



Sourcing and Automation of Political News and Information over Social Media in the United States, 2016-2018

Samantha Bradshaw , Philip N. Howard , Bence Kollanyi & Lisa-Maria Neudert

To cite this article: Samantha Bradshaw , Philip N. Howard , Bence Kollanyi & Lisa-Maria Neudert (2020) Sourcing and Automation of Political News and Information over Social Media in the United States, 2016-2018, Political Communication, 37:2, 173-193, DOI: [10.1080/10584609.2019.1663322](https://doi.org/10.1080/10584609.2019.1663322)

To link to this article: <https://doi.org/10.1080/10584609.2019.1663322>



© 2019 The Author(s). Published with license by Taylor & Francis Group, LLC.



[View supplementary material](#)



Published online: 26 Oct 2019.



[Submit your article to this journal](#)



Article views: 4949



[View related articles](#)




[View Crossmark data](#)



Citing articles: 2 [View citing articles](#)



Sourcing and Automation of Political News and Information over Social Media in the United States, 2016-2018

SAMANTHA BRADSHAW, PHILIP N. HOWARD,  BENCE KOLLANYI, and LISA-MARIA NEUDERT

Social media is an important source of news and information in the United States. But during the 2016 US presidential election, social media platforms emerged as a breeding ground for influence campaigns, conspiracy, and alternative media. Anecdotally, the nature of political news and information evolved over time, but political communication researchers have yet to develop a comprehensive, grounded, internally consistent typology of the types of sources shared. Rather than chasing a definition of what is popularly known as “fake news,” we produce a grounded typology of what users actually shared and apply rigorous coding and content analysis to define the phenomenon. To understand what social media users are sharing, we analyzed large volumes of political conversations that took place on Twitter during the 2016 presidential campaign and the 2018 State of the Union

Samantha Bradshaw is a DPhil candidate at the Oxford Internet Institute and works on the Computational Propaganda project as a Research Assistant. Prior to joining the COMPROP team, she worked at the Centre for International Governance Innovation in Waterloo, Canada, where she was a key member of a small team facilitating the Global Commission on Internet Governance. Samantha holds an MA in global governance from the Balsillie School of International Affairs, and a joint honours BA in political science and legal studies from the University of Waterloo. Philip N. Howard is the Director of the OII. He is a professor of sociology, information and international affairs and a statutory Professor of Internet Studies at the Oxford Internet Institute and Balliol College at the University of Oxford. He has courtesy appointments as a professor at the University of Washington’s Department of Communication and as a fellow at Columbia University’s Tow Center for Digital Journalism. Bence Kollanyi is a PhD candidate in Sociology at the Corvinus University of Budapest where his research focuses on the automation of social media accounts and the development of open-source bots for political purposes. He is a data scientist at the COMPROP research team at the Oxford Internet Institute and holds a MS from Georgia Institute of Technology. Lisa-Maria Neudert is a doctoral candidate at the Oxford Internet Institute and a core researcher at the Computational Propaganda Project, where her work is located at the nexus of political communication, technology studies and governance. Her current research is looking into the public and private governance of policy issues surrounding disinformation through governments and social media platforms. Lisa-Maria holds a MSc in Social Science of the Internet from the University of Oxford and a BA in Communication Science from the Ludwig-Maximilians-University in Munich. Selected as a Fulbright scholar she studied at the Georgetown University in Washington DC and the National University of Singapore.

Address correspondence to Lisa-Maria Neudert, Oxford Internet Institute, 1 St Giles, Ox1 3LW Oxford. E-mail: lisa-maria.neudert@oii.ox.ac.uk

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

address in the United States. We developed the concept of “junk news,” which refers to sources that deliberately publish misleading, deceptive, or incorrect information packaged as real news. First, we found a 1:1 ratio of junk news to professionally produced news and information shared by users during the US election in 2016, a ratio that had improved by the State of the Union address in 2018. Second, we discovered that amplifier accounts drove a consistently higher proportion of political communication during the presidential election but accounted for only marginal quantities of traffic during the State of the Union address. Finally, we found that some of the most important units of analysis for general political theory—parties, the state, and policy experts—generated only a fraction of the political communication.

Keywords Social media, political news and information, news sources, elections, junk news, US 2016 election, state of the union address

Introduction

The spread of misinformation over social media platforms has become a critical public interest issue. During the 2016 presidential election in the United States, a wide range of politically and economically motivated actors exploited social media platforms to spread propaganda, hate speech, and conspiracy theories to voters. Citizens were flooded with highly polarizing messages, hyper-partisan commentary, and troll rhetoric in a coordinated effort to manipulate public opinion. Foreign agents, partisan interest groups, and profit-driven actors homed in on numerous communities with sensational messages, “clickbait” content, and highly politicized influence campaigns. In a climate characterized by declining levels of trust in the media and public institutions, there is increasing concern about the impact misinformation is having on democracy.

Following the 2016 US presidential election, there has been a growing body of empirical work demonstrating how large volumes of misinformation can circulate over social media (Allcott & Gentzkow, 2017; Vicario et al., 2016; Vosoughi, Roy, & Aral, 2018). Scholars have argued that the spread of “computational propaganda” sustained by algorithms can negatively impact democratic discourse (Bradshaw & Howard, 2018a; Marchal & Neudert, 2019; Persily, 2017; Tucker, Theoharis, Roberts, & Barberá, 2017; Woolley & Howard, 2016, 2017). Indeed, both social network infrastructure and user behaviors provide capacities and constraints for the spread of misinformation (Bradshaw & Howard, 2018b; Flaxman, Goel, & Rao, 2016; Marwick & Lewis, 2017; Pariser, 2011; Wu, 2017). Yet the body of work that is devoted to conceptualizing the broader “information disorder” (Wardle & Derakhshan, 2017) has remained fragmentary, lacks a common vocabulary, and is increasingly weaponized by politically motivated actors (Neudert, Howard, & Kollanyi, 2017; Boyd, 2017; Fletcher, Cornia, Graves, & Nielsen, 2017).

Drawing on perspectives from political communication, this paper develops a grounded typology to understand the nature of information shared over social media. We collected 21.8 million tweets and examined the domains that users shared on Twitter during the 2016 US presidential election and the State of the Union (SOTU) address in January 2018. Based on a systematic analysis of 710 sources, we developed a grounded typology of news and information sources. Rather than assessing the accuracy of individual stories that users shared, our grounded typology distinguishes different types of news and information domains. This method allows for a comprehensive evaluation of web sources and their reporting.

We found an increasing prevalence of junk news, which we use as a distinct concept to describe news and information that fulfills at least three out of the following five criteria: (1) professionalism, which is where sources do not employ the standards and best

practices of professional journalism, including information about real authors, editors, and owners; (2) style, which is where emotionally driven language, ad hominem attacks, and mobilizing language and imagery are used; (3) credibility, which is where sources rely on false information or conspiracy theories and do not post corrections; (4) bias, which is where sources are highly biased, ideologically skewed, and publish opinion pieces as news; and (5) counterfeit, which is where sources mimic established news reporting by using certain fonts, having branding, and employing content strategies.

To help advance our theoretical and empirical understandings of contemporary political communication, we ask three questions:

- What are the salient qualities of political news and information shared over social media?
- How much of the content purporting to be political news and information during the presidential election in 2016 and the SOTU address in 2018 was junk news?
- How much junk news was organically shared by humans and how much was at least partially amplified during these two events?

