



Measuring **change** and **diversity** in Twitter accounts helps identify suspicious activity.

Twitter

- 368 million users
- 65 million users in US
- A 2017 study estimated that 9–15% of active Twitter accounts were bots



Dataset

- The datasets are generated through
 - Twitter API
 - Web Scraping
- Each dataset include
 - userIds.txt
 - tweets.jsons.gz

Account Name	Bot Type	Count	Human Type	Count
astroturf	political_Bot	585		
botwiki	bot	704		
cresci-17	bot-socialspam	4912	human	3474
	bot-fakefollower	3351		
	bot-traditionspam	2631		
verified			verified_account	2000
stock	financial_bot	18508	financial_account	7479
pronbots	Adult Content_scam	21964		
caverlee	bot	22179	Legit_human	19276
midterm-2018			political_account	8092
gregory_purchased	bot	1088		
gilani-17	bot	1130	human	1522
josh_political	bot	62		
kevin_feedback	bot	143	human	386
rtbust	bot	391	human	368
varol-icwsm	bot	826	human	1747
Total		75974		36844

users.txt

1	145510451	bot
2	45705944	human
3	70458968	bot
4	19480741	bot
5	105389441	bot
6	82055095	bot
7	35975529	human
8	25954484	human
9	41526294	human
10	38026425	bot
11	20825825	human
12	89955122	bot
13	123334275	bot
14	82683680	bot
15	83459702	human
16	80796893	human
17	8561122	human
18	82692395	bot
19	75108938	bot
20	21561557	bot
21	62892022	bot
22	12168052	human
23	103475229	bot
24	25754021	human
25	41050745	human

[tweets.json.gz](#)

```
1 2312443350 [{"contributors": null, "truncated": false, "text": "RT @WBridgeWtrMAGOP: Cruz, Lee Push Amendment to GOP Bill to Give Insurers M
2 1865011970 [{"contributors": null, "truncated": false, "text": "1 new unfollower and 4 new followers (hello! hello!) in the last week. Via h
3 701329338 [{"contributors": null, "truncated": false, "text": "RT @MAGOPProject: Dump Elizabeth Warren #MApoli #MAasen \n Always #VoteGOP #NoI
4 2318983652 [{"contributors": null, "truncated": false, "text": "Native Americans Demand Proof Of Lizzie Warren\u2019s Phony \u2020Heritage\u2020
5 1551216937 [{"contributors": null, "truncated": false, "text": "Meet John Hugo, December 13th in Holliston https://t.co/Wy2T1Icmjy https://t
6 2244754568 [{"contributors": null, "truncated": false, "text": "RT @PolToons: #GlobalWarming Is Not Rocket Science @afbranco #JunkScience #De
7 2214071510 [{"contributors": null, "truncated": false, "text": "RT @PolToons: Taking a Knee @afbranco #BoycottNFL #BoycottNFLSponsors #TakeA
8 2209122018 [{"contributors": null, "truncated": false, "text": "RT @MAGOPProject: Cherokee genealogist casts doubt on Liz Warren's claim of p
9 2163541567 [{"contributors": null, "truncated": false, "text": "RT @PolToons: #DemocratsBeingDemocrats #DrainTheSwamp #DemocratLiesMatter Al
10 2340420192 [{"contributors": null, "truncated": false, "text": "RT @YannisMAGOP: Trump Administration Not Taking Up 2020 Census Recommendat
11 625986322 [{"contributors": null, "truncated": false, "text": "https://t.co/por3BWxhiF", "is_quote_status": false, "in_reply_to_status_id":
12 2412859567 [{"contributors": null, "truncated": false, "text": "Right Path Forward #DrainTheSwamp #DemocratLiesMatter Always #VoteGOP #NoRIN
13 2267900545 [{"contributors": null, "truncated": false, "text": "Shiva Loses Case About Claim of Being The Inventor of Email #MApoli https://
14 2207395363 [{"contributors": null, "truncated": false, "text": "RT @PolToons: Still Pathetic #ObamasLegacy #DrainTheSwamp #DemocratLiesMatte
15 1135150676 [{"contributors": null, "truncated": false, "text": "RT @PolToons: Beware #GlobalWarming @Ramireztoons #ClimateChange #HappyNewYe
16 2335569980 [{"contributors": null, "truncated": false, "text": "Thank You David DeCoste for standing up for Advocates of Autism. https://t.co
17 2166582637 [{"contributors": null, "truncated": false, "text": "I gained 5 followers in the last week. You? Know your stats and grow with ht
18 821979836 [{"contributors": null, "truncated": false, "text": "RT @PolToons: Lyin' Lizzie is a Few Feathers Short of a Headdress @afbranco
19 626005201 [{"contributors": null, "truncated": false, "text": "I have a rising follower graph! 4 more followers in the past week. Get your
20 1135293558 [{"contributors": null, "truncated": false, "text": "RT @RealJamesWoods: \u2020If you like your doctor...\u2020id https://t.co/7y93
21 637314961 [{"contributors": null, "truncated": false, "text": "RT @FraminghamMAGOP: Fed Up: A New National Poll Delivers a Stark Warning to
22 910905116 [{"contributors": null, "truncated": false, "text": "RT @DiamondandSilk: Just so you know; We have not communicated nor have we ta
23 2180467796 [{"contributors": null, "truncated": false, "text": "IG Report Roasts Andrew McCabe https://t.co/c9Cojjsia7", "is_quote_status":
24 2203337701 [{"contributors": null, "truncated": false, "text": "RT @MAGOPProject: Boycott hits MSNBC & CNN \"hate\" shows #BoycottMSM #Fal
25 2244325512 [{"contributors": null, "truncated": false, "text": "4 new followers in the last week and it is more than just stats, I use it fo
26 2244229052 [{"contributors": null, "truncated": false, "text": "The House\u2020s Collusion Investigation Ends https://t.co/YuZnGWXGgn", "is_
27 2294603191 [{"contributors": null, "truncated": true, "text": "The state\u2020s top election official - a Democrat - went off the rails with
```

