South African bot behaviour post the July 2018 Twitter account cull

Coral Featherstone
The Council for Scientific and Industrial Research (CSIR)
Gauteng, South Africa
cfeatherstone@csir.co.za

***Abstract*—World wide, there is a concerning use of social media to sway public opinion through the use of disinformation campaigns. Elections, political decisions such as the Brexit referendum and topics such as vaccination have all been targets of what has come to be known as computational propaganda. Twitter maintainers finally acted on the problem and in July 2018 they suspended millions of fake accounts. Automated accounts, known as bots, are substantially easier to spot and analyse when they are newly created and in the "egg" stage. The culling of accounts therefore provided the perfect opportunity to look for common behaviour. This paper provides an account of the common behaviour seen on a sample of high profile South African focused Twitter accounts in the days following the action by Twitter maintainers to remove abusive accounts.**

***Index Terms*—social media, Twitter, micro-blogging, computational propaganda, bots, South Africa**

I. Introduction

World wide, there is a concerning use of social media to sway public opinion through the use of disinformation campaigns. Some of the recent prominent examples are the suggested allegations that Russian trolls used social media to sway the American, French and German elections [1], [2]. Another example is alleged Russian interference and the spreading of propaganda to influence the Brexit referendem [2], [3] and swaying decisions on complex topics such as vaccination [4] and anti-science topics such as climate change, and stem-cell research [5].

Social media is also used to amplify distorting information and draw attention to websites. It has been suggested that 9-15% of active Twitter accounts are automated or semi-automated, and that these accounts are behind a third of the links to mainstream websites [6]. A popular technique is to follow popular accounts around a theme; posting content, answering other users and in general impersonating human users in order to get followers so that the reach of the content is expanded [6]. Some automated accounts exists just to make other accounts look like they have followers. Many social media users are now aware that there is manipulation on these platforms, but the extend of manipulation may be far larger than many people appreciate.

These are only some examples, and manipulation has been found in over 60 countries so far, including Angola, Egypt, Kenya, Nigeria, South Africa, and Zimbabwe [7]. This manipulation has come to be known as, "Computational propaganda" [2]. Various social media platforms have been used, including Youtube [8], Facebook, Twitter, Instagram, WhatsApp, Tinder and others [7]. The use of groups of individuals paid to spread propaganda using social media has come to be known as troll farms [9]. Troll farms came to the South African public's attention when the public relations company Bell Pottinger used social media in a smear campaign which spread racially divisive content [10]. The campaign saw the subsequent arrival of the hashtag #paidTwitter.

Twitter was abused to the point where the maintainers were forced to act on the problem, and in July 2018 they actively shut down over 70 million automated, semi-automated and abusive accounts [11], [12]. Several well known South African celebrities lost tens of thousands of followers overnight, with one user (Karabo Mokgoko) losing 84 887 followers [12].

Automated accounts, known as (ro)bots, are substantially easier to spot and analyse when they are newly created. These bots started re-spawning in large numbers and were easily visible. The culling of accounts therefore provided the perfect opportunity to look for common behaviour. An unfortunate consequence of investigating the behaviour of bots is that we assist them in further disguising the behaviour [13]. There are techniques for looking for them, but the newer bots evolve as the research highlights the behaviour and are now also change behaviour over time to avoid detection [14].

II. Identifying bots

Firstly, a note on Twitter terminology; a tweet is a short text message of 280 characters or less, that can also contain images, videos, and links. A timeline is a collection of tweets, belonging to a particular Twitter account. A retweet is the action of one Twitter user sharing a tweet of another Twitter user. A Twitter bio, consists of a background image, a profile image, and a short biography, and is provided by the owner of the Twitter account. One user can indicate that they approve of another user's tweets by clicking an option to "Like" the other user's tweet.

There are several types of abusive accounts on Twitter. Some accounts are completely automated by computers. Not all bots are malicious, but many are. Other malicious accounts are run by humans. Some malicious accounts are operated by humans but have a large amounts of followers who are bots [9]. The purpose of the bot followers is to retweet the content, which is spread more widely as a result.

Evading techniques of bots include, faking followers, deleting tweets, mixing normal tweets between malicious content, posting the same message but altering words while retaining semantic meaning, and using other users profile photographs [14]. Evading techniques are getting more sophisticated over time.