To develop a grounded typology, first we situate our concept of junk news in the literature on political communication. Second, we discuss the foundational role of typology building when it comes to understanding new phenomena in communication research. Third, we outline our methodology for building a grounded typology of news and information and coding junk news sources during two recent political events in the United States, the 2016 presidential election and the 2018 SOTU address. Fourth, we analyze the overall trends across the two events and reflect on the changing patterns of political discourse online. We conclude with a discussion of our findings for public life and the future of political communication research.

Contemporary Political Communication and the Rise of Junk News

Since the Brexit referendum in the United Kingdom and the 2016 presidential election in the United States, scholars from a wide range of disciplines have become increasingly concerned with the spread of so-called fake news on social media. Although “fake news” is not a new phenomenon (Florida, 2016), it has engendered a new research agenda that examines the purveyors of “fake news,” how it interacts with human cognition and behavior, and the impact its consumption has on democratic decisions and polarization (De Keersmaecker & Roets, 2017; Hambrick & Marquardt, 2018; Tucker et al., 2017; Marchal & Neudert, 2019). Underlying these studies is a wide range of conceptualizations of “fake news” and other forms of poor information, including mis- and disinformation, hyper-partisan or alternative media, political satire, rumor and conspiracy, and so-called clickbait content (Wardle, 2017).

Given the definitional scope of “fake news,” scholars have identified a number of epistemological challenges surrounding its vocabulary, including the limitations of its binary nomenclature, as well as increasing appropriation of it by political elites (European Commission, 2018). Part of the complexity regarding defining types of problematic information stems from the variety of motivations and actors who purposefully deceive (disinformation) users compared to those who unintentionally share misleading information (misinformation). This complexity makes it difficult to disentangle the characteristics of “fake news” into limited categories since there are several overlapping features that define these sources.

In response to a growing need for a comprehensive definition that captures the variety of information being shared over social media, we analyzed 710 sources and identified a growing prevalence of junk news domains. Grounded in the literature on political communication, junk news is a distinct concept we developed to systematically analyze the characteristics of the different forms of problematic information online. Junk news satisfies at least three out of the five following criteria: (1) professionalism, (2) style, (3) credibility, (4) bias, and (5) counterfeit. Our coding decisions encapsulate the content of the domain as a whole (i.e., style, bias, credibility), information about authors and the organization (i.e., professionalism), and the layout and design of the domain itself (i.e., counterfeit). A criterion by itself is a necessary but not sufficient condition that confirms that junk news exists. Partisan news organizations that report from an ideological angle (bias) would not be considered junk news unless they also satisfied at least two of the other criteria. For example, the *Huffington Post* and *Fox News* are examples of where we found elements of bias, but content production generally adheres to journalistic standards and does not counterfeit major news brands, and reporting is generally credible. Our coding process for determining junk news domains is described in greater detail below.

Professionalism

Several scholars have noted the rise of low-quality news and information organizations that do not adhere to standards of journalistic content production. Lazer et al. (2018) describe a lack of professionalism and adherence to journalistic norms that are present in several information sources being shared online. Distinct from other forms of user-generated content and citizen journalism, junk news domains satisfy the professionalism criterion because they purposefully refrain from providing clear information about real authors, editors, publishers, and owners, and they do not publish corrections of debunked information. When evaluating sources for elements of professionalism, we systematically checked the about pages of domains for contact information, information about ownership and editors, and other information relating to professional standards. In order to find evidence of corrections, we reviewed whether the sources appeared in third-party fact-checking reports. Then we checked whether sources published corrections of fact-checked reporting. Examples of domains that fail on professionalism include zerohedge.com, conservative-fighters.org, and deepstatenation.news.

Counterfeit

Some sources mimic established news reporting by using certain fonts, having branding, and employing content strategies. For example, Tandoc, Lim, and Ling (2018) highlight how fake news websites use fabrication to draw in readers by counterfeiting legacy news organizations. This is because users who are unfamiliar with a source will often apply mental heuristics to determine its legitimacy (Flanagin & Metzger, 2007). As a result, junk news is stylistically disguised as professional news by the inclusion of references to news agencies and credible sources as well as headlines written in a news tone with date, time, and location stamps. In the most extreme cases, outlets will copy logos and counterfeit entire domains. One prominent example of counterfeiting was the *Denver Guardian*, which mimicked the legitimate local news organization *The Denver Post* to spread disinformation and conspiracy theories about Hillary Clinton (Lubbers, 2016). In order to identify the counterfeit criterion, we systematically reviewed organizational

information about the owner and headquarters by checking sources like Wikipedia, the WHOIS database, and third-party fact-checkers (like Politico or MediaBiasFactCheck) that have begun to compile lists of known counterfeiting websites. We also relied on country-specific expert knowledge of the media landscape in the US to identify counterfeiting websites. Two examples of websites engaged in counterfeiting are politicoinfo.com and NBC.com.co.

Style

While the counterfeit criterion has to do with the esthetic design of a domain, style is concerned with the literary devices and language used throughout news reporting. Satisfying the style criterion involves propaganda techniques to spread information that may or may not be true to rally public support and shape popular opinion (Born & Edgington, 2017). Designed to systematically manipulate users for political purposes, junk news sources deploy propaganda techniques to persuade users at an emotional, rather than cognitive, level and employ techniques that include using emotionally driven language with emotive expressions and symbolism, ad hominem attacks, misleading headlines, exaggeration, excessive capitalization, unsafe generalizations, logical fallacies, moving images and lots of pictures or mobilizing memes, and innuendo (Bernays, 1928; Jowette & O'Donnell, 2012; Taylor, 2003). Silverman has noted that the techniques of propaganda have become widespread in the digital age for purposes beyond political ideology and agenda setting (Silverman, 2015). As advertising has become the fundamental business model on the Internet, traditional propaganda techniques are also being used to drive traffic and generate profit. This is sometimes referred to as the “clickbait” economy or “the outrage industry” (Berry & Sobieraj, 2016). Stylistically, problematic sources will employ propaganda and clickbait techniques to varying degrees. As a result, determining style can be highly complex and context dependent. To begin to evaluate stylistically problematic sources more systematically, we examined at least five stories on the front page of each news source in depth at the time of our data collection during the US presidential campaign in 2016 and the SOTU address in 2018. We checked the headlines of the stories and the content of the articles for literary and visual propaganda devices. If three of the five stories systematically exhibited elements of propaganda, we considered the source stylistically problematic. Examples of junk news domains that satisfied the style criterion include 100percentfedup.com, barenakedislam.com, theconservativetribune.com, and dangerandplay.com.

Bias

Over the last decade, an extensive network of hyper-partisan media websites and blogs have emerged (Eldridge, 2017; Faris et al., 2017). These sources are highly biased, ideologically skewed, and publish opinion pieces as news. Basing their stories on the same events, these sources manage to convey strikingly different impressions of what actually transpired. It is such systematic differences in the mapping from facts to news reports that we call bias. Faris et. al. found that far-right news networks, especially those on Facebook, were responsible for spreading a large amount of misinformation about politics during the 2016 election. Indeed, the American far right has a history of exploiting new media to advance its ideological agenda (Marwick & Lewis, 2017). But bias exists on both sides of the political spectrum. Like determining style, determining bias

can be highly complex and context dependent. In order to evaluate the bias criterion, first we checked third-party sources that systematically evaluate media bias. If the domain was not evaluated by a third party, we examined the ideological leaning of the sources used to support stories appearing on the domain (as in Groseclose & Milyo, 2005) during the time of our data collection. We also looked for bias by evaluating the labeling of politicians (are there differences between the left and the right?) and aimed to identify bias created through the omission of unfavorable facts, or through writing that is falsely presented as being objective. Examples of junk news websites that satisfied the bias criterion include *breitbart.com*, *dailycaller.com*, *infowars.com*, and *truthfeed.com* (on the right), and *occupydemocrats.com*, *addictinginfo.com* and *bipartisanreport.com* (on the left).

