



UvA-DARE (Digital Academic Repository)

Political advertising exposed: tracking Facebook ads in the 2021 Dutch elections

Beraldo, D.; Milan, S.; de Vos, J.; Agosti, C.; Nadalic-Sotic, B.; Vliegenthart, R.; Kruikemeier, S.; Otto, L.P.; Vermeer, S.A.M.; Chu, X.; Votta, F.

Publication date

2021

Document Version

Final published version

License

CC BY

[Link to publication](#)

Citation for published version (APA):

Beraldo, D. (Author), Milan, S. (Author), de Vos, J. (Author), Agosti, C. (Author), Nadalic-Sotic, B. (Author), Vliegenthart, R. (Author), Kruikemeier, S. (Author), Otto, L. P. (Author), Vermeer, S. A. M. (Author), Chu, X. (Author), & Votta, F. (Author). (2021). Political advertising exposed: tracking Facebook ads in the 2021 Dutch elections. Web publication or website, Internet Policy Review. <https://policyreview.info/articles/news/political-advertising-exposed-tracking-facebook-ads-2021-dutch-elections/1543>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

Political advertising exposed: tracking Facebook ads in the 2021 Dutch elections

Davide Beraldo, Media Studies; ALEX project, University of Amsterdam, Netherlands, D.Beraldo@uva.nl

Stefania Milan, Media Studies; ALEX Project, University of Amsterdam, Amsterdam, Netherlands, s.milan@uva.nl

Jeroen de Vos, ALEX project, Media Studies, University of Amsterdam, Netherlands

Claudio Agosti, Tracking.exposed; ALEX project, University of Amsterdam, Netherlands

Bruno Nadalic Sotic, ALEX project, University of Amsterdam, Netherlands

Rens Vliegenthart, Communication science, University of Amsterdam, Netherlands

Sanne Kruikemeier, Communication science, University of Amsterdam, Netherlands

Lukas P Otto, Communication science, University of Amsterdam, Netherlands

Susan A. M. Vermeer, Communication science, University of Amsterdam, Netherlands

Xiaotong Chu, Communication science, University of Amsterdam, Netherlands

Fabio Votta, School of Communication Research, University of Amsterdam, Netherlands

PUBLISHED ON: 11 Mar 2021

Micro-targeting is the protagonist of political campaigning in COVID-19 times. In the absence of live campaign events such as rallies, political parties invest more than ever on intercepting their constituencies in their digital whereabouts. Not surprisingly, social media platforms are a central gear in the process. The Netherlands is no exception. Dutch citizens are called to elect the members of the House of Representatives (*Tweede Kamer*) on 15-17 March 2021. With the country in lockdown and the internet penetration at 96 percent, the transition of political campaigning to the digital invests social media platforms with an onerous stake in the game. According to the Reuters Institute Digital News Report 2020, 77 percent of Dutch adults consume news online, and 39 percent regularly use social media as news source, with Facebook leading the classification. Furthermore, unlike the United States, the Netherlands is a parliamentary democracy, in which multiple parties (as many as 37 in this round) take part in the elections. Many parties thus reach out to voters by means of targeted messages on Facebook. This bears the questions: how can we monitor political ads circulating on social media in the run-up to an election? How can we understand the impact of political micro-targeting on the shaping of political preferences?

The project *Politieke Advertenties Analyse van Digitale Campagnes* (Analysis of Political Ads in Digital Campaigns, henceforth PAADC) is tasked with monitoring and analysing Facebook sponsored content in relation to the Dutch general elections. Combining scraping Facebook posts with relevant survey data in the context of political elections, the project takes advantage of the synergy between digital methods and public opinion research to gain insights on the usage and effects of political micro-targeting, and the dynamics and popular perception of parties' advertising strategies. This essay introduces the challenge of studying political micro-targeting amidst platform corporate policies and illustrates the PAADC methodology and its implications for the study of political micro-targeting. It reflects on two open questions for the study of political micro-targeting, namely the assessment of social media's transparency initiatives and the methodological and political potential of collaborative, independent research centering on platform users.