Newly created Twitter accounts are known as "eggs". This is for historical reasons and describes the lack of profile picture or bio of recently created accounts (the default profile picture used to be an egg). Eggs make newly created accounts very easy to identify, which is why the cull of accounts provided a good opportunity to identify them. Table I presents attributes that are commonly used to manually identify automated accounts.

TABLE I
ATTRIBUTES COMMONLY USED TO MANUALLY IDENTIFY AUTOMATED ACCOUNTS

Attribute	Description
Age	The account may have been active for a long time, but appears to have never tweeted [15]. Accounts tend to be newer as older accounts that produce spam are likely to get suspended [16].
Likes	The account has multiple likes, but there are no tweets to like. In other words content has been deleted [6]. An example of an account with likes and no timeline can be seen in Figure 2.
Profile detail	Accounts de-prioritise profile pictures, biography text and decoration making them easy to spot [15]. An example of followers of an account just post the Twitter cull is shown in Figure 1.
Behaviour	Humans have no fixed agenda, but computer programs will perform the same very consistent, methodical behaviour, when the account is first created. The account activity also happens at all times during the day and night, on weekdays and weekends, and the same message is frequently repeated with minor changes [15].
Tweet count	Automated accounts can produce large numbers of tweets in a very short space of time [15], [16].
Following Follower ratios	Bots frequently don't follow many other Twitter accounts and have skewed ratios of follower to followee [15].
Geolocation	Most bots turn off location identification [15], [16].
Multiple accounts same topic	Multiple malicious accounts tweeting about the same topic at the same time. [16]

This investigation focused specifically on visually trying to identify common behaviour between, "eggs" that were showing the features that are typically used in the identification of bots. The identified accounts demonstrating Bot-like behaviour were also queried on Botometer[1] [6] – an online tool for bot detection – for further verification. The Twitter cull happened with very little warning and there was no time to automate the investigation, but a follow up of the accounts using programmatic methods could be useful. Botometer uses supervised machine learning together with a number of known bot detection methods, but due to a grey area between human and bot behaviour cannot always definitively determine whether an account is non-human [6].

[1]https://botometer.iuni.iu.edu

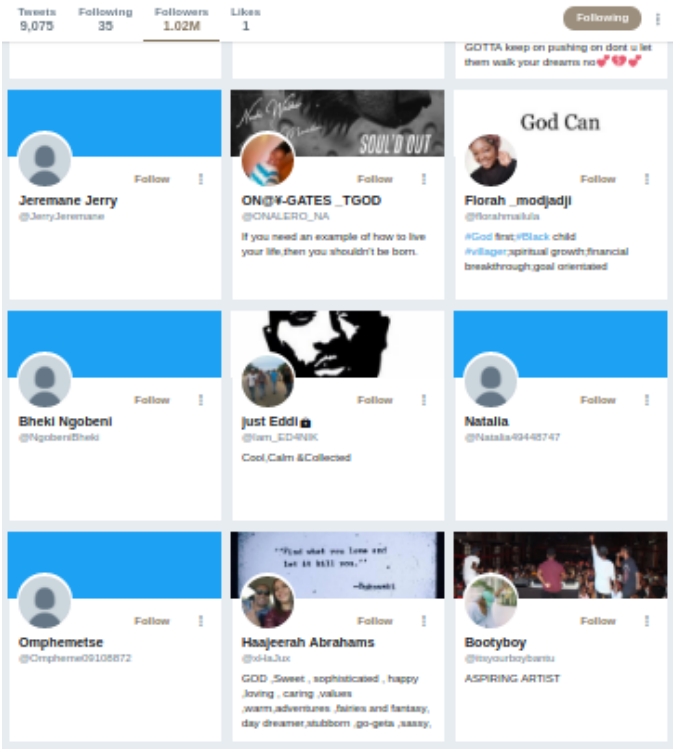

Fig. 1. An example of an account being followed by large numbers of "eggs". After the Twitter cull these "eggs" were easy to spot and assess for attributes likely to identify bots.

Fig. 2. The account @avodah4 has 211 likes but must have deleted the "liked" tweets because the timeline is empty. This is an attribute commonly found on automated and potentially malicious accounts.

Having described the known behaviour of bots and introducing the methods with which they can be identified, the rest of this paper is structured as follows; Section III presents the method and describes the characteristics of the accounts chosen for analysis. Section IV presents findings. Section V presents some of the accounts as they stand six months after the Twitter cull.