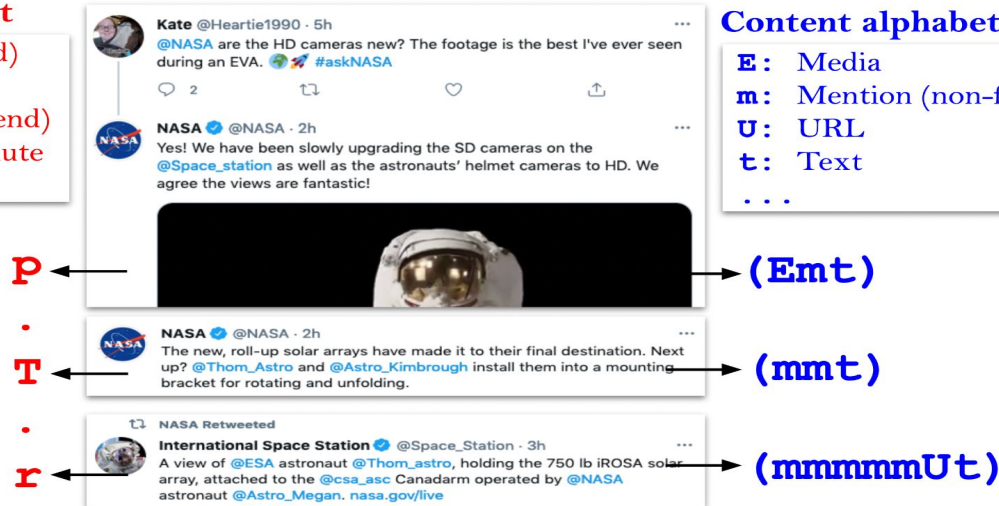
Behavioral Languages for Online Characterization (BLOC)

Action alphabet

p: Reply (non-friend)
T: Tweet
r: Retweet (non-friend)
.: Pause over a minute
...

Content alphabet

E: Media
m: Mention (non-friend)
U: URL
t: Text
...



Ref : BLOC repository

BLOC string - P.T.R (Emt).(mmt).(mmmmmmUt)