The challenge

Ever since Facebook was accused of steering voting preferences in the 2016 US Presidential campaign (Madrigal, 2017), the platform has been under the crossfire of policymakers, researchers, and organised civil society alike. As a result, Facebook Inc. has committed to increasing transparency in the functioning of its personalisation algorithms. In 2018 it launched Social Science One, an in-house research organisation hosted by Harvard University and aimed at providing selected research parties with (part of) the platform's goldmine of user and traffic data. In the same year, Facebook Inc. made available its political-focused ads archive, today integrated in the broader Ad Library service, which "provides advertising transparency by offering a comprehensive searchable collection of all ads currently running from across Facebook apps and services". In practice, the service allows to manually and programmatically inspect Facebook's vast collection of sponsored content, including (selected) details about sponsorships and placement logic (for an analysis of the quality of ad libraries see Leerseen et al., 2018). At the same time, however, Facebook Inc. has dramatically restricted the options to access its data through other established channels such as the Graph Application Programming Interface, or API (Puschmann & Ausserhofer, 2017). This controversial move was intended to limit abuse, but it ended up also restricting access to Facebook data for scholars committed to research in the public interest (Bruns, 2018). Meanwhile, the company has also actively pursued researchers investigating its political-ad-targeting practices, such as the New York University's *Ad Observatory*, on grounds that "scraping tools, no matter how well-intentioned, are not a permissible means of collecting information from us", as the *Wall Street Journal* reported.

To the company's supporters, initiatives of this kind testify to the good faith intentions of the social media giant to make amends for the unintentional misbehaviors of the Facebook / Cambridge Analytica scandal. To the detractors and the skeptical ones, these moves are seen rather as part of a carefully crafted 'open washing' strategy which calls for further vigilance. To be sure, relying on carefully edited company data sets bears some critical questions for research purposes. How complete and reliable are data made available through corporate-controlled channels? To what extent are they tailored to answer the pressing questions that emerge from users, researchers and other concerned stakeholders? And ultimately: who 'owns' social media-generated data, and is thus entitled to oversee collection, sharing, and repurposing?

The approach

Against this backdrop, PAADC intervenes offering a novel methodology of noninvasive user audit (Sandvig et al., 2014) to generate social media data sets on political ads by engaging volunteer users. PAADC is a Dutch interdisciplinary research collaboration involving researchers of the University of Amsterdam, respectively from the Algorithms Exposed (ALEX) team at the Department of Media Studies and the Amsterdam School of Communication Research (ASCOR), the audience research organisation I&O, and the daily newspaper *de Volkskrant*. The core of PAADC's innovative methodology is a browser extension, named PAADC-*fbtrex*, which repurposed the *fbtrex* browser extension developed by the ALEX team with the open source analysts of Tracking Exposed (Milan & Agosti, 2019). This tailor-made plugin allows for the collection of political ads classified as public posts as they appear on a user's Facebook timeline.

In January 2021, the PAADC-*fbtrex* extension was downloaded and installed in the computers of a controlled group of 588 voting citizens selected by I&O (see methodological note), who agreed to collect data during the election campaign and in the aftermath of the voting (January 15 to March 28). As detailed in the privacy policy of the project, only the so-called 'public posts' are collected, in line with the General Data Protection Regulation. Next to volunteering individual (public) Facebook data, the user group is asked to fill in surveys on political attitudes, voting preferences and media use at regular intervals, for a total of four surveys. In a nutshell, PAADC overcomes the limitations of aggregated, platform-generated data (e.g., provided by Facebook itself) and of self-reported data by linking user experience (i.e., what a user is served in her timeline) to survey data about perceptions and political preferences.

Our approach points to at least three novel directions of analysis which we explore next, including: the descriptive characterisation of political campaigning and micro-targeting; the exploratory investigation of the 'political ad-sphere'; and the explanatory analysis of the impact of online political advertising.

