Title
The Disinformation Landscape and the Lockdown of Social Platforms

Authors

Shawn Walker[a]*, Dan Mercea and Marco Bastos[b]

[a]*School of Social and Behavioral Sciences, Arizona State University, Phoenix, AZ, US*

[b]*Department of Sociology, City, University of London, London, UK*

Corresponding author
Shawn Walker, School of Social and Behavioral Sciences, Arizona State University, West Campus, Phoenix, Phoenix, AZ, UK 85069, US shawn.w@asu.edu

Shawn Walker is Assistant Professor at Arizona State University. He received his PhD in Information Science from the University of Washington Information School. He is a founding member of the Social Media (SoMe) Lab and a member of the DataLab at the University of Washington.

Dan Mercea is Reader in the Department of Sociology at City, University of London. He is the author of *Civic Participation in Contentious Politics: The Digital Foreshadowing of Protest*.

Marco Bastos is Senior Lecturer in Media and Communication in the Department of Sociology at City, University of London and an affiliate of Duke University's Network Analysis Center.

Abstract

This introduction to the special issue considers how independent research on mis/disinformation campaigns can be conducted in a corporate environment hostile to academic research. We provide an overview of the disinformation landscape in the wake of the Facebook–Cambridge Analytica data scandal and social platforms' decision to enforce access lockdowns and the throttling of Application Programming Interfaces (APIs) for data collection. We argue that the governance shift from user communities to social media algorithms, along with social platforms' intensive emphasis on generating revenue from user data, has eroded the mutual trust of

networked publics and opened the way for dis/misinformation campaigns. We discuss the importance of open, public APIs for academic research as well as the unique challenges of collecting social media data to study highly ephemeral mis/disinformation campaigns. The introduction concludes with an assessment of the growing data access gap that not only hinders research of public interest, but that may also preclude researchers from identifying meaningful research questions as activity on social platforms becomes increasingly more inscrutable and unobservable.

Introduction

This Special Issue addresses the question of how researchers can conduct independent, ethical research on dis/misinformation operations in a rapidly changing and hostile data environment. The escalating issue of data access we discuss is thrown into sharp relief by the strategic use of bots, trolls, fake news, strategies of false amplification, and a corporate environment favoring platform lockdowns and the restriction of access to Application Programming Interfaces (APIs). As social media platforms increase obstacles to independent scholarship by dramatically curbing access to APIs, researchers are faced with the stark choice of either limiting their use of trace data or developing new methods of data collection. Without a breakthrough, social media research may go the way of search engine research, in which only a small group of researchers who have direct relationships with search companies such as Google and Microsoft can access data and conduct research.

The reflections that follow highlight the current scholarly predicament of grappling with the time-sensitive nature of strategic and disruptive communication in highly ephemeral dis/misinformation campaigns. These campaigns unfold in increasingly polarized, hybrid media environments where news stories are written, disseminated, and interpreted within and across intricate digital networks. While researchers are developing more sophisticated multi-method research designs and rich multi-source data sets (Chadwick, Vaccari, & O'Loughlin, 2018; Möller, Trilling, Helberger, & van Es, 2018), this methodological progress is threatened by the active resistance of social media platforms to providing support for research on topics that require greater transparency and may negatively impact their bottom line (Dance, LaForgia, & Confessore, 2018).

The Disinformation Landscape

The set of articles presented in this issue take stock of the epochal changes triggered by the deployment of data-driven micro-targeting in political campaigns epitomized by the Cambridge Analytica data scandal and the ensuing data lockdown enforced by social media platforms. Digital trace data has been increasingly linked to disinformation, misinformation, and state propaganda across Western industrialized democracies and countries in the Global South, where state and non-state actors seek to strategically diffuse content that heightens partisanship and erodes the general trust in democratic institutions.

Influence operations weaponizing social media have been identified in elections worldwide, with prominent examples including the 2016 US elections and the 2017 general elections in

France (Bessi & Ferrara, 2016; Ferrara, 2017; Weedon, Nuland, & Stamos, 2017). This evolving disinformation landscape required the increasing adoption of specialized vocabulary associated with influence and disruptive operations to describe a set of media practices designed to exploit deep-seated tensions in liberal democracies (Bennett & Steven, 2018). The tactics documented are part of a concerted strategy to polarize voters and alienate them from the electoral process (Benkler, Farris, & Roberts, 2018). This scheme—infamously associated with Russian troll factories—took the form of misinformation, or information identified as inaccurate (Karlova & Fisher, 2013), and disinformation, or the intentional distribution of fabricated stories to advance political goals (Bennett & Steven, 2018).