Credibility

Websites that satisfy the credibility criterion typically report on unsubstantiated claims and rely on conspiratorial and dubious sources. Faris et al. (2017) note that several problematic websites “comb[ine] decontextualized truths, [and] repeat falsehoods ... to create a fundamentally misleading view of the world.” Although conspiracy theories have always been a part of public life, Marwick and Lewis (2017) note that these dubious sources are used to support narratives of distrust vis-a-vis “the government or the official stories of the media,” juxtaposing the established political and media systems. Junk news sources that satisfy the credibility criterion frequently fail to vet their sources, do not consult multiple sources, and do not fact-check. When coding for credibility, we examined at least five front page stories and reviewed the sources that were cited. We also reviewed pages to see if they included known conspiracy theories on issues such as climate change, vaccination, and “Pizzagate.” In addition, we checked third-party fact-checkers for evidence of debunked stories and conspiracy theories. Examples of websites that satisfy the credibility criterion include: *infowars.com*, *endingthefed.com*, *thegatewaypundit.com*, and *newspunch.com*.

Typologies and New Modes of Political Communication

Content typologies are important to the study of political communication, and for many years the broad categories and subcategories of political news and information have remained widely accepted by researchers, though of course there is debate over how transportable such traditional categories are to new media political communication (Earl, Martin, McCarthy, & Soule, 2004; Edgerly, Thorson, Bighash, & Hannah, 2016; Freelon & Karpf, 2015; Karlsson & Sjøvaag, 2016). Recent debates about the impact of junk news on the media ecosystem have forced researchers to re-evaluate the new forms, production models, and normative values of political news and information shared on social media (Jr, Lim, & Ling, 2018). However, the current scholarly discussion lacks a grounded and comparative framework that captures the different types of information that circulate on social media. We advance this debate with a rigorously composed typology based on a focused, cross-case comparison of key political events in the US.

Typology building is one of the most foundational tasks in political research and is especially important when it comes to investigating new phenomena, unexpected problems, or sudden changes in social systems (Aronovitch, 2012; Howard & Hussain, 2013; Swedberg, 2018). Understanding the diversity of information sources involves carefully constructing categories that accurately describe the features of such new political phenomena and serve as transportable concepts across several cases. Propaganda,

misinformation, and negative campaigning are certainly not new features of public life. But social media applications are new platforms for spreading political news and information, the speed of information transmission is significantly greater, and the use of an individual's data is used in the targeting formula for misinformation; these are three distinct components of this contemporary mode of political communication.

Typologies have been useful for framing analysis of the study of news and the organization of event-based datasets (Althaus, Edy, & Phalen, 2001; Erickson & Howard, 2007). Before social media, scholars used such methods to expose the ways in which sensational news organizations used human interest frames, while serious news organizations used responsibility and conflict frames (Semetko & Valkenburg, 2000). In order to understand what users were actually sharing over social media, we develop a typology of political news and information shared on Twitter. While Twitter provides access to a wealth of data on public news sharing in the United States, the user base is not fully representative (Blank, 2017). Nevertheless, Twitter remains a central source of news and is especially popular among journalists, politicians, and opinion leaders, who further disseminate information in nonpublic social media spaces such as Facebook and WhatsApp (Jungherr, 2016).

Methods and Sampling

We performed a real-time data collection of political news and information shared on Twitter during the US presidential election and the SOTU address. Both of these events cause critical public interest issues that generate large amounts of political debate and media coverage. While the presidential election captures political communication over an extended period of time, the debate surrounding the SOTU address is traditionally more momentous and issue based. We tested and developed the methodological approach and typology described below in several elections in Europe between 2016 and 2018 (Neudert, Howard, & Kollanyi, 2019).

Conducting a real-time social media analysis, we developed a grounded typology and categorized the different URLs Twitter users were sharing as news and information about politics. Our grounded typology distinguishes between different web sources of political news and information. This approach allows for a highly contextual evaluation that considers the reporting, design, and practices of the content production of a web source as a whole. This is especially relevant since we often observed junk news web sources mixing factual reporting and news agency wire service reports to generate legitimacy. Real-time data collection allowed us to capture content before it was deleted by a user or the non-permanent URLs became invalid. If a post was in violation of Twitter's Terms of Service Agreement, real-time data collection also allowed us to capture these tweets before they were removed. While data collection on Twitter has known limitations—the streaming API only collects around 1 percent of all public tweets related to a specific search query, and the representativeness of its random samples has been called into question (Morstatter, Pfeffer, Liu, & Carley, 2013)—it is nonetheless a widely accepted approach to studying the formation of “ad-hoc publics,” especially around those that relate to political events (Burgess & Bruns, 2012; Larsson & Moe, 2012).

Our methodology was carried out in four stages. First, we identified relevant hashtags about the presidential race and the SOTU address. These hashtags included the names of the presidential candidates, official campaign slogans, and salient political issues. For the SOTU address, we followed prominent hashtags in relation to issues that were likely to be raised in the address. After manually identifying relevant hashtags, we

detected additional salient hashtags based on frequency and co-use with our initial dataset. Selecting political hashtags allowed us to home in on conversations about the election and the SOTU address. Although they do not provide insight into *all* of the conversations taking place on Twitter (i.e., those sharing political content on Twitter and not using hashtags), they still provide us with data about the issues we are focusing on in this study. Our team reviewed a list of the 400 most co-used hashtags and the total number of shares for relevance. This sampling strategy allowed us to focus on the most central hashtags while excluding minor hashtags in relation to short-lived conversations about particular people or issues. For the presidential election, this approach yielded more than 45 hashtags related to the event. For the SOTU address, our team identified only four event-specific hashtags that generated substantial traffic on Twitter. This is consistent with an overall slower public debate over social media during the SOTU address, as compared to that relating to the presidential election. A complete list of the hashtags can be found in the notes for [Table 1](#).

Second, based on the selected hashtags, we collected tweets from Twitter's streaming API that (1) contained the selected hashtags; (2) contained a URL to a web source, such as a news article, where the URL or the title of the web source included a selected hashtag; (3) were retweets that contained a message's original text, wherein a selected hashtag was used either in the retweet or in the original tweet; and (4) were quote tweets where the original text was not included but Twitter used a URL to refer to the original tweet. Tweets with URLs that pointed toward another tweet were removed from our sample, as these tweets are generally generated automatically when someone quotes a tweet.

In total, we collected 19.53 million tweets about the presidential election and 2.26 million tweets about the SOTU address. The tweets were collected between November 1, 2016, and November 9, 2016, for the presidential election and between January 24, 2018, and January 30, 2018, for the SOTU address. We then identified all of the tweets containing a URL and extracted them as a subset for analysis. Overall, we extracted 3.26 million URLs about the presidential

Table 1

Sample frames for political news and information shared on Twitter, 2016–2018

	US Presidential Election	State of the Union Address
Sample Date Range	1/11/2016–9/11/2016	24/1/2018 - 30/1/2018
Tweets, Total Number	19.53m	2.26m
Number of Users	3.69m	532,000
Tweets Per User	5.3	4.23
Number of URLs	3.26m	145,000

Source: Authors' calculations based on data collected between 2016 and 2018.