III. METHOD

A small pre chosen sample of South Africa Twitter accounts were randomly picked based on the fact that they had a large number of followers and differed in character from each other.

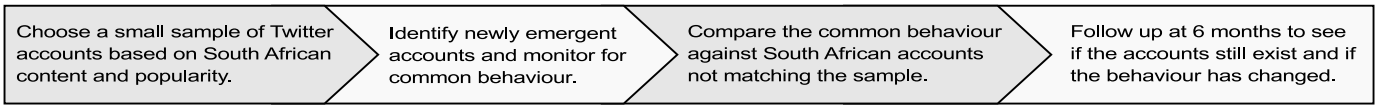

Fig. 3. The process that was followed to identify the common behaviour of potentially automated, semi-automated, or malevolent accounts in the wake of the July 2018 attempt by Twitter to bulk remove the more malicious accounts.

Popular accounts were chosen the literature indicates that they are a popular target. The accounts chosen are listed in Table 10. The follower counts (indicated in brackets behind the Twitter handle) are the counts as they were in July 2018.

TABLE II
THE STARTING SAMPLE OF HIGH PROFILE SOUTH AFRICA TWITTER ACCOUNTS.

Account name	Account handle	Count	Type
amaBhungane	@amaBhungane	64.5K	Investigative journalists
Chief of police Johannesburg Metro	@AsktheChiefJMPD	15K	Metro Police Department
eNCA	@enca	1.77M	News Channel
Helen Zille	@helenzille	1.32M	Politician
Herman Mashaba	@HermanMashaba	131K	Politician
Organisation Undoing Tax Abuse (OUTA)	@OutaSA	42.6K	Civil society group
Parks Tau	@Parks_Tau	14.6K	Politician
Snow Report	@SnowReportSA	15.7K	Weather service
The Gautrain	@thegautrain	271K	Gauteng Train
Thuli Madonsela	@ThuliMadonsela3	1.1M	Public figure
University of Pretoria	@UPTuks	321K	University
Wits University	@WitsUniversity	76.1K	University

The "eggs" on the accounts were the target of observation. An example of the ease with which they can be noticed in the early stages is shown in Figure 1. Only accounts created in July were considered, making them even easier to identify.

The resulting observations were also checked against a one or two random individuals (like the Eskom Whistle Blower, Bianca Goodson (@goodson_bianca 615 followers), as well as South African born Trevor Noah's account (@Trevornoah) and that of his talk show (international account), the Daily Show (@TheDailyShow), in order to ascertain whether the results generalised beyond South African accounts with a large following. This process is illustrated in Figure 3.

IV. FINDINGS

Accounts with features of automation, systematically started following the same list of South Africa politicians. The bots did not follow all the accounts, but the accounts followed were always from the same list. The accounts that were followed are listed in Table III.

The accounts also followed exactly three international politicians, namely: Hillary Clinton, Donald J. Trump (@realDonaldTrump) and Barack Obama (@BarackObama).

The accounts all proceeded to follow the Twitter accounts belonging to South African television, newspapers, journalists and radio stations. They consistently followed the same accounts. The list of folloed South African television, newspapers and radio stations accounts are in Table IV. The followed journalists are listed in Table V.

TABLE III
BOTS SYSTEMATICALLY STARTED FOLLOWING THE SAME LIST OF SOUTH AFRICA POLITICIANS.

Account name	Account handle	Type
Ayanda Dlodlo	@MinAyandaDlodlo	Politician
Bantu Holomisa	@BantuHolomisa	Politician
Dali Mpofu	@AdvDali_Mpofu	Politician
Floyd Shivambu	@FloydShivambu	Politician
Gwede Mantashe	@GwedeMantashe1	Politician
Fikile Mbalula	@MbalulaFikile	Politician
Helen Zille	@helenzille	Politician
Julius Sello Malema	@Julius_S_Malema	Politician
Lindiwe Sisulu	@LindiweSisuluSA	Politician
Mmusi Maimane	@MmusiMaimane	Politician
Patricia de Lille	@PatriciaDeLille	Politician
Paul Mashatile	@PaulMashatile	Politician
President Cyril Ramaphosa	@CyrilRamaphosa	President
Tito Mboweni	@tito_mboweni	Politician
Zwelinzima Vavi	@Zwelinzima1	Politician
Cope	@COPE_SA	Political party
Economic Freedom Fighters	@EFFSouthAfrica	Political party
Parliament of RSA	@ParliaamentofRSA	Political party
PresidencyZA	@PresidencyZA	Political party

An example of this behaviour can be seen in Figure 4.

