# Identifying and Tracking Disinformation during the May 2019 South Africa Elections: Analytical and Methodological Lessons from a Carter Center Pilot Mission

*Michael Baldassaro, Digital Threats Project Lead, The Carter Center*

*Prepared for The Workshop on Disinformation at Harvard University on October 4, 2019*

*Introduction & Overview*. The dissemination and spread of mis / disinformation and use of computational propaganda[1] are a source of growing concern around the world due to the threats they pose to the integrity of elections. According to international human rights law on democratic elections, voters must be able to form opinions independently, and to vote for and/or support candidates or issues without undue influence or manipulative interference.[2] Improving our ability to identify and track mis / disinformation in a rapid and efficient manner around elections is a critical first step towards assessing its impact on political participation and electoral integrity, and where possible, to mitigate negative impacts.

This paper draws on the work of a small technical election observation team deployed by The Carter Center (TCC) to assess the dissemination and spread of mis / disinformation ahead of the May 8, 2019 general elections in South Africa. The mission, composed of a small team of international and local researchers, experimented with a range of tools and methodological approaches to track mis / disinformation on multiple social media platforms. The goal of the mission was to work towards establishing a baseline methodological approach for identifying and tracking the spread of mis / disinformation and computational propaganda which could be adapted and expanded upon in future missions.

Although the mission took a multi-faceted approach to identifying mis / disinformation on multiple platforms, this paper focuses on TCC's efforts to identify and track mis / disinformation on Twitter usig a website domain classifier. The team scraped nearly 400,000 tweets from the "political mainstream" on Twitter, extracted the domain URLs from those tweets, and then applied a classifier developed by a credible South African NGO to identify which domains were classified as "dodgy," as a first step for identifying possible mis / disinformation. The paper details TCC's methodological approach, findings, and conclusions, as well as lessons learned for its future mis / disinformation monitoring activities.

*Context.* On May 8, 2019, South Africans voted in the sixth general elections since the end of Apartheid. While past elections have been largely credible, concerns over "fake news" were heightened in advance of the 2019 polls arising from two high-profile incidents: in 2016, the ruling African National Congress (ANC) sought to establish a clandestine "war room" to engage in pro-ANC counter-messaging campaigns on social media platforms.[3] In 2017, the wealthy

[1] "Automation, Algorithms, and Politics | Political Communication, Computational Propaganda, and Autonomous Agents", International Journal of Communication, S.C. Woolley & P.N. Howard, p.9, 2016

[2] General Comment 25 on Article 25 of the International Covenant on Civil and Political Rights stipulates that "Persons entitled to vote must be free to vote for any candidate for election and for or against any proposal submitted to referendum or plebiscite, and free to support or to oppose government, without undue influence or coercion of any kind which may distort or inhibit the free expression of the elector's will. Voters should be able to form opinions independently, free of violence or threat of violence, compulsion, inducement or manipulative interference of any kind." Paragraph 19 (1996)

[3] "South Africa's ANC allegedly tried—but failed—to deploy fake news to win the 2016 election", Quartz Africa, January 25, 2017 https://qz.com/africa/894364/south-africas-ruling-anc-allegedly-recruited-a-black-ops-team-to-disseminate-fake-news-during-the-2016-election-campaign/.

South African Gupta family paid a UK-based communications firm Bell Pottinger to create hundreds of fake accounts and spread fake news on Twitter to divert social media discussions away from critical coverage of their business enterprises.[4] In the wake of these incidents, Media Monitoring Africa (MMA), a South African NGO dedicated to promoting ethical journalism, launched a web client browser extension called KnowNews to help web users distinguish between "credible" and "dodgy" online news sources.[5]

*Method.* The Carter Center mission focused on three main goals: (1) identify possible mis / disinformation from "dodgy" news sources that entered the "political mainstream", (2) assess the lifespan and potential reach of "dodgy" news links, and where possible, (3) identify possible deployment of computational propaganda to amplify mis / disinformation.

Given the variety, volume, and velocity of social media,[6] it is neither possible nor practical to detect all mis / disinformation that could affect voter behavior. Although South Africans use Facebook (16M±) than Twitter (8M±), interviewees noted that Twitter was the predominant platform for political and electoral discourse. This is reinforced by the larger numbers of followers on Twitter of political parties and candidates on Twitter than on Facebook.  As a result, TCC focused on Twitter as the key platform where the bulk of political discourse takes place.

