Weaponising 280 characters

What 200,000 tweets and 4,000 bots tell us about state of Twitter in Sri Lanka

Sanjana Hattotuwa, Yudhanjaya Wijeratne and Raymond Serrato

Weaponising 280 characters

What 200,000 tweets and 4,000 bots tell us about state of Twitter in Sri Lanka

by Sanjana Hattotuwa, Yudhanjaya Wijeratne and Raymond Serrato

21 April 2018

Summary

Starting late March, *Groundviews* and other Twitter users in Sri Lanka began noticing a tsunami of Twitter accounts with no bios, no tweets and the default profile picture following them.

Immediately evident were interesting characteristics, not unlike the bots and fake followers anchored to account of Namal Rajapaksa. The bots had Sinhala, Muslim and Tamil sounding names. Many of the profiles were the default Twitter profile image, but many of the female accounts had photos lifted from public profiles of other individuals. Given the scale and scope of the infestation, *Groundviews*, for the first time in the Sri Lankan Twittersphere, took the step of making public its block list, which other users could import. Even this measure though was not enough to address the high frequency with which new, fake accounts were being created, attaching themselves to prominent Twitter users in Sri Lanka.

Preliminary analysis of 1,262 accounts, a subset of the larger dataset we were working with, indicated that the majority of suspicious accounts following Twitter users were bots. A visualisation of the number of accounts targeted by the bots revealed that leading diplomats, Ambassadors based in Sri Lanka, the official accounts of diplomatic missions, leading local politicians, the former President of the Maldives, media institutions, civil society organisations and initiatives, leading journalists, cricketers and other individuals were amongst those who had large bot numbers of bot followers.

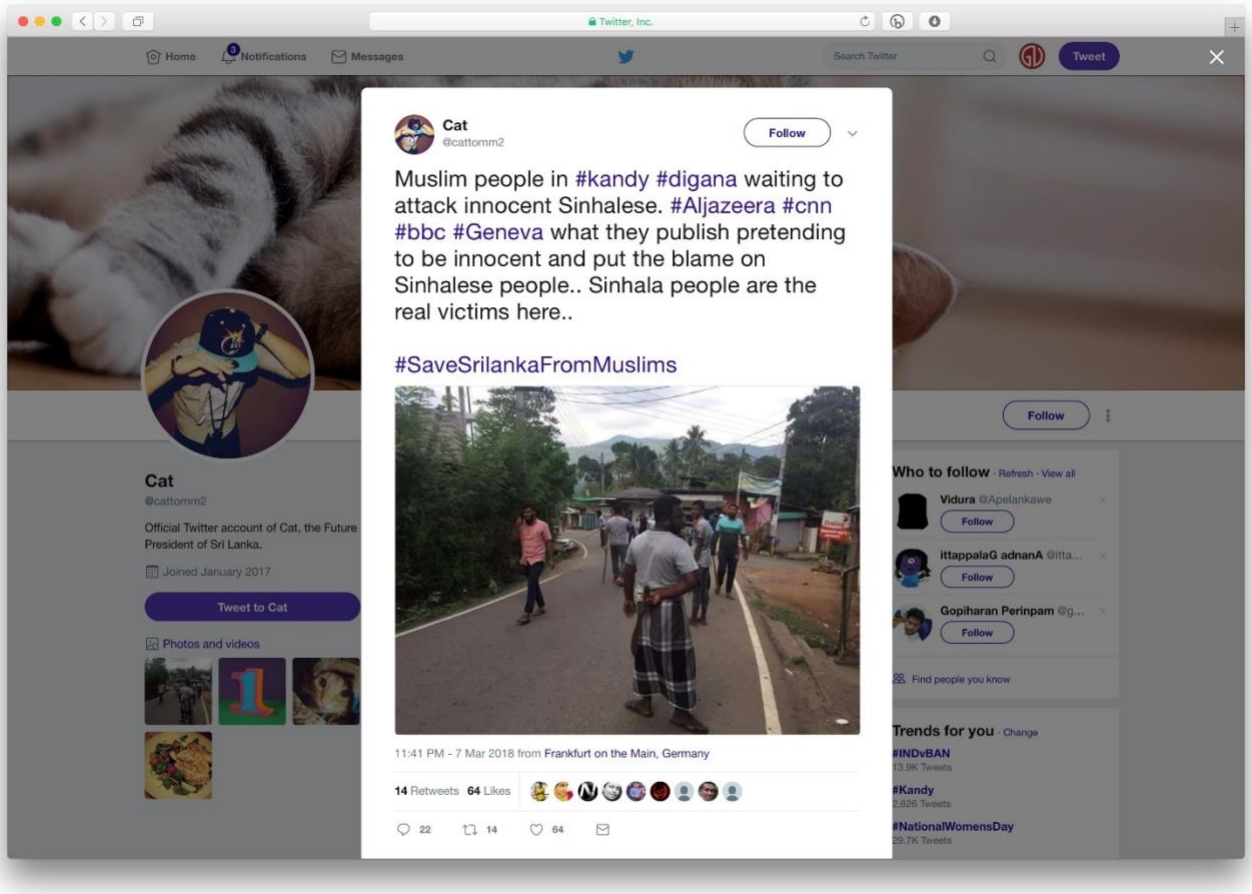
The authors scraped the details of these 2,229 people followed by the bots: their names, their locations, languages, public details, account creation dates, number of followers, amongst other details, and their last 100 tweets. Collectively, this gave us 214,013 items and around 71 variables.

Considering bots are now a permanent feature of Sri Lanka's Twitter landscape and will likely grow in scope and scale leading up to elections or a referendum, it is important to ask how to address the issue at scale, given the number of citizens – directly connected as well as influenced by those connected – involved.

Even with its limited scope and data, this report is a clear snapshot of the political landscape we now inhabit, and projects in the future real dangers that result from just the visible investments made around key social media platforms, which are today the key information and news vectors for a demographic between 18-34.

The context

During the height of the violence that gripped Digana and Kandy in March 2018, one tweet – amongst many others – prompted *Groundviews* to issue an unprecedented warning[1], on Twitter, against reacting or responding to content on the platform without first checking its veracity.



The tweet, still online, from 7 March 2018, had by the time *Groundviews* taken a screenshot been liked 64 times and retweeted 14 times. At the time of authoring this report, this has increased to 112 likes and 23 retweets. *Prima facie*, the tweet suggests that a gang of Muslim men are on the street, waiting to attack Sinhalese. Not a single other person on the ground, media, Police or government backed this claim or reported on the same issue or incident at the time, or even after. Tellingly, the tweet itself is published from Frankfurt, Germany, a fact many who saw and responded to the tweet on mobile or desktop would not have been attentive to. Though not impossible for someone out of Sri Lanka to report on what is going on inside the country, even a cursory reading of the account suggests it is not one that can be taken seriously – with tweets and retweets of textual gibberish and photos of kittens, teddy bears and food. A day before, another account, @askhitauk, published a tweet blaming the Muslims for engineering the violence against them[2]. This tweet, at the time of taking the screen, had generated 151 responses, 61 retweets and 322 likes. While the sentiment wasn't directly inciting

[1] https://twitter.com/groundviews/status/971943162062159872
[2] https://twitter.com/groundviews/status/971943162062159872

hatred when read out of context, it is important to recall that the tweet – no longer online – was published at the height of the violence in Digana. Upon further investigation, *Groundviews* discovered the profile image was fake. There were dozens of other tweets on similar lines.