Misinformation and disinformation pose a serious threat to objective decision-making by the voting public (Lewandowsky, Ecker, & Cook, 2017, p. 354). The effectiveness of dis/misinformation campaigns has in part (Benkler, Farris, et al., 2018) been attributed to the manner in which they have been able to take advantage of the biases (Comor, 2001; Innis, 1982) intrinsic to social media platforms, particularly the attention economy and the social media supply chain that relies on viral content (Jenkins, Ford, Green, & New, 2013) for revenue generation. In their response, social media platforms attempted to rekindle trust by appearing to reinforce individual privacy within a newly secured user community, a set of measures that also locked academic and non-profit researchers out from studying social platforms while preserving corporate and business access to social media users' data.

Infrastructural Transformation of the Networked Publics
This backdrop of influence operations and information warfare presents a considerable departure from years of euphoric rhetoric praising the democratization of public discourse brought by networking technology and social media platforms (Howard & Hussain, 2013). Early scholarship extolling the potential of social media for democratization and deliberation inadvertently reinforced a narrative championing communication and collaboration as expected affordances of social platforms (Loader & Mercea, 2011). By the end of the decade, however, the narrative surrounding social platforms increasingly turned to metaphors foregrounding polarization and division in a landscape marked by tribalism and information warfare (Benkler, Faris, & Roberts, 2018), enabled by a business model driven above all by the commodification of digital circulation and its capitalization on financial markets (Langley & Leyshon, 2017).

Scholarship on this hybrid media ecosystem (Chadwick, 2017) explored the technological affordances and ideological leanings that shape social media interaction, with a topical interest in the potential for civic engagement and democratic revitalization (Zuckerman, 2014). Bennett and Segerberg (2013) expanded on Olson's seminal work on the logic of collective action to explain the rise of digital networked politics where individuals would come together to address common problems. Similarly, Castells (2009) described a global media ecology of self-publication and scalable mobilization that advanced internet use and political participation (Castells, 2012).

Open platforms and unrestricted access offered the cornerstone of networked publics that reconfigured sociality and public life (Boyd, 2008). The relatively open infrastructure of networked publics was also explored in scholarship detailing how online social networks support gatewatching (Bruns, 2005) and practices in citizen journalism that are central to a diverse media ecosystem (Hermida, 2010), with citizens auditing the gatekeeping power of mainstream media and holding elite interests to account (Tufekci & Wilson, 2012). By most assessments, social network sites were welcoming challengers to the monopoly enjoyed by

the mass media (Castells, 2012), with only limited attention devoted to the opportunities offered to propagandists that could similarly coordinate and organize disinformation campaigns through decentralized and distributed networks (Benkler, Faris, et al., 2018).

These developments challenged the very idea of networked publics and Castells' (2012) depiction of the internet as universal commons. However, the transition from narratives emphasizing open communication to concerns about information warfare was neither immediate nor trivial. With mobile platforms slowly replacing desktop-based applications, open standards gave way to centralized communication systems epitomized by social media platforms, social technologies pivoted from a business model centered on software and services to the selling and reselling of user data. These changes endangered the openness of networked publics, with the debate underpinning networks in the late 90s being replaced by a focus on the affordances of mobile apps and social platforms, whose user base largely differs from living communities of users that would come together around common interests.

Also noticeable in the transition from networked publics to social platforms was the increased commercialization of previously public, open, and often collaborative spaces that were increasingly reduced to private property. This infrastructural transformation of the networked publics continues to drive anxieties about social media platforms in the aftermath of the Cambridge Analytica data scandal, including topical issues of digital privacy, data access, surveillance, microtargeting, and the growing influence of algorithms in society. Counterbalancing reactions to these developments, distributed networks services have started in the Fediverse such as Mastodon or Pleroma. However, on the most prevalent social media platforms, the existing networked publics are defined by the technological and market exigencies of the corporations who own them.

Algorithmization of Online Communities
Social platforms built much of their social infrastructure on the back of networked publics and the community organization that shaped internet services in the early 90s. The drive towards community formation remains an important component of social media platforms, notwithstanding the growing trend towards data access restrictions and end-to-end encryption of their services. Indeed, on 27 February 2017, at a time when investigations into disruptive communication in the previous year's US elections were still in their infancy, the CEO of Facebook, Mark Zuckerberg, wrote in a note published on his Facebook page titled "Building Global Community:"

> "History is the story of how we've learned to come together in ever greater numbers—from tribes to cities to nations. At each step, we built social infrastructure like *communities*, media, and governments to empower us to achieve things we couldn't on our own… Today we are close to taking our next step. Our greatest opportunities are now global… Progress now requires humanity coming together not just as cities or nations, but also as a *global community*." (Zuckerberg, 2017, emphasis added)

Zuckerberg's vision simultaneously highlighted and projected the end of open networked publics. With increasing government pressure before the end of that same year, Zuckerberg would yield and testify before the US Congress in an investigation on Russian interference in the 2016 elections; the company would see its share price plunge amid revelations that commercial third-parties and foreign arms-length agencies were able to harness Facebook to micro-target voters (Neate, 2018). In the attempt to reverse its fortunes, Facebook launched

multiple measures, including restricting third-party access to its Pages Application Programming Interface (API) that provided access to posts, comments, and metadata associated with communication on public Facebook pages (Schroepfer, 2018).

