



**Selected Papers of #AoIR2020:
The 21st Annual Conference of the
Association of Internet Researchers**
Virtual Event / 27-31 October 2020

‘COORDINATED INAUTHENTIC BEHAVIOUR’ AND OTHER ONLINE INFLUENCE OPERATIONS IN SOCIAL MEDIA SPACES

Tobias R. Keller
Digital Media Research Centre, Queensland University of Technology

Timothy Graham
Digital Media Research Centre, Queensland University of Technology

Daniel Angus
Digital Media Research Centre, Queensland University of Technology

Axel Bruns
Digital Media Research Centre, Queensland University of Technology

Nahema Marchal
Oxford Internet Institute, University of Oxford

Lisa-Maria Neudert
Oxford Internet Institute, University of Oxford

Rolf Nijmeijer
LUISS Guido Carli

Kristoffer Laigaard Nielbo
Aarhus University

Marie Damsgaard Mortensen
Aarhus University

Suggested Citation (APA): Keller, Tobias, Graham, Tim, Angus, Dan, Bruns, Axel, Marchal, Nahema, Neudert, Lisa-Maria, Nijmeijer, Rolf, Nielbo, Kristoffer Laigaard, Mortensen, Marie Damsgaard, Bechmann, Anja, Rossini, Patricia, Stromer-Galley, Baptista, Erica Anita, & Veiga de Oliveira, Vanessa. (2020, October 28-31). *‘Coordinated Inauthentic Behaviour’ and Other Online Influence Operations in Social Media Spaces*. Panel presented at AoIR 2020: The 21th Annual Conference of the Association of Internet Researchers. Virtual Event: AoIR. Retrieved from <http://spir.aoir.org>.

Anja Bechmann
Aarhus University / University of Antwerp

Patrícia Rossini
University of Liverpool

Jennifer Stromer-Galley
Syracuse University

Erica Anita Baptista
Federal University of Minas Gerais

Vanessa Veiga de Oliveira
Federal University of Minas Gerais

Panel Description

Responding to criticism of their role in the dissemination of hyperpartisan political propaganda (Faris et al. 2017), ‘fake news’, and other malinformation (Wardle & Derakhshan 2017), leading social media platforms such as Facebook and Twitter have recently made a number of major public announcements about their efforts to identify and mitigate the use of human and automated accounts in online influence operations that often pursue overtly political aims.

In a significant move, Facebook has increasingly adopted the term “coordinated inauthentic behaviour” (Gleicher 2018) to describe such activities; this shift in corporate language points to the fact that malinformation campaigns are becoming more and more sophisticated in their use of large numbers of accounts that individually may not appear to be especially aberrant in their activities, but in combination interfere substantially with ordinary user activities. This development is not limited to Facebook alone: other social media platforms have similarly highlighted such coordinated activity.

In spite of efforts by platform providers to address the misuse of their platforms by domestic as well as foreign, commercial as well as political operators, however, independent scholarly research continues to produce evidence that such operations persist unabated. Such research, too, must increasingly come to terms with coordinated inauthentic behaviour: it must advance beyond approaches and tools that examine only the activity patterns of *individual accounts* for evidence of automated behaviour, and pay increasing attention to the signs of coordination *across a large number of such accounts*. Such work is further complicated by the often precarious and unreliable nature of data access to the various major social media platforms, however (Bruns 2019).

In pursuit of the conceptual and methodological innovation required to address the challenge of coordinated inauthentic behaviour, this panel brings together a selection of recent studies that advance the methods available for the forensic, mixed-methods, in-depth and large-scale analysis of inauthentic information operations, and present

important new findings from current analyses of such activities across platforms and countries.

Paper 1, “#ArsonEmergency: Climate Change Disinformation during the Australian Bushfire Season 2019-2020”, conducts a forensic analysis of coordinated inauthentic behaviour by Twitter accounts during the recent catastrophic bushfires in Australia. The paper uses computational tools to examine the activity patterns of a population of accounts that sought to promote the debunked claim that arson rather than climate change was the root cause for the severity of the fires, and retraces the trajectory of the #ArsonEmergency hashtag from its first emergence to being amplified by Donald Trump Jr.’s tweets.

Paper 2, “Investigating Visual Media Shared Over Twitter During the 2019 European Parliamentary Elections”, shifts our focus to the 2019 European Parliament election, and investigates the content and distribution strategies for memes and other visual content on Twitter during the campaign. It finds that images from traditional political actors remained prominent during the campaign, and that anti-European and populist imagery, often also expressed in a humorous fashion, accounted for a substantial amount of posts but largely remained disconnected from mainstream debate.

Paper 3, “Different Narratives for Different People? Understanding Deviant Semantic Effects across Countries in the Political Campaigns of the European Parliament Election 2019”, complements the previous study by exploring whether Facebook’s microtargeting advertising functionality allows political parties to promote conflicting narratives to different groups of people, without making such divergent campaign messaging sufficiently transparent. It draws on the Facebook Ad Library to identify deviant semantics in 2019 European Parliament election advertising, and identifies the factors that predict such campaign strategies.

Finally, paper 4, “Explaining Dysfunctional Information Sharing on WhatsApp and Facebook in Brazil”, turns our attention also to a critical emerging space for the fight against malinformation: WhatsApp. It examines the respective roles that WhatsApp and Facebook in the accidental as well as intentional spreading of political malinformation in Brazil, finding especially that more engaged users are also more likely to have shared mis- and disinformation, and that social correction of such information is more likely to occur on WhatsApp than Facebook.

In combination, these contributions advance the study of online influence operations by providing key new methodological and empirical impulses for malinformation research across a range of social media platforms. In particular, they enable an independent assessment of the extent and impact of “coordinated inauthentic behaviour” beyond the occasional glimpses that platform providers’ press releases can offer. In doing so, these studies provide pointers on how to move forward in the ongoing arms race between those who initiate covert malinformation operations and those who seek to detect and counter them.