To efficiently identify possible mis / disinformation on Twitter, between April 8 and May 8, 2019, TCC scraped 379,877 tweets from the "political mainstream" using the Twitter API to extract domains, and using the KnowNews domain classifier, created and maintained by MMA, to identify "dodgy" domains that could suggest possible mis / disinformation.  To approximate the "political mainstream," TCC scraped tweets that were shared by the official accounts of the three main political parties (specifically ANC, Democratic Alliance, and Economic Freedom Fighters), or which mention political parties and leaders of the main parties, as well as tweets tagged with prominent election-related hashtags (e.g. #Elections2019, #SAElections2019, #MzansiVotes, etc.).

*Findings.* Among tweets scraped from the "political mainstream," TCC found 32,202 tweets containing 884 distinct domains and evaluated those domains using KnowNews.  Of the 884 distinct domains, TCC found that 10 of these domains were classified as "dodgy" sources. The team used the Twitter API to search for tweets by domain, collected an additional 13,373 tweets, and identified 608 unique links from the previously identified "dodgy" sources.

With user follower counts as a proxy for reach, TCC found that the 608 unique links were retweeted 161,206 times and had a potential reach of up to 10,756,201 Twitter users.[7] While the vast majority of links from "dodgy" news sources gained little to no traction beyond an initial

[4] "The Guptas, Bell Pottinger and the fake news propaganda machine" Sunday Times, September 4, 2017. timeslive.co.za/news/south-africa/2017-09-04-the-guptas-bell-pottinger-and-the-fake-news-propaganda-machine

[5] The KnowNews domain classifier labels sources as "credible" or "dodgy" based on an evolving set of criteria that can be found here: https://newstools.co.za/page/knownews. For more details about the methodology, see "Innovator Q&A: Media Monitoring Africa's Thandi Smith on KnowNews", Medium, October 2, 2017 https://medium.com/jamlab/innovator-q-a-media-monitoring-africas-thandi-smith-on-newscred-1eb0d8fdcbf9

[6] "Can Democracy Survive the Internet?", Journal of Democracy 63, Nathaniel Persily, 2017

[7] It is likely that there is some overlap among followers between accounts however for the purposes of this limited scope pilot mission, a proxy count based on absolute followers was used.

tweet or retweet, some (15) links gained significant traction, each with a potential reach of more than 100,000 Twitter users.

**Table 1. 10 "dodgy" domains, number of links, tweets, & followers reached through shares**

| Domain | Links | Tweets | Followers |
|---|---|---|---|
| briefly.co.za | 251 | 1209 | 2962092 |
| southafricatoday.net | 144 | 151378 | 3858198 |
| newsoweto.co.za | 73 | 4643 | 1984551 |
| africanews24-7.co.za | 63 | 66 | 275065 |
| news360.co.za | 27 | 3327 | 763298 |
| uncensoredopinion.co.za | 20 | 381 | 775698 |
| xpouzar.com | 13 | 76 | 57709 |
| blackopinion.co.za | 9 | 121 | 64914 |
| sa-news.com | 4 | 0 | 4973 |
| weeklyxpose.co.za | 4 | 5 | 9703 |
| **Total** | **608** | **161206** | **10756201** |

Due to the limited scale and scope of the mission, TCC did not have the time and resources to assess the accuracy of every possible link that was shared. Nor did the team assess links to news articles that were also shared by "credible" news sources. Although the TCC team was unable to verify the accuracy of each news article on the "dodgy" links, the team found that most of the links from these "dodgy" sources seemed to be generally credible or at least ignorable and/or apolitical. However, TCC researchers suspected that 21 links appeared to contain possible mis / disinformation that could distort public opinion. Links from "dodgy" news sources that played on racial divisions, which pervade the political landscape,[8] were often identified as containing potential mis / disinformation and seemingly gained the most traction.