Descriptive characterisation of political campaigning and micro-targeting

The data collected through the PAADC-*fbtrex* extension can be aggregated to produce real-time overviews of the micro-targeting strategies of parties and candidates. Content analysis can inform reports exploring themes and focus areas of digital campaigns. Furthermore, matching the collected ads with user survey data allows us to dig into the logic of micro-targeting itself. It is worth noting that, whereas similar findings could be produced by interrogating the aforementioned Facebook Ad Library, the latter only returns information about individual ads and pre-defined, abstract targeting categories. Our methodology, instead, provides data on actual impressions as they appeared on the timelines of real users, cross-referencing the

data with detailed user-level data. A real-time overview of the data collection can be found here. Figure 1 shows the daily impressions per party over a fortnight in the heat of the electoral campaign.

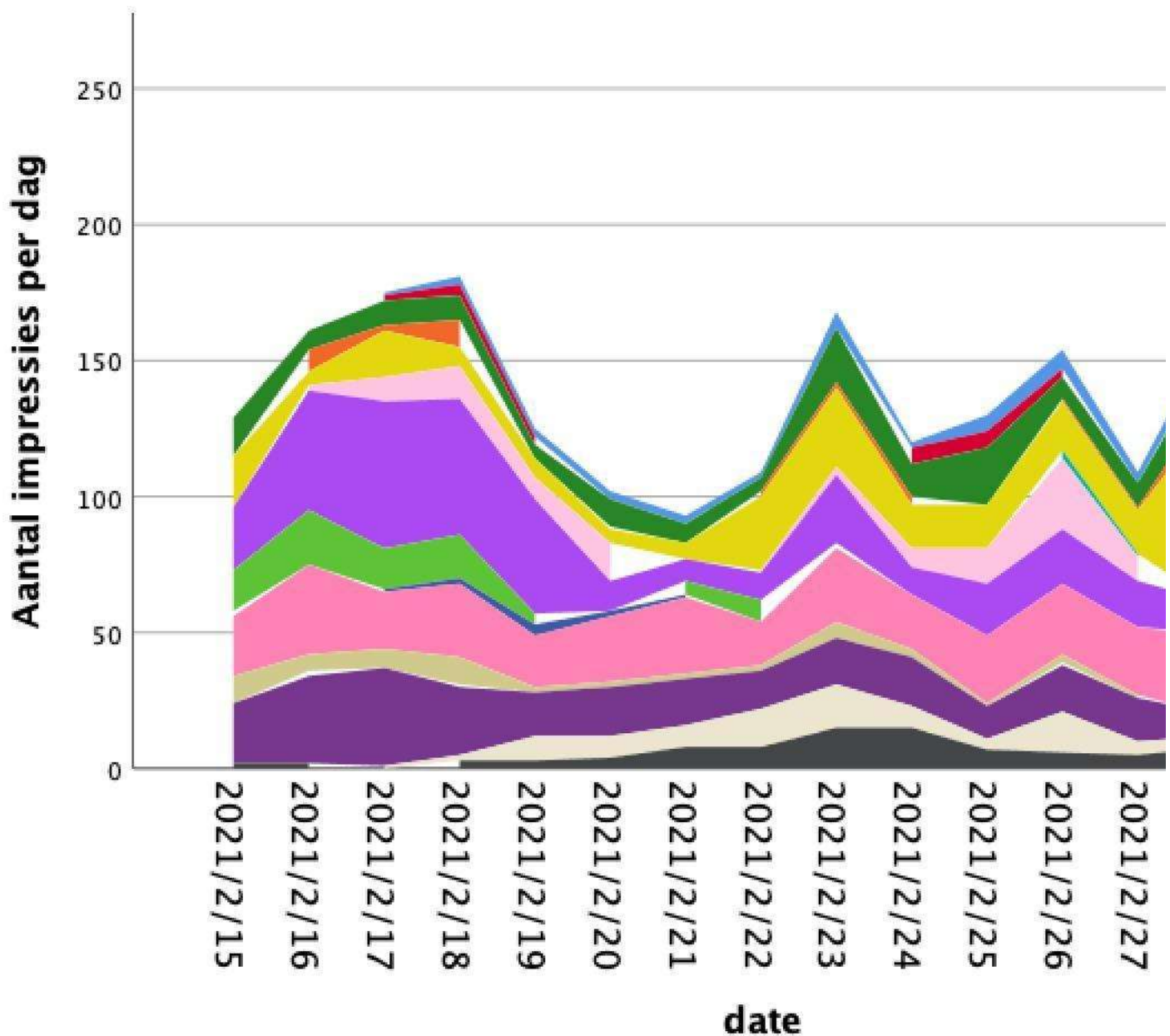


Figure 1: Number of daily impressions over the period 15 February-1 March 2021 per political party (PAADC project)

Exploratory investigation of the 'political ad-sphere'

Once we know to which (anonymised) users a certain ad has been served, it is possible to reconstruct the network of co-occurrence among different political advertisers targeting different users. In other words, our methodology shows which combination of ads users are exposed to on Facebook. Once linked to survey data, this allows us to gain insights on patterns of co-targeting, answering questions like: which parties tend to target the same user base? Which parties tend to be 'omnivorous' (i.e., aimed at all voters indistinctly) versus 'specialised' (i.e., aimed at a certain segment of the population) when it comes to targeting? Which advertisers tend to target users already aligned with their party, and which use advertising to target competing parties and thus trying to move votes?

Explanatory analysis of the impact of online political advertising

The possibility to poll users' political opinion in a longitudinal fashion—repeating a range of questions multiple times throughout the election campaign—together with data on which ads have concretely appeared on their timelines, paves the way for assessing the impact of political micro-targeting on voting behaviour and political opinions at large.

