

Bradshaw, Samantha

Article

Disinformation optimised: gaming search engine algorithms to amplify junk news

Internet Policy Review

Provided in Cooperation with:

Alexander von Humboldt Institute for Internet and Society (HIIG), Berlin

Suggested Citation: Bradshaw, Samantha (2019) : Disinformation optimised: gaming search engine algorithms to amplify junk news, Internet Policy Review, ISSN 2197-6775, Alexander von Humboldt Institute for Internet and Society, Berlin, Vol. 8, Iss. 4, pp. 1-24,
<https://doi.org/10.14763/2019.4.1442>

This Version is available at:

<https://hdl.handle.net/10419/214101>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/3.0/de/legalcode>



Disinformation optimised: gaming search engine algorithms to amplify junk news

Samantha Bradshaw

Oxford Internet Institute, United Kingdom, samantha.bradshaw@oii.ox.ac.uk

Published on 31 Dec 2019 | DOI: 10.14763/2019.4.1442

Abstract: Previous research has described how highly personalised paid advertising on social media platforms can be used to influence voter preferences and undermine the integrity of elections. However, less work has examined how search engine optimisation (SEO) strategies are used to target audiences with disinformation or political propaganda. This paper looks at 29 junk news domains and their SEO keyword strategies between January 2016 and March 2019. I find that SEO — rather than paid advertising — is the most important strategy for generating discoverability via Google Search. Following public concern over the spread of disinformation online, Google’s algorithmic changes had a significant impact on junk news discoverability. The findings of this research have implications for policymaking, as regulators think through legal remedies to combat the spread of disinformation online.

Keywords: Disinformation, Junk news, Computational propaganda, Search engine optimisation, Digital advertising

Article information

Received: 02 Jul 2019 **Reviewed:** 26 Nov 2019 **Published:** 31 Dec 2019

Licence: Creative Commons Attribution 3.0 Germany

Funding: The author is grateful for support in the form of a Doctoral fellowship from the Social Science and Humanities Research Council (SSHRC). Additional support was provided by the Hewlett Foundation [2018-7384] and the European Research Council grant “Computational Propaganda: Investigating the Impact of Algorithms and Bots on Political Discourse in Europe”, Proposal 648311, 2015–2020, Philip N. Howard, Principal Investigator.

Competing interests: The author has declared that no competing interests exist that have influenced the text.

URL:

<http://policyreview.info/articles/analysis/disinformation-optimised-gaming-search-engine-algorithms-amplify-junk-news>

Citation: Bradshaw, S. (2019). Disinformation optimised: gaming search engine algorithms to amplify junk news. *Internet Policy Review*, 8(4). DOI: 10.14763/2019.4.1442

This paper is part of Data-driven elections, a special issue of Internet Policy Review guest-edited by Colin J. Bennett and David Lyon.

INTRODUCTION

Did the Holocaust *really* happen? In December 2016, Google’s search engine algorithm determined the most authoritative source to answer this question was a neo-Nazi website peddling holocaust denialism (Cadwalladr, 2016b). For any inquisitive user typing this question into Google, the first website recommended by Search linked to an article entitled: “Top 10 reasons why the Holocaust didn’t happen”. The third article “The Holocaust Hoax; IT NEVER HAPPENED” was published by another neo-Nazi website, while the fifth, seventh, and ninth recommendations linked to similar racist propaganda pages (Cadwalladr, 2016b). Up until Google started demoting websites committed to spreading anti-Semitic messages, anyone asking whether the Holocaust actually happened would have been directed to consult neo-Nazi websites, rather than one of the many credible sources about the Holocaust and tragedy of World War II.

Google’s role in shaping the information environment and enabling political advertising has made it a “de facto infrastructure” for democratic processes (Barrett & Kreiss, 2019). How its search engine algorithm determines authoritative sources directly shapes the online information environment for more than 89 percent of the world’s internet users who trust Google Search to quickly and accurately find answers to their questions. Unlike social media platforms that tailor content based on “algorithmically curated newsfeeds” (Golebiewski & boyd, 2019), the logic of search engines is “mutually shaped” by algorithms — that shape access — and users — who shape the information being sought (Schroeder, 2014). By facilitating information access and discovery, search engines hold a unique position in the information ecosystem. But, like other digital platforms, the digital affordances of Google Search have proved to be fertile ground for media manipulation.

Previous research has demonstrated how large volumes of mis- and disinformation were spread on social media platforms in the lead up to elections around the world (Hedman et al., 2018; Howard, Kollanyi, Bradshaw, & Neudert, 2017; Machado et al., 2018). Some of this disinformation was micro-targeted towards specific communities or individuals based on their personal data. While data-driven campaigning has become a powerful tool for political parties to mobilise and fundraise (Fowler et al., 2019; Baldwin-Philippi, 2017), the connection between online advertisements and disinformation, foreign election interference, polarisation, and non-transparent campaign practices has caused growing anxieties about its impact on democracy.

Since the 2016 presidential election in the United States, public attention and scrutiny has largely focused on the role of Facebook in profiting from and amplifying the spread of disinformation via digital advertisements. However, less attention has been paid to Google, who, along with Facebook, commands more than 60% of the digital advertising market share. At the same time, a multi-billion-dollar search engine optimisation (SEO) industry has been built around understanding how technical systems rank, sort, and prioritise information (Hoffmann, Taylor, & Bradshaw, 2019). The purveyors of disinformation have learned to exploit social media platforms to engineer content discovery and drive “pseudo-organic engagement”.¹ These websites — that do not employ professional journalistic standards, report on conspiracy theory, counterfeit professional news brands, and mask partisan commentary as news — have been referred to as “junk news” domains (Bradshaw, Howard, Kollanyi, & Neudert, 2019).

Together, the role of political advertising and the matured SEO industry make Google Search an interesting and largely underexplored case to analyse. Considering the importance of Google

Search in connecting individuals to news and information about politics, this paper examines how junk news websites generate discoverability via Google Search. It asks: (1) How do junk news domains optimise content, through both paid and SEO strategies, to grow discoverability and grow their website value? (2) What strategies are effective at growing discoverability and/or growing website value; and (3) What are the implications of these findings for ongoing discussions about the regulation of social media platforms?

To answer these questions, I analysed 29 junk news domains and their advertising and search engine optimisation strategies between January 2016 and March 2019. First, junk news domains make use of a variety of SEO keyword strategies in order to game Search and grow pseudo-organic clicks and grow their website value. The keywords that generated the highest placements on Google Search focused on (1) navigational searches for known brand names (such as searches for “breitbart.com”) and (2) carefully curated keyword combinations that fill so-called “data voids” (Golebiewski & Boyd, 2018), or a gap in search engine queries (such as searches for “Obama illegal alien”). Second, there was a clear correlation between the number of clicks that a website receives and the estimated value of the junk news domains. The most profitable timeframes correlated with important political events in the United States (such as the 2016 presidential election, and the 2018 midterm elections), and the value of the domain increased based on SEO optimised — rather than paid — clicks. Third, junk news domains were relatively successful at generating top-placements on Google Search before and after the 2016 US presidential election. However, their discoverability abruptly declined beginning in August 2017 following major announcements from Google about changes to its search engine algorithms, as well as other initiatives to combat the spread of junk news in search results. This suggests that Google can, and has, measurably impacted the discoverability of junk news on Search.

This paper proceeds as follows: The first section provides background on the vocabulary of disinformation and ongoing debates about so-called fake news, situating the terminology of “junk news” used in this paper in the scholarly literature. The second section discusses the logic and politics of search, describing how search engines work and reviewing the existing literature on Google Search and the spread of disinformation. The third section outlines the methodology of the paper. The fourth section analyses 29 prominent junk news domains to learn about their SEO and advertising strategies, as well as their impact on content discoverability and revenue generation. This paper concludes with a discussion of the findings and implications for future policymaking and private self-regulation.

THE VOCABULARY OF POLITICAL COMMUNICATION IN THE 21ST CENTURY

“Fake news” gained significant attention from scholarship and mainstream media during the 2016 presidential election in the United States as viral stories pushing outrageous headlines — such as Hillary Clinton’s alleged involvement in a paedophile ring in the basement of a DC pizzeria — were prominently displayed across search and social media news feeds (Silverman, 2016). Although “fake news” is not a new phenomenon, the spread of these stories—which are both enhanced and constrained by the unique affordances of internet and social networking technologies — has reinvigorated an entire research agenda around digital news consumption and democratic outcomes. Scholars from diverse disciplinary backgrounds — including psychology, sociology and ethnography, economics, political science, law, computer science,

journalism, and communication studies — have launched investigations into circulation of so-called “fake news” stories (Allcott & Gentzkow, 2017; Lazer et al., 2018), their role in agenda-setting (Guo & Vargo, 2018; Vargo, Guo, & Amazeen, 2018), and their impact on democratic outcomes and political polarisation (Persily, 2017; Tucker et al., 2018).