The immediate implications of this step for the—already remarkably limited—ability of independent academic researchers to form a systematic understanding of social interaction on private social media platforms (Driscoll & Walker, 2014; Skeggs & Yuill, 2015) were soon highlighted in an open letter by scholars (Bruns et al., 2018). However, at a deeper and more abstract level, one can unpick the discourse of Zuckerberg's manifesto to discern why his appropriation of the trope and exhortation of the value of community exposed, inadvertently, both the ills that plagued his platform and the steps to redress them, which were lamented as foreclosing democratic accountability (Bruns et al., 2018).

Social media platforms cannot be separated from the user communities that populate them. The business model of social platforms extracts data from community interactions (i.e. transactions among members) that can be monetized (Dijck, 2013; Fernback, 2007, p. 64) by a lucrative advertising business. The latter governs group interaction and individual experience alike through a set of intricate learning algorithms (Bucher, 2017). These algorithms rely on users as "affective processors" who interpret and help govern the communities through shares, likes, retweets, and pins (Gehl, 2011; Lomborg & Kapsch, 2019). However opaque to users, algorithms generate knowledge about users beyond their immediate interactions, thereby triggering further interactions and "imaginaries of interaction," i.e., user theories about what the algorithm is and ought to be (Bucher, 2017).

The algorithmization of communities championed by social platforms was a milestone that instantaneously rendered networked publics into a profitable source of users' interactions. Transferring community governance from users to algorithms removed a key basis for mutual trust, opening the way for large-scale disinformation campaigns that conspicuously plagued election cycles, ethnic relations, and civic mobilization from 2016 onwards (Apuzzo & Santariano, 2019). By Facebook's own account (Weedon et al., 2017), its advertising algorithms were harnessed to segment users into belief communities that could be micro-targeted with materials that amplified their intimate political preferences. This repurposing of intimate knowledge and networked interaction for revenue-making remained the corollary of commercial social media enterprises, including the individuals and academics involved in the notorious and now defunct political consultancy firm Cambridge Analytica (Rosenberg, 2018).

The operations executed by Cambridge Analytica may have violated Facebook's terms of service (Rosenberg, 2018), but they were broadly similar to Facebook's business model that extracts commercial value from users' data. As such, they did not go against the grain of the platform's business model. On the contrary, the consultancy cynically maximized Facebook's political utility by monetizing social-psychological user traits and using granular trace data to micro-target political advertisements on the platform (House of Commons Digital, 2019). In an attempt to reassert the integrity of the user community, Facebook rolled out in 2018 a tighter data management regime that, in some assessments, equally protects users' privacy and safeguards the company's advertising business by closely guarding user data for its own corporate use only (Tufekci, 2019).

Facebook's community governance will likely continue to be contested, not least because of what it represents: a corporate hegemony that is far removed from networked publics which

used to provide a counterweight to state and corporate power over individuals. Policymakers, in democratic countries, have demanded with some success more accountability from social platforms in respect to their efforts to arrest disruptive communication and preserve the intimacy of users, with the German legislation being portrayed as an exemplar (Bundeskartellamt, 2019; Volpicelli, 2019). Academics, on the other hand, have mostly been unsuccessful in their appeal to social platforms to open themselves up to a level of public scrutiny that permits investigations of its user communities by scholars. In response to Facebook's public API throttling, several prominent academic researchers stressed the following:

> "Platform providers—and the research advisors they collaborate with—cannot be allowed to position themselves as the gatekeepers for the research that investigates how their platforms are used. Instead, we need far more transparent data access models that clearly articulate to platform users who may be accessing their data, and for what purposes." (Bruns et al., 2018)

Why Public APIs Matter
Studies in this Special Issue add to a growing body of purposeful attempts to generate meaningful, valid, and reliable results with proprietary data to which access is limited, selective, and often opaque (Driscoll & Walker, 2014). This introduction and Bruns' article, specifically, spotlight the corporate response of social media platforms—the most stringent of which has been that of Facebook, which drastically restricted access to its public APIs—to justified alarm regarding the use of personal trace data in disruptive communication (House of Commons Digital, 2019).

