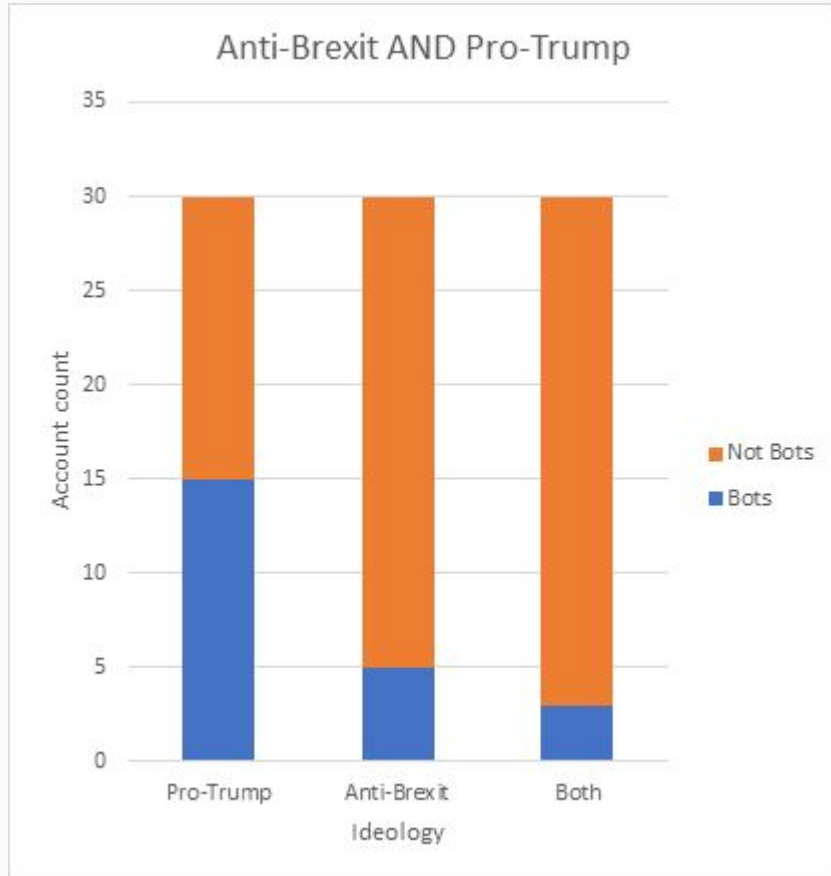
- Both: 23% are bots
- Anti-Trump: 17% are bots
- Anti-Brexit: 17% are bots

Result: Pro-Brexit + Anti-Trump



- Both: 23% are bots
- Pro-Brexit: 66% are bots
- Anti-Trump: 16% are bots

Result: Anti-Brexit + Pro-Trump



- Both: 10% are bots
- Pro-Trump: 50% are bots
- Anti-Brexit: 16% are bots

Botometer®

An OSoMe project (bot•o•meter)



- Botometer agreed with our manual analysis 78% of the time.
- When it failed it invariably thought that an account was a human, while in fact it was a bot.

What does this all mean?

- Pro-Brexit and Pro-Trump accounts are driven heavily by bots
- Bots push strange combinations of ideologies (may imply reuse of accounts for different campaigns? Needs additional research)
- Humans hold multiple ideologies instead of singular focuses?
- The good bots are getting more sophisticated and harder to find (so we may be underestimating what we are seeing)

Twitter needs to be more responsible ahead of election, Gould says



By Shruti Shekar @shruti_shekar JUN 7, 2019 | 9:10 AM EDT | 0 COMMENTS

Canada Declaration on Electoral Integrity Online

Social media and other online platforms play a meaningful role in promoting a healthy and resilient democracy provide a forum for sharing ideas, encouraging innovation, enabling Canadians to engage in dialogue about important issues and facilitating economic opportunity and civic engagement. However, these platforms have also been used to spread disinformation in an attempt to undermine free and fair elections and core democratic institutions and aggravate existing societal tensions.

Social media and other online platforms and the Government of Canada will help safeguard this fall's election and to support healthy political discourse by working together to address these challenges head-on.

To this end, we commit to working together in the lead-up to the October 21 election to ensure transparency and authenticity, subject to Canadian laws and consistent with the principles of electoral integrity.

Integrity

Platforms will:

1. Intensify efforts to combat disinformation to promote transparency and efforts to safeguard the Internet ecosystem.

VANCOUVER

Bots, hackers and trolls: How the digital misinformation war is being waged ahead of the federal election



By Perrin Grauer Star Vancouver
Melanie Green Star Vancouver
Wed., April 10, 2019 | 5 min. read

Facebook Won't Remove Doctored Content During The Canadian Election

Election authorities could still investigate false or misleading content like doctored videos, but the social media giant said it's not its role to decide where the "line" is.