References

- Bruns, A. (2019). After the 'APIcalypse': Social Media Platforms and Their Fight against Critical Scholarly Research. *Information, Communication & Society*, 22(11), 1544–1566. <https://doi.org/10.1080/1369118X.2019.1637447>
- Faris, R., Roberts, H., Etling, B., Bourassa, N., Zuckerman, E., & Benkler, Y. (2017). *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election* (Berkman Klein Center Research Publication No. 2017–6). Social Science Research Network. <https://papers.ssrn.com/abstract=3019414>
- Gleicher, N. (2018, December 6). Coordinated Inauthentic Behavior Explained. *Facebook Newsroom*. <https://newsroom.fb.com/news/2018/12/inside-feed-coordinated-inauthentic-behavior/>
- Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making* (DGI(2017)09). Council of Europe. <https://shorensteincenter.org/wp-content/uploads/2017/10/Information-Disorder-Toward-an-interdisciplinary-framework.pdf>

PAPER 1

#ARSONEMERGENCY: CLIMATE CHANGE DISINFORMATION DURING THE AUSTRALIAN BUSHFIRE SEASON 2019-2020

Tobias R. Keller

Digital Media Research Centre, Queensland University of Technology

Timothy Graham

Digital Media Research Centre, Queensland University of Technology

Daniel Angus

Digital Media Research Centre, Queensland University of Technology

Axel Bruns

Digital Media Research Centre, Queensland University of Technology

Introduction

The 2019-2020 Australian bushfire season was devastating (UN Environment, 2020): an estimated 18 million hectares burned, millions of animals died, and public health was threatened when the national capital Canberra measured the worst air quality index of any major city worldwide. During this bushfire season, scientists pointed to the link between the bushfires and climate change (Abram, 2019): increasing temperatures in Australia mean that bushfires will become ever more extreme and difficult to contain.

Yet with social media platforms including Twitter becoming the major news source for Australians (*Digital News Report*, 2019), citizens did not only encounter evidence-based information regarding the major issues in the 2019-2020 bushfire season online. These

platforms have also become a fertile habitat for bots and trolls that spread propaganda, conspiracy theories, and disinformation (Shao et al., 2018). Initial research into the public debate around the 2019-2020 bushfire season identified the hashtag #ArsonEmergency as a particular focus of such activity (Graham, 2020). The hashtag was used especially to spread the - false (ABC News, 2020) - claim that the majority of the bushfires had been caused by arson, and were therefore due to human activity rather than an increasingly extreme climate.

Our exploratory research therefore focusses on two questions:

RQ1: What is the history of the hashtag #ArsonEmergency?

RQ2: What is the role of bots and trolls in the hashtag #ArsonEmergency?

Methods

Using *twint*, we collected all tweets including the hashtag #ArsonEmergency for 1 to 6 January 2020 (1,340 tweets from 315 accounts). To generate comparable data on the number of bots and trolls within the dataset, we also collected all tweets including two other popular hashtags for debating the Australian bushfires, using the same tool and for the same period: for the hashtag #BushfireAustralia, *twint* returned 1,309 tweets from 1,106 unique accounts, and for the hashtag #AustraliaFire it retrieved 4,592 tweets from 3,966 unique accounts. Finally, we also collected every (Australian) news article including the words 'arson*' and 'bushfires' for 2019 and 2020, and traced the hashtag #ArsonEmergency back to its first appearance on Twitter.

To determine the history of the hashtag #ArsonEmergency, we collected the search behaviour of Australian Google Search users for the terms 'arson' and 'bushfires' via Google Trends. We synchronised peak search behaviour with the publication of tweets and news articles to explore which tweets or news articles may have sparked Australian users' interest to search for the alleged links between arson and the bushfires.

We then analysed all detected Twitter accounts with two tools designed to spot bots and trolls on Twitter: Botometer is a machine learning algorithm that compares the presentation and behaviour of a Twitter account with a set of known bots and calculates a score that describes its resemblance to a bot (Varol et al., 2018). BotSentinel, too, is a machine learning algorithm and was trained on trollbots (BotSentinel, 2020): these accounts distinguish themselves from bots in that they may not be fully automated and aim to be highly active and/or to hinder rational debate.

Results

We identified four critical elements that sparked an increased interest for Australians to search for arson and the bushfires in 2019-2020 (see Figure 1):