Fig. 4. The account @Mihlali64923709 which has the attributes of a bot follows the same politicians, newspapers and journalists as other suspicious accounts.

The bot @Bakang12934859, followed 37 South African news sites. The account was subsequently suspended, confirming that the account was demonstrating malicious behaviour. @Natalia49448747 followed a disproportionate number of politicians (11 accounts) and 8 of the South African news accounts, before starting to follow random users.

TABLE IV
BOTS SYSTEMATICALLY STARTED FOLLOWING THE SAME TELEVISION, NEWSPAPERS AND RADIO STATIONS.

Account name	Account handle	Type
Beeld	@Beeld_Nuus	Newspaper
News 24	@News24	Newspaper
The Citizen	@TheCitizen_News	Newspaper
Daily Maverick	@dailymaverick	Newspaper
eNCA	@eNCA	Newspaper
SABC News Online	@SABCNewsOnline	Newspaper
Sunday Times	@SundayTimesZA	Newspaper
The Star news	@TheStar_news	Newspaper
Independant Online	@IOL	Newspaper
The Sunday Times	@SundayTimesZA	Newspaper
Sowetan Live	@SowetanLive	Newspaper
Eye Witness News Reporter	@ewnreporter	Newspaper
Eye Witness News	@ewnupdates	Newspaper
Pretoria news	@pretorianews	Newspaper
SAfm news	@SAfmnews	Newspaper
Rand Daily Mail	@rdm_za	Newspaper
Mail & Guardian	@mailandguardian	Newspaper
Times Live	@TimesLive	Newspaper
City Press	@City_Press	Newspaper
Jacaranda News	@JacaNews	Newspaper
Power FM 98.7	@Powerfm987	Radio Station
Radio 702	@Radio702	Radio station
Jacaranda FM	@jacarandafm	Radio station
MetroFM SABC	@METROFMSA	Radio station
Tuks FM	@TuksFM1072	Radio station
amaBhungane	@amaBhungane	Investigative journalists
SABC 1	@Official_SABC1	TV
SABC 2	@SABC_2	TV
SABC 3	@SABC3	TV
DST	@DStv	TV
ETV	@etv)	TV

TABLE V
BOTS SYSTEMATICALLY STARTED FOLLOWING THE SAME JOURNALISTS AND RADIO STATION HOSTS.

Account name	Account handle	Type
Annika Larsen	@AnnikaLarsen1	Journalist
Tanya Neft	@TanyaNeft	Journalist
Leigh-Anne Jansen	@LA_JANSEN	Journalist
Max du Preez	@MaxduPreez	Journalist
Adriaan Basson	@AdriaanBasson	Journalist
Ferial Haffajee	@ferialhaffajee	Journalist
Karyn Maughan	@karynmaughan	Journalist
John Robbie	@John_C_Robbie	Radio host
Gareth Cliff	@GarethCliff	Radio host
Redi Tlhabi	@RediTlhabi)	Radio host
Xolani Gwala	@gwalax)	Radio host

Each bot then chose at least one theme, such as amazing nature (Figure 5), CEO's, records (Figure 6), motivational quotes, operating systems, and religion, and followed ten or more Twitter accounts with that theme. The bot @Natalia49448747 followed 13 accounts with the theme "Amazing Nature", some of which didn't have any tweets and show markers of being bots (Figure 5). Some of the "Amazing Nature" bots pre-dated the July suspension of accounts. The bot @Zuko06435633 had the themes CEO's and Operating Systems.

After following the same list of politicians, news, and media accounts, the bot would start to follow random individuals. This is the point where it became increasingly difficult to identify the accounts. This is because the followed accounts start to show together and it becomes harder to see the themes, or compare the behaviour between bots. The occasional bot followed a non-English account (Hebrew or Chinese).