This was new. Four years ago, during the violence in Aluthgama, *Groundviews* was amongst just a handful of others which reported on the incident over the web and social media. We used Twitter extensively. The weaponisation of Twitter at the time was extremely limited. One example was of an image from Myanmar, and the Rakhine Province, paraded as one that was from, and a result of the violence in Aluthgama. A simple Google Reverse Image search put a stop to the spread of that rumour, and we also at the time flagged how other users on Twitter and social media could verify images before they retweeted, liked, shared, forwarded or reacted.

The weaponisation of social media in the lead up to and during Digana was at a different scale. A lot of the focus has to date been on Facebook. As the *New York Times* in a report published on 21 April 2018 notes[3],

> Time and again, communal hatreds overrun the newsfeed — the primary portal for news and information for many users — unchecked as local media are displaced by Facebook and governments find themselves with little leverage over the company. Some users, energized by hate speech and misinformation, plot real-world attacks. A reconstruction of Sri Lanka's descent into violence, based on interviews with officials, victims and ordinary users caught up in online anger, found that Facebook's newsfeed played a central role in nearly every step from rumour to killing. Facebook officials, they say, ignored repeated warnings of the potential for violence, resisting pressure to hire moderators or establish emergency points of contact.

Twitter escaped scrutiny at the time, and in fact, became after the government blocked Facebook, the primary platform for the dissemination of news, information and opinion on the violence, including read time updates from the ground. In fact, though through VPNs access to Facebook even during the time it was officially blocked by the government picked up, @groundviews and @vikalpavoices (in Sinhala), along with the personal account of one of the authors, (@sanjanah), became on Twitter some of the most mentioned and retweeted sources of analysis, updates, news and information on the violence[4] followed by accounts belonging to seasoned journalists like @tingilye as well as activists like @rukitweets and @thyagir. Twitter, clearly, has established itself firmly as a key platform – on desktop and mobile - of political communication and contestation.

Perhaps because of this, things started to get very strange, very quickly. Starting late March, *Groundviews* and other Twitter users in Sri Lanka began noticing a tsunami of Twitter accounts with no bios, no tweets and the default profile picture following them. It's usually just someone setting up a Twitter account, but soon complaints started pouring in from activists and the few heavy Twitter users in the country. They reported gains of 20-40 bots *a day.* Early accounts that faced this onslaught of new followers included @anushwij, @azambm, @drac, @iromip, @sachp, @ThyagiR & @vindib_. The accounts of two of the authors, @sanjanah and @yudhanjaya were also severely impacted.

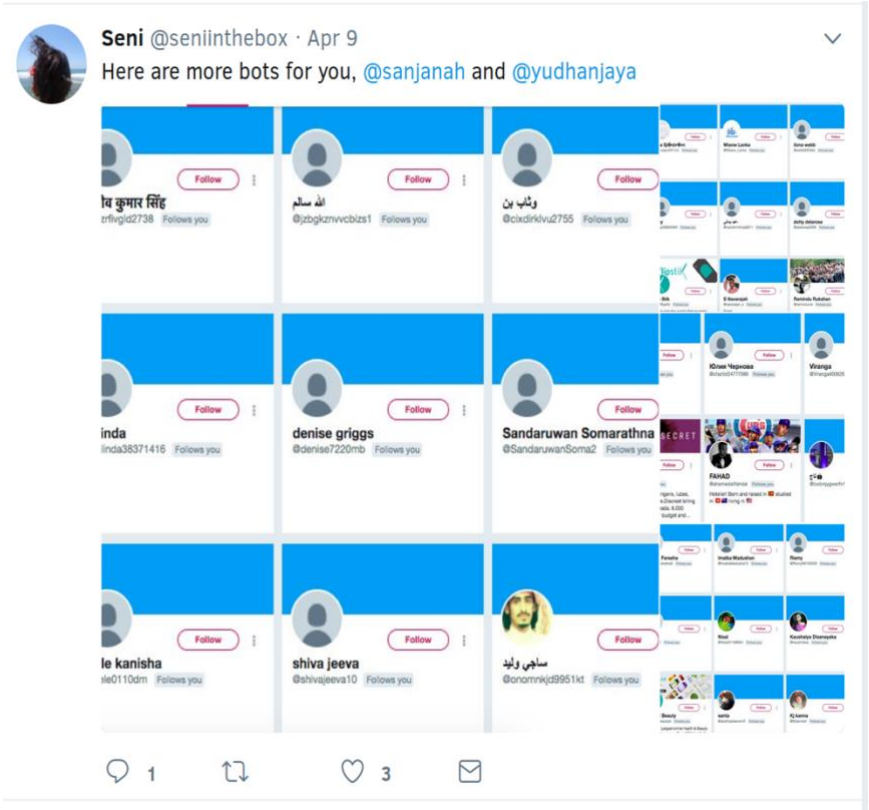
---

[3]Where Countries Are Tinderboxes and Facebook Is a Match, https://www.nytimes.com/2018/04/21/world/asia/facebook-sri-lanka-riots.html by Amanda Taub and Max Fisher, accessed 22 April 2018.

[4] http://truthy.indiana.edu/tools/networks/#?hashtag=%23digana&network_type=rm&start_date=3-1-2018&end_date=3-25-2018

Groundviews tweeted on 1 April that it used the Twitter Audit tool to purge around 30,000 fake followers from its account, accumulated since the time it was setup around 10 years ago, but a very large number from 2018. At the same time, *Groundviews* alerted Twitter to this unprecedented phenomenon through a tweet thread that flagged examples of bots, the dangers they presented and some tips on how to identify and get rid of them[5].

Immediately evident were interesting characteristics, not unlike the bots and fake followers anchored to account of Namal Rajapaksa[6]. The bots had Sinhala, Muslim and Tamil sounding names. Many of the profiles were the default Twitter profile image, but many of the female accounts had photos lifted from public profiles of other individuals[7]. Given the scale and scope of the infestation, *Groundviews*, for the first time in the Sri Lanka Twittersphere, took the step of making public its block list, which other users could import. Even this measure though was not enough to address the high frequency with which new, fake accounts were being created, attaching themselves to prominent Twitter users in Sri Lanka.



Unlike the bots that have made headlines elsewhere and are featured in academic studies on the harmful effects of autonomous agents on social media networks, these bots were curious: they did and said nothing whatsoever. They just followed, *en masse*. Screenshots followed, and on the heels of that people began to report that they had begun blocking these bots - perhaps because the effect is like being stared at by a silent crowd. It is unnerving, and one does not know what to expect, or when the crowd will turn ugly. Many over the following weeks expressed their frustration that just dealing with the bots was extremely time consuming, and that they just couldn't keep up since Twitter's own tools to deal with fake accounts are archaic, unwieldy and not geared to deal with this scale of fake follower generation.

[5] https://twitter.com/groundviews/status/980239708964843520

[6] Namal Rajapaksa, bots and trolls: New contours of digital propaganda and online discourse in Sri Lanka, http://groundviews.org/2018/01/24/namal-rajapaksa-bots-and-trolls-new-contours-of-digital-propaganda-and-online-discourse-in-sri-lanka/

[7] This follows academic research on how bots have used, in the past and elsewhere, attractive photos from hotornot.com to be particularly effective at gaining influence.

The interest of the authors was piqued by what at first blush appeared to be a phenomenon hitherto unseen on Twitter in Sri Lanka, and at a disturbing velocity, over just this one vector, that gave rise to the fear that sometime in the future, the bots could be used to increase the volume of opinions detrimental to democracy.