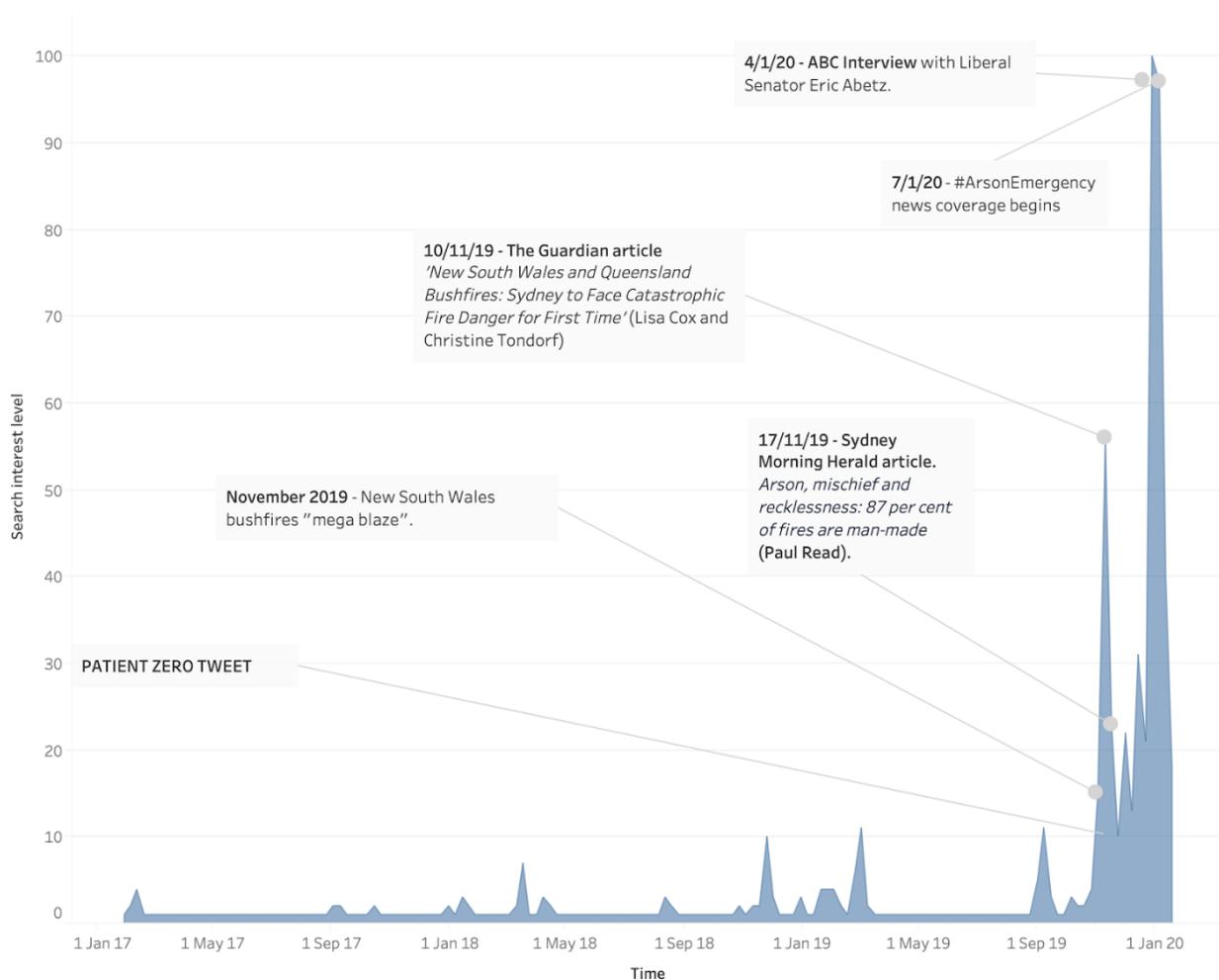
1. Sydney declared a catastrophic fire danger for the first time in November 2019, and Australians' interest in the possible links between arson and the bushfires increased (*The Guardian*, 10 Nov. 2019).
2. A *Sydney Morning Herald* article on 17 Nov. 2019 directed attention to the alleged links between arson and the bushfires by claiming in its title "Arson [...]:

87 percent of fires are man-made”. Some Twitter users cited this article as an argument to get the new #ArsonEmergency hashtag trending.

3. One such account tweeted on 21 November 2019: “With the vast majority of fires being deliberately lit, a better hashtag for the bushfires instead of #ClimateEmergency would be #ArsonEmergency”, and received over 500 likes and 130 retweets (“Patient Zero Tweet” in Figure 1).
4. When prominent international and domestic politicians such as Donald Trump Jr. (on Twitter, 8 Jan. 2020) or the arch-conservative Liberal Party senator Eric Abetz (to the national broadcaster ABC, on 4 Jan. 2020) repeated these claims, Australians’ search for this link peaked.

Thus, the idea of an ‘arson emergency’ can be identified as a counter-narrative to concerns about climate change, relying on deconstructed news (i.e. disinformation) and supported by prominent climate change deniers to national and international audiences.

Google Trends search interest for “arson + bushfires” (past three years)



The plot of sum of F2 for Category: All categories Day. The data is filtered on Category: All categories Day, which keeps 162 of 262 members. The view is filtered on sum of F2, which keeps non-Null values only.

Figure 1. Google Search history for Australians searching for “arson + bushfires” from January 2017 to January 2020, linked with crucial events for the hashtag #ArsonEmergency.

We relied on Botometer to compare the number of fully automated, bot-like accounts spreading #ArsonEmergency to those spreading #BushfireAustralia and #AustraliaFire (see Figure 2a). Overall, the distribution of bot-likeness scores between the three hashtags is almost identical. On average, accounts that contributed to #ArsonEmergency received a score of 0.18 (SD = 0.14), which is statistically significantly lower than those for accounts contributing to #BushfireAustralia ($t = 5.3$, $df = 704$, $p < 0.001$) and #AustraliaFire ($t = 8.8$, $df = 462$, $p < 0.001$).

However, analysis using BotSentinel shows a different picture (see Figure 2b): While those accounts contributing to #AustraliaFire (mean = 0.08, SD = 0.12) and #BushfireAustralia (mean = 0.08, SD = 0.1) have a very similar distribution of trollbot-like scores, those accounts contributing to #ArsonEmergency showed significantly higher scores on average ($t = 12$, $df = 268$, $p < 0.001$ to #AustraliaFire and $t = 14$, $df = 324$, $p < 0.001$ to #BushfireAustralia).

Although contributors to #ArsonEmergency did not resemble fully automated *bots*, then, they clearly exhibited a behaviour that was similar to that of known *trollbots*, in BotSentinel’s definition. That is, these accounts aim to hinder rational debate, are highly active, and seek to sow discord among ordinary users.