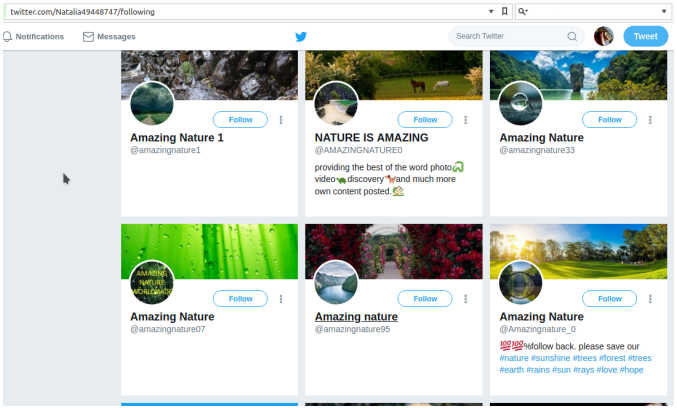

Fig. 5. The account @Natalia49448747 which has the attributes of a bot follows the theme, "Amazing nature". At the time of screenshot the account was following 13 "Amazing nature" accounts.

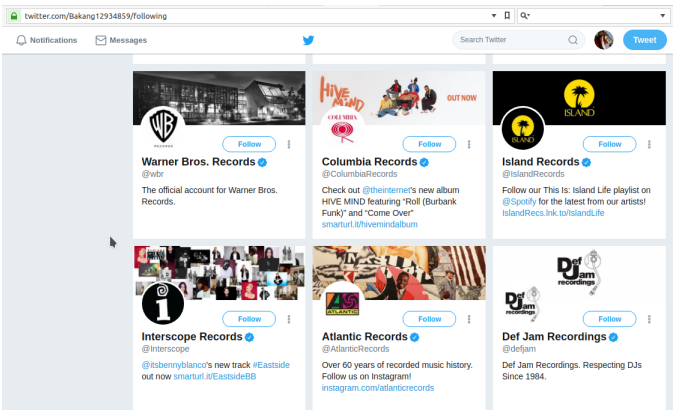

Fig. 6. The account @Bakang12934859 which had the attributes of a bot follows the theme, "Records". @Bakang12934859's account has subsequently been suspended.

The accounts could be seen to focus on controversial topics for a day and then the tweets were deleted and the account would retweet the accounts from the chosen theme. This can be seen in the Figures 7 and 8 which show @Bakang12934859's timeline two days apart. On the 21st July @Bakang12934859 is retweeting the accounts themes (proverbs and bible related content). Two days later these retweets have been deleted and the account is retweeting tweets about land, colonialism, Trump, and Putin. It was difficult to catch the days on which non-themed content was tweeted.

Innocuous accounts, like the Gautrain Twitter account, and the various universities, were also the target of the same behaviour. The large amount of non-human accounts following Gautrain can be seen on Botometer (Figure 9). The shown

result is for the first twenty accounts Botometer queries. The entries showing, “There was an error”, are Twitter accounts with no timeline. Viewing them revealed the typical likes-but-no-tweets behaviour. This trend is the same for all high profile accounts and matches the large numbers of emergent new profiles seen shortly after the Twitter cull.

Random accounts such as the whistle blower Bianca Goodson showed the same emergent bots as the other accounts, but on a smaller scale. This despite her small amount of followers. Trevor Noah’s personal account showed the same behaviour as the other high profile local accounts; however, the Daily Show Twitter account broke the trend. The bots emergent on this account (Such as @acollins417) still followed a high proportion of political and news accounts, but this time the politicians were American and the news sites were international. One account is not sufficient to generalise that the same programmed behaviour is adjusted for international accounts – implying the same group of malicious people behind the bots – but an investigation of themes may be useful.