The Method

Twitter allows any user to export a list of users that they've blocked. No names are given in these lists - merely an internal account ID - but it's more convenient than typing down names from screenshots. Working quickly, one of the authors, Hattotuwa, issued a call (over Twitter itself) for block lists and made them part of the *Groundviews* archive, ending up with a list of 40,239 blocked users from both *Groundviews* and 17 other leading Twitter users who were willing share with a larger public their private likes and dislikes this way.

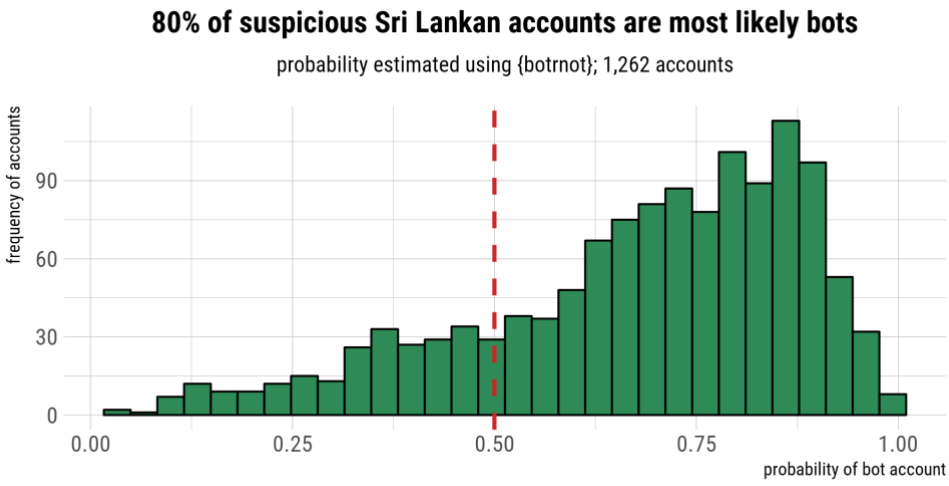
This list, once parsed through the Twitter API, resulted in further information - names, creation dates, descriptions, URLs, verification tags, picture IDs. As it turns out, these 40,239 blocks were in reality 5,235 unique accounts.

The challenge was then to differentiate between bots and accounts belonging to individuals. Various technical and academic literature exists on the detection of Twitter bots. Some of these were explored in the article dealing with the bot generation and following by a prominent politician in Sri Lanka[8]. There are more complex approaches using machine-learning (ML) models. However, in this case, we traded a more technical and complicated analysis for a no less robust but far more end-user friendly analysis, laying down the following 'rules':

1. A bot must have an unverified account
2. A bot must have an empty 'bio'
3. A bot must have less than five tweets over their entire lifetime

We are aware that this is an extremely conservative approach: we've detected thousands of bots with non-empty bios, and no doubt there are many bots that actually do tweet. We were also dealing with *a very small sample space.* There are, by some rough estimates, some 80,000-90,000 Sri Lankan Twitter users. This entire analysis, for reasons of privacy, was done only using a base dataset from just 17 of them. Until we can convince a substantial percentage of people to share their private blocklists with us, we can only act as a lighthouse, warning of a greater catastrophe ahead for Twitter as a company and its users, if the signs go unheeded[9].

Preliminary analysis of 1,262 accounts, a subset of the larger dataset we were working with, indicated that the majority of suspicious accounts following Twitter users were bots.



Another revealing visualisation of the number of accounts targeted by the bots revealed that leading diplomats, Ambassadors based in Sri Lanka, the official accounts of diplomatic missions, leading local politicians, the former President of the Maldives, media institutions, civil society organisations and initiatives, leading journalists, cricketers

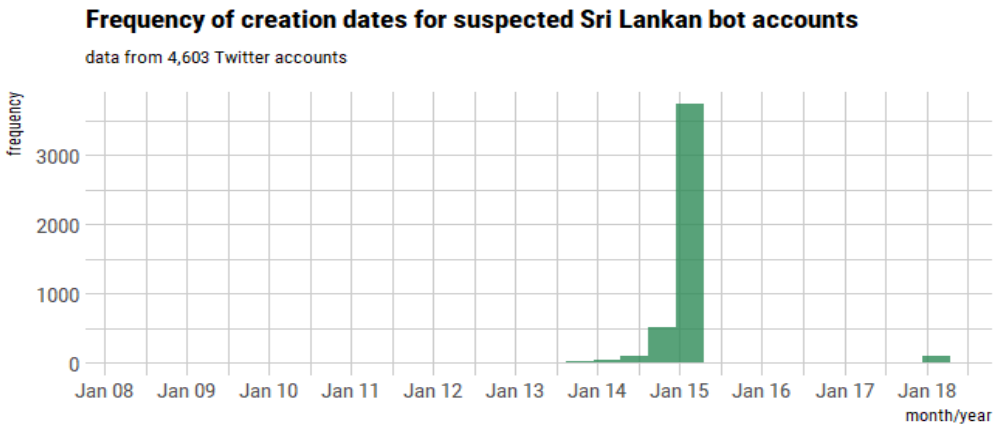
[8] Ibid

[9] At the time of writing, based on Twitter Audit's (https://www.twitteraudit.com) automated and algorithmic analysis, accounts where there are more bots following them than real users belong to the Prime Minister, several senior MPs of the UNP, leading journalists, civil society institutions and others.

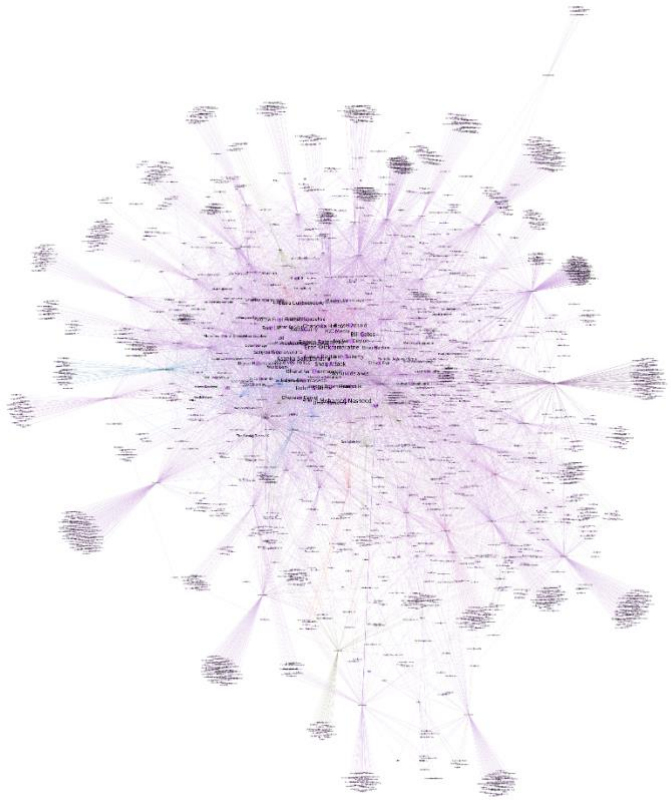
and other individuals were amongst those who had large bot numbers of bot followers[10]. A more detailed account of this follows.

More granular detail was needed to proceed. A bit of data wrangling ensued, and soon this approach left us with 4,603 Twitter accounts. These accounts have some interesting characteristics. The largest of these accounts has 49 followers but follows over 2,000 people. The next has 47 followers and follows 1,351 people. With this data and more in hand, we set about to visualising the scale and scope of what we saw.