Our approach comes with three limitations to account for and, possibly, overcome. One issue relates to the size and representativity of the sample. For the Dutch elections 2021, the partnership with I&O allowed us to work with a fairly large-sized, diverse and controlled sample of the voting population. Whereas the sample of users who agreed to install the PAADC-fbrefx extension does not fully satisfy the standards for statistical representativity, its size and variability provide us with the possibility to derive informed inferences on the relation between collected ads and survey variables. Future iterations, however, could aim at involving statistically representative samples to allow for solid generalisations. A second limitation has to do with the fact that our browser extension currently works only on the desktop version of Facebook and not on the mobile app version. However, mobile access is the most popular

on certain segments of the population: Dutch Facebook users prefer the mobile in the 60.7 percent of the cases. Harvesting the same information from the mobile app would improve the representativity of our data. One solution to this problem is the use of ‘mobile experience sampling’, a complementary data collection strategy implemented on a subset of 120 respondents, who are uploading screenshots of ads as they appear on their Facebook timeline and answering short surveys. This will allow for cross-validation between desktop-based and mobile-based data. Finally, the process of metadata extraction at the level of the parser, as well as the function for distinguishing sponsored from non-sponsored content, have presented a number of challenges that required meticulous, continuous intervention, testing, and maintenance at the code level. Our experience exposes how certain metadata seem to be effectively shielded by the platform. We do not believe this problem to be a purely technical or methodological issue, but to a great extent as a political one—which we explore next.

Open questions in the study of political micro-targeting

The PAADC design allows us to hone in two large open questions when it comes to the analysis of digitally personalised political campaigning, namely: (1) the assessment of social media’s transparency initiatives, and (2) the methodological *and political* potential of collaborative, independent research centering on platform users.

(1) Next to exploring online political advertising, a corollary to our research relates to the meta-theme of algorithmic auditing and platform politics. Specifically, it contributes to the development of methods for the assessment of platform’s transparency assessment, and to the study of the politics of technical obfuscation.