However, scholars at the forefront of this research agenda have continually identified several epistemological and methodological challenges around the study of so-called “fake news”. A commonly identified concern is the ambiguity of the term itself, as “fake news” has come to be an umbrella term for all kinds of problematic content online, including political satire, fabrication, manipulation, propaganda, and advertising (Tandoc, Lim, & Ling, 2018; Wardle, 2017). The European High-Level Expert Group on Fake News and Disinformation recently acknowledged the definitional difficulties around the term, recognising it “encompasses a spectrum of information types...includ[ing] low risk forms such as honest mistakes made by reporters...to high risk forms such as foreign states or domestic groups that would try to undermine the political process” (European Commission, 2018). And even when the term “fake news” is simply used to describe news and information that is factually inaccurate, the binary distinction between what is true and what is false has been criticised for not adequately capturing the complexity of the kinds of information being shared and consumed in today’s digital media environment (Wardle & Derakhshan, 2017).

Beyond the ambiguities surrounding the vocabulary of “fake news”, there is growing concern that the term has begun to be appropriated by politicians to restrict freedom of the press. A wide range of political actors have used the term “fake news” to discredit, attack, and delegitimise political opponents and mainstream media (Farkas & Schou, 2018). Certainly, Donald Trump’s (in)famous use of the term “fake news”, is often used to “deflect” criticism and to erode the credibility of established media and journalist organisations (Lakoff, 2018). And many authoritarian regimes have followed suit, adopting the term into a common lexicon to legitimise further censorship and restrictions on media within their own borders (Bradshaw, Neudert, & Howard, 2018). Given that most citizens perceive “fake news” to define “partisan debate and poor journalism”, rather than a discursive tool to undermine trust and legitimacy in media institutions, there is general scholarly consensus that the term is highly problematic (Nielsen & Graves, 2017).

Rather than chasing a definition of what has come to be known as “fake news”, researchers at the Oxford Internet Institute have produced a grounded typology of what users actually share on social media (Bradshaw et al., 2019). Drawing on Twitter and Facebook data from elections in Europe and North America, researchers developed a grounded typology of online political communication (Bradshaw et al., 2019; Neudert, Howard, & Kollanyi, 2019). They identified a growing prevalence of “junk news” domains, which publish a variety of hyper-partisan, conspiracy theory or click-bait content that was designed to look like real news about politics. During the 2016 presidential election in the United States, social media users on Twitter shared as much “junk news” as professionally produced news about politics (Howard, Bolsover, Kollanyi, Bradshaw, & Neudert, 2017; Howard, Kollanyi, et al., 2017). And voters in swing-states tended to share more junk news than their counterparts in uncontested ones (Howard, Kollanyi, et al., 2017). In countries throughout Europe — in France, Germany, the United Kingdom and Sweden — junk news inflamed political debates around immigration and amplified populist voices across the continent (Desiguid, Howard, Kollanyi, & Bradshaw, 2017; Kaminska, Galacher, Kollanyi, Yasseri, & Howard, 2017; Neudert, Howard, & Kollanyi, 2017).

According to researchers on the Computational Propaganda Project junk news is defined as

having at least three out of five elements: (1) *professionalism*, where sources do not employ the standards and best practices of professional journalism including information about real authors, editors, and owners (2) *style*, where emotionally driven language, *ad hominem* attacks, mobilising memes and misleading headlines are used; (3) *credibility*, where sources rely on false information or conspiracy theories, and do not post corrections; (4) *bias*, where sources are highly biased, ideologically skewed and publish opinion pieces as news; and (5) *counterfeit*, where sources mimic established news reporting including fonts, branding and content strategies (Bradshaw et al., 2019).

In a complex ecosystem of political news and information, junk news provides a useful point of analysis because rather than focusing on individual stories that may contain honest mistakes, it examines the domain as a whole and looks for various elements of deception, which underscores the definition of disinformation. The concept of junk news is also not tied to a particular producer of disinformation, such as foreign operatives, hyper-partisan media, or hate groups, who, despite their diverse goals, deploy the same strategies to generate discoverability. Given that the literature on disinformation is often siloed around one particular actor, does not cross platforms, nor integrate a variety of media sources (Tucker et al., 2018), the junk news framework can be useful for taking a broader look at the ecosystem as a whole and the digital techniques producers use to game search engine algorithms. Throughout this paper, I use the term “junk news” to describe the wide range of politically and economically motivated disinformation being shared about politics.

THE LOGIC AND POLITICS OF SEARCH

Search engines play a fundamental role in the modern information environment by sorting, organising, and making visible content on the internet. Before the search engine, anyone who wished to find content online would have to navigate “cluttered portals, garish ads and spam galore” (Pasquale, 2015). This didn’t matter in the early days of the web when it remained small and easy to navigate. During this time, web directories were built and maintained by humans who often categorised pages according to their characteristics (Metaxas, 2010). By the mid-1990s it became clear that the human classification system would not be able to scale. The search engine “brought order to chaos by offering a clean and seamless interface to deliver content to users” (Hoffman, Taylor, & Bradshaw, 2019).

Simplistically speaking, search engines work by crawling the web to gather information about online webpages. Data about the words on a webpage, links, images, videos, or the pages they link to are organised into an index by an algorithm, analogous to an index found at the end of a book. When a user types a query into Google Search, machine learning algorithms apply complex statistical models in order to deliver the most “relevant” and “important” information to a user (Gillespie, 2012). These models are based on a combination of “signals” including the words used in a specific query, the relevance and usability of webpages, the expertise of sources, and other information about context, such as a user’s geographic location and settings (Google, 2019).

Google’s search rankings are also influenced by AdWords, which allow individuals or companies to promote their websites by purchasing “paid placement” for specific keyword searches. Paid placement is conducted through a bidding system, where rankings and the number of times the advertisement is displayed are prioritised by the amount of money spent by the advertiser. For example, a company that sells jeans might purchase AdWords for keywords such as “jeans”,

“pants”, or “trousers”, so when an individual queries Google using these terms, a “sponsored post” will be placed at the top of the search results. ² AdWords also make use of personalisation, which allow advertisers to target more granular audiences based on factors such as age, gender, and location. Thus, a local company selling jeans for women can specify local female audiences — individuals who are more likely to purchase their products.

The way in which Google structures, organizes, and presents information and advertisements to users is important because these technical and policy decisions embed a wide range of political issues (Granka, 2010; Introna & Nissenbaum, 2000; Vaidhynathan, 2011). Several public and academic investigations auditing Google’s algorithms have documented various examples of bias in Search or problems with the autocomplete function (Cadwalladr, 2016a; Pasquale, 2015). Biases inherently designed into algorithms have been shown to disproportionately marginalise minority communities, women, and the poor (Noble, 2018).

At the same time, political advertisements have become a contentious political issue. While digital advertising can generate significant benefits for democracy, by democratising political finance and assisting in political mobilisation (Fowler et al., 2019; Baldwin-Philippi, 2017), it can also be used to selectively spread disinformation and messages of demobilisation (Burkell & Regan, 2019; Evangelista & Bruno, 2019; Howard, Ganesh, Liotsiou, Kelly, & Francois, 2018). Indeed, Russian AdWord purchases in the lead-up to the 2016 US election demonstrate how foreign states actors can exploit Google Search to spread propaganda (Mueller, 2019). But the general lack of regulation around political advertising has also raised concerns about domestic actors and the ways in which legitimate politicians campaign in increasingly opaque and unaccountable ways (Chester & Montgomery, 2017; Tufekci, 2014). These concerns are underscored by the rise of the “influence industry” and the commercialisation of political technologies who sell various ‘psychographic profiling’ technologies to craft, target, and tailor messages of persuasion and demobilisation (Chester & Montgomery, 2019; McKelvey, 2019; Bashykarla, 2019). For example, during the 2016 US election, Cambridge Analytica worked with the Trump campaign to implement “persuasion search advertising”, where AdWords were bought to strategically push pro-Trump and anti-Clinton information to voters (Lewis & Hilder, 2018).

Given growing concerns over the spread of disinformation online, scholars are beginning to study the ways in which Google Search might amplify junk news and disinformation. One study by Metaxa-Kakavouli and Torres-Echeverry examined the top ten results from Google searches about congressional candidates over a 26-week period in the lead-up to the 2016 presidential election. Of the URLs recommended by Google, only 1.5% came from domains that were flagged by PolitiFact as being “fake news” domains (2017). Metaxa-Kakavouli and Torres-Echeverry suggest that the low levels of “fake news” are the result of Google’s “long history” combatting spammers on its platform (2017). Another research paper by Golebiewski and boyd looks at how gaps in search engine results lead to strategic “data voids” that optimisers exploit to amplify their content (2018). Golebiewski and boyd argue that there are many search terms where data is “limited, non-existent or deeply problematic” (2018). Although these searches are rare, if a user types these search terms into a search engine, “it might not give a user what they are looking for because of limited data and/or limited lessons learned through previous searches” (Golebiewski & boyd, 2018).