The first observation is that these bots were created in two lumps: one in 2015, the other in 2018. 105 of the accounts (as of the time of analysis) were from 2018.



Without any tweets or bios, there's very little that these bots can tell us on their own. But they do, collectively, happen to follow 2,229 people. Our theory was that if we analyse their information, and showed where these follow networks meet, we'd be able to figure out the logic behind this bot network. We scraped the details of these 2,229 people followed by the bots: their names, their locations, languages, public details, account creation dates, number of followers, amongst other details, and their last 100 tweets. Collectively, this gave us 214,013 items and around 71 variables.



[10] Access high resolution image of this visualisation here http://groundviews.org/wp-content/uploads/2018/04/image002.png

Admittedly, this is more art than science. This ink-blot like image is the network of people being followed by the 2018 bot wave: all follows visualised are one-directional - from the bots to the users. The reason it so small is to render the names. This had to be a 5000x5000-pixel image[11]. For the sake of non-visual analysis, we also computed network characteristics - measures of how which users had the most number of inbound connections, and so on.

The tiny, hyper-connected points are our bots. All bots follow a seemingly random mix of accounts. Singers, songwriters, actors, celebrities - possibly done to prevent Twitter from detecting them automatically. Right at the centre are where the bot's tastes begin to overlap. There's a measure called Eigenvector centrality that can rank these 'victims' by their importance to this directed network:

Id	Eigencentrality
Mohamed Nasheed	1
Eran Wickramaratne	0.972222
Bharatha Thennakoon	0.944444
Sagala Ratnayaka	0.944444
Chandika Hathu	0.930556
Daren Sammy	0.930556
Bill Gates	0.930556
Asanka Sahabandu ♠	0.930556
Shaq Attack	0.916667
Robin Sharma	0.916667
ICC Media	0.902778
NewsCurry	0.888889
James Dauris	0.888889
Nathan Ceylon	0.861111
Presidency Maldives	0.847222
Russel Arnold	0.833333
Udara Dharmasena	0.833333
Chevaan Daniel	0.833333
Maldives Police	0.819444
Dilrini	0.819444
shanuki de alwis	0.805556
Namal Rajapaksa	0.791667
Roar LK	0.791667
Manthri.lk	0.791667
Sanjiva Weerawarana	0.777778
Otara Gunewardene	0.763889
Joanne Doornewaard	0.763889
Umar Mansoor	0.763889
Bryce Hutchesson	0.75
Andrew Fidel Fernando	0.736111
Ritsu Nacken	0.708333
Elmo Leon	0.708333

[11] See the original at http://groundviews.org/wp-content/uploads/2018/04/Untitled.png

Ethics Eye	0.694444
Mohandas Menon	0.680556
World Bank	0.597222
NA	0.583333
Mahela Jayawardena	0.583333
TEDxColombo	0.583333
Amalan Dhananjayan	0.555556
Kumar Sangakkara	0.527778
Amrit S	0.527778
Shanaka Amarasinghe	0.527778
Dihan De Silva	0.527778
LKI	0.513889

The Eigenvector measure confirms what the data visualisation[12] also flagged. Here we seem to have an almost overwhelmingly Sri Lankan mix - except for Bill Gates, Robin Sharma, Shaq and the World Bank. Bill Gates, we suspect, is one of those default follow options that Twitter shows to you when you create a new account. The rest are a scattershot mix of celebrities, politicians, new media outfits, prominent journalists and people who air outspoken views on social media.

It is impossible to draw firm conclusions from this motley mix of accounts. But we can draw out some observations, by looking at a few of the accounts from above. Joanne Doornewaard, Bryce Hutchesson and James Dauris are senior diplomats in Sri Lanka, so one can at first blush argue the bots are targeting these accounts because they are obviously influential within the country and amongst political circles. Ritsu Nacken is the UNFPA Sri Lanka Representative and Maldives Country Director, so the attraction to a bot may well be, *inter alia*, the connection to the UN. On similar lines, Mohamed Nasheed would be influential as a political authority, though not on Sri Lanka even though he is present in the country. Interestingly, the bots do seem to go after Maldivian accounts - from the Police to the Presidency. Chandika Hathurusingha, Russel Arnold, Mahela Jayawardena and Kumar Sangakkara are clearly huge influencers over and on any (social media) network they are part of. One can extend this to Daren Sammy, who isn't Sri Lankan, but a popular cricketer. Shanaka Amarasinghe as a leading sports commentator would also be an obvious target, as would be Andrew Fernando, a cricket writer on ESPN Cricinfo. Mohandas Menon is a legend in the world of cricketing statistics. Chevaan Daniel and the other journalists in this list are unsurprising targets for each of their respective spheres of influence online, and in the case of some, offline, extending into the domains of mainstream politics. Namal Rajapaksa, Sagala Ratnayaka and Eran Wickramaratne are politicians and Members of Parliament. Media accounts like Ethics Eye, Roar and Manthri are unsurprising targets as well - and keep in mind that at the time of data analysis, *Groundviews* had already purged around 30,000 fake accounts. TEDxColombo, by virtue of its association with TED and its reputation, is also an easy target for bots in search of influential accounts. Amalan Dhananjayan is a coder, and interestingly, has also tweeted about 400 bot accounts that had followed his account over just 8 days[13]. Asanka Sahabandu seems to be a popular entertainment personality. Udara Dharmasena is a marketer at Neo@Ogilvy, and LKI is the Lakshman Kadirgamar Institute (LKI), a leading think-tank.

[12] See http://groundviews.org/wp-content/uploads/2018/04/image002.png
[13] https://twitter.com/batzeee/status/983287759371583491