Note: Sample collection for the US presidential election involved the following hashtags: #AmericaFirst, #benghazi, #CrookedHillary, #DrainTheSwamp, #lockherup, #maga3x, #MAGA, #MakeAmericaGreatAgain, #NeverHillary, #PodestaEmails, #projectveritas, #riggedelection, #tcot, #Trump2016, #Trump, #TrumpPence16, #TrumpTrain, #VoterFraud, #votetrump, #wakeupamerica, #Clinton, #ClintonKaine16, #democrats, #dems, #dnc, #dumptrump, #factcheck, #hillary2016, #Hillary, #HillaryClinton, #hillarysupporter, #hrc, #ImWithHer, #LastTimeTrumpPaidTaxes, #NeverTrump, #OHHillYes, #p2, #strongertogether, #trumtape, #uniteblue, #Election2016, #Elections2016, #uselections, #uselection, #earlyvote, #iVoted, #Potus.

Note: Sample collection for the State of the Union address included the following hashtags and keywords: #SOTU, #stateoftheunion, #SOTU2018, #stateoftheunionaddress.

election and 145,000 about the SOTU address. If Twitter users shared more than one URL in their tweet, only the first URL was analyzed.

Third, we prepared the URLs for semiautomated analysis by creating a simple spreadsheet that listed the base URL, random examples of full URL users shared, and the frequency of the base URL shares, as well as additional metadata about the Twitter accounts. For our study, the base URLs were constructed by only keeping the domain and subdomain parts of the URLs. We also trimmed off the www part from the beginning of the URLs to merge the different versions of the links pointing to the same domain or subdomain. However, we refer to the domain as the part of the URL that only contains the domain name. For example, for the <https://www.media.journal.com/article1.html> URL, the base URL is “media.journal.com” and the domain is “journal.com.” URLs that were shortened using a link shortener (such as bit.ly) were unwrapped and added to the spreadsheet for analysis. We also unwrapped URLs to social media posts that were sharing URLs (i.e., a tweet about a public Facebook post linking to a news story).

Finally, our team developed a grounded typology in an iterative process and manually cataloged the base URLs, and this process is described in detail below. Our team labeled a total of 710 base URLs following the described method. Using a simple Python script, the coding decisions for the base URLs were applied to individual links that users shared during the political events. By applying this method, we were able to analyze a large volume of tweets and precisely categorize 710 web sources that users shared on Twitter.

Building a Grounded Typology of Political News and Information on Social Media

The process of labeling content involved multiple stages in line with best practices for both concept formation in typology building and content analysis (Collier, LaPorte, & Seawright, 2012; Earl et al., 2004). We labeled data in an iterative way and for each of the events went through multiple rounds of labeling by coders who worked independently. In the first stage of this process, we tested our assumptions and definitions derived from political communication research on real data, which we then refined over multiple rounds of coding. This allowed us to improve the typology relevancy and the representativeness of the kinds of political news and information shared by the people discussing US politics.

The first iteration of our grounded typology was tested on a pilot dataset based on a sample of URLs shared in Michigan during the 2016 presidential election. Michigan was selected for the initial study as the state is traditionally considered a swing state and polls predicted a close race between the two presidential candidates. We collected tweets based on our sampling strategy described above and extracted a subset of tweets from users who provided location information as being from the state of Michigan through the manual input of a city or state in the location field of their profiles. Initially, our team elaborated four broad categories of sources of political news and information: (1) professional news outlets, (2) established political actors, (3) organizations producing polarizing and conspiratorial content, and (4) other sources of political news and information.

Subsequently, we developed a system of subtypes of sources within those parent categories. Our coding process usually involved a full-team discussion of subcategory labels, definitions, and individual coding decisions. To determine whether a source satisfied

the criteria of the junk news category, coders browsed the domain for information about authors, owners, and funding sources and corroborated this information with third-party sources like Wikipedia. This helped the team to decide which sources satisfied the Professionalism and Counterfeit criterion. Coders also searched for information about the domain on the website of other third-party fact-checkers, such as Politifact, Snopes, and Media Bias Fact Check, to improve the certainty of criteria such as credibility and bias. Decisions regarding the style criterion were made based on visual and esthetic cues relating to the kinds of images, words, and videos used. Overall, this process allowed the team to develop a comprehensive typology of sources of political news and information, described below, involving over 1,000 hours of coding, six training sessions, and biweekly review meetings. The typology reflects 22 months of iterative coding procedures, and during this time the typology was exported to other country contexts: France, the UK, Germany, Mexico, Sweden, and Brazil (Glowacki et al., 2018; Hedman et al., 2018; Machado, Kira, Hirsch, & Marchal, 2018; Neudert et al., 2019).

To train our team of US experts to categorize sources of political news and information according to our testing grounded typology, we established a rigorous training system. For the analysis of the presidential election and the SOTU address, we worked with a team of six and five coders respectively. Coding and training regarding the shared dataset from the US achieved an intercoder reliability score of Krippendorff's $\alpha = 0.89$, signaling good concept formation and a high adeptness of our method. To measure intercoder reliability, our team categorized web sources from our full list of base URLs in sets of random 50 sources into the different subcategories of our grounded typology. Individual cases for which the experts did not reach a consensus were re-evaluated in face-to-face team meetings. One core team member from each of the two teams reviewed final source classifications and made consecutive coding decisions about highly ambiguous cases.

Ultimately, this system of analysis yielded valuable distinctions across major categories of political news and information and important nuances in subcategories. The major category of *Professional News and Information* were defined largely by the professional reputation of the organization that owns the source and has four subcategories. *Major News Brands* included large outlets that displayed the qualities of professional journalism, had known fact-checking operations and credible standards of production, and provided clear information about real authors, editors, publishers, and owners. Content from *Local News* came from local and regional news domains that displayed evidence of organization, resources, and professionalized output and distinguished between fact-checked news and commentary. The category *New Media and Start-ups* encompassed new media, digitally native, and digital first outlets. Yellow press publications on sex, crime, astrology, and celebrities were labeled *Tabloid*.

Professional Political Sources are produced by traditional political actors whose bulletins, working papers, websites, and reports provide evidence during debates among politicians and citizens. The subcategory *Political Parties or Candidates* comprised of domains that were produced by an official political party or candidate campaign. There were also links to *Government* websites and reports from public agencies and intergovernmental organizations. Finally, sources in the subcategory *Experts* took the form of white papers, policy papers, or academic writing by researchers based at universities, think tanks, or other professional research organizations.

The third category, *Divisive and Conspiracy Content*, included various forms of unreliable and deceptive sources. The most important subcategory here was labeled *Junk News and Information*. A source was labeled as junk news when it satisfied at

least three out our five criteria (professionalism; style; credibility; bias; and counterfeit). This subcategory also included URLs linking to *Wikileaks* since the page repeatedly promoted conspiracy theories about Hillary Clinton and the Democratic Party. The subcategory *Russia* was constituted exclusively of links to the government-funded *Russia Today*, *Sputnik*, and the Russian-hacker *Guccifer 2.0*. Russian sources were included in the category *Divisive and Conspiracy Content* due to their known slanted reporting and application of propaganda techniques (Helmus, 2018).