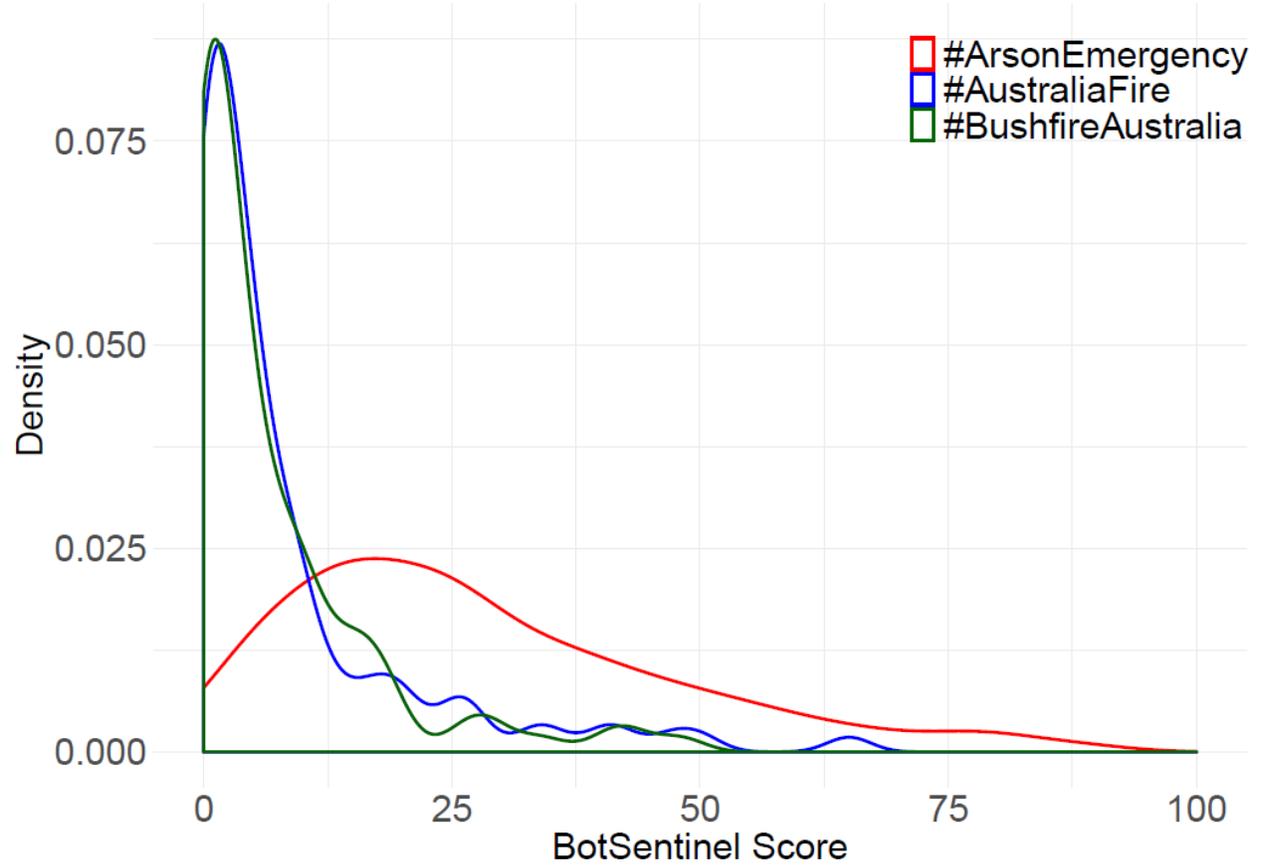
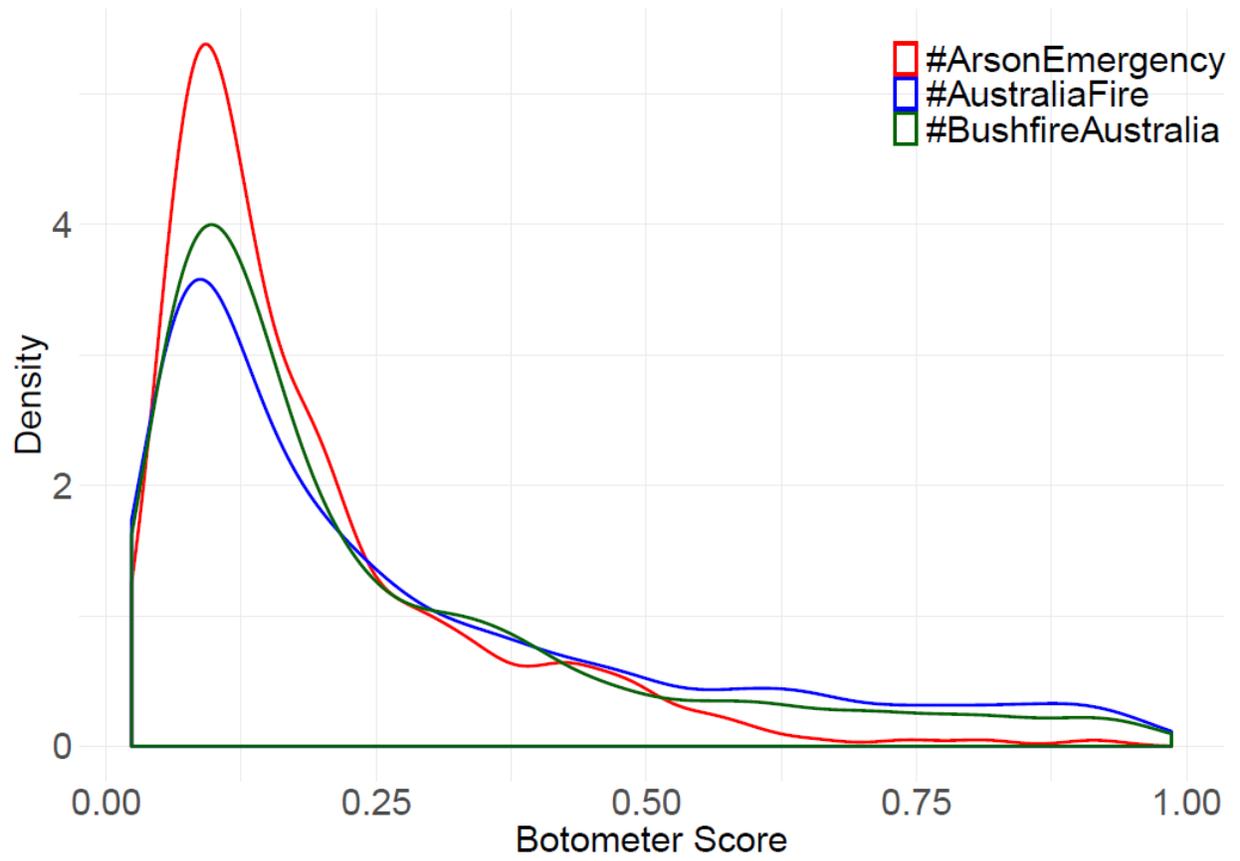