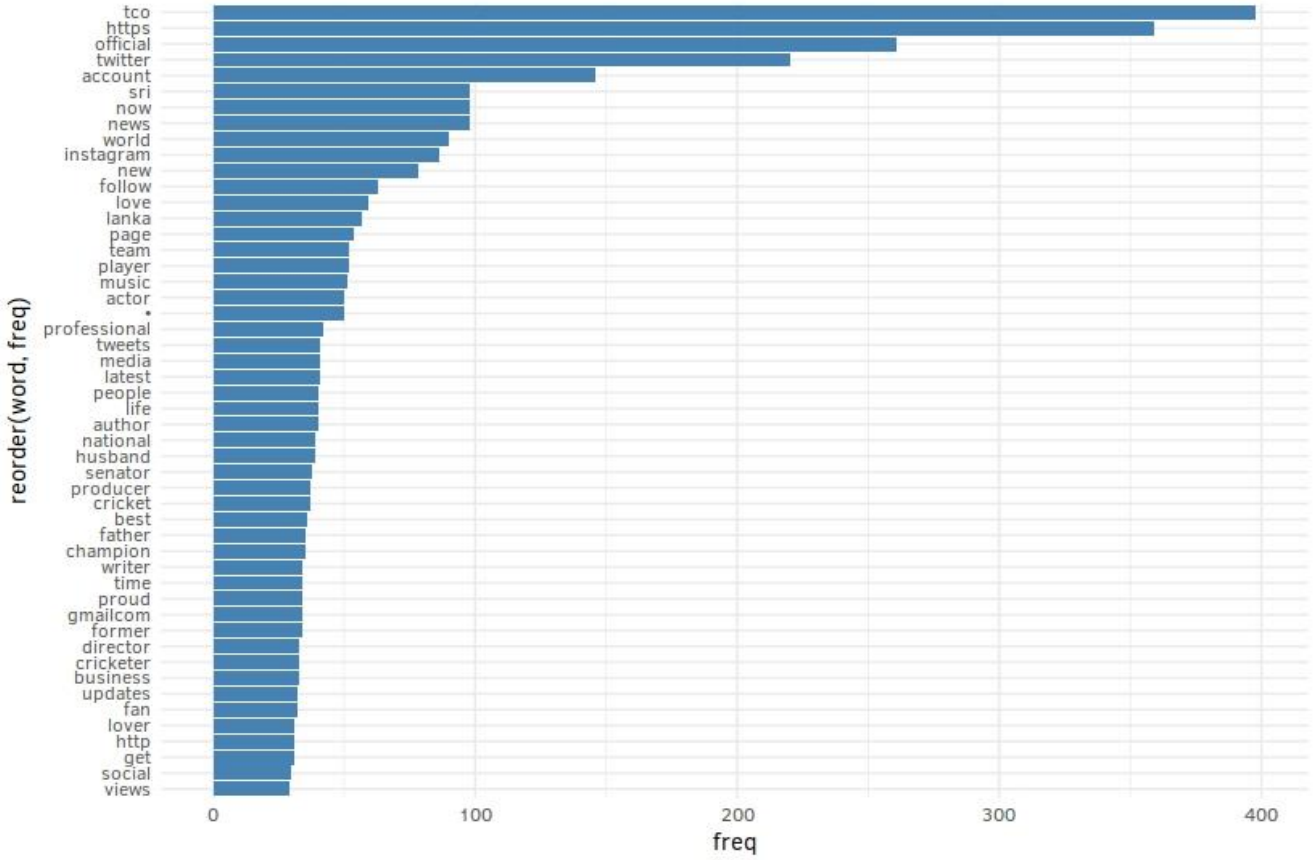
From a purely algorithmic selection perspective, the targeting of these accounts makes sense. Entertainment, gaming, sports, cricket, politics, academia, media and diplomacy are safe bets for bots to attach themselves to as nodes in a network more influential than others, based on certain guiding metrics hard coded into them. And if the bots were created in Sri Lanka (with a Sri Lankan IP address) it may be the case that popular cricketers and personalities are offered as accounts to follow. The same goes for the World Bank and Bill Gates.

And yet, the Twitter account of Shanuki de Alwis - who tweets about dogs, animals and English theatre in the main - also attracts a lot of bots, which gives rise the suspicion these bots are algorithmically driven to latch themselves on to accounts which have a certain content production and engagement signature, or follow a certain pattern or blueprint (pegged to influencer metrics and possibly semantics), no matter which domain they belong to, what they are in the main interested in or tweet about. In other words, the bots seem to be at present spreading in an automated fashion, as opposed to any precise targeting of individuals or specific accounts in Sri Lanka. They are gender, location and political bias agnostic - latching on to non-Sri Lankan accounts with as much gusto as those that belong to foreigners.

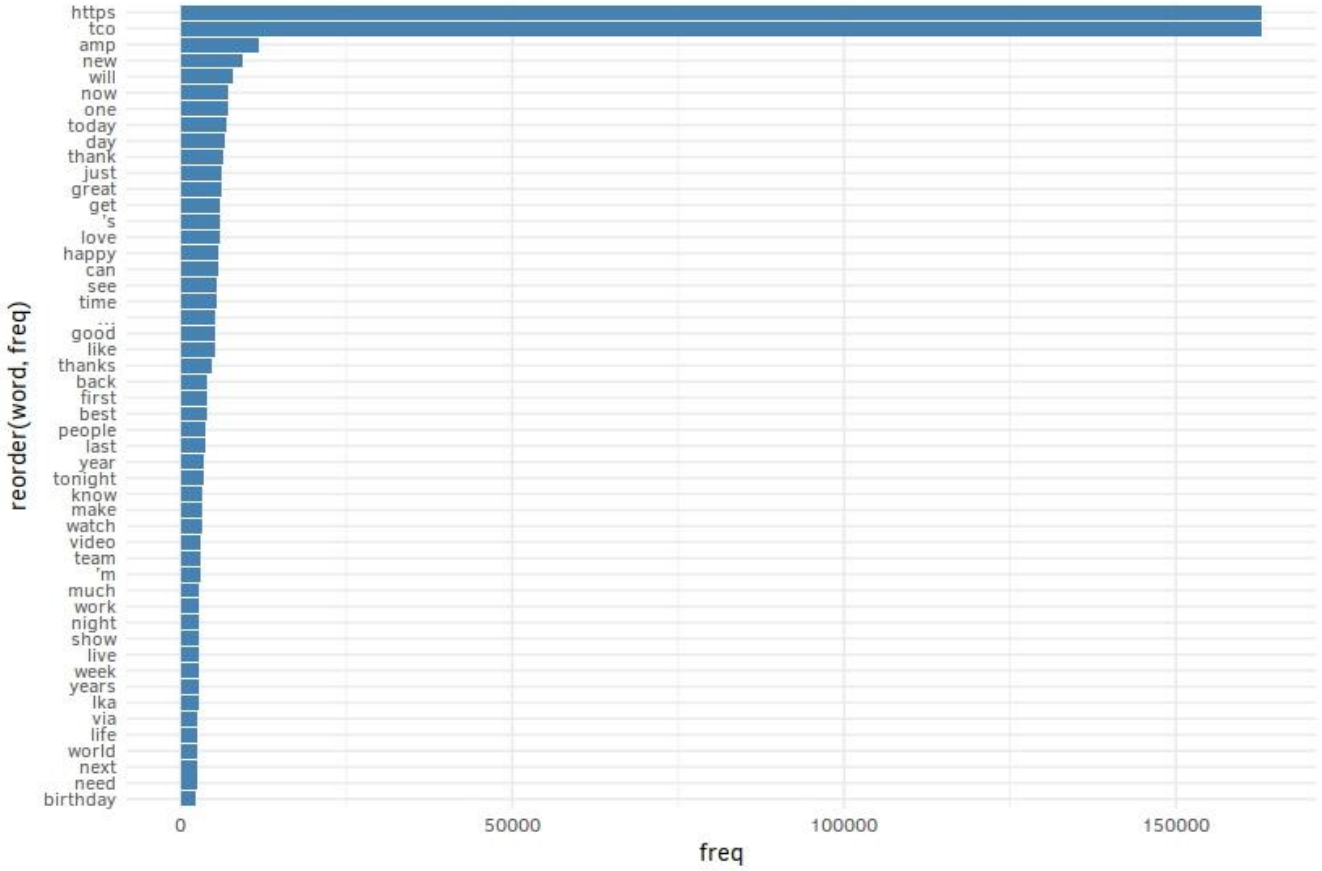
This is comforting and disturbing at the same time. Comforting, because it clearly is not the case that the spread and scope at present indicate, to the best we can say, they are specifically targeting accounts critical of the present government or former regime. Disturbing, because the speed, spread and scope of the bot account generation and targeting suggests that one central aim may be, at some point in the future, to activate these bots in a way that through multi-nodal, multi-lingual, multi-media content production, dissemination and echoing, at scale, overwhelms the discussion of specific issues or actors, and overall, serves to strategically confuse, misdirect, misinform and undermine trust in Twitter as a whole.

Other interpretations are also possible, and in large part why this report is also being released to the public domain.

Wijeratne and Serrato, the two authors of this report who are experienced data scientists, conducted further data analysis also help see what the targeted accounts are talking about. The resulting scraped datasets contain the bios of 2,165 people. By lumping it all together and splitting words apart we can analyse what words most often appear in these bios.