Jane Lytvynenko
BuzzFeed News Reporter



Alex Boutilier
Toronto Star Reporter



Craig Silverman
BuzzFeed News Reporter

Posted on May 28, 2019, at 5:00 p.m. ET

Post Election research



Marc Owen Jones

@marcowenjones

[Thread] 1/ I have compiled a lot of data over the past month on various Trudeau hashtags. In this analysis I compiled samples from the past 1.5 months. In the following analysis, you will see pro-Trump accounts still dominating anti-Trudeau hashtags [#cdnelxn2019](#) [#cdnpoli](#)

5:08 PM · Oct 13, 2019 · [Twitter Web Client](#)

1K Retweets 1.3K Likes

FIRST
DRAFT

First Draft ✓

@firstdraftnews

Follow

NEW: Hundreds of newly created and pro-Trump Twitter accounts rallied behind Boris Johnson on Thursday in activity deemed "suspicious" by a number of experts familiar with online manipulation.



Pro-Trump Twitter accounts boost Boris Johnson Brexit message

Hundreds of newly created and pro-Trump Twitter accounts rallied behind Boris Johnson on Thursday in activity deemed "suspicious" by a number of experts familiar with online manipulation.

firstdraftnews.org

5:29 AM - 19 Oct 2019

China - Canada tweets

Huawei Chief Financial Officer - Meng Wanzhou

Brainspace | theresa_may_1549656163_ | Select a Focus | Select CMMML Classifier | Analytics | Notebooks | Supervised Learning

Q Concept Search

M chinese C 孟晚舟

Dashboard Cluster Wheel Communications Conversations

People TO CC BCC IN OUT # of Recipients

Search People Merge Add to Query

Lovegaomeng

UNIQUE SENT 11

UNIQUE RECEIVED 0

SENDS MOST TO

Justintrudeau TO: 10 CC: 10 BCC: 0 20

Voachinese TO: 1 CC: 1 BCC: 0 2

Grcrcmpolice TO: 0 CC: 1 BCC: 0 1

Casaacae TO: 0 CC: 1 BCC: 0 1

Mclaudelibeau TO: 0 CC: 1 BCC: 0 1

RECEIVES MOST FROM

TOP TERMS

中国公民孟晚舟 > 9

中国公民 > 9

程放 > 9

晚舟 > 9

公民 > 9

无条件 > 4

@justintrudeau > 9

中国 > 10

@justintrudeau @peterschiefke @casa... > 1

Grccrcmpolice

Casaacae

Jeanhenrycant

Toymountain

Emmanueldubourg

Peterschiefke

Mclaudelibeau

Clvtoronto

Lovegaomeng

Justintrudeau

392 RESULTS

RELEVANCY DISTRIBUTION

TRIM

ORIGINAL

No Title

January 13, 2019 · 《建民论推增452》华为波兰经理

爱梦 @Lovegaomeng

Replying to @JustinTrudeau

立即释放中国公民孟晚舟!

Translated by Google

Immediate release of Chinese citizen Meng night boat!

3:07 AM · Dec 21, 2018 · Twitter for Android

December 12, 2018 · 美国驻加拿大大使Kelly Craft: “中共官员把孟晚舟被捕事件报道成是大学与华盛顿一手策划的政治阴谋，这绝对是对孟晚舟1孟的逮捕与引渡本来是独立司法程序的一部分，连我都不能随便评论” | https://...

ORIGINAL

No Title

January 14, 2019 · 【人质外交很流氓】之前中共在港、东南亚、澳及加等国越境执法，绑架通缉国公民，今习外交又有创新，开始在党国境内假法律之名抓捕外国公民做人质来报复西方民主国家。孟晚舟案发生后，中共抓捕了至少13名加国公民...

ORIGINAL

SHOWING CONNECTIONS 10 OF 10

Possible threat actors/ Attribution?

China, Russia, Iran, India, Private Companies, Political parties, Wakanda, etc. (so massive we don't even know)

Threat = capability + intent. Intent = everyone. Capability = gets easier every year. Capabilities can be purchased with Propaganda As A Service (PaaS)

Attribution is difficult without backend data

Russia in Africa Angle?



Russia Tests New Disinformation Tactics in Africa to Expand Influence

Facebook said it removed three Russian-backed influence networks aimed at African countries. The activity by the networks suggested Russia's approach was evolving.



Bad News



Carole Cadwalladr @carolecadwalla · Nov 4

NEW: @guardian reveals Boris Johnson has now **BLOCKED** release of report into Russian interference ahead of election. Chair of committee below explains he has done so on basis of a lie. Just what exactly does he have to hide? theguardian.com/politics/2019/...