Figure 2a+b. Density plot of Botometer results (top) and BotSentinel results (bottom) for a sample of 100 random accounts for each of the hashtags #BushfireAustralia (green) and #AustraliaFire (blue), and for all 315 accounts spreading #ArsonEmergency (red).

Discussion

In a crisis such as the 2019-2020 Australian bushfire season, citizens need to be able to find reliable information on social media platforms such as Twitter. With the rise of bots and trolls on these platforms, these platforms are polluted and may even prevent citizens from forming their opinions on the ongoing crisis.

Our investigation shows that it was not fully automated bots, but human-curated trollbots, that sought to kick-start a counter-narrative to replace the scientifically proven link between climate change and the bushfires with a non-existent crisis, specifically by claiming that arsonists were the main cause of the 2019-2020 bushfire emergency. This disinformation campaign relied on decontextualised and deconstructed news, amplification from prominent and influential public personas around the globe, and media organisations that provided such climate change deniers with a mainstream media platform.

Such deliberate disinformation campaigns are likely to become ever more common in a news environment that now heavily relies on social media as distribution channels; they require constant vigilance from media organisations, journalists, politicians, public figures, scholars, and ordinary users. Focussing on a crucial case study, our study provides an important blueprint for further research to investigate similar such disinformation campaigns: drawing on readily available tools, it combines a forensic analysis of the dynamics of a campaign with a method for detecting the presence of bots and trolls among the accounts promoting it. This enables researchers to disentangle how alternative narratives seek to ingratiate themselves into mainstream public debates.

References

- ABC News (2020). Peter Dutton Says 250 Have Been Charged with Arson. But the Data Tells a Different Story. *ABC News*, 18 Feb. 2020. <https://www.abc.net.au/news/2020-02-18/fact-check-peter-dutton-arson-250-charged/11971454>
- Abram, Nerilie (2019). Australia's Angry Summer: This Is What Climate Change Looks Like. <https://blogs.scientificamerican.com/observations/australias-angry-summer-this-is-what-climate-change-looks-like/>
- BotSentinel (2020). Bot Sentinel. <https://botsentinel.com/>
- Digital News Report (2019). Reuters Institute - Digital News Report 2019. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/inline-files/DNR_2019_FINAL.pdf
- Graham, T. (2020). Twitter bots and trolls promote conspiracy theories about Australian bushfires. <https://www.zdnet.com/article/twitter-bots-and-trolls-promote-conspiracy-theories-about-australian-bushfires/>
- Shao, C., Ciampaglia, G. L., Varol, O., Yang, K. [Kaicheng], Flammini, A., & Menczer, F. (2018). *The spread of low-credibility content by social bots*. Retrieved from <http://arxiv.org/pdf/1707.07592v4>

UN Environment (2020). Ten impacts of the Australian bushfires.

<https://www.unenvironment.org/news-and-stories/story/ten-impacts-australian-bushfires>

PAPER 2

INVESTIGATING VISUAL MEDIA SHARED OVER TWITTER DURING THE 2019 EUROPEAN PARLIAMENTARY ELECTIONS

Nahema Marchal
Oxford Internet Institute, University of Oxford

Lisa-Maria Neudert
Oxford Internet Institute, University of Oxford

Introduction

From the rise of selfies, memes¹ and animated GIFs in digital culture to the surging popularity of visual-centric platforms like Instagram, Snapchat, and TikTok that reach billions of monthly users across the world – we live in an age of visual communication. Visual content in the form of photos, videos, infographics and user-generated images are becoming central to our day-to-day interactions, informing how we present ourselves (Thomson & Greenwood, 2020), communicate and understand the world around us (Pearce et al. 2020).

Most contemporary political communication research into the spread of disinformation and computational propaganda on social media relies heavily on text-based evidence and methodologies. Yet political disinformation is increasingly taking on visual forms, such as memes, gifs, short-format video, synthetic deep fakes and other user- and machine-generated visuals. In modern attention economies, visual forms of communication offer advantages over text; they are easier to process, elicit emotions and are effective at capturing viewers' attention (Fahmy et al., 2014). In recent years, however, they have also emerged as catalysts of mis- and disinformation (Bradshaw & Howard, 2019; Guy, 2017). It is therefore vital to better understand how social media images are strategically employed during critical moments of public life as well as their potential consequences on information flows and audiences.

Despite the ubiquity and potency of manipulative visuals as vehicles of both political and emotional information, their use during political campaigns and information operations remains critically understudied (Weller et al., 2014). In this paper we conduct a multi-lingual, cross-case comparative content and thematic analysis of Twitter images posted

¹Memes typically consist of 'digital objects that riff on a given visual, textual or auditory form and are then appropriated, re-coded, and slotted back into the internet infrastructures they came from' (Nooney & Portwood-Stacer, 2014, p.249).

by users in six different European language spheres—English, French, German, Italian, Spanish and Swedish—during a two-week-long period leading up to the 2019 EU Parliamentary elections. Specifically, we investigate images on Twitter in the context of two conversations: one surrounding the EU elections in general, and one surrounding the more contentious issue of membership to the EU. Three main research questions drive our analysis:

RQ1: What salient formats and modes of visual content were users in Europe sharing over Twitter during the 2019 EU Parliamentary election campaign?