We can also do the same for what they've been talking about. We harvested the last 100 tweets of 2,165 of these people - the rest seem to be private accounts. With Twitter's API limits, we scraped slightly under the actual figure of 2165 x 100 - but 214,013 tweets should be enough to figure out what words this community of users put out.

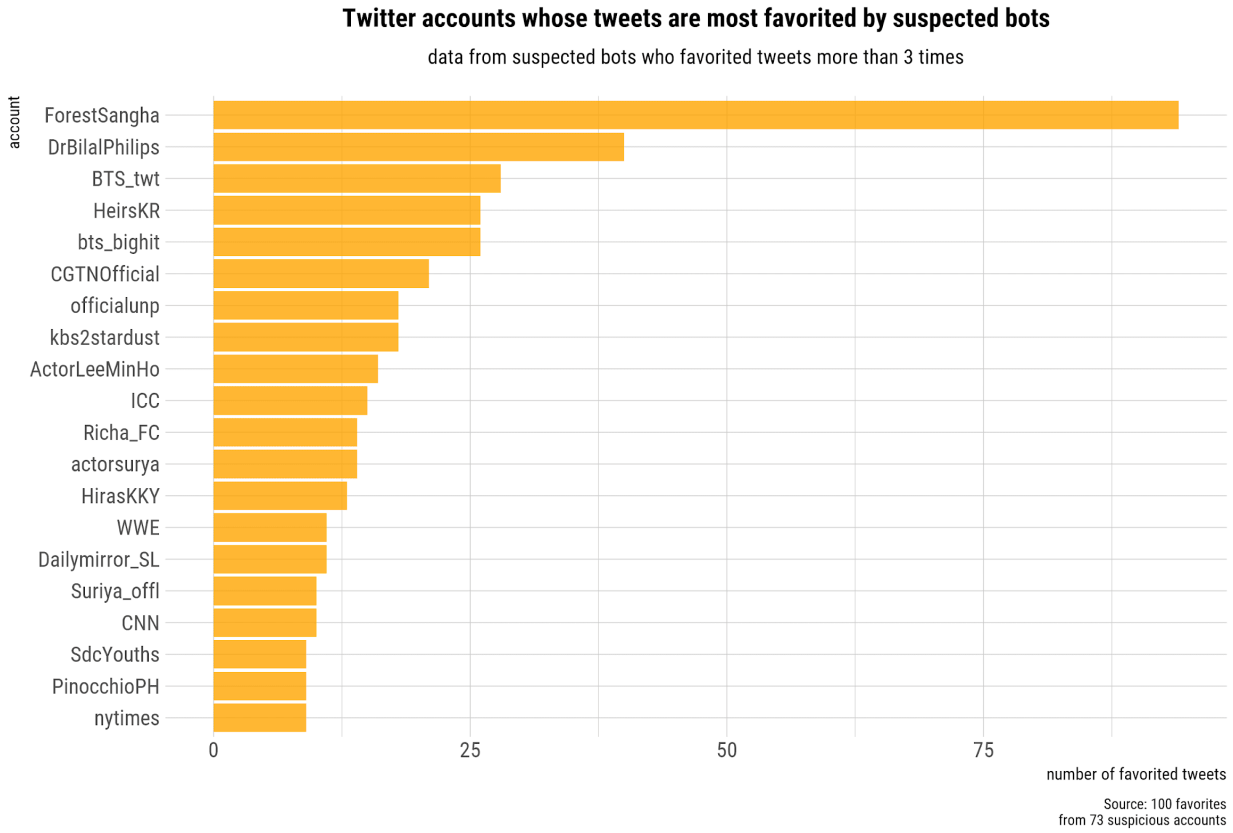


Individually, neither of these charts are useful (and one might be forgiven for thinking that the collective Twitter output adds up to just sheer nonsense). Put together though, they show a pattern: in both bios and tweets, https, tco and amp show prominence over everything else.

https signifies that these people are sharing links to web articles or content. Tco is the processed version of t.co, the URL that signifies something's been through the default Twitter domain shortening service. Twitter applies this to URLs when you go over the character limit. And amp is the precursor of Accelerated Mobile Pages links championed by Google - which practically every major media operation has switched to: load a page on mobile and it'll look like amp.whateveritis.com.

Simply put, these bots seem to be targeting people who share a lot of web content. We are not sure why: perhaps because these people are seen as more likely to be influencers over Twitter in Sri Lanka?

As an additional layer of analysis, we looked at whose content was being favourited the most. We know that the more favourites and retweets a status has, the more credibility it gains in the minds of most



Twitter users – and boosts engagement rates, of course. However, these bots seem to be largely inactive with regard to manipulating content this way: only a small subset of around 73 accounts had performed more than three favourites.

This top 20 list seems strange (including accounts very clearly not from Sri Lanka) and may be something to re-examine further down the line in light of what is at the time of publication is increasingly being reported as a Twitter bot plague affecting other countries in South and South East Asia[14].

Keep in mind that this is data from 17 people, from whom over 4,000 bots produced in 2015 and a much smaller handful in 2018 were identified. It is from this handful the authors have analysed 214,000+ tweets. While this seems like a lot, it's in fact a small sample space. The larger group of active 2015 bots have not been touched at all. We were hamstrung by Twitter API limits, which require us to wait 15 minutes in between rate limits[15]. We would need priority or privileged access to the Twitter

[14] Twitter doesn't care that someone is building a bot army in Southeast Asia, https://techcrunch.com/2018/04/20/twitter-doesnt-care-that-someone-is-building-a-bot-army-in-southeast-asia/
[15] Twitter killed off its API whitelisting in 2011, https://readwrite.com/2011/02/11/twitter_kills_the_api_whitelist_what_it_means_for/

firehose as well as over 50 times the amount of data gathered for this study to repeat this analysis. And that's still for 17 people. If a tenth of the Sri Lankan Twitter base gave us their block lists, we would be sitting in front of datasets for a very long time.

But the indications are clear. The 4,000 bots we've detected may be anywhere from half of the network to a tenth of the network: what we have shown here is only the tip of the iceberg.

Discussion points

Our report makes it evident there is a clear strategic investment around the production of bots. We may be the first or one of the very few to analyse a phenomenon that is now being widely reported across our region.

"Journalists, academics, activists, entrepreneurs and UN officials who tweet about or live in Myanmar, Vietnam, Cambodia and other south-east Asian countries have been followed by hundreds — in some cases thousands — of new accounts since late March. The mass following phenomenon is also occurring in China, Taiwan, Hong Kong and Sri Lanka."[16]

Not unlike Facebook's historical and tragic reticence to engage with civil society and do something meaningful to stop its platforms from being used to spread hate and violence, attempts to get Twitter to acknowledge the issue have been met with surprising silence. Not a single tweet or email sent by the authors has generated even a canned response from senior company representatives or official accounts. The silence can mean the company itself sees but doesn't know what to do with the increasing number of bots on its platform, which is something the media have repeatedly flagged[17].

We have multiple theories, that complement those published online by others who have also explored the bot explosion in Sri Lanka based on their own data analysis[18]:

[16] https://www.ft.com/content/e59dba5a-421d-11e8-803a-295c97e6fd0b
[17] Twitter's huge bot problem is out of the bag, https://www.cnet.com/news/twitter-bot-fake-followers-problem-out-of-the-bag/
[18] *The curious case of Tikiri*, https://medium.com/throwaway-thoughts/the-curious-case-of-tikiri-1c7550cba74f by Drac

1. These bots could be engineered by political factions in order to mask their investments in Twitter accounts. *Groundviews* has been openly and repeatedly taunted by certain twitter accounts (who are known trolls) to check the President's and PM's accounts, which is interesting, because it is almost as if they know what some the key targets are of the bot accounts.