The existence of biases, disinformation, or gaps in authoritative information on Google Search matters because Google directly impacts what people consume as news and information. Most of the time, people do not look past the top ten results returned by the search engine (Metaxas,

2010). Indeed, eye-tracking experiments have demonstrated that the order in which Google results are presented to users matters more than the actual relevance of the page abstracts (Pan et al., 2007). However, it is important to note that the logic of higher placements does not necessarily translate to search engine advertising listings, where users are less likely to click on advertisements if they are familiar with the brand or product they are searching for (Narayanan & Kalyanam, 2015).

Nevertheless, the significance of the top ten placement has given rise to the SEO industry, whereby optimisers use digital keyword strategies to move webpages higher in Google's rankings and thereby generate higher traffic flows. There is a long history of SEO dating back to the 1990s when the first search engine algorithms emerged (Metaxas, 2010). Since then, hundreds of SEO pages have published guesses about the different ranking factors these algorithms consider (Dean, 2019). However, the specific signals that inform Google's search engine algorithms are dynamic and constantly adapting to the information environment. Google makes hundreds of changes to its algorithm every year to adjust the weight and importance of various signals. While most of these changes are minor updates designed to improve the speed and performance of Search, sometimes Google makes more significant changes to its algorithm to elude optimisers trying to game the system.

Google has taken several steps to combat people seeking to manipulate Search for political or economic gain (Taylor, Walsh, & Bradshaw, 2019). This involves several algorithmic changes to demote sources of disinformation as well as changes to their advertising policies to limit the extent to which users can be micro-targeted with political advertisements. In one study, researchers interviewed SEO strategists to audit how Facebook and Google's algorithmic changes impacted their optimisation strategies (Hoffmann, Taylor, & Bradshaw, 2019). Since the purveyors of disinformation often rely on the same digital marketing strategies used by legitimate political candidates, news organisations, and businesses, the SEO industry can offer unique, but heuristic, insight into the impact of algorithmic changes. Hoffmann, Taylor and Bradshaw (2019) found that despite more than 125 announcements over a three-year period, the algorithmic changes made by the platforms did not significantly alter digital marketing strategies.

This paper hopes to contribute to the growing body of work examining the effect of Search on the spread of disinformation and junk news by empirically analysing the strategies — paid and optimised — employed by junk news domains. By performing an audit of the keywords junk news websites use to generate discoverability, this paper evaluates the effectiveness of Google in combatting the spread of disinformation on Search.

METHODOLOGY

CONCEPTUAL FRAMEWORK: THE TECHNO-COMMERCIAL INFRASTRUCTURE OF JUNK NEWS

The starting place for this inquiry into the SEO infrastructure of junk news domains is grounded conceptually in the field of science and technology studies (STS), which provides a rich literature on how infrastructure design, implementation, and use embeds politics (Winner, 1980). Digital infrastructure — such as physical hardware, cables, virtual protocols, and code — operate invisibly in the background, which can make it difficult to trace the politics embedded in technical coding and design (Star & Ruhleder, 1994). As a result, calls to study internet infrastructure has engendered digital research methods that shed light on the less-visible areas

of technology. One growing and relevant body of research has focused on the infrastructure of social media platforms and the algorithms and advertising infrastructure that invisibly operate to amplify or spread junk news to users, or to micro-target political advertisements (Kim et al., 2018; Tambini, Anstead, & Magalhães, 2017). Certainly, the affordances of technology — both real and imagined — mutually shape social media algorithms and their potential for manipulation (Nagy & Neff, 2015; Neff & Nagy, 2016). However, the proprietary nature of platform architecture has made it difficult to operationalise studies in this field. Because junk news domains operate in a digital ecosystem built on search engine optimisation, page ranks, and advertising, there is an opportunity to analyse the infrastructure that supports the discoverability of junk news content, which could provide insights into how producers reach audiences, grow visibility, and generate domain value.

JUNK NEWS DATA SET

The first step of my methodology involved identifying a list of junk news domains to analyse. I used the Computational Propaganda Project's (COMPROP) data set on junk news domains in order to analyse websites that spread disinformation about politics. To develop this list, researchers on the COMPROP project built a typology of junk news based on URLs shared on Twitter and Facebook relating to the 2016 US presidential election, the 2017 US State of the Union Address, and 2018 US midterm elections.³ A team of five rigorously trained coders labelled the domains contained in tweets and on Facebook pages based on a grounded typology of junk news that has been tested and refined over several elections around the world between 2016 and 2018.⁴ A domain was labelled as junk news when it failed on three of the five criteria of the typology (style, bias, credibility, professionalism, and counterfeit, as described in section one). For this analysis, I used the most recent 2018 midterm election junk news list, which is comprised of the top-29 most shared domains that were labelled as junk news by researchers. This list was selected because all 29 domains were active during the 2016 US presidential election in November 2016 and the 2017 US State of the Union Address, which provides an opportunity to comparatively assess how both the advertising and optimisation strategies, as well as their performance, changed overtime.

SPYFU DATA AND API QUERIES

The second step of my methodology involved collecting data about the advertising and optimisation strategies used by junk news websites. I worked with *SpyFu*, a competitive keyword research tool used by digital marketers to increase website traffic and improve keyword rankings on Google (SpyFu, 2019). *SpyFu* collects, analyses and tracks various data about the search optimisation strategies used by websites, such as organic ranks, paid keywords bought on Google AdWords, and advertisement trends.

To shed light onto the optimisation strategies used by junk news domains on Google, *SpyFu* provided me with: (1) a list of historical keywords and keyword combinations used by the top-29 junk news that led to the domain appearing in Google Search results; and (2) the position the domain appeared in Google as a result of the keywords. The historical keywords were provided from January 2016 until March 2019. Only keywords that led to the junk news domains appearing in the top-50 positions on Google were included in the data set.

In order to determine the effectiveness of the optimisation and advertising strategies used by junk news domains to either grow their website value and/or successfully appear in the top positions on Google Search, I wrote a simple python script to connect to the *SpyFu* API service. This python script collected and parsed the following data from *SpyFu* for each of the top-29 junk news domains in the sample: (1) the number of keywords that show up organically on

Google searches; (2) the estimated sum of clicks a domain receives based on factors including organic keywords, the rank of keyword, and the search volume of the keyword; (3) the estimated organic value of a domain based on factors including organic keywords, the rank of keywords, and the search volume of the keyword; (4) the number of paid advertisements a domain purchased through Google AdWords; and (5) the number of paid clicks a domain received from the advertisements it purchased from Google AdWords.

DATA AND METHODOLOGY LIMITATIONS

There are several data and methodology limitations that must be noted. First, the junk news domains identified by the Computational Propaganda Project highlights only a small sample of the wide variety of websites that peddle disinformation about politics. The researchers also do not differentiate between the different actors behind the junk news websites — such as foreign states or hyper-partisan media — nor do they differentiate between the political leaning of the junk news outlet — such as left-or-right-leaning domains. Thus, the outcomes of these findings cannot be described in terms of the strategies of different actors. Further, given that the majority of junk news domains in the top-29 sample lean politically to the right and far right, these findings might not be applicable to the hyper-partisan left and their optimisation strategies. Finally, the junk news domains identified in the sample were shared on social media in the lead-up to important political events in the United States. A further research question could examine the SEO strategies of domains operating in other country contexts.

When it comes to working with the data provided by SpyFu (and other SEO optimisation tools), there are two limitations that should be noted. First, the historical keywords collected by SpyFu are only collected when they appear in the top-50 Google Search results. This is an important limitation to note because news and information producers are constantly adapting keywords based on the content they are creating. Keywords may be modified by the source website dynamically to match news trends. Low performing keywords might be changed or altered in order to make content more visible via Search. Thus, the SpyFu data might not capture *all* of the keywords used by junk news domains. However, the collection strategy will have captured many of the most popular keywords used by junk news domains to get their content appearing in Google Search. Second, because SpyFu is a company there are proprietary factors that go into measuring a domain's SEO performance (in particular, the data points collected via the API on the estimated sum of clicks and the estimated organic value). Nevertheless, considering that Google Search is a prominent avenue for news and information discovery, and that few studies have systematically analysed the effect of search engine optimisation strategies on the spread of disinformation, this study provides an interesting starting point for future research questions about the impact SEO can have on the spread and monetisation of disinformation via Search.