Public and open APIs allow researchers to retrieve large-scale data and curate databases associated with sociologically meaningful events. Without them, web interfaces have to be scraped to access the data (Freelon, 2018), which is labor-intensive and drastically limits the amount of information that can be collected and processed. Locking researchers out of the APIs constrains them to human-intensive means of data collection that cannot produce large or representative samples of real-world events, such as social movements, elections, let alone state and non-state sponsored disinformation campaigns.

Illustratively, Twitter operates three well-documented, public APIs (Twitter, 2019) in addition to its premium and enterprise offerings. Twitter's relative accessibility leads it to being vastly over-represented in social media research (Blank, 2016). Public and open APIs such as that of Twitter are an exception in the social media ecosystem. By contrast, Facebook's Public Feed API (Facebook, 2019) is restricted to a limited set of media publishers. It is against this backdrop that researchers have suggested that restrictions on data access may lead to the consideration of alternative methods (Venturini & Rogers, 2019); and, in the "post API Age", the increased use of data collection methods that may run counter to platform terms of service such as web scraping (Freelon, 2018).

These suggestions offer a roadmap to resources researchers may leverage to implement their studies, but they underestimate the central role of APIs in providing scalable and reproducible access to data. This line of thought is epitomized in references to a "post API Age" (Freelon, 2018) asserting that APIs are in the process of being retired. It would be more accurate to refer to a "post-public-API age," as Application Programming Interfaces (APIs) remain a central component of mobile and cloud-based technologies that are central to the infrastructure of social media platforms. Indeed, it is difficult to see how cloud-based

business development, and web applications in general, could perform operations requiring personalization and scalability without resorting to APIs.

Perhaps more importantly, these APIs and alternative methods of data collection are not drop-in replacements. The volume, type (text, image, videos, interface, etc.), fidelity, timeliness, platform filtering, and amount of metadata vary considerably across these methods. High-volume data retrieved from APIs cannot directly replace low-volume web scraping data. Even if the volume of data collected using web scraping or APIs were identical, the metadata available via API requests is considerably different from metadata that is visible on the user-facing portions of a social media platform's website used for web scraping.

Social platform restrictions to public API access are only one aspect of the multi-faceted challenge involved in collecting digital trace data. Data access is the first in a number of steps researchers have to take as they collect, process, validate, interpret, share, and archive the data. These steps often require robust technical skills, as API endpoints for data collection were designed for programmers building application software that adds to the services offered by social platforms. As such, APIs were envisioned for purposes that differ from reproducible scientific research, a problem compounded by the significant differences between data retrieved from APIs and data visible to users on the websites and mobile apps of social platforms (Venturini & Rogers, 2019).

Locked, Instable, and Ephemeral
Asymmetries between information retrieved from APIs and that to which users have accessed on the web are due to social media sites being simultaneously a platform and an infrastructure (Plantin, Lagoze, Edwards, & Sandvig, 2016). Their underlying features and structures offer a rigid set of affordances, or entry points, constraining the ability to access, query, format, and collect data. These entry points take two forms: (1) interfaces for human-consumption (e.g. Facebook.com, Twitter.com, and mobile applications) and (2) software interfaces designed for consumption by computer programs called Application Programming Interfaces, with prominent examples including the Facebook Graph API, Twitter Streaming API, and Instagram API (Helmond, 2015). Social media sites have offered these website interfaces on the open web to extend their reach, decentralize data production, and centralize data collection and processing (Gerlitz & Helmond, 2013).

The underlying features of social platforms impinge on research designs and data collection, as one cannot ask questions of data that is not possible to collect. While data access is a perennial problem in social science research, a topic extensively examined in relation to survey response and privileged access to interviews (Babbie, 2015; Biernacki & Waldorf, 1981; Weller & Kinder-Kurlanda, 2016), the implications of observing dynamic content at a particular or arbitrary point in time and issues of preservation of social media data are the object of limited attention and rarely discussed in research publications. Social media posts and their accompanying metadata are fundamentally ephemeral, a term largely used as a shorthand for instability: data is constantly changing, being updated, or deleted.

The everchanging nature of social media data make it difficult for any given two researchers to collect the same exact dataset in real-time. It also makes it virtually impossible for disparate research teams to collect the same dataset retrospectively via the purchase of data from a reseller or by scraping social media websites (Burgess & Bruns, 2015). In addition to that, and perhaps more worryingly, researchers are often forbidden from sharing full datasets by the terms of service of many platforms. While some platforms, such as Twitter, allow for

the sharing of each post's unique identification number, this still requires researchers to programmatically "rehydrate" social media posts, if still accessible. While this solves the issue of informing other researchers which posts were included in the study, and also provides a method for comparing posts in different datasets, it creates at least three major issues across many social media platforms.