Methods for assessing platforms’ transparency assessment

Following the infamous data abuse scandals, Facebook Inc. has promoted an alleged ‘new era’ of transparency and collaboration with researchers, favouring initiatives nominally devoted to guaranteeing user privacy while allowing for investigating the impact of algorithms on the democratic process. While these moves open up promising lines of research (see Venturini & Rogers, 2019), a question lingers behind: how can we assess... these transparency assessments? The PAADC data set, consisting of actual posts being served to real users, can ground the base for auditing... the Facebook transparency initiatives themselves. In other words: does the Facebook Ads Library return, as claimed, any political ad being served by the platform, and does it provide truthful information about the target audience? Given the current regulations and contractual power structures, the answer to this question might come from two sources: unconditional faith in Facebook Inc. or a large data set of actual ads served to real users. We have the latter. And, we argue, society needs more of these data sets.

The politics of technical obfuscation

However, similar objections of data set (in)completeness and data reliability can be advanced with regards to PAADC data as well. But this open question touches upon an important issue worth exploring further: that of the politics of obfuscation set into place by platforms in order to restrain malicious actors and independent researchers alike from collecting data. Developing scraping tools today is literally a cat-and-mouse situation, given that, as our experience shows, Facebook is putting active efforts to mislead parsers by obfuscating metadata and unexpectedly (and repeatedly) restructuring HTML codes (see Mancosu & Vegetti, 2020). We hope PAADC will serve as a case study for contributing to the methodological and political level of analysis of scraping proprietary platforms.

(2) The restructuring of Facebook data access policy following data breach scandals has paradoxically corresponded to a ‘platformisation’ (see Poell et al., 2019) of its research affordances, further centralising its control over data. Whereas it is desirable that loopholes such as those that allowed for the Facebook / Cambridge Analytica abuses are amended for, these moves have dramatically restricted the possibilities for (resource-poor) researchers to conduct analyses in the public interest. However, there exists a third way between unrestricted access and complete centralisation: distributed, volunteered, privacy-preserving data donation infrastructures. This aspect touches upon two points associated with what we called the contentious politics of data (Beraldo & Milan, 2019) applied to the research design: i) the political claim of breaking Terms of Services in the public interest, and ii) the question of public participation and user awareness through engagement.

Breaking Terms of Service

Scraping the content of a proprietary platform happens in a legal ‘grey area’, as it often formally violates a platform’s Terms of Service (ToS). This might occasionally result in legal contention as a study of the music streaming platform Spotify has exposed (Eriksson et al., 2018). Platform resistance to scraping, which *Facebook.tracking.exposed* was met with in the past (as denounced by the UK digital rights organisation Open Rights Group), has legit ethical foundations when it comes to preserving users’ privacy from being violated by third parties against their will. However, this justification does not hold true when users have explicitly provided consent for *their* data to be contributed to a certain research project. Ultimately, this boils down to a simple question: are data produced by the users who originate them, or by the platform that hosts them? No matter what the fine prints variably consciously ‘agreed’ upon by users say, users should have a legitimate right to transfer their data to parties they trust and whose aims they deem valuable, as experimented in the framework of data sharing/donation projects (e.g., Ohme et al., 2020). To be sure, the user base mobilised by the PAADC project has no responsibility in the (potential) violation of the ToS on the part of the researchers. However, future initiatives like ours should articulate this tension further, prefiguring organised cases of digital civil disobedience or strategic litigation. A promising direction is to legally establish the user’s right to keep a personal copy of the data in its quality of unique combination of content served by the platform to a specific individual, which could then potentially be shared with selected others.

Awareness through engagement

This challenge invites a reflection upon the potential political and transformative element associated with initiatives similar to the PAADC project, and any future iterations of it. Doing research based on data consciously contributed by users can potentially mean to do research *not* much on, but rather *with* them (Kazansky et al., 2019). The collective analysis of algorithmic recommendations around COVID-19 is a case in point: organised by *Tracking Exposed*, it results in the first crowdsourced analysis of YouTube’s recommended system during the pandemic. Further emphasising the collaborative, voluntary, and conscious nature of users’ recruitment to these initiatives can conciliate the research goals of these projects with the societal need of awareness raising and data literacy. Our hope is that consciously contributing their experience to our research goals might have stimulated our participants’ own personal reflection on the hurdles of political micro-targeting.