ANALYSIS: OPTIMIZING DISINFORMATION THROUGH KEYWORDS AND ADVERTISING

JUNK NEWS ADVERTISING STRATEGIES ON GOOGLE

Junk news domains rarely advertise on Google. Only two out of the 29 junk news domains (infowars.com and cnsnews.com) purchased Google advertisements (See [Figure 1: Advertisements purchased vs. paid clicks](#)). The advertisements purchased by infowars.com were all made prior to the 2016 election in the United States (from the period of May 2015 to March

2016). cnsnews.com made several advertisement purchases over the three-year time period.

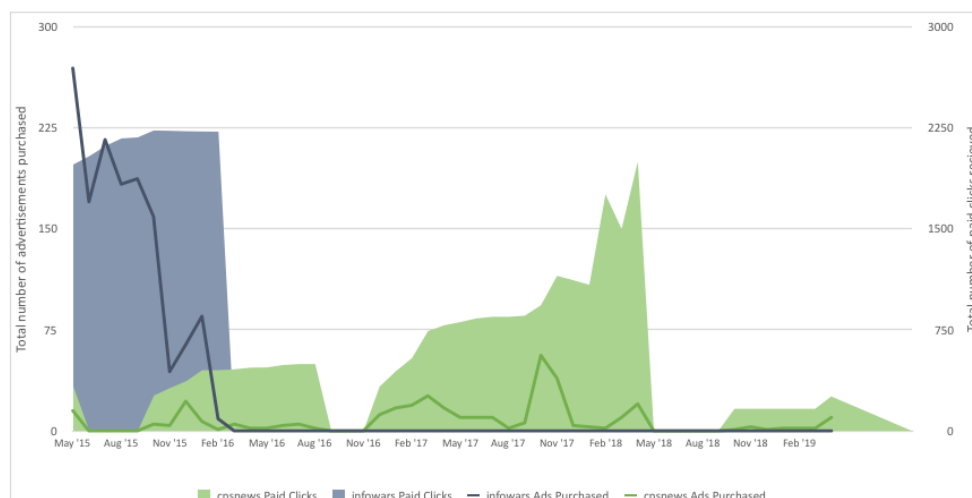


Figure 1: Advertisements purchased vs. paid clicks received: infowars.com and cnsnews.com (May 2015-March 2019)

Looking at the total number of paid clicks received, junk news domains generated only a small amount of traffic using paid advertisements. Infowars on average, received about 2000 clicks as a result of their paid advertisements. cnsnews.com peaked at approximately 1800 clicks, but on average generated only about 600 clicks per month over the course of three years. By comparing the number of clicks that are paid versus those that were generated as a result of SEO keyword optimisation, there is a significant difference. During the same time period, cnsnews.com and infowars.com were generating on average 146,000 and 964,000 organic clicks respectively (See Figure 2: Organic vs. paid clicks (cnsnews.com and infowars.com)). Although it is hard to make generalisations about how junk news websites advertise on Google based on a sample of two, the lack of data suggests that advertising on Google Search might not be as popular as advertising on other social media platforms. Second, the return on investment (i.e., paid clicks generated as a result of Google advertisements) was very low compared to the organic clicks these junk news domains received for free. Factors other than advertising seem to drive the discoverability of junk news on Google Search.

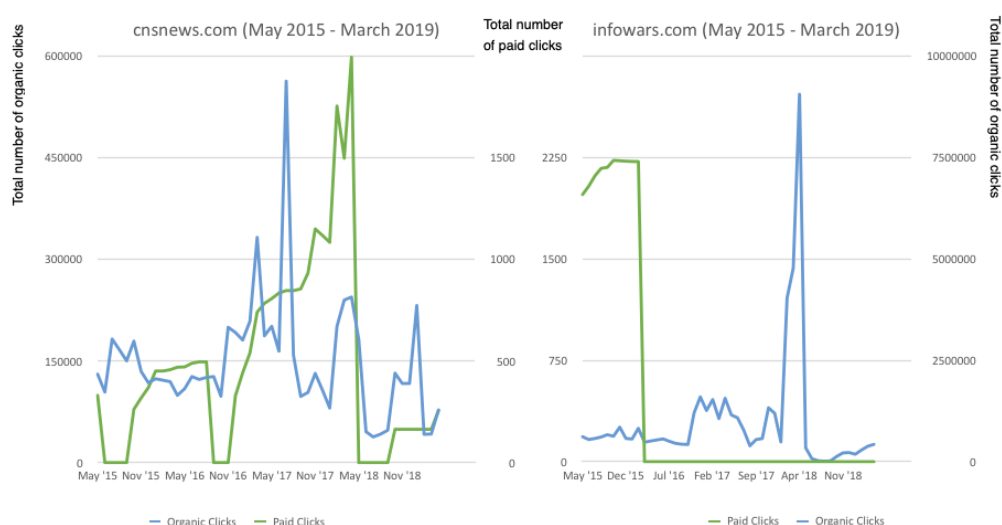


Figure 2: organic vs. paid clicks (cnsnews.com and infowars.com)

JUNK NEWS KEYWORD OPTIMISATION STRATEGIES

In order to assess the keyword optimisation strategies used by junk news websites, I worked with SpyFu, which provided historical keyword data for the 29 junk news domains, when those keywords made it to the top-50 results in Google between January 2016 and March 2019. In total, there were 88,662 unique keywords in the data set. Given the importance of placement on Google, I looked specifically at keywords that indexed junk news websites on the first — and most authoritative — position. Junk news domains had different aptitudes for generating placement in the first position (See Table 1: Junk news domains and number of keywords found in the first position on Google). Breitbart, DailyCaller and ZeroHedge had the most successful SEO strategies, respectively having 1006, 957 and 807 keywords lead to top placements on Google Search over the 39-month period. In contrast, six domains (committedconservative.com, davidharrisjr.com, reverbpress.news, thedailydigest.org, thefederalist.com, thepoliticalinsider.com) had no keywords reach the first position on Google. The remaining 20 domains had anywhere between 1 to 253 keywords place between the 2-10 positions on Google Search over the same timeframe.

Table 1: Junk news domains and number of keywords found in the first position on Google

Domain	Keywords reaching position 1
breitbart.com	1006
dailycaller.com	957
zerohedge.com	807
infowars.com	253
cnsnews.com	228
dailywire.com	214
thefederalist.com	200
rawstory.com	199
lifenews.com	156
pjmedia.com	140
americanthinker.com	133
thepoliticalinsider.com	111
thegatewaypundit.com	105
barenakedislam.com	48
michaelsavage.com	15
theblacksphere.net	9
truepundit.com	8
100percentfedup.com	5
bigleaguepolitics.com	3
libertyheadlines.com	2

Domain	Keywords reaching position 1
ussanews.com	2
gellerreport.com	1
truthfeednews.com	1

Different keywords also generate different kinds of placement over the 39-month period. Table 2 (see Appendix) provides a sample list of up to ten keywords from each junk news domain in the sample when the keyword reached the first position.

First, many junk news domains appear in the first position on Google Search as a result of “navigational searches” whereby a user entered a query with the intent of finding a website. A search for a specific brand of junk news could happen naturally for many users, since the Google Search function is built into the address bar in Chrome, and sometimes set as the default search engine for other browsers. In particular, terms like “infowars” “breitbart” “cnsnews” and “rawstory” were navigational keywords users typed into Google Search. The performance of brand searches over time consistently places junk news webpages in the number one position (see Figure 3: Brand-related keywords over time). This suggests that brand-recognition plays an important role for driving traffic to junk news domains.



Figure 3: the performance of brand-related keywords overtime: top-5 junk news websites (January 2016-March 2019)

There is one outlier in this analysis, where keyword searches for “breitbart” drops to position two: in January 2017 and September 2017. This drop could have been a result of mainstream media coverage of Steve Bannon assuming (and eventually leaving) his position as the White House Chief Strategist during those respective months. The fact that navigational searches are one of the main drivers behind generating a top ten placement on Search suggests that junk news websites rely heavily on developing a recognisable brand and a dedicated readership that actively seeks out content from these websites. However, this also demonstrates that a complicated set of factors go into determining what keywords from what websites make the top placement in Google Search, and that coverage of news events from mainstream professional news outlets can alter the discoverability of junk news via Search.