Carole Cadwalladr @carolecadwalla · Nov 2

Wow. Extraordinary interview with Conservative MP Dominic Grieve on report Boris Johnson has blocked on Russian interference. 'What Number 10 has said is not true. It's a lie'



298

5.7K

8.6K





We've made the decision to stop all political advertising on Twitter globally. We believe political message reach should be earned, not bought. Why? A few reasons... 🧵

4:05 PM · Oct 30, 2019 · [Twitter for iPhone](#)

104.5K Retweets **430.5K** Likes

🕒 This article is more than **3 years old**

Twitter deletes 125,000 Isis accounts and expands anti-terror teams

Building confidence in its anti-terror policies, the social media company is expanding its specialist teams in the US and Ireland to monitor extremist content



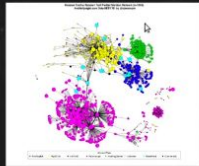
- Evidence management with Hunchly
- Use dedicated VM like Buscador for investigation: protects your OPSEC, makes workflow easier/faster. Build your own tools for automation.
- Use clean Twitter sock puppets for investigations. My account got blocked by a few targets. You want twitter to recommend you other susicious accounts based off what AI magic happens with their recommendation algo.
- Malicious actors are constantly evolving their TTPs. They read the news too. Don't expect the job to ever be done.



Other capabilities we have but did not use for this project

More analysis of Clemson Dataset

- <https://twitter.com/Capn00k/status/1025386967081977856> (machine learning using QUID)
- <https://twitter.com/csmarcum/status/1025069589053497345> (nice relationship mapping)
- https://russian-troll-tweets-dataset.com/russian-troll-tweets-debc91f/tweets?facet=account_category&facet=author_facet_region (nice search interface)
- <https://www.fromrussiaythetroll.com/>
- <http://beebe-west.com/viz/russian-trolls/> (interesting visualization)



Generally public efforts have EITHER Visualization OR Search but not both.

Walkthrough 2 - Finding Russian Twitter Bots with Brainspace

1 year ago | More



Immunity Videos

PRO

+ Follow

More from Immunity Videos

Autoplay next video

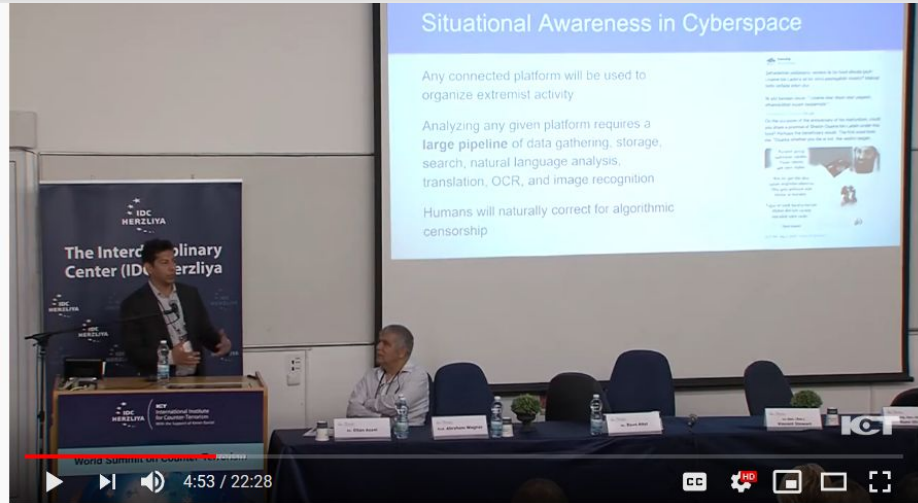


Wal

Cyxtera Blog: Brainspace

Written by Dave Aitel | Acknowledgment: William Grayson Hilliard - Cyxtera Data Engineering on August 08, 2018

Data Carving the Internet Research Agency Tweets



Mr. Dave Aitel - Workshop - The Steven E. Stern workshop on Cyber-Terrorism.

40 views • Oct 23, 2019

1 0 SHARE SAVE ...

SUBSCRIBE

Other good research to check out

[bellingcat.com](https://www.bellingcat.com)

[bellingcat.com/news/2019/09/03/twitter-analysis-identifying-a-pro-indonesian-propaganda-bot-network/](https://www.bellingcat.com/news/2019/09/03/twitter-analysis-identifying-a-pro-indonesian-propaganda-bot-network/)

firstdraftnews.org/

twitter.com/thegrugq

cyber.fsi.stanford.edu/

<https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/hunting-threats-on-twitter>

Contact me

Slides will be up on zeshanaziz.com in a few days

Twitter: [@LockheedMartini](https://twitter.com/LockheedMartini)

My email: me@zeshanaziz.com

[Linkedin.com/in/nahsez](https://www.linkedin.com/in/nahsez)

Or pull me aside at this conference