2. Alternatively, someone, likely a merchant or third party, is using compromised IP addresses in South or South East Asia to register new, fraudulent accounts, probably for resale at a later date.

3. We do not believe in the theory that these bots are used for monitoring what target accounts in particular are publishing, and what the (Sri Lankan) Twittersphere in general is talking about. Any operation competent enough to set up thousands of bots will find it easier to scrape data the way we did here.

4. It is possible, indeed even probable, that these accounts will be weaponised in the future, perhaps to control/shape political discourse around key electoral processes or referenda. Right now, they are just dormant accounts, and increasingly lazily constructed. As noted earlier, whereas some have names that are immediately associated with the Sinhala, Tamil or Muslim communities, there are now dozens if not more accounts with just bizarre names in Arabic, or long alphanumeric strings.

5. It is unclear who the primary targets of this wave of bot accounts is - perhaps further and more robust network analysis, which much more data, can determine this by way of those hardest hit. The follower structure may be more sophisticated than what we hypothesize. As noted earlier, Shanuki de Alwis, who appears in that list of people hardest hit, tweets more about drama and dogs than anything political, and doesn't use any #lka or #srilanka hashtag. However, she follows many accounts that do promote political news and opinion. The criteria we've unearthed may not be the only thing these bots are running on, latching on to and are algorithmically programmed to be alive or alert to.

6. Since there is no echo chamber effect yet, it is impossible to determine the discursive, disruptive or even democratic dynamics of these fake accounts.

Implications and projections

Existing literature on bots for social networking sites (SNS) in general and Twitter in particular do not match the kind of patterns we are seeing in Sri Lanka. In the US, Mexico and other countries, bots on Twitter have featured in political communication or public discourse: but they have invariably involved the production of content or the retweeting of existing content. Their volume is tightly controlled in the manner they latch on to, seek to drown out through noise, rally each other around, and then almost at the same time, leave a topic or account. The bots created since March do none of this, at present.

The danger is that if they activate, the bots may have disturbing implications, even beyond Twitter. There is general agreement that bots influence the discursive landscape around politics, whenever they are a discernible feature on Twitter. The limited number of characters on Twitter, even with the doubling of the limit in recent months, lends itself to content production by bots which do not need to have a high command of sentence structure, grammar or the primary language lexicon in order to at first blush or with just cursory scrutiny, appear to be a real user. The user interface of Twitter on desktop or mobile also doesn't lend itself to a scrutiny around where a tweet was published. The nature of the platform usage, which is to engage often on the basis of a single tweet as opposed to any critical appreciation of the account writ large, also lends itself to the episodic high frequency dissemination of content that is geared for virality.

However, what some researchers have tellingly discovered is that while the cost of bots may be low, their deployment with a view to shape conversations is in fact not a simple exercise.

> *We quickly realised that high amount of capital – cultural, economic, temporal and social – were needed to influence political discourse and that using bots in this way required some methodological sophistication that we had not anticipated. This finding is interesting in itself, as it indicates that those without significant financial resources and pre-existing bot knowledge are not able to readily influence political discourse[19].*

This lends itself to our primary hypothesis, looking at the data available to us, that this isn't the work of a hobbyist, or occasional interlocutor in political communication, social media or Twitter in particular. There is strategy and intent behind this bot infestation. Similar efforts on social media have been undertaken by governments like Russia and China through what's called 'sockpuppet' accounts[20]. Though the authors stress that disruptive bot behaviour at scale is not something Sri Lanka has experienced to date, there are disturbing reports[21] around how bots deploy different but complementary tactics to achieve their intended outcomes,

- Sockpuppets (part-human/part-bot or, simply, cyborgs) initiate the conversation, seeding new ideas and driving discussion[22] in online communities.
- The ideas are then amplified by as many as tens of thousands of automated accounts that we call "amplifier bots," repurposing, retweeting, and republishing the same language.
- "Approval bots" engage with specific tweets or comments, "liking," "retweeting," or "replying" to enhance credibility, and give legitimacy to an idea.
- In hotly contested topic areas, bots are often used to harass and attack[23] individuals and organizations in an attempt to push them out of the conversation.

---

[19] *Automation, Algorithms, and Politics| Bots and Political Influence: A Sociotechnical Investigation of Social Network Capital*, Murthy, Power, Tinati et al, http://link.galegroup.com.ezproxy.otago.ac.nz/apps/doc/A478974418/AONE?u=otago&sid=AONE&xid=1f166680

[20] The Bots That Are Changing Politics, https://motherboard.vice.com/en_us/article/mb37k4/twitter-facebook-google-bots-misinformation-changing-politics

[21] Ibid

[22] *Sockpuppets, Secessionists, and Breitbart*, Jonathan Morgan, https://medium.com/data-for-democracy/sockpuppets-secessionists-and-breitbart-7171b1134cd5, Data for Democracy

Twitter as company and platform is also to blame. Some have flagged that what contributes to the power of bots is the 'black box algorithms' of social media companies, which both allow for bots and doing so, essentially allow their platforms to be fertile ground for abuse. In addition to Devumi which was flagged in a recent pathbreaking article in the *New York Times* looking at the global and growing bot industry[24], there's MonsterSocial, which offers just for Twitter bots that can[25],

- Write a tweet with all your accounts
- Comment on a tweet with all your accounts
- Retweet a tweet with all your accounts
- Follow a specific profile with all your accounts
- Follow profiles that posted recently using a specific word
- Follow profiles that are following your profiles
- Follow profiles that are following a specific profile
- Unfollow a specific profile with all your accounts
- Unfollow profiles that are following your profiles
- Unfollow profiles you are following
- Unfollow profiles that are not following you back
- Favourite a specific tweet with all your accounts
- Favourite tweets that were recently posted using a specific word
- Favourite multiple tweets on a specific profile with all your accounts
- Favourite multiple tweets on profiles that recently posted using a specific word
- Clear your favourites
- Exclude specific profiles from following or unfollowing

There are many other companies that do much the same. The bot industry is global in nature and local in focus, contributing to the ease through which any interested party can tap into established technologies, context-specific experience, as well as accounts on and for specific social media platforms. As the BBC reported in early March[26], all this for a very low cost in comparison to the volume of content generated over the lifetime of the bot, and the number of bots that can be deployed. Twitter seems to be aware of the problem at least since January 2018[27] when it deleted a large number of bots globally[28], but the silence and inaction over the latest bot outbreak seems to suggest they are overwhelmed by the scale of the production.

Linked to this is the problem that as a research endeavour, it may not be possible to discern the most sophisticated botnets – or at the meta-level, the control architectures of individual accounts – because the platforms do not allow for robust data collection through policies that govern API use and data hoovering (particularly post Cambridge Analytica). Our data collection was painful because of API limits:

---

[23] *Effort to Expose Russia's 'Troll Army' Draws Vicious Retaliation*, https://www.nytimes.com/2016/05/31/world/europe/russia-finland-nato-trolls.html?_r=1, New York Times

[24] *The Follower Factory*, https://www.nytimes.com/interactive/2018/01/27/technology/social-media-bots.html

[25] https://monstersocial.net/features/twitter-bot/

[26] http://www.bbc.com/news/blogs-trending-43218939

[27] https://twitter.com/TwitterComms/status/957316490889347072

[28] https://mashable.com/2018/01/29/twitter-bots-purge-influencers-accounts/#Enc_lEaG5iqh

what we've done here by brute force and obstinacy is ideally better done by a well-funded team that can fork out hard money for API access (which isn't cheap[29]).