Second, many keywords that made it to the top position in Google Search results are what Golebiewski and boyd (2018) would call terms that filled “data voids”, or gaps in search engine queries where there is limited authoritative information about a particular issue. These keywords tended to focus on conspiratorial information especially around President Barack Obama (“Obama homosexual” or “stop Barack Obama”), gun rights (“gun control myths”), pro-life narratives (“anti-abortion quotes” or “fetus after abortion”), and xenophobic or racist content (“against Islam” or “Mexicans suck”). Unlike brand-related keywords, problematic search terms do not achieve a consistently high placement on Google Search over the 39-week period. Keywords that ranked in number one for more than 30-weeks include: “vz58 vs. ak47”, “feminizing uranium”, “successful people with down syndrome”, “google ddrive”, and “westboro[sic] Baptist church tires slashed”. This suggests that, for the most part, data voids are either being filled by more authoritative sources, or Google Search has been able to demote websites attempting to generate pseudo-organic engagement via SEO.

THE PERFORMANCE OF JUNK NEWS DOMAINS ON GOOGLE SEARCH

After analysing what keywords are used to get junk news websites in the number one position, the next half of my analysis looks at larger trends in SEO strategies overtime. What is the relationship between organic clicks and the value of a junk news website? How has the effectiveness of SEO keywords changed over the past 48 months? And have changes made by Google to combat the spread of junk news on Search had an impact on its discoverability?

JUNK NEWS, ORGANIC CLICKS, AND THE VALUE OF THE DOMAIN

There is a close relationship between the number of clicks a domain receives and the estimated value of that domain. By comparing figure 4 and 5, you can see that the more clicks a website receives, the higher its estimated value. Often, a domain is considered more valuable when it generates large amounts of traffic. Advertisers see this as an opportunity, then, to reach more people. Thus, the higher the value of a domain, the more likely it is to generate revenue for the operator. The median estimated value of the top-29 most popular junk news was \$5,160 USD during the month of the 2016 presidential election, \$1,666.65 USD during the 2018 State of the Union, and \$3,906.90 USD during the 2018 midterm elections. Infowars.com and Breitbart.com were the two highest performing junk news domains — in terms of clicks and domain value. While Breitbart.com maintained a more stable readership, especially around the 2016 US presidential election and the 2018 US State of the Union Address, its estimated organic click rate has steadily decreased since early 2018. In contrast, infowars.com has a more volatile readership. The spikes in clicks to infowars.com could be explained by media coverage of the website, including the defamation case against Alex Jones in April 2018 who claimed the shooting at Sandy Hook Elementary School was “completely fake” and a “giant hoax”. Since then, several internet companies — including Apple, Twitter, Facebook, Spotify, and YouTube — banned *Infowars* from their platforms, and the domain has not been able to regain its clicks nor value since. This demonstrates the powerful role platforms play in not only making content visible to users, but also controlling who can grow their website value — and ultimately generate revenue — from the content they produce and share online.

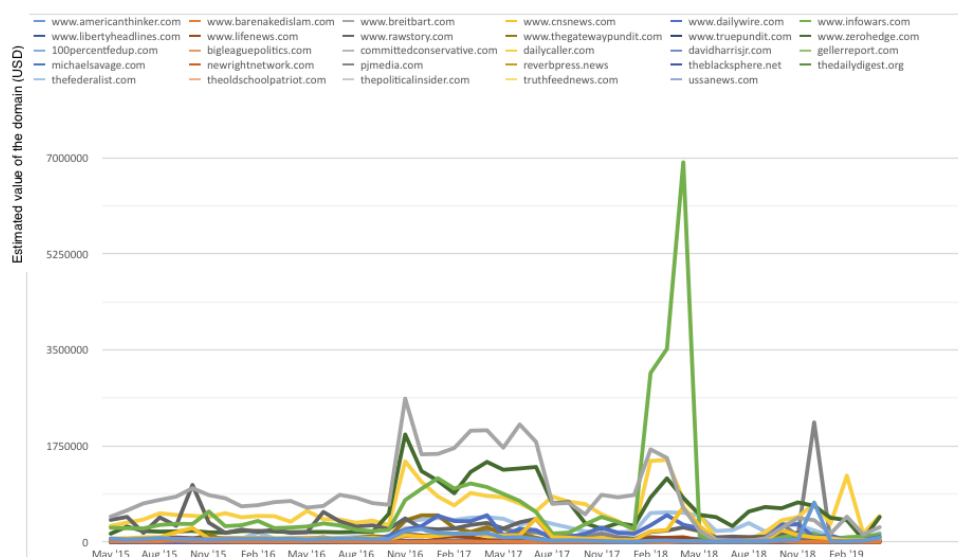


Figure 4: Estimated organic value for the top 29 junk news domains (May 2015 – March 2019)

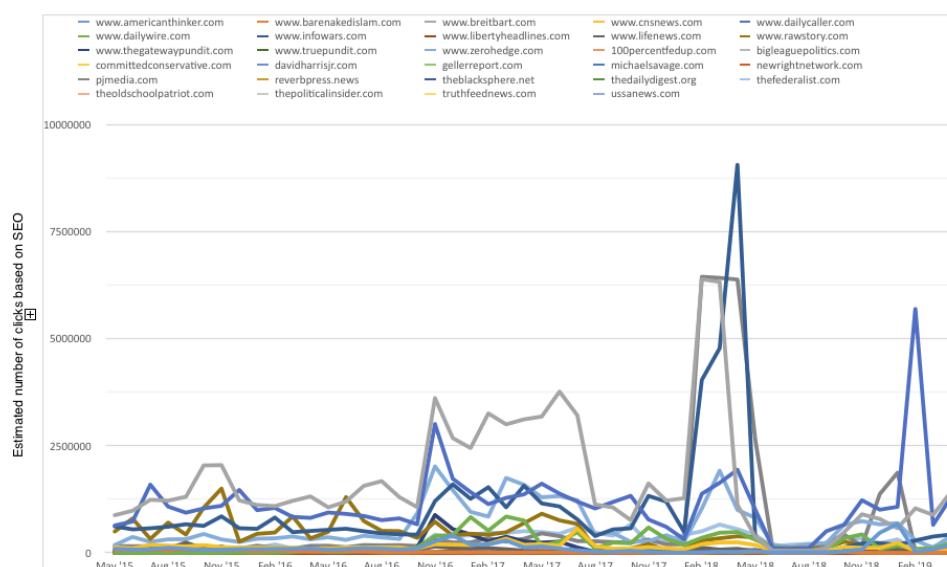


Figure 5: Estimated organic clicks for the top 29 junk news domains (May 2015–April 2019)

JUNK NEWS DOMAINS, SEARCH DISCOVERABILITY AND GOOGLE'S RESPONSE TO DISINFORMATION

Figure 6 shows the estimated organic results of the top 29 junk news domains overtime. The estimated organic results are the number of keywords that would organically appear in Google searches. Since August 2017, there has been a sharp decline in the number of keywords that would appear in Google. The four top-performing junk news websites (infowars.com, zerohedge.com, dailycaller.com, and breitbart.com) all appeared less frequently in top-positions on Google Search based on the keywords they were optimising for. This is an interesting finding and suggests that the changes Google made to its search algorithm did indeed have an impact on the discoverability of junk news domains after August 2017. In comparison, other professional news sources (washingtonpost.com, nytimes.com, foxnews.com, nbcnews.com, bloomberg.com, bbc.co.uk, wsj.com, and cnn.com) did not see substantial drops in their search visibility during

this timeframe (see Figure 7). In fact, after August 2017 there has been a gradual increase in the organic results of mainstream news media.

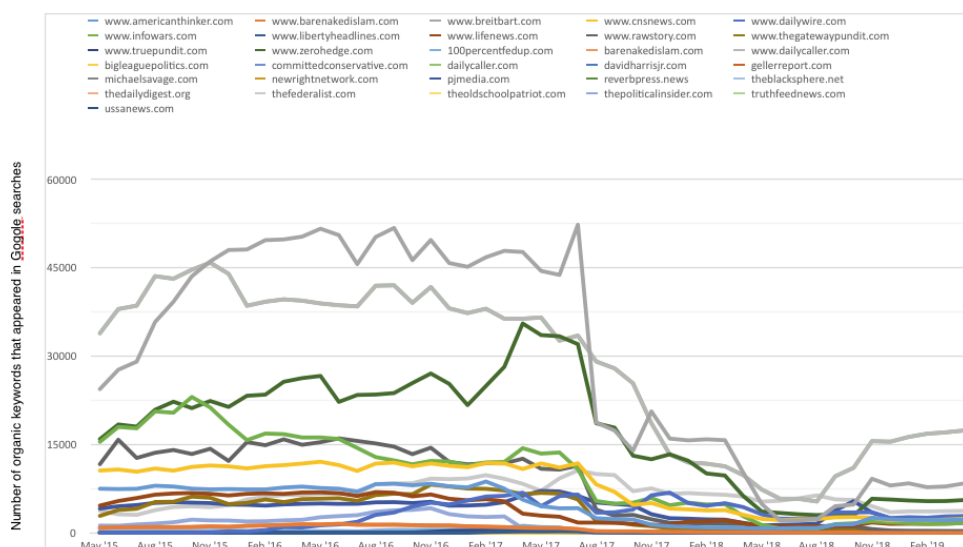


Figure 6: Estimated organic results for the top 29 junk news domains (May 2015- April 2019)