The fourth category included a host of *Other Political News and Information* sources. These sources included many kinds of political actors we would expect to generate political news and information during an election—other than formally chartered political parties and declared candidates. The subcategory *Citizen, Civic, and Civil Society* described sources produced by independent citizens, civic groups, or civil society organizations, watchdog organizations, interest groups, and lobby groups representing specific political interests or agendas. It also included blogs and websites dedicated to citizen journalism, personal activism, and other forms of civic expression that displayed originality and creation beyond curation or aggregation. *Humor and Entertainment* was used to label sources that produced political jokes, sketch comedy, and political art. *Video/Image Sharing and Content Subscriptions* included links to music and video streaming portals, as well as to image-sharing services. We also captured sources in the category *Fundraising and Petitions* that included links to citizen-generated petitions or fundraising pages. The category *Lifestyle and Special Interest* captured publications like women's and men's magazines and content focused on arts and sports. *Religion* referred to faith-based sources with distinctly religious themes. *Online Portals, Search Engines, and Aggregators* included sources that do not have editorial policies, like AOL and Yahoo. Web-hosting services were labeled *Cloud*. We sparingly used a subcategory called *Other Political Content* to capture myriad other kinds of political sources, such as information about polling stations and registering to vote.

Large numbers of captured links referred to *Other* content. Links labeled as *Social Media Platforms* referred to other social media platforms, such as Facebook posts, unless we were able to attribute them to other sources. The category *Shopping, Services, and Applications* was used for e-commerce and content monetization tools. *Link Shorteners* were coded as such, unless we were able to access the linked source. *Other Nonpolitical* sources were active sources that did not appear to be providing information about politics, even though they were shared in tweets using election-related hashtags. Spam was also included in this category. We also created a *Language* subcategory for links that led to content in a foreign language that could not be interpreted or evaluated by the coding team. Finally, the *Not Available* category included subcategories of sources that were inaccessible after repeated attempts to find them. However, if a source was available on the Wayback Machine during the dates we collected data, it was coded.

Findings and Analysis

Sources of Political News and Information

What are the salient qualities of political news and information shared over social media and how much of the content purporting to be political news and information during the two events was junk news? Based on our grounded typology, [Table 2](#) breaks down the categorical proportions and total counts of sources that citizens shared on Twitter. The category *Other Political News and Information* included many different kinds of content, thus accounting for

Table 2
Political news and information shared on Twitter, 2016–2018

Type of Source	Presidential Election		State of the Union	
	No. of Tweets	Percent	No. of Tweets	Percent
Professional News Outlets				
Major News Brands	398,258	14.3	23,807	20.5
Local News	47,329	1.7	318	0.3
New Media and Start-ups	116,051	4.2	9,855	8.5
Tabloid	56,022	2.0	983	0.8
Subtotal	617,660	22.1	34,963	30.0
Professional Political Sources				
Political Party	44,068	1.6	90	0.1
Government	2,123	0.1	13,054	11.2
Experts	4,583	0.2	144	0.1
Subtotal	50,774	1.8	13,288	11.4
Divisive and Conspiracy Sources				
Junk News and Information	486,446	17.4	21,071	18.1
WikiLeaks	148,366	5.3	3	0.0
Russia	39,457	1.4	145	0.1
Subtotal	674,269	24.1	21,219	18.2
Other Political News and Information				
Citizen, Civic, and Civil Society	151,589	5.4	1,707	1.5
Political Humor	13,680	0.5	736	0.6
Video/Image Sharing	505,560	18.1	21,715	18.7
Fundraising and Petitions	12,270	0.4	3,101	2.7
Lifestyle and Special Interest	6,275	0.2	127	0.1
Religion	3,110	0.1	0	0.0
Portals, Search Engines, and Aggregators	66,449	2.4	3,190	2.7
Cloud	12,304	0.4	130	0.1
Other Political	6,321	0.2	28	0.0
Subtotal	777,558	27.8	30,734	26.4
Other				
Social Media Platforms	128,113	4.6	3,345	2.9
Not Available	107,686	3.9	729	0.6
Shopping, Services, and Apps	290,596	10.4	3,211	2.8
Link Shorteners	85,688	3.1	8,124	7.0
Other Nonpolitical	58,853	2.1	759	0.7
Subtotal	670,936	24.0	16,168	13.9
Total	2,792,760	100.0	116,380	100.0

the largest number of shares during the election and the second largest number of shares during the SOTU address, proportionally. The category *Other* comprised a myriad of content that was not obviously related to political issues but nevertheless was shared using political hashtags. For both events, the subcategory *Shopping, Services, and Apps* generated the

highest numbers of shares, which can be attributed to the fact that advertisers freeride on popular hashtags to increase their visibility.

There are three main findings that should be noted across categories. First, the ratio of *Polarizing and Conspiracy Content* to *Professional News Outlets* during the US election was roughly 1:1. During the SOTU address, the proportion decreased to roughly 1:1.7 as the proportion of professional sources grew from 22.1 percent to 30.0 percent. The increase in tweets from “Professional News Outlets” may be related to the “Trump bump” in subscriptions and traffic that several prominent professional outlets reported in the first year of Trump’s time in office. Within the *Polarizing and Conspiracy Content* category, the overwhelming majority of tweets that users shared came from *Junk News and Information*. And while sources originating from *WikiLeaks* accounted for 5.3 percent of all the URLs shared during the election, it was not relevant during the SOTU address.

Second, only a small number of domains operated by known Russian agents were shared by users in the lead-up to both the presidential election (1.4 percent) and the SOTU address (.01 percent). This reflects the Russians’ strategy of immersive propaganda, which made use of a variety of media—such as memes, images, videos, fake accounts, pages, posts, and comments—to disseminate their messages on social media (US Department of Justice, 2018).

Third, the category *Professional Political Sources* was the smallest category of sources shared during the election and the SOTU address, accounting for 1.8 percent and 11.4 percent respectively of the total URLs shared. This category included many of the key traditional sources for political communication research and political theory, including content from parties, the government, and policy experts. Proportionally, the share of tweets in this category increased by 9.6 percent from the election to the SOTU address. Although many have claimed we are in a so-called post-truth era, this finding might suggest that social media users are actively seeking out better sources of information to inform their political opinions. Together with the increase in sources from *Professional News Outlets*, the share of professional voices in the political discourse increased by 17.5 percent across categories from the US presidential election in 2016 to the SOTU address in 2018. This increase could also suggest that traditional gatekeepers of political information have been at least partially reinstalled in the political discourse. [Tables 3](#) and [4](#) show the number of tweets per day to professional news sources and divisive and

Table 3

Professional news outlets and polarizing and conspiracy sources shared on Twitter overtime, US presidential election

Date	Professional News Outlets		Divisive and Conspiracy Sources	
	No. of Tweets	Percent	No. of Tweets	Percent
Nov 1, 2016	64,665	10.5	92,764	13.8
Nov 2, 2016	64,606	10.5	68,326	10.1
Nov 3, 2016	62,899	10.2	81,357	12.1
Nov 4, 2016	64,300	10.4	88,019	13.1
Nov 5, 2016	56,614	9.2	71,473	10.6
Nov 6, 2016	69,258	11.2	79,171	11.7
Nov 7, 2016	73,199	11.8	104,574	15.5
Nov 8, 2016	95,749	15.5	53,421	7.9
Nov 9, 2016	66,507	10.8	35,107	5.2
Total	617,797	100	674,212	100

Table 4

Professional news outlets and polarizing and divisive and conspiracy sources shared on Twitter during the State of the Union Address, 2018

Date	Professional News Outlets		Divisive and Conspiracy Sources	
	No. of Tweets	Percent	No. of Tweets	Percent
Jan 24, 2018	303	0.9	15	0.1
Jan 25, 2018	285	0.8	22	0.1
Jan 26, 2018	1,486	4.3	2,702	12.9
Jan 27, 2018	978	2.8	2,692	12.8
Jan 28, 2018	1,247	3.6	2,073	9.9
Jan 29, 2018	4,776	13.7	2,420	11.5
Jan 30, 2018	25,665	73.9	11,060	52.7
Total	34,740	100	20,984	100

conspiracy sources that were generated in the days leading up to the event and shortly after. No conclusive pattern emerges.