Dealing with bots

There is a real challenge around bot detection and deletion. In Sri Lanka, innovative solutions created by Sri Lankans like Malinthe Samarakoon (https://projects.malinthe.com/quickblock/index.php) join established platforms like Twitter Audit (https://www.twitteraudit.com[30]. Both invariably result in false positives. At scale and given frequency of bot production, there is no better alternative to these tools, especially since Twitter itself provides only the most rudimentary protection against bots in one of three ways – by sequentially blocking bot accounts after first checking them manually, by importing a block list generated by another account or user or by making an account private, which makes follower requests a process that is manually curated instead of automatically approved. None of these are ideal or real solutions to the explosion of bot accounts.

In response to the growing problem of bots, Twitter changed its API rules earlier this year[31]. Worth noting is the fact that the current bot explosion in Sri Lanka, and more broadly, South and South East Asia is occurring despite these new restrictions in play, since 23 March 2018. Worth recalling is a commitment by Twitter in June 2017 to investigate bots and misinformation[32]. It is unclear what Twitter has actually done since.

What is clear and known is that bots involved in elections have influenced or disrupted the discursive landscape online. This includes the Mexican general election as far back as 2010, anti-Kremlin dissent in Russia and notably, the US Presidential Election in 2016, where some media reports suggest Russian bots retweeted the now incumbent American President's Twitter account 470,000 times[33]. More recently Reuters reports that bots flooded Twitter in Malaysia, just ahead of polls, with pro-government messages[34]. As quoted in that report, a researcher at the Digital Forensic Research (DFR) Lab of the Washington-based Atlantic Council think tank said over 17,000 bots tweeted content related to the Malaysian election over a week. Extant academic research suggests that most users are unable to detect a bot or fake account from a real one. A report released in March 2017 by the University of Southern California and Indiana University suggest that at the time of publication up to 15% of Twitter accounts globally were fake – or around 48 million individual accounts[35].

---

[29] https://techcrunch.com/2017/11/14/twitter-launches-lower-cost-subscription-access-to-its-data-through-new-premium-apis/

[30] There are many others online. For a good overview, see https://www.poynter.org/news/how-tell-if-you-have-fake-twitter-followers-and-how-remove-them

[31] https://blog.twitter.com/developer/en_us/topics/tips/2018/automation-and-the-use-of-multiple-accounts.html

[32] https://blog.twitter.com/official/en_us/topics/company/2017/Our-Approach-Bots-Misinformation.html

[33] https://www.motherjones.com/politics/2017/10/twitter-bots-distorted-the-2016-election-including-many-controlled-by-russia/ and https://www.bloomberg.com/news/articles/2018-01-26/twitter-says-russian-linked-bots-retweeted-trump-470-000-times

[34] https://www.reuters.com/article/us-malaysia-election-socialmedia/ahead-of-malaysian-polls-bots-flood-twitter-with-pro-government-messages-idUSKBN1HR2AQ

[35] https://www.cnbc.com/2017/03/10/nearly-48-million-twitter-accounts-could-be-bots-says-study.html

Though this was the case globally, the discernible increase in bot production post-Digana in March 2018 seems to follow another heightened phase in 2015. The episodic nature suggests interest and investment, instead of accident or coincidence. Bot production is a tap, and someone, somewhere, is controlling the flow.

Finally, the bots may exist to work in tandem with trolls, as we have seen in the Maldives. Techniques like 'smoke screening' or 'hijacking the hashtag' are phenomena that *Groundviews* over Twitter has engaged with in the Maldives in February[36], where clear violations of Twitter guidelines resulted in many accounts being blocked in the single largest takedown of accounts the country has seen. It is unclear what lasting impact this has had on political communications within the country. What is clear is that strategically deployed, bots do make an impact. In a field report written in 2011 involving real bots deployed for research purposes, the authors of the report speculated that bots will have the ability to determine "the general interests of other users". Other studies note that bots will be able to "subtly shape and influence targets across much larger user networks, driving them to connect with targets, or to share opinion and shape consensus in a particular direction"[37].

What this all means is that the more bots there on Twitter, the greater their ability to be networked and used to corrupt the discursive landscape online, which can have serious implications around key historical moments or leading up to them.

Considering bots are now a permanent feature of Sri Lanka's Twitter landscape and will likely grow in scope and scale leading up to elections or a referendum, it is important to ask how to address the issue at scale, given the number of citizens – directly connected as well as influenced by those connected – involved. It is likely that the first response of political actors will be to invest in counter-bot strategies, resulting in a multifaceted cyberwar for the control of the most persuasive and pervasive bot network. This is in addition to the threats posed by bots in general independent of country or political context, which includes hashtag spamming, the creation of artificial trends, smear campaigns, death-threat campaigns and political propaganda. Academic literature points to the challenge of algorithmic political communication, of a world where computational power (or agency) is used to identity, target and misdirect, in real time, conversations that are deemed inconvenient to any party that has access to the technology to harvest social media content, for parochial gain.

Even with its limited scope and data, this report is a clear snapshot of the political landscape we now inhabit, and projects in the future real dangers that result from just the visible investments made around key social media platforms, which are today the key information and news vectors for a demographic between 18-34. The risks have been communicated to individuals in government, but the challenge of responding lies with Twitter and its users as well. The company's silence to date on what is now a phenomenon that is serious enough to be reported by international wire news services suggests that social media companies in Silicon Valley have lost control of platforms created in a culture and context far removed from where they are being used the most today. And end users should also be concerned, in particular that so many today – on Twitter and social media in general - respond or react to what's online without robust verification and questioning veracity. The data we have presented is a

[36] https://twitter.com/groundviews/status/963377862345678848

[37] Social bots: Voices from the front, Pearce Hwang, M. Nanis, *Social Mediator*, 2012.

clear warning to check before you retweet and think before you react. We cannot only blame government inaction and Twitter's silence around a corrupted online conversation that we have contributed to as well, through our own negligence, ignorance or misplaced trust.

Authors

Sanjana Hattotuwa is the founding editor of *Groundviews*, and a Senior Researcher at the Centre for Policy Alternatives, where he also heads the Civic Media Team. He is also an Advisor at the ICT4Peace Foundation. He is currently a PhD student at the University of Otago, New Zealand, engaged in research on social media, violence and politics.

Yudhanjaya Wijeratne is a science fiction author and Big Data researcher at LIRNE*asia*, a think tank with research operations across multiple SAARC and ASEAN countries. He studies social networks and information flow. He's run news operations, designed games and fallen off cliffs (most of these things by accident), but he's known in his native Sri Lanka for bringing data analysis to political commentary via the *Icaruswept* blog. Among other things, he's analysed influencers in the 2015 Sri Lankan general elections, and led a corporate team analysing the 2016 US Election using hundreds of news reports and up to a million tweets per day. His published novels include the critically acclaimed *Numbercaste*, and he's currently working with HarperCollins on a book about using Big Data for the greater public good.

Raymond Serrato is a Senior Programme Officer for Asia and Governance and Innovation Expert at Democracy Reporting International. He manages DRI's work in Myanmar and technology and democracy and recently led a project to monitor social media during the German elections. His research on social media, hate speech, and elections has appeared in The Guardian, CNN, and The Irrawaddy.