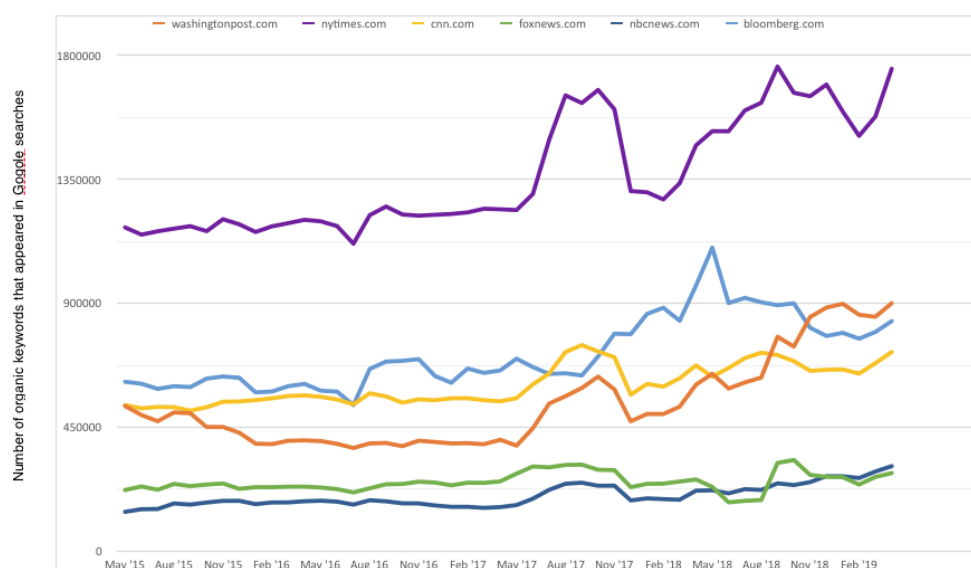


Figure 7: Estimated organic results for mainstream media websites in the United States (May 2015- April 2019)

After almost a year, the top-performing junk news websites have regained some of their organic results, but the levels are not nearly as high as they were leading up to and preceding the 2016 presidential election. This demonstrates the power of Google's algorithmic changes in limiting the discoverability of junk news on Search. But it also shows how junk news producers learn to adapt their strategies in order to extend the visibility of their content. In order to be effective at limiting the visibility of bad information via search, Google must continue to monitor the keywords and optimisation strategies these domains deploy — especially in the lead-up to elections — when more people will be naturally searching for news and information about politics.

CONCLUSION

In conclusion, the spread of junk news on the internet and the impact it has on democracy has certainly been a growing field of academic inquiry. This paper has looked at a small subset of this phenomenon, in particular the role of Google Search in assisting in the discoverability and monetisation of junk news domains. By looking at the techno-commercial infrastructure that junk news producers use to optimise their websites for paid and pseudo-organic clicks, I found:

1. Junk news domains do not rely on Google advertisements to grow their audiences and instead focus their efforts on optimisation and keyword strategies;
2. Navigational searches drive the most traffic to junk news websites, and data voids are used to grow the discoverability of junk news content to mostly small, but varying degrees.
3. Many junk news producers place advertisements on their websites and grow their value particularly around important political events; and
4. Overtime, the SEO strategies used by junk news domains have decreased in their ability to generate top-placements in Google Search.

For millions of people around the world, the information Google Search recommends directly impacts how ideas and opinions about politics are formulated. The powerful role of Google as an information gatekeeper has meant that bad actors have tried to subvert these technical systems for political or economic game. For quite some time, Google's algorithms have come under attack by spammers and other malign actors who wish to spread disinformation, conspiracy theories, spam, and hate speech to unsuspecting users. The rise of "computational propaganda" and the variety of bad actors exploiting technology to influence political outcomes has also led to the manipulation of Search. Google's response to the optimisation strategies used by junk news domains has had a positive effect on limiting the discoverability of these domains over time. However, the findings of this paper are also showing an upward trend, as junk news producers find new ways to optimise their content for higher search rankings. This game of cat and mouse is one that will continue for the foreseeable future.

While it is hard to reduce the visibility of junk news domains when individuals actively search for them, more can be done to limit the ways in which bad actors might try to optimise content to generate pseudo-organic engagement, especially around disinformation. Google can certainly do more to tweak its algorithms in order to demote known disinformation sources, as well as identify and limit the discoverability of content seeking to exploit data voids. However, there is no straightforward technical patch that Google can implement to stop various actors from trying to game their systems. By co-opting the technical infrastructure and policies that enable search, the producers of junk news are able to spread disinformation — albeit to small audiences who might use obscure search terms to learn about a particular topic.

There have also been growing pressures for regulators to take steps that force social media platforms to take greater actions that limit the spread of disinformation online. But the findings of this paper have two important lessons for policymakers. First, the disinformation problem — through both optimisation and advertising — on Google Search is not as dramatic as it is sometimes portrayed. Most of the traffic to junk news websites are by users performing navigational searches to find specific, well-known brands. Only a limited number of placements — as well as clicks — to junk news domains come from pseudo-organic engagement generated by data voids and other problematic keyword searches. Thus, requiring Google to take a heavy-handed approach to content moderation could do more harm than good, and might not reflect the severity of the problem. Second, the reason why disinformation spreads on Google are

reflective of deeper systemic problems within democracies: growing levels of polarisation and distrust in the mainstream media are pushing citizens to fringe and highly partisan sources of news and information. Any solution to the spread of disinformation on Google Search will require thinking about media and digital literacy and programmes to strengthen, support, and sustain professional journalism.

REFERENCES

- Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives*, 31(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>
- Barrett, B., & D. Kreiss (2019). Platform Transience: changes in Facebook’s policies, procedures and affordances in global electoral politics. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1446>
- Bradshaw, S., Howard, P., Kollanyi, B., & Neudert, L.-M. (2019). Sourcing and Automation of Political News and Information over Social Media in the United States, 2016–2018. *Political Communication*. <https://doi.org/10.1080/10584609.2019.1663322>
- Bradshaw, S., & Howard, P. N. (2018). *Why does Junk News Spread So Quickly Across Social Media? Algorithms, Advertising and Exposure in Public Life* [Working Paper]. Miami: Knight Foundation. Retrieved from https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/142/original/Topos_KF_White-Paper_Howard_V1_ado.pdf
- Bradshaw, S., Neudert, L.-M., & Howard, P. (2018). *Government Responses to the Malicious Use of Social Media*. NATO.
- Burkell, J., & Regan, P. (2019). Voting Public: Leveraging Personal Information to Construct Voter Preference. In N. Witzleb, M. Paterson, & J. Richardson (Eds.), *Big Data, Privacy and the Political Process*. London: Routledge.
- Cadwalladr, C. (2016a, December 4). Google, democracy and the truth about internet search. *The Observer*. Retrieved from <https://www.theguardian.com/technology/2016/dec/04/google-democracy-truth-internet-search-facebook>
- Cadwalladr, C. (2016b, December 11). Google is not ‘just’ a platform. It frames, shapes and distorts how we see the world. *The Guardian*. Retrieved from <https://www.theguardian.com/commentisfree/2016/dec/11/google-frames-shapes-and-distorts-how-we-see-world>
- Chester, J. & Montgomery, K. (2019). The digital commercialisation of US politics—2020 and beyond. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1443>
- Dean, B. (2019). Google’s 200 Ranking Factors: The Complete List (2019). Retrieved April 18, 2019, from Backlinko website: <https://backlinko.com/google-ranking-factors>
- Desiguid, C., Howard, P. N., Kollanyi, B., & Bradshaw, S. (2017). *Junk News and Bots during the French Presidential Election: What are French Voters Sharing Over Twitter In Round Two?* [Data Memo No. 2017.4]. Oxford: Project on Computational Propaganda, Oxford University. Retrieved May 19, 2017, from <http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/05/What-Are-French-Voters-Sharing-Over-Twitter-Between-the-Two-Rounds-v7.pdf>
- European Commission. (2018). *A multi-dimensional approach to disinformation: report of the independent high-level group on fake news and online disinformation*. Luxembourg: European Commission.