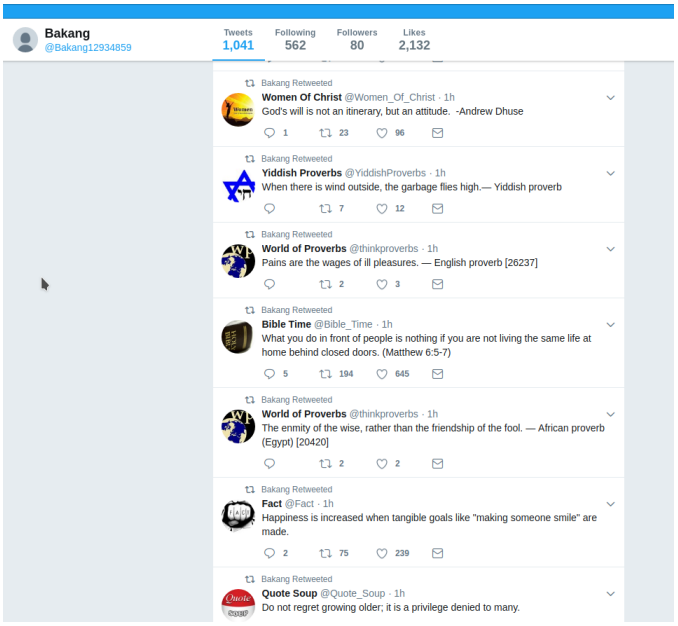

Fig. 7. On the 21 July the account @Bakang12934859 retweets the themes (proverbs and bible related accounts).

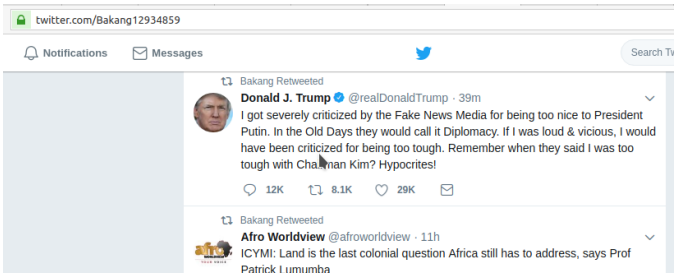

Fig. 8. On the 23rd of July the proverbs and bible related retweets have been deleted, and @Bakang12934859 is retweeting tweets about land, colonialism, Trump, and Putin.

Fig. 9. According to Botometer a large number of Gautrain followers have a high probability of being bots.

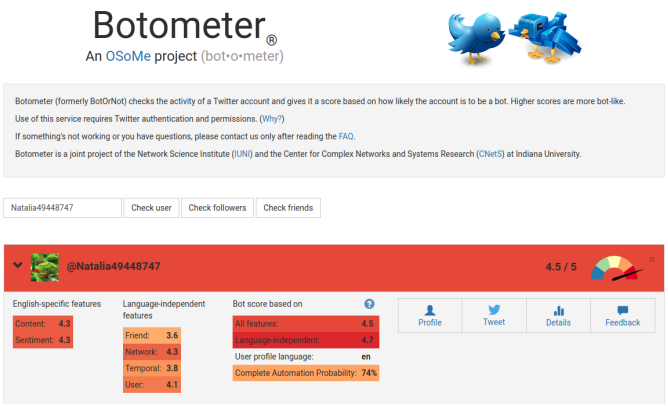

Fig. 10. @Natalia49448747 registers with a high probability of being a bot when queried against the Botometer tool.

V. Follow up on findings

Some of the accounts have been suspended by the Twitter maintainers, further supporting the fact that the accounts were behaving outside of Twitter’s agreed upon policy usage terms. Many still identify as malicious on Botometer. @avodah4 still has 211 likes and no timeline. Others accounts, @Natalia49448747 for example, still exist and are no longer obviously identifiable as bots. @Natalia49448747 follows 1263 users as of January 2019. The account has acquired a bible verses theme, an inspirational quotes theme,

a baby animal theme, and a Tupac theme. Some themes are slightly conspicuous, mostly due to the large amount of accounts with the same theme. The "Amazing Nature" theme still exists, but is harder to spot among the large amount of followed accounts. The originally followed newspaper, radio and political accounts are still being followed, but you have to scroll through thousands of accounts to see them, and the accounts are surrounded by other newspapers of international origin. @Natalia49448747 registers with a high probability of being a bot when queried against the Botometer tool (Figure 10).

A visual comparison between two bots no longer clearly shows the correlated behaviour. Bianca Goodson is no longer on Twitter, but there are bot accounts using her name.

VI. Conclusion

The new "eggs" of the accounts chosen for observation all showed the same behaviour. They followed a predefined set of newspapers, politicians and journalists, before following any other accounts. They then followed an unnaturally large number of accounts of a particular theme, before starting to follow (what appeared to be) random accounts. There was some indication that they may alternate between retweeting the theme on some days and sharing more controversial content occasionally.

While this study observed only a small group of accounts the consistent behaviour is concerning. The behaviour described is unsophisticated; however there is enough evidence to indicate that popular South African Twitter accounts are targets.