**Table 2: 21 links suspected of containing possible mis / disinformation, highest to lowest**

| Link | Domain | Tweets | Followers |
|---|---|---|---|
| eskom white contractors are destroying our substations to get jobs to repair them | newsoweto.co.za | 4317 | 1747656 |
| multichoices boss is the son in law of the ancs gwede mantashe | southafricatoday.net | 81580 | 503383 |
| farm murder farmer gunned down attackers wearing police gear hoopstad | southafricatoday.net | 15360 | 335365 |
| 13 farm attacks 3 farm murders over the easter weekend in south Africa | southafricatoday.net | 2812 | 219224 |
| destruction of south africa protests looting arson killings pretoria | southafricatoday.net | 41 | 149333 |
| ancs own idiotic racist policies to blame for division and not the old sa flag | southafricatoday.net | 2429 | 101974 |
| barbaric uncivilised violence farm attacks and the savage | southafricatoday.net | 2741 | 78774 |
| multichoice supports killing of whites by eff and blf but not an afrikaans singer | southafricatoday.net | 4984 | 62318 |
| malema asked standard bank to fire only white workers report | newsoweto.co.za | 1 | 47754 |

[8] "South Africa's Politics is Still a Racial Minefield", Keith Gottschalk, University of Western Cape, Quartz, May 3, 2019 https://qz.com/africa/1611658/south-africas-politics-is-still-a-racial-minefield/

| | | | |
|---|---|---|---|
| blf and other black leaders call for war on white south africans | southafricatoday.net | 1 | 44624 |
| blf mngxitama fires threat current looting and riots in sa is a dress rehearsal for civil war against whites | news360.co.za | 87 | 22700 |
| shivambu got less than vbs loot malema asks why fuss | briefly.co.za | 143 | 19767 |
| how zumas nuclear deal could have saved sa from power crisis | briefly.co.za | 2 | 19749 |
| ntsiki mazwai joins call christian holidays removed | briefly.co.za | 1 | 15046 |
| julius malema claims cyril ramaphosa offered a cabinet position | briefly.co.za | 1 | 14717 |
| south africa will be renamed azania after 8 may eff | news360.co.za | 3 | 12660 |
| nothing could stop us from expropriating the white residential suburbs without compensation andile mngxitama | xpouzar.com | 4 | 9264 |
| if you didnt believe whites created cyclone idai you are under their spell | newsoweto.co.za | 2 | 5186 |
| land grab in progress yet president cyril ramaphosa of south africa once again said that land grabbing has not happened in south Africa | sa-news.com | 1 | 3314 |
| anc councillor linked to dead people listed on special vote list in Limpopo | newsoweto.co.za | 1 | 2754 |
| blf leader celebrates the farm murder of 2 elderly couple shot while sleeping | newsoweto.co.za | 1 | 2400 |

For example, one “dodgy” news link alleging white contractors were sabotaging state-run electrical stations to obtain repair contracts gained the most traction after it was retweeted by a senior ruling party official (ANC Director of Elections and former Deputy Minister of Police Fikile Mbalula) to his more than 1.6 million followers.[9] The “dodgy” news link, which cited unnamed sources, was tweeted more than 4,000 times between April 9 and May 6 (two days before Election Day), with an estimated possible reach of nearly 1.8 million Twitter followers. While there may have been others that the mission missed, this was the only instance TCC found where a possible mis / disinformation link was shared by a political figure and thus had some legitimacy presumably conferred upon it.

Another example highlights evidence of computational propaganda. Links from the “dodgy” news source southafricatoday.net, which researchers identified as a hyper partisan pro-Afrikaner news source, had a suspiciously high number of retweets disproportionate to its follower base. The source’s Twitter account had just 8,543 followers, but its 144 links were collectively retweeted 151,378 times.[10] TCC sampled 557 of the 8,543 Twitter user profiles that shared southafricatoday.net links and using a bot detection algorithm, identified 364 of the 557 users as “likely bots” with a 60 percent or greater probability.[11]