- Evangelista, R., & F. Bruno. (2019) WhatsApp and political instability in Brazil: targeted messages and political radicalization. *Internet Policy Review*, 8(4).
<https://doi.org/10.14763/2019.4.1435>
- Farkas, J., & Schou, J. (2018). Fake News as a Floating Signifier: Hegemony, Antagonism and the Politics of Falsehood. *Journal of the European Institute for Communication and Culture*, 25(3), 298–314. <https://doi.org/10.1080/13183222.2018.1463047>
- Gillespie, T. (2012). The Relevance of Algorithms. In T. Gillespie, P. J. Boczkowski, & K. Foot (Eds.), *Media Technologies: Essays on Communication, Materiality and Society* (pp. 167–193). Cambridge, MA: The MIT Press. Retrieved from
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.692.3942&rep=rep1&type=pdf>
- Golebiewski, M., & Boyd, D. (2018). *Data voids: where missing data can be easily exploited*. Retrieved from Data & Society website: https://datasociety.net/wp-content/uploads/2018/05/Data_Society_Data_Voids_Final_3.pdf
- Google. (2019). How Google Search works: Search algorithms. Retrieved April 17, 2019, from <https://www.google.com/intl/en/search/howsearchworks/algorithms/>
- Granka, L. A. (2010). The Politics of Search: A Decade Retrospective. *The Information Society*, 26(5), 364–374. <https://doi.org/10.1080/01972243.2010.511560>
- Guo, L., & Vargo, C. (2018). “Fake News” and Emerging Online Media Ecosystem: An Integrated Intermedia Agenda-Setting Analysis of the 2016 U.S. Presidential Election. *Communication Research*. <https://doi.org/10.1177/0093650218777177>
- Hedman, F., Sivnert, F., Kollanyi, B., Narayanan, V., Neudert, L. M., & Howard, P. N. (2018, September 6). *News and Political Information Consumption in Sweden: Mapping the 2018 Swedish General Election on Twitter* [Data Memo No. 2018.3]. Oxford: Project on Computational Propaganda, Oxford University. Retrieved from
<https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/09/Hedman-et-al-2018.pdf>
- Hoffmann, S., Taylor, E., & Bradshaw, S. (2019, October). The Market of Disinformation. [Report]. Oxford: Oxford Information Labs; Oxford Technology & Elections Commission, University of Oxford. Retrieved from <https://oxtec.oii.ox.ac.uk/wp-content/uploads/sites/115/2019/10/OxTEC-The-Market-of-Disinformation.pdf>
- Howard, P., Ganesh, B., Liotsiou, D., Kelly, J., & Francois, C. (2018). *The IRA and Political Polarization in the United States, 2012–2018* [Working Paper No. 2018.2]. Oxford: Project on Computational Propaganda, Oxford University. Retrieved from
<https://comprop.oii.ox.ac.uk/research/ira-political-polarization/>
- Howard, P. N., Bolsover, G., Kollanyi, B., Bradshaw, S., & Neudert, L.-M. (2017). *Junk News and Bots during the U.S. Election: What Were Michigan Voters Sharing Over Twitter?* [Data Memo No. 2017.1]. Oxford: Project on Computational Propaganda, Oxford University. Retrieved from
<http://comprop.oii.ox.ac.uk/2017/03/26/junk-news-and-bots-during-the-u-s-election-what-we-re-michigan-voters-sharing-over-twitter/>
- Howard, P. N., Kollanyi, B., Bradshaw, S., & Neudert, L.-M. (2017). *Social Media, News and Political Information during the US Election: Was Polarizing Content Concentrated in Swing*

States? [Data Memo No. 2017.8]. Oxford: Project on Computational Propaganda, Oxford University. Retrieved from <http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2017/09/Polarizing-Content-and-Swing-States.pdf>

Introna, L., & Nissenbaum, H. (2000). Shaping the Web: Why the Politics of Search Engines Matters. *The Information Society*, 16(3), 169–185. <https://doi.org/10.1080/01972240050133634>

Kaminska, M., Galacher, J. D., Kollanyi, B., Yasseri, T., & Howard, P. N. (2017). *Social Media and News Sources during the 2017 UK General Election*. [Data Memo No. 2017.6]. Oxford: Project on Computational Propaganda, Oxford University. Retrieved from <https://www.oii.ox.ac.uk/blog/social-media-and-news-sources-during-the-2017-uk-general-election/>

Kim, Y. M., Hsu, J., Neiman, D., Kou, C., Bankston, L., Kim, S. Y., ... Raskutti, G. (2018). The Stealth Media? Groups and Targets behind Divisive Issue Campaigns on Facebook. *Political Communication*, 35(4), 515–541. <https://doi.org/10.1080/10584609.2018.1476425>

Lakoff, G. (2018, January 2). Trump uses social media as a weapon to control the news cycle. Retrieved from <https://twitter.com/GeorgeLakoff/status/948424436058791937>

Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>

Lewis, P. & Hilder, P. (2018, March 23). Leaked: Cambridge Analytica's Blueprint for Trump Victory. *The Guardian*. Retrieved from: <https://www.theguardian.com/uk-news/2018/mar/23/leaked-cambridge-analyticas-blueprint-for-trump-victory>

Machado, C., Kira, B., Hirsch, G., Marchal, N., Kollanyi, B., Howard, Philip N., ... Barash, V. (2018). *News and Political Information Consumption in Brazil: Mapping the First Round of the 2018 Brazilian Presidential Election on Twitter* [Data Memo No. 2018.4]. Oxford: Project on Computational Propaganda, Oxford University. Retrieved from http://blogs.oii.ox.ac.uk/comprop/wp-content/uploads/sites/93/2018/10/machado_et_al.pdf

McKelvey F. (2019). Cranks, Clickbaits and Cons: On the acceptable use of political engagement platforms. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1439>

Metaxa-Kakavouli, D., & Torres-Echeverry, N. (2017). Google's Role in Spreading Fake News and Misinformation. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3062984>

Metaxas, P. T. (2010). Web Spam, Social Propaganda and the Evolution of Search Engine Rankings. In J. Cordeiro & J. Filipe (Eds.), *Web Information Systems and Technologies* (Vol. 45, pp. 170–182). https://doi.org/10.1007/978-3-642-12436-5_13

Nagy, P., & Neff, G. (2015). Imagined Affordance: Reconstructing a Keyword for Communication Theory. *Social Media + Society*, 1(2). <https://doi.org/10.1177/2056305115603385>

Narayanan S., & Kalyanam K. (2015). Position Effects in Search Advertising and their Moderators: A Regression Discontinuity Approach. *Marketing Science*, 34(3), 388–407.

<https://doi.org/10.1287/mksc.2014.0893>

Neff, G., & Nagy, P. (2016). Talking to Bots: Symbiotic Agency and the Case of Tay. *International Journal of Communication*, 10, 4915–4931. Retrieved from <https://ijoc.org/index.php/ijoc/article/view/6277>

Neudert, L.-M., Howard, P., & Kollanyi, B. (2017). *Junk News and Bots during the German Federal Presidency Election: What Were German Voters Sharing Over Twitter?* [Data Memo 2 No. 2017.2]. Oxford: Project on Computational Propaganda, Oxford University. Retrieved from <http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/89/2017/03/What-Were-German-Voters-Sharing-Over-Twitter-v6-1.pdf>

Nielsen, R. K., & Graves, L. (2017). “News you don’t believe”: Audience perspectives on fake news. Oxford: Reuters Institute for the Study of Journalism, University of Oxford. Retrieved from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2017-10/Nielsen&Graves_factsheet_1710v3_FINAL_download.pdf

Noble, S. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press.

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., & Granka, L. (2007). In Google We Trust: Users’ Decisions on Rank, Position, and Relevance. *Journal of Computer-Mediated Communication*, 12(3), 801–823. <https://doi.org/10.1111/j.1083-6101.2007.00351.x>

Pasquale, F. (2015). *The Black Box Society*. Cambridge: Harvard University Press.

Persily, N. (2017). The 2016 U.S. Election: Can Democracy Survive the Internet? *Journal of Democracy*, 28(2), 63–76. <https://doi.org/10.1353/jod.2017.0025>

Schroeder, R. (2014). Does Google shape what we know? *Prometheus*, 32(2), 145–160. <https://doi.org/10.1080/08109028.2014.984469>

Silverman, C. (2016, November 16). This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook. *Buzzfeed*. Retrieved July 25, 2017 from <https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook>

SpyFu. (2019). SpyFu - Competitor Keyword Research Tools for AdWords PPC & SEO. Retrieved April 19, 2019, from <https://www.spyfu.com/>

Star, S. L., & Ruhleder, K. (1994). Steps Towards an Ecology of Infrastructure: Complex Problems in Design and Access for Large-scale Collaborative Systems. *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, 253–264. New York: ACM.