Future investigations could automate the collection of data for Twitter accounts and target a larger sample. Automation would more accurately pick up the behaviour and themes of the accounts. It could be useful to figure out why some accounts are active and others are dormant and try to identify the networks between accounts. It is clear that popular South African Twitter accounts are the target of manipulative behaviour. The unknown is why, who, and for what purpose? The limited nature of the programmatic access to Twitter makes it difficult to figure what propaganda is being spread, since keywords are generally pre-chosen for the collection of the data.

Obviously, this was a very small, manual investigation; however, it did definitively show the presence of large quantities of suspicious accounts targeting popular South African Twitter accounts. It also highlighted a potential target of these accounts for South African news, journalist, and politically associated accounts, which could be investigated further.

Highlighting these abusive accounts, may be useful for social media users, but may also assist the programmers of the malicious activity in further disguising the activity of the programmed bots.

References

[1] A. Badawy, E. Ferrara, and K. Lerman, "Analyzing the digital traces of political manipulation: The 2016 russian interference twitter campaign," *arXiv preprint arXiv:1802.04291*, 2018.

[2] W. Williamson and J. Scrofani, "Trends in detection and characterization of propaganda bots," in *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.

[3] C. Llewellyn, L. Cram, A. Favero, and R. L. Hill, "Russian troll hunting in a brexit Twitter archive," in *Proceedings of the 18th ACM/IEEE on Joint Conference on Digital Libraries*. ACM, 2018, pp. 361–362.

[4] D. A. Broniatowski, A. M. Jamison, S. Qi, L. AlKulaib, T. Chen, A. Benton, S. C. Quinn, and M. Dredze, "Weaponized health communication: Twitter bots and russian trolls amplify the vaccine debate," *American Journal of Public Health*, vol. 108, no. 10, pp. 1378–1384, 2018.

[5] M. Erbschloe, "Extremist propaganda in social media: A threat to homeland security," *Taylor & Francis*, 2018.

[6] K. Yang, O. Varol, C. A. Davis, E. Ferrara, A. Flammini, and F. Menczer, "Arming the public with artificial intelligence to counter social bots," *Human Behavior and Emerging Technologies*, vol. 1, no. 1, pp. 48–61, 2019.

[7] S. Bradshaw and P. N. Howard, "Challenging truth and trust: A global inventory of organized social media manipulation," *The Computational Propaganda Project*, 2018.

[8] M. N. Hussain, S. Tokdemir, N. Agarwal, and S. Al-Khateeb, "Analyzing disinformation and crowd manipulation tactics on YouTube," in *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 2018, pp. 1092–1095.

[9] R. Gorwa and D. Guilbeault, "Unpacking the social media bot: A typology to guide research and policy," *Policy & Internet*, 2018.

[10] M. Thamm. (2017, Jul) Analysis: Bell pottinger more than just spin, its political interference in sovereign states. [Online]. Available: https://www.dailymaverick.co.za/article/2017-07-05-analysis-bell-pottinger-more-than-just-spin-its-political-interference-in-sovereign-states

[11] V. Gadde. (2018, Jul) Confidence in follower counts. [Online]. Available: https://blog.twitter.com/official/en_us/topics/company/2018/Confidence-in-Follower-Counts.html

[12] J. de Villiers. (2018, Jul) Helen Zille and Julius Malema lose 50,000 followers in Twitter purge. [Online]. Available: https://www.businessinsider.co.za/sa-celebrity-twitter-follower-count-plunges-as-platform-rolls-out-blackout-of-inactive-accounts-2018-7

[13] S. Cresci, R. Di Pietro, M. Petrocchi, A. Spognardi, and M. Tesconi, "The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race," in *Proceedings of the 26th International Conference on World Wide Web Companion*. International World Wide Web Conferences Steering Committee, 2017, pp. 963–972.

[14] C. Yang, R. Harkreader, and G. Gu, "Empirical evaluation and new design for fighting evolving twitter spammers," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 8, pp. 1280–1293, 2013.

[15] P. G. Efthimion, S. Payne, and N. Proferes, "Supervised machine learning bot detection techniques to identify social Twitter bots," *SMU Data Science Review*, vol. 1, no. 2, p. 5, 2018.

[16] M. Piras, "Detection of suspicious users posting claims about cancer on twitter," Ph.D. dissertation, University of Illinois, 2018.