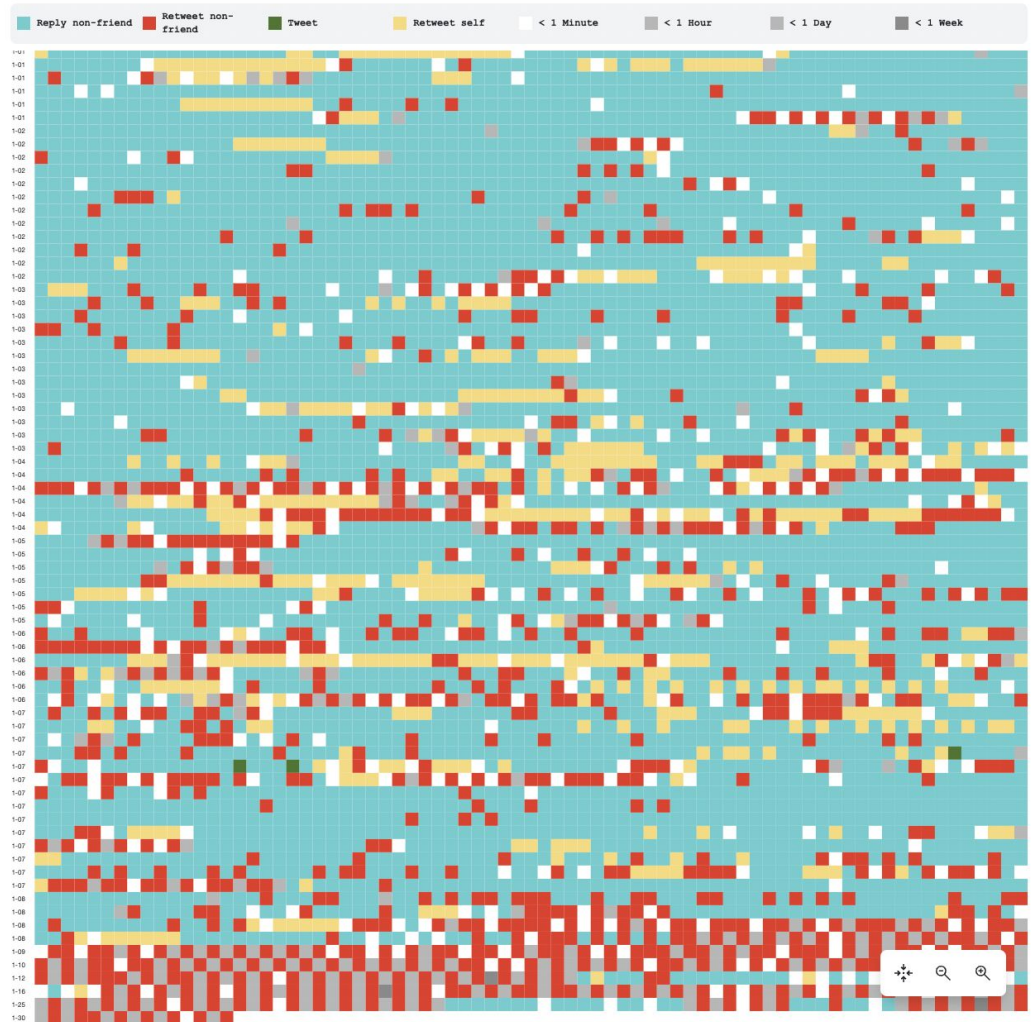
Low Change

- CNN News twitter account
- Ref - [agwagram](#)
- Similar Behaviour
- Bot-like behaviour
- Automated for do specific task



High Change

- @CassidyRae1414, user id:2648502412
- Ref - [agwagram](#)
- Behaviour changes frequently
- US presidential day(2018) has a high change