First, deleted posts and posts from deleted accounts cannot be retrieved from the API thereby generating orphaned data. Researchers studying misinformation are particularly interested in posts that have been deleted by users or platforms (Bastos & Mercea, 2019). Second, modified posts and modified post metadata are not flagged by APIs or web interfaces, so researchers cannot determine if a post or its metadata was changed, updated, or corrected since it was posted. Third, large datasets are difficult and time-consuming to rehydrate due to API request limits. The Twitter REST API is currently rate-limited to 150 requests per hour, returning a maximum of 100 tweets per request. While it is possible to get around these limitations by using multiple accounts simultaneously, doing so increases the technical complexity of the rehydration process.

The post IDs themselves fall under the definition advanced by Gray, Szalay, Thakar, and Stoughton (2002) of ephemeral data, since in most instances once collected these datasets cannot be reconstructed. When a post changes or disappears, it may end up being a research opportunity lost forever (Lynch, 2008) or present a false account of the phenomena under study. In the end, such obstacles to collecting and sharing datasets make it difficult and often impossible for researchers to validate or replicate studies using social media data (Felt, 2016). It additionally prevents researchers from repurposing previously collected data or expanding on research using the same dataset.

Overcoming the Data Access Gap
Perennial issues of research replication have been compounded by social platforms' API throttling, simultaneously reducing their public accountability and increasing their opaqueness (Bastos & Mercea, 2018). Shortly after the decision to drastically limit API access, Facebook sought to counter researchers' concerns by vowing (Schrage & Ginsberg, 2018) to help the academic community gain access to social media data of public interest, starting with elections. In partnership with leading academics, public bodies, and established funding organizations, Facebook sought to collaborate with a centralized data management scheme overseen by Social Science One, an initiative that would invite and filter applications for access to datasets (for details, see commentaries of Bruns and Puschmann in this issue).

The predicament in which academics now seeking data access find themselves is evocative of the longer-standing transformation of the mission of public universities (Gumport, 1997; Walton, 2011). The erstwhile conception of universities as self-governed entities committed to the formulation, testing, and dissemination of scientific knowledge is undermined by their dependence on private enterprise and corporate activity that control data access and, to a growing extent, use. As the foundation for not only productive innovation but also new regulation, scientific knowledge could act as a restraint on corporate control (Gauchat, 2012, p. 183). However, its ability to perform this role is restricted, inter alia, by the enforcement of intellectual property rights on various aspects of research including, as described above, inputs such as data and outputs such as publications. More immediately, the creation of Social Science One as a gatekeeping body governing the relationship between Facebook and academics exemplifies a governance model that may widen the gap between data-rich

industry researchers with connections to social platforms and independent researchers working outside corporations.

This divide has been characterized as the gap between "big data rich researchers," who have access to proprietary data and might be working in the interests of the company employing them, and the "big data poor", or the broad universe of academic researchers whose findings may be of public interest but may ultimately be critical of social media platforms (Boyd & Crawford, 2012). The data access gap not only hinders research that is peripheral to commercial interests, but it may also preclude researchers from identifying sociologically meaningful research questions because activity on social media is becoming increasingly more inscrutable and unobservable.

Alongside this apparent erosion of academic self-governance (Walton, 2011), researchers have to contend with the likelihood that social media users may not be readily open to the idea that their public communication become the object of scientific research (Fiesler & Proferes, 2018). Research entailing the automatic collection of large datasets covering long periods of time must meet expectations of non-exploitation and minimum risk/maximum benefit to informed and consenting users. To reduce the chances of causing harm to users, researchers need to find ways to obscure the presence or remove traces of any identifiable user from the data such that users are not unwillingly identified as members of an unintended community (e.g. of political dissidents).

Halavais' article (this issue) addresses some of these issues by proposing a framework of ethical distributed research involving the pragmatic partnership between users and researchers, thereby bypassing platform owners. The article suggests ways this could be accomplished and argues that the balance of power between the social media industry on one side and the users and researchers on the other has become dangerously skewed. Another innovative method to study mis/disinformation is presented by Acker and Donavan (this issue), who show how data craft through metadata manipulation and keyword squatting play a prominent role in attracting audiences. The proposed method explores data archives of disinformation offered by social media platforms and shows that such sanctioned archives prevent researchers from examining organic contexts of manipulation.