RQ2: How does this differ across two conversations with varying degrees of contention?

RQ3: What were the most common themes embedded in different modes of visual communication?

Methodology

Our data collection proceeded in four stages. We first identified a set of relevant hashtags in English, French, German, Italian, Polish, Spanish, and Swedish intended to capture Twitter traffic. Using a set of 84 hashtags, our team then collected a total of 3,620,701 tweets in real time between 13 May and 26 May 2019 through Twitter's Streaming API. From this initial dataset we extracted tweets that contained static visuals in their metadata fields. Next, we identified tweets that had been removed or deleted, resulting in a final dataset of 307,951 tweets with visual content of which 256,204 related to the EU election and 3,164 related to EU exit. Finally, we selected a random sample of 599 tweets for what we refer to as the 'General' sample and 505 tweets for the 'Exit' sample.

Our team developed a grounded typology to classify visuals into categories based on their mode and format. Here, format describes the type of media shared by users, based on their constitutive elements (e.g. *photographs, screen-captures, posters, infographics, composites, and quotations*), while the concept of mode captures the ways in which political information is being transmitted (e.g. *official campaign communication, campaign event, political humor, satellite campaign material, news media reporting* etc.), based on an image's content and its apparent provenance. Intercoder reliability was then determined using Krippendorff's alpha on two random sub-samples achieving high scores ($\alpha = 0.843$ for format, $\alpha = 0.865$ for mode). Finally, we performed a thematic analysis of images, inductively identifying recurrent themes and patterns of meaning as they emerged in the data through semantic and visual symbols.

Findings & Discussion

We find that visuals originating from traditional political actors prevailed among both Twitter conversations. While users shared substantial amounts of anti-European, populist and, to a lesser extent, extremist images, this content remained largely disjointed from the mainstream public debate. Finally our data revealed political humor to be as a strong vector for anti-establishment and Eurosceptic themes, especially in discussions critical of the European project. Our findings underscore that visual media played a central role in the Twitter political discourse ahead of the 2019 European Parliamentary Elections, both as a conduit for official campaigning and candidate communications and for novel forms of political expression and user-generated political

content. Our analysis provides both a terminology and conceptual framework for future studies of the strategic spread of social media visuals and their impact.

References

- Bradshaw, S., & Howard, P. N. (2019). *The Global Disinformation Order 2019 Global Inventory of Organised Social Media Manipulation* (Working Paper 2019.2). Project on Computational Propaganda.
- DiResta, R., Shaffer, K., Ruppel, B., Sullivan, D., & Matney, R. (2018). *The Tactics & Tropes of the Internet Research Agency*. New Knowledge.
- Fahmy, S., Bock, M. A., & Wanta, W. (2014). *Visual Communication Theory and Research*. Palgrave Macmillan US.
- Guy, H. (2017, October 17). *Why We Need to Understand Misinformation Through Visuals*. First Draft.
- Pearce, W., Özkula, S. M., Greene, A. K., Teeling, L., Bansard, J. S., Omena, J. J., & Rabello, E. T. (2020). Visual Cross-Platform Analysis: Digital Methods to Research Social Media Images. *Information, Communication & Society*, 23(2), 161–180. <https://doi.org/10.1080/1369118X.2018.1486871>
- Thomson, T. J., & Greenwood, K. (2020). Profile Pictures across Platforms. In S. Josephson, J. D. Kelly, & K. Smith (Eds.), *Handbook of Visual Communication* (1st ed., pp. 349–363). Routledge.
- Weller, K., Bruns, A., Burgess, J., Mahrt, M., & Puschmann, C. (2014). *Twitter and Society*. Peter Lang Publishing.

PAPER 3

DIFFERENT NARRATIVES FOR DIFFERENT PEOPLE? UNDERSTANDING DEVIANT SEMANTIC EFFECTS ACROSS EU COUNTRIES IN POLITICAL CAMPAIGNS OF THE 2019 EUROPEAN PARLIAMENT ELECTION

Rolf Nijmeijer
LUISS Guido Carli, Italy

Kristoffer Laigaard Nielbo
Aarhus University, Denmark

Marie Damsgaard Mortensen
Aarhus University, Denmark

Anja Bechmann
Aarhus University, Denmark / University of Antwerp, Belgium

Introduction

Facebook has been heavily criticized for allowing political parties in their campaigns to potentially advertise with conflicting voices to different groups of people through their microtargeting advertising interface without being transparent about it. This has

potential implications for deliberative democracy because lack of transparency can create voter discrimination where certain voters are targeted either heavily due to their behavioral profile, voter region (available in the ad interface) and/or with different campaign messages (Bechmann, 2019). Furthermore, the lack of overview of the amount of placed advertisements prevent an understanding of the potential correlation between the Facebook campaign of any candidate and the subsequent election results.

Based on data from the Facebook ad library established in the wake of European Commission pressure (Buning et. al., 2018) this paper will analyze whether we find conflicting semantics in the political campaigns of candidates running for the 2019 European Parliament Election (EPE) across European Union (EU) countries and whether such conflicting semantics are to be found in certain countries. The paper will analyze all campaign items available in the archive from May 2019 across political parties and all EU countries. We will measure if deviant semantics (see also Bechmann & Nielbo, 2018) within a political party campaign correlate with factors such as country, left-center-right leaning, and which factors are the strongest predictor for deviant semantics.