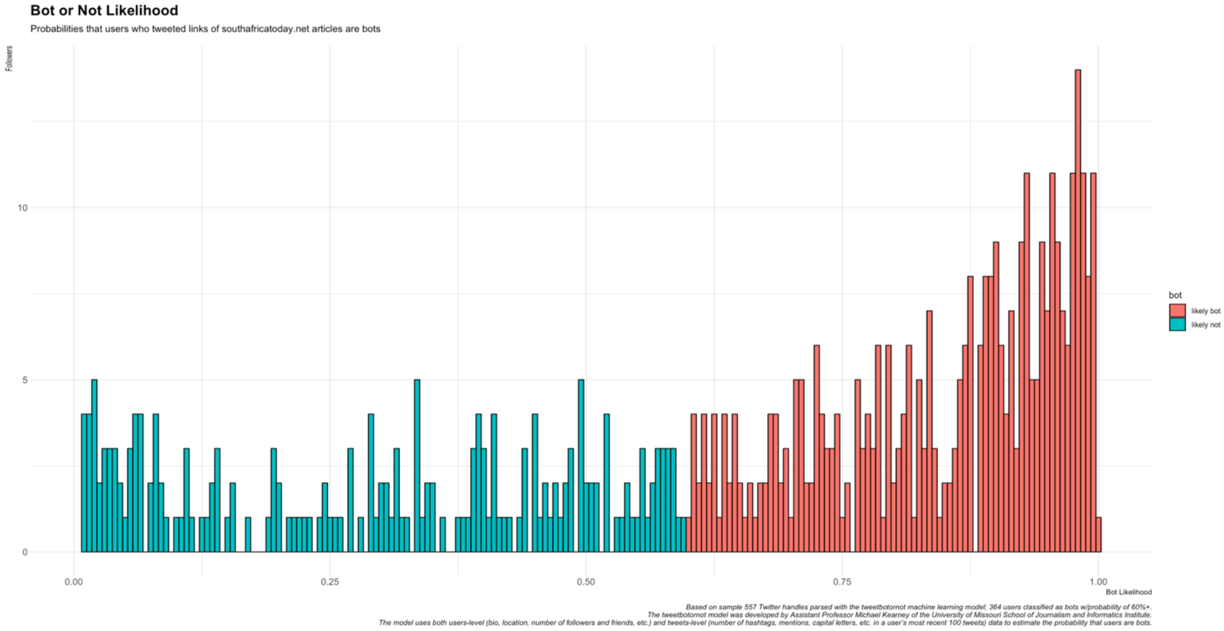
Due to time and resource limitations, TCC was unable to investigate and verify how many accounts were “actual bots” however the high probability scores attributed to most identified bots strongly suggests that many “likely bots” are indeed actual bots.

[9] “Eskom: White contractors are destroying our substations to get jobs to repair them” Newsoweto.co.za, April 9, 2019 https://newsoweto.co.za/eskom-white-contractors-are-destroying-our-substations-to-get-jobs-to-repair-them/

[10] For comparison, the team scraped the last 100 tweets containing by the New York Times, two mainstream South African news sources – The Independent Online (IOL) and News24 – and South Africa Today. Here is the average ratio of tweets per follower for each news links: New York Times 1:4,017; IOL 1:864; News24 1:1,424; South Africa Today: 1:2.2.

[11] Researchers used the “sampler” R package to draw a random sample of unique Twitter accounts that tweeted any link that contained a southafricatoday.net domain: https://www.rdocumentation.org/packages/sampler/versions/0.2.4. And researchers used the “tweetborornot” model from the eponymous R package developed and maintained by Professor Michael Kearney at the University of Missouri School of Journalism: https://tweetbotornot.mikewk.com/.

**<u>Table 3: Histogram of “bot” probabilities (n=557;bot  ≥ 60% = 364) of users tweeting southafricatoday.net links.</u>**


Bot or Not Likelihood

Probabilities that users who tweeted links of southafricatoday.net articles are bots

*Based on sample 557 Twitter handles parsed with the tweetbotornot machine learning model; 364 users classified as bots w/probability of 60%+.*
*The tweetbotornot model was developed by Assistant Professor Michael Kearney of the University of Missouri School of Journalism and Informatics Institute.*
*The model uses both users-level (bio, location, number of followers and friends, etc.) and tweets-level (number of hashtags, mentions, capital letters, etc. in a user's most recent 100 tweets) data to estimate the probability that users are bots.*

*Conclusions.* The TCC misison’s approach highlighted the obvious benefits of using a domain classifier to facilitate the identification of possible political and election-related mis / disinformation. As noted above, most of the links from sources that were previously identified as “dodgy” were considered to be generally credible or ignorable (i.e. apolitical content).  However, starting with a base of suspected “dodgy” content greatly facilitated the effort; it would have been far more difficult to detect mis /disinformation absent having the pre-existing domain classifier. The use of a domain classifier allowed for a sharper focus on content from possible mis / disinformation sources and, in turn, allowed more time to devote to investigating spread. Related, the research also attempted to distinguish between mis / disinformation that is amplified and legitimized by political figures (presumably unintentionally), versus mis /disinformation that is amplified nefariously vis-à-vis computational propaganda (intentionally).