The remainder of the articles in this Special Issue tackle issues surrounding data access and ethical dilemmas in studying mis/disinformation or offer a roadmap to studying the disinformation landscape with limited or fragmented data. Bruns (this issue) outlines the societal implications of the "APIcalypse" and reviews potential options available for researchers studying hate speech, trolling, and disinformation campaigns. The piece likewise offers a critical evaluation of Facebook's partnership with the Social Science One initiative. A response to Bruns' article is presented by Puschmann (this issue), who argues that current models of data access for social media research are also fraught with problems and pose significant risks to user privacy. The articles of Bruns and Puschmann offer opposing views of the partnerships between academics and industry seeking to address structural issues of data access.

Lastly, the Issue includes three case studies of mis/disinformation that successfully overcame the data access gap. Xia et al. (this issue) present an in-depth analysis of how the team behind an IRA Twitter account crafted the persona "Jenna Abrams" across multiple platforms over time and describe the techniques employed to perform personal authenticity and cultural competence. Proferes and Summers (this issue) rely on a novel web archiving and scraping

approach for data collection to analyze the Wikileaks' release of John Podesta's e-mails. The article details how the serialized release of batches of e-mails together with the strategic use of sequential hashtags allowed Wikileaks to game Twitter trending topics locally, nationally, and eventually worldwide. Giglietto et al. (this issue) explore partisan engagement with political news through the analysis of Twitter and Facebook interactions in the period leading up to the 2018 Italian general election. They show that polarization is not hard-wired even into highly partisan networked publics, which may engage strategically with news sources covering their favorites.

To conclude, a drive to formulate standards for public-interest research through extensive involvement of multiple stakeholders—from corporations to governments, political representatives, academic institutions, non-governmental and citizen organizations from the Global South (Milan & Treré, 2019) and North—would represent a more durable and equitable basis on which to build an alternative data governance regime to the current one. It would have to strike a balance between the accountability of all parties, their interests and rights while providing effective mechanisms to exercise a check on the power asymmetries that led to the sudden closure of the Facebook Pages API.

In that way, the data regime would have to reconcile the universalizing commercial impetus that has propelled the expansion of social platforms with the plurality of cultural, political and social communities that populate them, through a more democratic power settlement (see Laclau, 2001). It is hard to imagine, however, how the current direction of regulation that places the onus of privacy protection enforcement in the hands of social platforms (e.g. the European General Data Protection Directive, see Puschmann in this issue) may lead to such an outcome. Instead, skeptics argue that it will further consolidate their hegemony due to the high cost of privacy enforcement (Doctorow, 2019).

References

Apuzzo, M., & Santariano, A. (2019, 12 May 2019). Russia is Targetting Europe's Elections. So Are Far-Right Copycats. *New York Times*. Retrieved from https://www.nytimes.com/2019/05/12/world/europe/russian-propaganda-influence-campaign-european-elections-far-right.html

Babbie, E. R. (2015). *The practice of social research*: Nelson Education.

Bastos, M. T., & Mercea, D. (2018). The public accountability of social platforms: lessons from a study on bots and trolls in the Brexit campaign. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. doi:10.1098/rsta.2018.0003

Bastos, M. T., & Mercea, D. (2019). The Brexit Botnet and User-Generated Hyperpartisan News. *Social Science Computer Review, 37*(1), 38-54. doi:10.1177/0894439317734157

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*: Oxford University Press.

Benkler, Y., Farris, R., & Roberts, H. (2018). *Network Propaganda*. New York, US: Oxford University Press.

Bennett, W. L., & Segerberg, A. (2013). *The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics*. Cambridge: Cambridge University Press.

Bennett, W. L., & Steven, L. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication, 33*(2), 122-139. doi:10.1177/0267323118760317

Bessi, A., & Ferrara, E. (2016). Social bots distort the 2016 US Presidential election online discussion. *First Monday, 21*(11).

Biernacki, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling. *Sociological methods & research, 10*(2), 141-163.

Blank, G. (2016). The Digital Divide Among Twitter Users and Its Implications for Social Research. *Social Science Computer Review, 35*(6), 679-697. doi:10.1177/0894439316671698

Boyd, D. (2008). *Taken out of context: American teen sociality in networked publics.* University of California, Berkeley, Berkeley.

Boyd, D., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society, 15*(5), 662-679. doi:10.1080/1369118x.2012.678878

Bruns, A. (2005). *Gatewatching: Collaborative online news production* (Vol. 26): Peter Lang Pub Inc.