The Potential Deviant Ad Semantics and the Associated Effects Across Countries

Studies have shown that known agents of disinformation dissemination, such as the Russian Internet Research Agency (IRA), adjusted the contents of their 2015-2016 Facebook advertising campaigns in the United States (US) based on the groups they intended to target. Advertisements targeting conservative voters, for example, preyed on existing prejudices against 'others', whereas advertisements aimed at African-Americans and Mexican-Americans tried to increase distrust of politics in the US (Howard et. al., 2018). This lends credence to the assumption that Facebook political advertisements and their intended effects are based on the concept of conflicting semantics.

When this concept is applied to the 2019 European Parliament Election, it can be hypothesized that the different cultural, political and social spheres in each member state influence the content of the targeted advertisements shown to voters on Facebook. A major advantage of studying this election is the multi-country campaign run by the same candidates/political groups, which allows for a comparative analysis. The percentage of the population that uses Facebook varies significantly per country, affecting any potential success for advertisements on this platform. Differences in media and political systems (Hallin & Mancini, 2004) further increase the likelihood that Facebook advertisements spread a different narrative in each country.

It is the objective of this paper to identify deviant semantics, if any, between microtargeting Facebook advertisements from different European countries. Research has shown that the spread of deliberately false narratives through social media is observed more among political parties on the far-right of the spectrum (Lance Bennett & Livingston, 2018). However, this may vary between countries. The anti-establishment Five Star Movement in Italy, for example, is more difficult to fit within that frame, but has a history of constructing emotional narratives on social media (Fusaro, 2018). It is the assumption of this study that there will be an observable semantic distance between

microtargeting Facebook advertising campaigns used in the 2019 EU election, where far-right politicians/parties are the most divergent.

Methods & Dataset

In order to examine whether we see different semantics, we use the Ad Library API from Facebook (<https://www.facebook.com/ads/library/api/>), comprising advertisements posted on Facebook during the 2019 EU election isolating the dataset to the run-up period of May 2019. The library API contains the text, images and any urls included in the advertisement, impression count (range provided) as well as data on who viewed the advertisement (7 age categories and 3 gender categories) and who paid for it.

We identify parties through national EPE lists and code them using the Chapel Hill Expert Survey scores as politically left, center or right-leaning. We then sort our dataset accordingly and run word count vectorization, term frequency–inverse document frequency, document clustering and paragraph embedding to identify semantic differences. These different levels of analysis should give a comprehensive overview of any potential textual deviations within the advertisements, and the political entities that used them. This is done for each EU member state individually, and the respective results will be compared to identify any major differences in the use of political microtargeting on Facebook across Europe.

Results, Discussion and Conclusion

The analysis is currently being processed and the results will be presented and discussed at the conference in October 2020 as an add on to the framework presented in this extended abstract.

References

- Bechmann, A. (2019). Data as Humans: Representation, Accountability, and Equality in Big Data. In *Information Policy Series: Vol. 16. Human Rights in the Age of Platforms* (pp. 73–94). Cambridge, MA: MIT Press.
- Bennett, W. L., & S. Livingston (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. In *European Journal of Communication*, Vol. 33 (2), 122–139. <https://doi.org/10.1177/026732311876031>.
- Bechmann, A., & Nielbo, K. L. (2018). Are We Exposed to the Same “News” in the News Feed? *Digital Journalism*, 6(8), 990–1002. <https://doi.org/10.1080/21670811.2018.1510741>.
- Buning et. al., M. d. C. (2018). *A Multi-dimensional Approach to Disinformation: Report of the independent high-level group on fake news and online disinformation* (pp. 1–40). Retrieved from EU Commission website: <https://ec.europa.eu/digital-single-market/en/news/final-report-high-level-expert-group-fake-news-and-online-disinformation>.
- Fusaro, C. (2018). Misinformation, an old issue in a new context. The state of the art in Italy. Conference: *Misinformation in Referenda*. Retrieved from <https://www.sipotra.it/old/wp-content/uploads/2018/09/Misinformation-an-old-issue-in-a-new-context.-The-state-of-the-art-in-Italy.pdf>.
- Hallin, D. C., & Mancini, P. (2004). *Comparing Media Systems: Three Models of Media and Politics*. Cambridge University Press.

Howard, et. al., P. N. (2018), The IRA, Social Media and Political Polarization in the United States, 2012-2018. *The Computational Propaganda Project*. Retrieved from <https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/12/IRA-Report.pdf>.
Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, 22(1), 79–86. <https://doi.org/10.1214/aoms/1177729694>

PAPER 4

EXPLAINING DYSFUNCTIONAL INFORMATION SHARING ON WHATSAPP AND FACEBOOK IN BRAZIL

Patrícia Rossini
University of Liverpool

Jennifer Stromer-Galley
Syracuse University

Erica Anita Baptista
Federal University of Minas Gerais

Vanessa Veiga de Oliveira
Federal University of Minas Gerais

Introduction

In the run-up to the 2018 Brazilian elections, false and misleading information was widely circulated through the mobile instant messaging service WhatsApp (First Draft, 2019). Researchers estimated that roughly half of all images circulating through the service were likely altered or distorted to convey false information (Tardáguila et al., 2018). Another study reported by *The Guardian* sampled WhatsApp messages prior to the election and found evidence of a politically right-leaning coordinated campaign to spread misinformation and bolster Jair Bolsonaro, who ultimately won the election (Avelar, 2019). Similar concerns have been raised in other democratic countries, including India and Indonesia, about the use of WhatsApp to spread false and misleading information in an effort to affect public opinion and alter election outcomes.

Mobile instant messaging services (MIMS), such as WhatsApp, Snapchat and Facebook Messenger, are increasingly being used for more than casual communication (Gil de Zúñiga et al., 2019; Valeriani & Vaccari, 2017). These private messaging applications became important venues for people to talk about political issues and news, access information, and communicate with businesses. The Reuters Digital News Report has been consistently capturing this trend: while the use of Facebook is declining worldwide, the use of messaging apps is on the rise (Nielsen et al., 2019).