Tambini, D., Anstead, N., & Magalhães, J. C. (2017, June 6). Labour’s advertising campaign on Facebook (or “Don’t Mention the War”) [Blog Post]. Retrieved April 11, 2019, from Media Policy Blog website: <http://blogs.lse.ac.uk/mediapolicyproject/>

Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining “Fake News”: A typology of scholarly definitions *Digital Journalism*, 6(2). <https://doi.org/10.1080/21670811.2017.1360143>

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D., & Nyhan, B. (2018, March). *Social Media, Political Polarization, and Political Disinformation: A Review of*

the Scientific Literature [Report]. Menlo Park: William and Flora Hewlett Foundation. Retrieved from <https://eprints.lse.ac.uk/87402/1/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf>

Vaidhynathan, S. (2011). *The Googlization of Everything*: (First edition). Berkeley: University of California Press.

Vargo, C. J., Guo, L., & Amazeen, M. A. (2018). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New Media & Society*, 20(5), 2028–2049. <https://doi.org/10.1177/1461444817712086>

Bashyakarla, V. (2019). Towards a holistic perspective on personal data and the data-driven election paradigm. *Internet Policy Review*, 8(4). Retrieved from <https://policyreview.info/articles/news/towards-holistic-perspective-personal-data-and-data-driven-election-paradigm/1445>

Wardle, C. (2017, February 16). Fake news. It's complicated. *First Draft News*. Retrieved July 20, 2017, from <https://firstdraftnews.com:443/fake-news-complicated/>

Wardle, C., & Derakhshan, H. (2017). *Information Disorder: Toward and interdisciplinary framework for research and policy making* [Report No. DGI(2017)09]. Strasbourg: Council of Europe. Retrieved from <https://rm.coe.int/information-disorder-report-november-2017/1680764666>

Winner, L. (1980). Do Artifacts have Politics. *Daedalus*, 109(1), 121–136. Retrieved from <http://www.jstor.org/stable/20024652>

APPENDIX 1

Junk news seed list (Computational Propaganda Project's top-29 junk news domains from the 2018 US midterm elections).

www.americanthinker.com, www.barenakedislam.com, www.breitbart.com, www.cnsnews.com, www.dailywire.com, www.infowars.com, www.libertyheadlines.com, www.lifenews.com, www.rawstory.com, www.thegatewaypundit.com, www.truepundit.com, www.zerohedge.com, 100percentfedup.com, bigleaguepolitics.com, committedconservative.com, dailycaller.com, davidharrisjr.com, gellerreport.com, michaelsavage.com, newrightnetwork.com, pjmedia.com, reverbpnews.com, theblacksphere.net, thedailydigest.org, thefederalist.com, ussanews.com, theoldschoolpatriot.com, thepoliticalinsider.com, truthfeednews.com.

APPENDIX 2

Table 2: A sample list of up to ten keywords from each junk news domain in the sample when the keyword reached the first position.

100percentfedup.com	dailywire.com	theblacksphere.net	
gruesome videos	6 states bankrupt	22 black sphere	28
snopes exposed	5 ms 13 portland oregon	15 dwayne johnson gay	10
gruesome video	4 the gadsen flag	12 george soros private security	1

teendreamers	2	f word on tv	12	bombshell barack	1
bush cheney inauguration	2	against gun control facts	10	madame secretary	1
americanthinker.com		end of america 90	9	head in vagina	1
medienkritic	23	racist blacks	8	mexicans suck	1
problem with taxes	22	associates clinton	8	obama homosexual	1
janet levy	19	diebold voting machine	8	comments this	1
article on environmental protection	18	diebold machines	8	thefederalist.com	
maya angelou criticism	18	gellerreport.com		the federalist	39
supply and demand articles 2011	17	geller report	1	federalist	30
ezekiel emanuel complete lives system	16	infowars.com		gun control myths	26
articles on suicide	12	www infowars	39	considering homeschooling	23
American Thinker Coupons	11	infowars com	39	why wont it work technology	22
truth about obama	10	info wars	39	debate iraq war	21
barenakedislam.com		infowars	39	lesbian children	20
berg beheading video	11	www infowars com	39	why homeschooling	19
against islam	11	al-qaeda 100 pentagon run	38	home economics course	18
beheadings	10	info war today	35	iraq war debate	17
iraquis beheaded	10	war info	34	thegatewaypundit.com	
muslim headgear	8	infowars moneybomb	34	thegatewaypundit.com	39
torture clips	7	feminizing uranium	33	civilian national security force	10
los angeles islam pictures	7	libertyheadlines.com		safe school czar	8
beheaded clips	7	accusers dod	2	hillary clinton weight gain 2011	8
berg video	7	liberty security guard bucks country	1	RSS Pundit	7
hostages beheaded	6	lifeneews.com		hillary clinton weight gain	7
bigleaguepolitics.com		successful people with down syndrome	39	all perhaps hillary	4
habermans	1	life news	35	hillary clinton gained weight	4
fbi whistleblower	1	lifeneews.com	35	london serendip i tea camp	4
ron paul supporters	1	fetus after abortion	26	whoa it	4
breitbart.com		anti abortion quotes	21	thepoliticalinsider.com	
big journalism	39	pro life court cases	17	obama blames	19
big government breitbart	39	rescuing hug	16	michael moore sucks	14
breitbart blog	39	process of aborting a baby	15	marco rubio gay	11
www.breitbart.com	39	different ways to abort a baby	14	weapons mass destruction iraq	10
big hollywood	39	adoption waiting list statistics	14	weapons of mass destruction found	10
breitbart hollywood	39	michaelsavage.com		wmd iraq	10
breitbart.com	39	www michaelsavage com	19	obama s plan	9
big hollywood blog	39	michaelsavage com	19	chuck norris gay	9
big government blog	39	michaelsavage	18	how old is bill clinton	8
breitbart big hollywood	39	michael savage com	18	stop barack obama	7
cnsnews.com		michaelsavage radio	17	truepundit.com	
cns news	39	michael savage	17	john kerrys daughter	8
cnsnews	39	savage nation	15	john kerrys daughters	5
conservative news service	39	michael savage nation	14	sex email	2
christian news service	21	michael savage savage nation	13	poverty warrior	2

cns	20 the savage nation	12 john kerry daughter	1
major corporations	20 pjmedia.com	RSS Pundit	1
billy graham daughter	18 belmont club	39 whistle new	1
taxing the internet	17 belmont club blog	39 pay to who	1
pashtun sexuality	15 pajamas media	39 truthfeednews.com	
record tax	15 dr helen	38 nfl.comm	5
dailycaller.com	instapundit blog	38 ussanews.com	
the daily caller	37 instapundit	33 imigration expert	2
vz 58 vs ak 47	33 pj media	33 meabolic syndrome	1
condition black	28 instapundit.	32 zerohedge.com	
patriot act changes	26 google ddrive	28 zero hedge	33
12 hour school	25 instapundits	27 unempolymnet california	24
common core stories	25 rawstory.com	hayman capital letter	24
courtroom transcript	23 the raw story	39 dennis gartman performance	24
why marijuana shouldnt be legal	22 raw story	39 the real barack obama	23
why we shouldnt legalize weed	22 rawstory	39 meredith whitney blog	22
why shouldnt marijuana be legalized	22 rawstory.com	39 weaight watchers	22
	westboro baptist church tires slashed	35 ohedge	22
	the raw	25 doug kass predictions	19
	mormons in porn	22 usa hyperinflation	17
	norm colemans teeth	19	
	xe services sold	18	
	duggers	17	

FOOTNOTES

1. Organic engagement is used to describe authentic user engagement, where an individual might click a website or link without being prompted. This is different from "transactional engagement" where a user engages with content through prompting via paid advertising. In contrast, I use the term "pseudo-organic engagement" to capture the idea that SEO practitioners are generating clicks through the manipulation of keywords that move websites closer to the top of search engine rankings. An important aspect of pseudo-organic engagement is that these results are indistinguishable from those that have "earnt" their search ranking, meaning, users may be more likely to treat the source as authoritative despite the fact their ranking has been manipulated.

2. It is important to note that AdWord purchases can also be displayed on affiliate websites. These "display ads" appear on websites and generate revenue for the website operator.

3. For the US presidential election, 19.53 million tweets were collected between 1 November 2016, and 9 November 2016; for the State of the Union Address 2.26 million tweets were collected between 24 January 2018, and 30 January 2018; and for the 2018 US midterm elections 2.5 million tweets were collected between 21-30 September 2018 and 6,986 Facebook groups between 29 September 2018 and 29 October 2018. For more information see Bradshaw et al., 2019.

4. Elections include: 2016 United States presidential election, 2017 French presidential election, 2017 German federal election, 2017 Mexican presidential election, 2018 Brazilian presidential election, and the 2018 Swedish general election.