Bruns, A., Bechmann, A., Burgess, J., Chadwick, A., Clark, L. S., Dutton, W. H., . . . Hofmann, J. (2018). Facebook shuts the gate after the horse has bolted, and hurts real research in the process. *Internet Policy Review.*

Bucher, T. (2017). The algorithmic imaginary: exploring the ordinary affects of Facebook algorithms. *Information, Communication & Society, 20*(1), 30-44. doi:10.1080/1369118X.2016.1154086

Bundeskartellamt. (2019). Bundeskartellamt prohibits Facebook from combining user data from different sources [Press release]. Retrieved from https://www.bundeskartellamt.de/SharedDocs/Meldung/EN/Pressemitteilungen/2019/07_02_2019_Facebook.html

Burgess, J., & Bruns, A. (2015). Easy Data, Hard Data: The Politics and Pragmatics of Twitter Research after the Computational Turn. In G. Elmer, J. Langlois, & J. Redden (Eds.), *Compromised Data From Social Media to Big Data* (pp. 93-111). London: Bloomsbury Academic.

Castells, M. (2009). *Communication power.* Oxford: Oxford University Press.

Castells, M. (2012). *Networks of Outrage and Hope: Social Movements in the Internet Age.* Cambridge: Polity Press.

Chadwick, A. (2017). *The hybrid media system: Politics and power.* Oxford: Oxford University Press.

Chadwick, A., Vaccari, C., & O'Loughlin, B. (2018). Do tabloids poison the well of social media? Explaining democratically dysfunctional news sharing. *New Media & Society*, 1461444818769689. doi:10.1177/1461444818769689

Comor, E. (2001). Harold Innis and 'The Bias of Communication'. *Information, Communication & Society, 4*(2), 274-294. doi:10.1080/713768518

Dance, G. J. X., LaForgia, M., & Confessore, N. (2018). As Facebook Raised a Privacy Wall, It Carved an Opening for Tech Giants.

Dijck, v., J. A. (2013). *The Culture of Connectivity: A Critical History of Social Media.* Oxford: Oxford University Press.

Doctorow, C. (2019, 6 June 2019). Regulating Big Tech makes them stronger, so they need competition instead. *The Economist.*

Driscoll, K., & Walker, S. (2014). Working Within a Black Box: Transparency in the Collection and Production of Big Twitter Data. *International Journal of Communication, 8*, 1745-1764.

Facebook. (2019). Public Feed API. Retrieved from https://developers.facebook.com/docs/public_feed/

Felt, M. (2016). Social media and the social sciences: How researchers employ Big Data analytics. *Big Data & Society, 3*(1), 2053951716645828.

Fernback, J. (2007). Beyond the diluted community concept: a symbolic interactionist perspective on online social relations. *New Media & Society, 9*(1), 49-69. doi:10.1177/1461444807072417

Ferrara, E. (2017). Disinformation and social bot operations in the run up to the 2017 French presidential election. *2017*. doi:10.5210/fm.v22i8.8005

Fiesler, C., & Proferes, N. (2018). "Participant" Perceptions of Twitter Research Ethics. *Social Media + Society, 4*(1), 2056305118763366. doi:10.1177/2056305118763366

Freelon, D. (2018). Computational Research in the Post-API Age. *Political Communication*, 1-4. doi:10.1080/10584609.2018.1477506

Gauchat, G. (2012). Politicization of Science in the Public Sphere: A Study of Public Trust in the United States, 1974 to 2010. *American Sociological Review, 77*(2), 167-187. doi:10.1177/0003122412438225

Gehl, R. W. (2011). The archive and the processor: The internal logic of Web 2.0. *New Media & Society, 13*(8), 1228-1244. doi:10.1177/1461444811401735

Gerlitz, C., & Helmond, A. (2013). The like economy: Social buttons and the data-intensive web. *New Media & Society, 15*(8), 1348-1365.

Gray, J., Szalay, A. S., Thakar, A. R., & Stoughton, C. (2002). *Online scientific data curation, publication, and archiving.* Paper presented at the Astronomical Telescopes and Instrumentation.

Gumport, P. J. (1997). Public Universities as Academic Workplaces. *Daedalus, 126*(4), 113-136.

Helmond, A. (2015). The platformization of the web: Making web data platform ready. *Social Media+ Society, 1*(2), 2056305115603080.

Hermida, A. (2010). Twittering The News: The emergence of ambient journalism. *Journalism Practice, 4*(3), 297-308. doi:10.1080/17512781003640703

House of Commons Digital, C., Media and Sport Committee. (2019). *Disinformation and Fake News: Final Report*. Retrieved from https://publications.parliament.uk/pa/cm201719/cmselect/cmcumeds/1791/1791.pdf

Howard, P. N., & Hussain, M. M. (2013). *Democracy's Fourth Wave?: Digital Media and the Arab Spring*. Oxford: Oxford University Press.