While TCC’s approach is demonstrably feasible for identifying possible mis / disinformation and can be replicated, there are theoretical and practical limitations that make it suboptimal. Foremost, by itself, this approach does not provide a contextual lens through which to understand possible mis / disinformation within the broader political media ecosystem – or even within sub-ecosystems in which specific narrative frames may pervade.[12] Seeking to identify possible mis / disinformation in isolation, however useful, eschews more rigorous and valuable analysis of intent or significance, i.e., whether information was circulated within an ideological network, whether it reinforced pre-existing narrative frames, etc.

[12] See <u>Network Propaganda</u>, Benkler, Faris, & Roberts (2018). From Chapter 2, p 45: “*To understand media and politics, we must understand the entire ecosystem: the outlets and influencers who form networks, the structure of networks, and the flow of information in networks…Some patterns of information flow emerge from organic, decentralized processes, and some are caused by intentional manipulation and marketing by centralized actors – most prominently political campaigns and state propaganda.”*

Moreover, a binary domain classification typology – "dodgy" v. "not-dodgy" -- is of limited utility given that it does not distinguish among "dodgy" sources in any meaningful way that could indicate motivation (i.e. satire, clickbait, hyper-partisan journalism, etc.). While a such a binary classifier is useful for basic digital literacy purpose for the general population, a classifier with a more diverse and nuanced typology would be more useful and beneficial to facilitate identification of possible mis / disinformation in an electoral context. A multi-class typology developed based on a rigorous methodology and criteria, such as the "Junk News Aggregator" developed by Oxford Internet Institute Computational Propaganda project researchers, may be more appropriate.[13]

The development of a multi-class typology domain classifier and identification of sub-ecosystems could be mutually reinforcing exercises. A carefully designed domain classifier that takes the broader political media ecosystem into consideration when developing a typology could enable the identification of sub-ecosystems for monitoring purposes. Typological designations of domains as hyper-partisan journalism could be sub-categorized in accordance with editorial biases and, in turn, be used to identify ideological sub-ecosystems for monitoring purposes. Conversely, if ideological sub-ecosystems are first identified, typological designations of domains could be denoted in accordance with the sub-ecosystems in which they are propagated.

As part of efforts to develop a baseline methodological approach and associated tools, future TCC monitoring missions will continue to test approaches for identifying and analyzing possible mis / disinformation through the contextual lens of the political media ecosystem and sub-ecosystems, facilitated by the use of a multi-class typology domain classifier. TCC will seek to first assess the broader political media ecosystem, identify sub-ecosystems, and monitor information flow within sub-ecosystems. This will involve extracting domains shared within these sub-ecosystems and categorizing them using a more robust classification methodology and typology. The goal here is to determine the feasibility of analyzing the motivation and significance of possible mis / disinformation spread within or across sub-ecosystems.

More broadly, this work will contribute to the larger and more challenging objectives that TCC and other election groups are working toward regarding approaches and techniques to allow analysts and observers to assess the degree to which mis / disinformation impact overall electoral integrity and effective political participation. Unlike other problems where impact is easier to isolate and quantify—such as voter fraud, ballot stuffing, manipulation of vote counts, and laws that restrict the right to stand, etc.—it will be extraordinarily difficult to measure the impact of mis /disinformation in a meaningful way, given the complex dynamics involved in determining the degree to which any specific piece of information affects voter behavior. Further dialogue regarding ways to frame possible approaches are needed. In the short term, observers are starting with approaches that flag and count incidents, and violations of laws or regulations.

---

*Michael Baldassaro is the Digital Threats Project Lead at The Carter Center with more than 15 years of experience designing, managing and implementing election integrity projects in Africa, Asia, Europe, and the Middle East. Prior to The Carter Center, Michael was the Director of Research, Evidence and Data at Democracy International and Elections Program Manager at the National Democratic Institute.*

[13] The Oxford Internet Institute Computational Propaganda Junk News Aggregator methodology can be found here: https://newsaggregator.oii.ox.ac.uk/methodology.php. A detailed overview of the typology can be found here: https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/05/EU-Parliamentary-Elections-Supplement.pdf.