Amplified Political Communication across Two Political Events in the US

How did amplification support the volume of political communication in each category of political news and information on Twitter? Our second research question was concerned with the role of so-called amplifier accounts in the political communication surrounding the presidential election and the SOTU address. Amplifier accounts are defined by some scholars as accounts that deliberately seek to increase the number of voices speaking about or the attention being paid to particular messages (Mckelvey & Dubois, 2017; Woolley & Guilbeault, 2017). They can include automated, semiautomated, and highly active human-curated accounts on social media. We analyzed messaging by amplifiers on Twitter and compared the activity of these accounts across the presidential election and the SOTU address. Table 5 provides a breakdown of the total and proportional shares of sources from “Professional News Outlets” and “Divisive and Conspiracy Sources” that were attributed to amplifiers.

Table 5

Activity of amplifier accounts during the presidential election and the State of the Union Address, 2016–2018

	US Presidential Election	State of the Union Address
Number of Accounts	4,165	15
Number of Tweets Generated	3.49m	6,945
Number of Tweets with a URL	755,128	1,478
Professional News Content	95,385	17
Divisive and Conspiracy Content	180,337	14

We identified amplifier accounts as those that posted 50 times a day or more using one of the selected hashtags during the data collection period. The detection of fully automated, or so-called bot accounts remains contested in the field of computational social science, and techniques for optimizing identification are currently being debated. Our metric (volume) was simple, but more complex methods using machine learning yield comparative numbers of positives. Despite this, we have identified very few human users that tweeted more than 49.5 times on average per day using the sets of identified political hashtags. The threshold for identifying accounts that employ some kind of automation was chosen to reflect this finding. The small number of users that produce content above this threshold are also included in this category.

It is important to note that we cannot attribute with certainty amplifier accounts to a particular political actor based on the data that is publicly available on Twitter's APIs. Consequently, we cannot draw any definite conclusions to determine whether amplifier accounts were run by a campaign or candidate or by external actors interfering in political communication.

During the presidential election, amplifier accounts drove high volumes of the total Twitter traffic. We classified 4,165 amplifier accounts that generated a total of 3.5 m tweets containing 755,000 URLs over the course of our sample period. Next, we calculated the total number of links to particular sources of political news and information that were generated by these accounts. Amplifier accounts accounted for 95,385 shares of content from *Professional News Outlets*, which constituted 15.5 percent of all shares in that category. For *Polarizing and Conspiracy Content*, the number of total shares almost doubled to 180,337 shares generating 27.8 percent of all tweets in this category. This reflects other scholarly findings that have established a link between amplifiers and the circulation of misinformation (Mele et al., 2017; Vosoughi et al., 2018).

For the SOTU address, it appears that only a marginal level of amplifier accounts were tweeting, driving only 0.3 percent of the total traffic. Overall, we identified 15 such accounts. A manual analysis of the accounts concluded that the majority were spam bots that jumped on popular hashtags to promote goods and services. Because of the very low volumes of automated messages, the role of highly automated accounts in spreading links to sources of political news and information is unsubstantial. The activity of accounts using high levels of automation was drastically reduced compared to the presidential election to the point where traffic was marginal. While we identified only a small number of hashtags related to the SOTU address, this very low level of automation is the lowest across all country contexts we have studied in the past, including analysis that included similarly small numbers of hashtags. Twitter has taken several steps to combat fake accounts—including amplifiers—that exist on its platform (Taylor, Walsh, & Bradshaw, 2018). This could suggest that the steps the company has taken had an impact on political discourse from 2016 to 2018.

Junk News Sourcing during the US Election and the State of the Union Address

How did the quantities of particular types of political news and information change over time? To understand the patterns of junk sharing, we performed a frequency analysis of daily traffic across the categories of *Professional News Outlets* and *Polarizing and Conspiracy Content*. Our comparative analysis shows that during the presidential election, the level of both professional and junk sources remained relatively consistent over the course of our sampling period. On a slow day, users shared 9.2 percent of all links to *Professional News Outlets*, and this rose to 15.5 percent on

the day of the election. Given that users share professional information about the outcome of electoral races, it is not surprising that the largest amount of professionally produced information was shared on the day of the election. In contrast, the daily sharing of *Polarizing and Conspiracy Content* was much more volatile, starting at 5.2 percent and peaking at 15.5 percent (or 104,574 URLs) on the day before the election. This finding has implications for the health of our digital public sphere, as undecided voters trying to find professional information about candidates might see more junk news being shared on social media.

For the SOTU address, the total number of shares of content from *Professional News Outlets* and content from the *Polarizing and Conspiracy* category peaked on the day of the event, January 30, 2018. Across both categories, traffic on that day alone accounted for more than the total traffic of all the other days included in the analysis combined. The topical conversation around the SOTU address was evidently more short-lived than the one concerning the election. Over the course of our sampling period, *Polarizing and Conspiracy Content* was shared at a higher frequency, whereas daily traffic on *Professional News Outlets* remained lower than 5 percent of total daily shares except for on the day of the SOTU address and the day before it.

Conclusion

Although social media remains an important source of news and information, it is difficult to know how much political learning occurs on these platforms or how to model the impact a tweet can have on voter turnout or decisions. Nevertheless, social media is an important part of contemporary political communication in the United States and other advanced democracies. The nature of political news and information has certainly evolved over time, and a comprehensive, grounded, internally consistent typology of the types of sources shared by social media users is required before we can advance our understanding of what impact such content has on voters.

We developed the concept of junk news to refer to sources that deliberately publish or aggregate misleading, deceptive, or incorrect information packaged as real news about politics, economics, or culture. First, we found a 1:1 ratio of junk news to professionally produced news and information shared by Twitter users during the US election in 2016, a ratio that had improved by the time of the SOTU address in 2018. Second, we discovered that high levels of automation drove a consistently higher proportion of political communication during the presidential election but declined significantly during the SOTU address. Finally, we learned that some of the most important units of analysis for traditional political communication research and general political theory—political parties, the state, and policy experts—generate only a tiny fraction of the political communication that is used in public life over social media.

In our grounded typology, we evaluated and tested professionalism, style, credibility, bias, and counterfeit criteria as part of an internally consistent analytical frame with good intercoder reliability. We found that users shared substantial amounts of junk news, and although the quantities and qualities of such polarizing and conspiratorial content varied across different countries, this finding reflects the changing nature of political communication: information from traditional gatekeepers is shared much less on social media than content from other, nonprofessional organizations. As patterns in how political news and information are shared over social media—and the forms of content themselves—reshape our contemporary political communication environment, researchers need to be flexible in applying the traditional definitions we have

for what counts as political content when interpreting the new types of sources and content. Policymakers and citizens themselves also need to recognize the changing information landscape, as some of the traditional heuristics for determining information credibility are evolving alongside the information landscape.

Typology building is a complex analytical exercise. But some of the dramatic political events and trends from the last few years illustrate that scholars of political communication have to be ready to retool on short notice. Typology building is a necessary analytical exercise if researchers are to proceed with sophisticated modeling techniques, with qualitative study of particular election systems, and with critical interrogation of how political discourse is evolving. More importantly, rigorous typology building in political communication is necessary before considering any policy intervention and or imagining the design choices that might improve the quality of political news and information circulating in public life.