Yet, political communication scholars have largely researched social networking sites (SNSs), and less attention has been paid to these burgeoning private messaging apps.

One challenge for the study of private messaging apps is that they are end-to-end encrypted and by design are not broadly visible for easy study by scholars. Nevertheless, as concerns spread globally around political propaganda and the intentional spread of politically false information, scholars are challenged to identify and measure the extent and effects of misinformation in democracies, in both public and private social applications.

In light of these concerns, this study aims to address this challenge by examining political misinformation on WhatsApp in Brazil. Data compiled by "We Are Social"ⁱ places WhatsApp in third place among all social platforms, with 1.5 billion monthly active users, only behind Facebook and YouTube. In Brazil, WhatsApp is the second most popular social media application, with over 120 million users in 2017. The concerns around misinformation on WhatsApp have led government institutions, media outlets, and NGOs to create fact-checking services to mitigate the spread of false and misleading information in Brazil, and the platform itself to fund social science research on this problem, including this study.

One important dimension of research is understanding who are more likely to share misinformation on these private messaging apps. In spite of concerns with coordinated disinformation efforts and computational propaganda (Jamieson, 2018; Wooley & Howard, 2018), regular users are responsible for spreading misinformation in their own networks — a behavior that Chadwick, Vaccari and O'Loughlin (2018) have described as "democratically dysfunctional news sharing". Although motivations for sharing news on social media may be varied, and we cannot assume that those who engage with and share misinformation have the intention to trick or troll others, understanding the types of users and behaviors associated with dysfunctional news sharing is an important step towards devising strategies to combat the spread of misinformation online.

Methods

We adopt a comparative approach to examine dysfunctional sharing on WhatsApp and Facebook, as semi-public platforms have been consistently scrutinized for facilitating the spread of misinformation because of algorithmic curation and an engagement-driven news feed (Guess et al., 2019). Considering the different affordances in these two platforms, comparing misinformation sharing dynamics on WhatsApp to Facebook may help understand the differences between private and semi-public venues. In this study, we present survey data of a representative sample of internet users in Brazil (N = 1,615) to examine dysfunctional sharing on Facebook and WhatsApp by examining both accidental misinformation sharing as well as intentional misinformation sharing, in which people recognize the information is incorrect and choose to share it anywayⁱⁱ. Specifically, we investigate the relationship between dysfunctional sharing and 1) frequency of political talk; 2) cross-cutting exposure; 3) social corrections (experiencing, witnessing and performing).

Results

Our findings provide further evidence of a participation vs. misinformation paradox: those who are more engaged in political talk are significantly more likely to have shared misinformation in the platform they use to discuss politics, and also significantly more likely to disinform — in the latter, the effects are cross-platform, as political talk on

WhatsApp is associated with intentional misinformation sharing on Facebook and vice-versa. We also find that instead of tempering the spread of false information, exposure to cross-cutting political views is positively associated with both types of dysfunctional sharing. Those who share misinformation are more likely to experience a social correction and to witness other being corrected, suggesting that false information does not go unnoticed in peoples' networks. Finally, we find that people are significantly more likely to experience, perform, and witness social correction on WhatsApp than on Facebook, suggesting that the closer social ties maintained through WhatsApp might provide a sense of safety that supports these behaviors.

References

- Avelar, D. (2019, October 30). WhatsApp fake news during Brazil election 'favoured Bolsonaro.' *The Guardian*.
<https://www.theguardian.com/world/2019/oct/30/whatsapp-fake-news-brazil-election-favoured-jair-bolsonaro-analysis-suggests>
- First Draft. (2019, June 27). *What 100,000 WhatsApp messages reveal about misinformation in Brazil*. First Draft. <https://firstdraftnews.org:443/latest/what-100000-whatsapp-messages-reveal-about-misinformation-in-brazil/>
- Gil de Zúñiga, H., Ardèvol-Abreu, A., & Casero-Ripollés, A. (2019). WhatsApp political discussion, conventional participation and activism: exploring direct, indirect and generational effects. *Information, Communication & Society*, 0(0), 1–18.
<https://doi.org/10.1080/1369118X.2019.1642933>
- Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, 5(1), eaau4586. <https://doi.org/10.1126/sciadv.aau4586>
- Jamieson, K. H. (2018). *Cyberwar: How Russian Hackers and Trolls Helped Elect a President What We Don't, Can't, and Do Know*. Oxford University Press.
- Nielsen, R. K., Newman, N., Fletcher, R., & Kalogeropoulos, A. (2019). *Reuters Institute Digital News Report 2019*. Reuters Institute for the Study of Journalism.
<http://www.digitalnewsreport.org/survey/2019>
- Tardáguila, C., Benevenuto, F., & Ortellado, P. (2018, October 17). Opinion | Fake News Is Poisoning Brazilian Politics. WhatsApp Can Stop It. *The New York Times*.
<https://www.nytimes.com/2018/10/17/opinion/brazil-election-fake-news-whatsapp.html>
- Valeriani, A., & Vaccari, C. (2017). Political talk on mobile instant messaging services: a comparative analysis of Germany, Italy, and the UK. *Information, Communication & Society*, 0(0), 1–17. <https://doi.org/10.1080/1369118X.2017.1350730>
- Wooley, S., & Howard, P. N. (Eds.). (2018). *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*. Oxford University Press.

Notes:

ⁱ <https://wearesocial.com/global-digital-report-2019> (Last access: August 20, 2019).

ⁱⁱ Scholars such as Jack (2017) have defined purposeful and coordinated efforts of sharing false or misleading information as "disinformation". We do not adopt this term to refer to what we call "intentional misinformation sharing" because we are not explicitly studying coordinated efforts of misinformation sharing, and we also do not know the motives behind the intentional sharing behaviors we investigate.