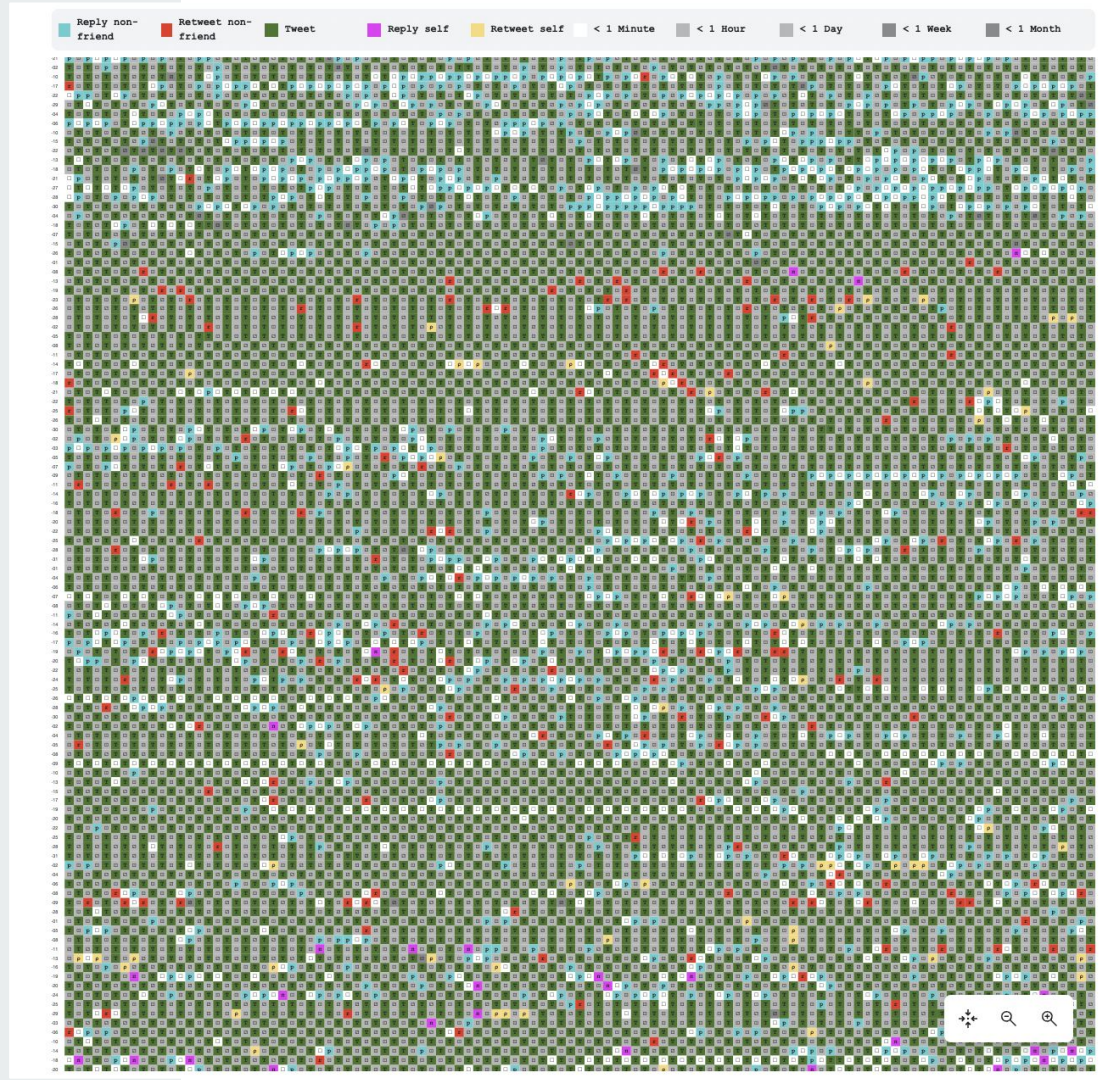
low Diversity

- CNN News twitter account
- Ref - [agwagram](#)
- Number of behaviour changes are low



High Diversity

- @TEN_GOP (Tennessee)
- Ref - [agwagram](#)
- Number of behaviour changes are high





Research Questions

- Identifying the correlation between high change and low change with the bot-like behavior.
- Identifying the correlation between high diversity and low diversity with the bot-like behavior.

Thanks!



Extra Slides

- Segmented bloc string -
<https://docs.google.com/document/d/1gXZpO706hGwevQ8-IR2Cky15nBRz6qbMgFrZZnrWB1o/edit?usp=sharing>



User Account

```
"bloc": {  
  "action": "TTTppppppp | TTTTTTTTTppp | TT | Tp | T | Tppp | T | T | T | T | T | T | T | r ",  
  "content_syntactic": "(t)(Ht)(Ht)(HHt)(t)(t)(t)(Ht)(Ht) |  
(Ht)(Ht)(mt)(mt)(mt)(t)(t)(Ht)(Ht)(Ht)(mmmt) | (t)(Ht) | (mt)(Hmt) | (mt) | (t)(t)(t)(mt) | (mt) | (mt) |  
(HmUt) | (mmt) | (mt) | (mt) | (m\u03c6) | (mqt) | (Hqt) "  
},
```




Features

Changes in actions = [a1, a2, a3, a4, a5]

Changes in content = [c1, c2, c3, c4, c5]

Features

- Total change of actions/content
- Mean change of actions/content
- Standard deviation of actions/content
- Highest change between weeks for actions/content
- Lowest change between weeks for actions/content
- Highest number of actions/content per week
- Lowest number of actions/content per week
- Average number of actions/content per week

Target Variable

- Bot/Human



Examples

- Low change - CNN news account
- High change - russian troll account (presidential election)
- Low Diversity - CNN news account
- High Diversity - russian troll account (change its behavior on specific time periods)