Figuring out what news is shared on social media, and whether fake news even holds as a concept, was an immense analytical project. One of the most important conclusions we drew from this experience was that “fake news” is not a useful concept, nor is it easily operationalized. Oftentimes, the low-quality, deceptive information, and other forms of harmful information, that people share online have multiple characteristics (i.e., mixing fact with fiction or news with commentary), and isolating each of these features is an impossible task. Rather than chasing a granular definition of the popular “fake news” term, we offer an alternative typology that captures the poor, unorthodox qualities of these sources of news and information.

Our grounded concept of junk news advances contemporary political communication by providing a theoretical and practical framework that is relevant for both policy intervention and future research. By conceptualizing junk news, we equip policymakers and social media companies with an analytical model for evaluating content on social media, measuring its spread, and assessing various efforts to reduce the proliferation of problematic content over online networks. Junk news offers a grounded, conceptual starting point for further research on understanding the effects, diffusion, and exposure of politically manipulative content in the social sciences.

Our conceptualization offers a nuanced perspective for evaluating the diffusion of content on social media. Often, junk news is not illegal and does not breach the terms of service of many social networks. It relies on an amalgam of a manipulative style, counterfeit activity, bias, a lack of professionalism, and enough credibility to deceive, and it freerides on social media algorithms to generate attention. Junk news sources thrive on social media because they are designed to exploit human attention logics, cognition, and emotion, which have been manifested in flawed social media algorithms that are designed to cater to human wants and needs. Junk news performs so well on social media because their systems have been designed to allow it to do so. Countering junk news will require interventions that address how social media algorithms work systematically by rewarding virality over veracity.

Disclosure Statement

The concept of junk news is based on a grounded typology tested in several scientific studies conducted by the COMPROP research group. It reflects our scientific opinion and evaluation, and does not necessarily reflect the views of our funders or the University of

Oxford. The basis of opinion is described in our methodology. No conflicts of interest were reported by the authors.

Funding

This research was funded by the European Research Council through the grant: “Computational Propaganda: Investigating the Impact of Algorithms and Bots on Political Discourse in Europe,” Proposal 648311, 2015–2020, Philip N. Howard, Principal Investigator. For supporting our research in the United States we are grateful to the Ford Foundation [0170-0774] and the Hewlett Foundation [2018-7384].

Supplementary Material

Supplemental data for this article can be accessed on the publisher’s website at <https://doi.org/10.1080/10584609.2019.1663322>.

ORCID

Philip N. Howard  <http://orcid.org/0000-0003-3380-821X>

References

- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*. doi:10.1257/jep.31.2.211
- Althaus, S. L., Edy, J. A., & Phalen, P. F. (2001). Using substitutes for full-text news stories in content analysis: which text is best? *American Journal of Political Science*, 45(3), 707–723. doi:10.2307/2669247
- Aronovitch, H. (2012). Interpreting Weber’s ideal-types. *Philosophy of the Social Sciences*, 42(3), 356–369. doi:10.1177/0048393111408779
- Bernays, E. L. (1928). *Propaganda*. Retrieved from http://archive.org/details/Propaganda1928ByEdwardL.Bernays_201608
- Berry, J. M., & Sobieraj, S. (2016). *The outrage industry: Political opinion media and the new incivility (Reprint edition)*. Oxford, UK: OUP USA.
- Blank, G. (2017). The digital divide among twitter users and its implications for social research. *Social Science Computer Review*, 35(6), 679–697. doi:10.1177/0894439316671698
- Born, K., & Edgington, N. (2017). *Analysis of philanthropic opportunities to mitigate the disinformation/propaganda problem*. Retrieved from <https://www.hewlett.org/wp-content/uploads/2017/11/Hewlett-Disinformation-Propaganda-Report.pdf>
- Boyd, D. (2017). *Google and Facebook can’t just make fake news disappear*. Retrieved from <https://www.wired.com/2017/03/google-and-facebook-cant-just-make-fake-news-disappear/>
- Bradshaw, S., & Howard, P. N. (2018a). *Challenging truth and trust: A global inventory of organized social media manipulation*. Working Paper 2018.1 No. Working Paper 2018.1; p. 26, Project on Computational Propaganda, Oxford Internet Institute, Oxford University, Oxford, England. Retrieved from <http://comprop.oii.ox.ac.uk/2017/07/17/troops-trolls-and-trouble-makers-a-global-inventory-of-organized-social-media-manipulation/>
- Bradshaw, S., & Howard, P. N. (2018b). Why does junk news spread so quickly across social media? Algorithms, advertising and exposure in public life. *Knight Foundation Working Paper*. Retrieved from https://kf-site-production.s3.amazonaws.com/media_elements/files/000/000/142/original/Topos_KF_White-Paper_Howard_V1_ado.pdf

- Burgess, J., & Bruns, A. (2012). Twitter archives and the challenges of “big social data” for media and communication research. *M/C Journal*, 15(5). Retrieved from <http://journal.media-culture.org.au/index.php/mcjournal/article/view/561>
- Collier, D., LaPorte, J., & Seawright, J. (2012). Putting typologies to work: Concept formation, measurement, and analytic rigor. *Political Research Quarterly*, 65(1), 217–232. doi:10.1177/1065912912437162
- De Keersmaecker, J., & Roets, A. (2017). ‘Fake news’: Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. *Intelligence*, 65, 107–110. doi:10.1016/j.intell.2017.10.005
- Earl, J., Martin, A., McCarthy, J. D., & Soule, S. A. (2004). The use of newspaper data in the study of collective action. *Annual Review of Sociology*, 30, 65–80. doi:10.1146/annurev.soc.30.012703.110603
- Edgerly, S. L., Thorson, K., Bighash, L., & Hannah, M. (2016). Posting about politics: Media as resources for political expression on Facebook. *Journal of Information Technology and Politics*, 13(2), 108–125. doi:10.1080/19331681.2016.1160267
- Eldridge, S. A. (2017). *Online journalism from the periphery: Interloper media and the journalistic field*. Abingdon-on-Thames, UK: Routledge.
- Erickson, K., & Howard, P. N. (2007). A case of mistaken identity? News accounts of hacker, consumer, and organizational responsibility for compromised digital records. *Journal of Computer-Mediated Communication*, 12(4), 1229–1247. doi:10.1111/jcmc.2007.12.issue-4
- European Commission. (2018). *A multi-dimensional approach to disinformation: Report of the independent high-level group on fake news and online disinformation*. Luxembourg: Author.
- Faris, R. M., 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 for Internet & Society Research Paper*. Retrieved from https://dash.harvard.edu/bitstream/handle/1/33759251/2017-08_electionReport_0.pdf
- Flanagin, A. J., & Metzger, M. J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information - Andrew J. Flanagin, Miriam J. Metzger, 2007. *New Media & Society*. Retrieved from <https://journals.sagepub.com/doi/abs/10.1177/1461444807075015>
- Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1), 298–320. doi:10.1093/poq/nfw006
- Fletcher, R., Cornia, A., Graves, L., & Nielsen, R. K. (2017). *Measuring the reach of “fake news” and online disinformation in Europe* (p. 10). Oxford, UK: Reuters Institute for Digital Journalism.
- Floridi, L. (2016, November 29). Fake news and a 400-year-old problem: We need to resolve the ‘post-truth’ crisis. *The Guardian*. Retrieved from <https://www.theguardian.com/technology/2016/nov/29/fake-news-echo-chamber-ethics-infosphere-internet-digital>
- Freelon, D., & Karpf, D. (2015). Of big birds and bayonets: Hybrid Twitter interactivity in the 2012 Presidential debates. *Information, Communication & Society*, 18(4), 390–406. doi:10.1080/1369118X.2014.952659
- Glowacki, M., Narayanan, V., Maynard, S., Hirsch, G., Kollanyi, B., Neudert, L. M., ... Barash, V. (2018). News and political information consumption in Mexico: Mapping the 2018 Mexican Presidential Election on Twitter and Facebook. *Data Memo 2018.2*. Oxford, UK: Project on Computational Propaganda. (No. 2018.2). Oxford, UK: Oxford University.
- Groseclose, T., & Milyo, J. (2005). A measure of media bias. *The Quarterly Journal of Economics*, 120(4), 1191–1237. doi:10.1162/003355305775097542
- Hambrick, D. Z., & Marquardt, M. (2018). *Cognitive ability and vulnerability to fake news*. Retrieved from <https://www.scientificamerican.com/article/cognitive-ability-and-vulnerability-to-fake-news/>
- Hedman, F., Sivnert, F., Kollanyi, B., Narayanan, V., Neudert, L. M., & Howard, P. N. (2018). News and political information consumption in Sweden: Mapping the 2018 Swedish general election on Twitter (No. 2018.3). Oxford, UK: Oxford University.