Innis, H. A. (1982). *The Bias of Communication*. Toronto: University of Toronto Press.

Jenkins, H., Ford, S., Green, J., & New, C. (2013). *Spreadable Media: Creating Value and Meaning in a Networked and London: New York University Press*.

Karlova, N. A., & Fisher, K. E. (2013). A social diffusion model of misinformation and disinformation for understanding human information behaviour. *Information Research, 18*(1).

Laclau, E. (2001). Democracy and the Question of Power *Constellations (Oxford, England), 8*(1), 3-14.

Langley, P., & Leyshon, A. (2017). Platform capitalism: The intermediation and capitalization of digital economic circulation. *Finance and Society, 3*(1), 11-31. doi:10.2218/finsoc.v3i1.1936

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era. *Journal of Applied Research in Memory and Cognition, 6*(4), 353-369. doi:https://doi.org/10.1016/j.jarmac.2017.07.008

Loader, B. D., & Mercea, D. (2011). Networking Democracy? *Information, Communication & Society, 14*(6), 757-769. doi:10.1080/1369118x.2011.592648

Lomborg, S., & Kapsch, P. H. (2019). Decoding algorithms. *Media, Culture & Society, 0*(0), 0163443719855301. doi:10.1177/0163443719855301

Lynch, C. (2008). Big data: How do your data grow? *Nature, 455*(7209), 28-29.

Milan, S., & Treré, E. (2019). Big Data from the South(s): Beyond Data Universalism. *Television & New Media, 20*(4), 319-335. doi:10.1177/1527476419837739

Möller, J., Trilling, D., Helberger, N., & van Es, B. (2018). Do not blame it on the algorithm: an empirical assessment of multiple recommender systems and their impact on content diversity. *Information, Communication & Society*, 1-19. doi:10.1080/1369118X.2018.1444076

Neate, R. (2018). Over $119bn wiped off of Facebook's market cap after growth shock. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2018/jul/26/facebook-market-cap-falls-109bn-dollars-after-growth-shock

Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2016). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*. doi:10.1177/1461444816661553

Rosenberg, M. (2018). Professor Apologizes for Helping Cambridge Analytica Harvest Facebook Data. *New York Times*. Retrieved from https://www.nytimes.com/2018/04/22/business/media/cambridge-analytica-aleksandr-kogan.html

Schrage, E., & Ginsberg, D. (2018). Facebook Launches New Initiative to Help Scholars Assess Social Media's Impact on Elections [Press release]. Retrieved from https://newsroom.fb.com/news/2018/04/new-elections-initiative/

Schroepfer, M. (2018). An update on our plans to restrict data access on Facebook | Facebook Newsroom.   Retrieved from https://newsroom.fb.com/news/2018/04/restricting-data-access/

Skeggs, B., & Yuill, S. (2015). The methodology of a multi-model project examining how values and value are made through Facebook relations. *Information Communication and Society, 19*(10), 1356-1372.

Tufekci, Z. (2019, 7 March 2019). Zuckerberg's So-Called Shift Toward Privacy. *New York Times*. Retrieved from https://www.nytimes.com/2019/03/07/opinion/zuckerberg-privacy-facebook.html

Tufekci, Z., & Wilson, C. (2012). Social media and the decision to participate in political protest: Observations from Tahrir Square. *Journal of Communication, 62*(2), 363-379.

Twitter. (2019). Documents.   Retrieved from https://developer.twitter.com/en/docs

Venturini, T., & Rogers, R. (2019). "API-Based Research" or How can Digital Sociology and Journalism Studies Learn from the Facebook and Cambridge Analytica Data Breach. *Digital Journalism*, 1-9. doi:10.1080/21670811.2019.1591927

Volpicelli, G. (2019). Passing laws to force Facebook to fix fake news is asking for trouble. *Wired*. Retrieved from https://www.wired.co.uk/article/dcms-report-fake-news-report

Walton, J. K. (2011). The Idea of the University. In M. Bailey & D. Freedman (Eds.), *The Assault on Universities* (pp. 15-26). London: Pluto Press.

Weedon, J., Nuland, W., & Stamos, A. (2017). *Information Operations and Facebook*. Retrieved from

Weller, K., & Kinder-Kurlanda, K. E. (2016). *A manifesto for data sharing in social media research.* Paper presented at the Proceedings of the 8th ACM Conference on Web Science.

Zuckerberg, M. (Producer). (2017, 03.06.2019). Building Global Community. Retrieved from https://www.facebook.com/notes/mark-zuckerberg/building-global-community/10154544292806634/

Zuckerman, E. (2014). New Media, New Civics? *Policy & Internet, 6*(2), 151-168. doi:10.1002/1944-2866.poi360