- Helmus, T. C. (2018). *Russian social media influence: Understanding Russian propaganda in Eastern Europe*. Santa Monica, CA: RAND Corporation.
- Howard, P. N., & Hussain, M. M. (2013). *Democracy's fourth wave? Digital media and the Arab Spring*. New York, NY: Oxford University Press.
- 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?* Oxford, UK: Internet Institute, Oxford University.
- Jowette, G. S., & O'Donnell, V. (2012). *Propaganda and persuasion* (5th ed.). London, UK: Sage Publications.
- Jungherr, A. (2016). Twitter use in election campaigns: A systematic literature review. *Journal of Information Technology & Politics*, 13(1), 72–91. doi:10.1080/19331681.2015.1132401
- Karlsson, M., & Sjøvaag, H. (2016). Content analysis and online news. *Digital Journalism*, 4(1), 177–192. doi:10.1080/21670811.2015.1096619
- Larsson, A. O., & Moe, H. (2012). Studying political microblogging: Twitter users in the 2010 Swedish election campaign. *New Media & Society*, 14(5), 729–747. doi:10.1177/1461444811422894
- 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. doi:10.1126/science.aao2998
- Lubbers, E. (2016). There is no such thing as the Denver Guardian, despite that Facebook post you saw. *The Denver Post*. Retrieved from <https://www.denverpost.com/2016/11/05/there-is-no-such-thing-as-the-denver-guardian/>
- Machado, C., Kira, B., Hirsch, G., & Marchal, N. (2018). *News and political information consumption in Brazil: Mapping the 2018 Brazilian presidential election on Twitter*. Oxford: Internet Institute, Oxford University. Retrieved from <https://comprop.oii.ox.ac.uk/research/brazil2018/>
- Marchal, N., & Neudert, L. M. (2019). *Polarisation and the use of technology in political campaigns and communication*. Retrieved from European Parliament website: [http://www.europarl.europa.eu/RegData/etudes/STUD/2019/634414/EPRS_STU\(2019\)634414_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2019/634414/EPRS_STU(2019)634414_EN.pdf)
- Marwick, A., & Lewis, R. (2017). Media manipulation and disinformation online. *Data & Society Research Institute*. Retrieved from https://datasociety.net/pubs/oh/DataAndSociety_Media_ManipulationAndDisinformationOnline.pdf
- Mckelvey, F., & Dubois, E. (2017). *Computational propaganda in Canada: The use of political bots* (No. 2017.6; p. 32). Oxford, United Kingdom: Oxford Internet Institute, University of Oxford.
- Mele, N., Lazer, D., Baum, M., Grinberg, N., Friedland, L., Joseph, K., ... Mattsson, C. (2017). *Combating fake news: An agenda for research and action*. Retrieved from <https://shorenstein-center.org/wp-content/uploads/2017/05/Combating-Fake-News-Agenda-for-Research-1.pdf>
- Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013). Is the sample good enough? comparing data from Twitter's streaming API with Twitter's firehose. *ArXiv:1306.5204 [Physics]*. Retrieved from <http://arxiv.org/abs/1306.5204>
- 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?* 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>
- Neudert, L.-M., Howard, P., & Kollanyi, B. (2019). Sourcing and automation of political news and information during three European elections. *Social Media + Society*, 5(3), 2056305119863147. doi:10.1177/2056305119863147
- Pariser, E. (2011). *The filter bubble: How the new personalized web is changing what we read and how we think*. London, UK: Penguin Books.
- Persly, N. (2017). The 2016 U.S. election: Can democracy survive the internet? *Journal of Democracy*, 28(2), 63–76. doi:10.1353/jod.2017.0025

- Semetko, H. A., & Valkenburg, P. M. V. (2000). Framing European politics: A content analysis of press and television news. *Journal of Communication*, 50(2), 93–109. doi:10.1111/j.1460-2466.2000.tb02843.x
- Silverman, C. (2015). *Lies, damn lies, and viral content. How news websites spread (and debunk) online rumors, unverified claims, and misinformation* (pp. 164). New York, NY: Tow Center for Digital Journalism.
- Swedberg, R. (2018). How to use Max Weber's ideal type in sociological analysis. *Journal of Classical Sociology*, 18(3), 181–196. doi:10.1177/1468795X17743643
- Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining "Fake News": A typology of scholarly definitions. *Digital Journalism*, 6(2). doi:10.1080/21670811.2017.1360143
- Taylor, E., Walsh, S., & Bradshaw, S. (2018). *Industry responses to the malicious use of social media*. Riga: NATO.
- Taylor, P. M. (2003). *Munitions of the mind: A History of propaganda* (3rd ed.). Manchester, UK: Manchester University Press.
- Tucker, J. A., Theocharis, Y., Roberts, M. E., & Barberá, P. (2017). From liberation2 to turmoil: Social media and democracy. *Journal of Democracy*, 28(4), 46–59. doi:10.1353/jod.2017.0064
- US Department of Justice. (2018). *U.S. v. Internet research agency* (p. 37). Retrieved from <https://www.justice.gov/file/1035477/download>
- Vicario, M. D., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., ... Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3), 554–559. doi:10.1073/pnas.1517441113
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. doi:10.1126/science.aap9559
- Wardle, C. (2017, February 16). *Fake news. It's complicated*. Retrieved from <https://firstdraftnews.com:443/fake-news-complicated/>
- Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward and interdisciplinary framework for research and policy making* (p. 108). Retrieved from <https://rm.coe.int/information-disorder-report-november-2017/1680764666>
- Woolley, S. C., & Howard, P. N. (2016). Automation, algorithms, and politics: Political communication, computational propaganda, and autonomous agents — Introduction. *International Journal of Communication*, 10, 9.
- Woolley, S. C., & Howard, P. N. (2017). Introduction: Computational propaganda worldwide. In P. Howard (Ed.), *Computational propaganda: Political parties, politicians, and political manipulation on social media* (pp. 3–21). Oxford, United Kingdom: Oxford Internet Institute, University of Oxford.
- Woolley, S. C., & Guilbeault, D. (2017). *Computational propaganda in the United States of America: Manufacturing consensus online*. Working Paper 2017.5 No. 2017.5; p. 37. Oxford, United Kingdom: Project on Computational Propaganda, Oxford Internet Institute, Oxford University.
- Wu, T. (2017). *The attention merchants: The epic struggle to get inside our heads*. New York, NY: Knopf.