Looking forward

As we approach the voting window, Dutch parties get increasingly active in online advertising to steer votes. Our data will help voters and policymakers understand the role and reach of political micro-targeting in the country, and the extent to which it influences the election outcome. It can also provide informed insights to political parties and candidates concerning how Facebook users consume their sponsored content. In the future, our methodology can be repurposed to explore micro-targeting during elections in other countries as well.

We believe that the number of directions of analysis presented in this piece can be boiled down to one fundamental *leitmotiv*: the urge of bringing to light practices otherwise invisible to the public eye. We argue that algorithmic personalisation cannot solely and adequately be investigated by means of aggregated, decontextualised data. On the contrary, one needs to engage with user-level, actual impressions that only a natural experiment yields. One should also not audit algorithms via the data that platforms themselves (strategically) disclose: any platform-originated transparency assessment initiative needs to be assessed in itself. And finally: one cannot do research with user data without directly intervening in the *politics of data*: researching with social media users, bypassing the intermediation of platforms, is a political statement in itself and, potentially, a transformative process for society as a whole.

Methodological note

The users involved in PAADC have been recruited by I&O research using sampling procedures typical of survey research, and are compensated for their participation as per standards in the field. Since the sample required the active use of Facebook from a web browser, it was not possible to obtain a representative sample. The initial sample of 781 respondents has an

overrepresentation of male respondents (65.5 percent) and older age categories (40.5 percent older than 65). Among this initial sample, 588 participants have agreed to install the PAADC-fbtrex browser extension. This limitation forces us to be cautious about generalisations to the whole population of Dutch Facebook users.

Acknowledgements

This project has received funding from the *Stimuleringsfonds voor de Journalistiek*; the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 825974-ALEX and No 639379-DATACTIVE), the University of Amsterdam's Research Priority Area Amsterdam Centre for European Studies (ACES) and the Amsterdam School of Communication Research.

References

- Beraldo, D., & Milan, S. (2019). From data politics to the contentious politics of data. *Big Data & Society*, 6(2). <https://doi.org/10.1177/2053951719885967>
- Bruns, A. (2018). Facebook Shuts the Gate after the Horse Has Bolted, and Hurts Real Research in the Process. *Internet Policy Review*. <https://policyreview.info/articles/news/facebook-shuts-gate-after-horse-has-bolted-and-hurts-real-research-process/786>
- Eriksson, M., Fleisher, R., Johansson, A., Snickars, P., & Vonderau, P. (2018). *Spotify Teardown. Inside the Black Box of Streaming Music*. MIT Press.
- Kazansky, B., Torres, G., van der Velden, L., Wissenbach, K. R., & Milan, S. (2019). Data for the social good: Toward a data-activist research agenda. In A. Daly & M. Mann (Eds.), *Good Data* (pp. 244–259). Institute of Network Cultures. <https://data-activism.net/wordpress/wp-content/uploads/2019/02/data-activist-research.pdf>
- Leerseen, P., Ausloos, J., Zarouali, B., Helberger, N., & De Vreese, C. H. (2018). Platform Ad Archives: Promises and Pitfalls. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1421>
- Madrigal, A. C. (2017, October 12). What Facebook Did to American Democracy. *The Atlantic*. <https://www.cs.yale.edu/homes/jf/MadrigalFeb2018-2.pdf>
- Mancosu, M., & Vegetti, F. (2020). What You Can Scrape and What Is Right to Scrape: A Proposal for a Tool to Collect Public Facebook Data. *Social Media + Society*, 6(3). <https://doi.org/10.1177/2056305120940703>
- Milan, S., & Agosti, C. (2019). Personalisation algorithms and elections: Breaking free of the filter bubble. *Internet Policy Review*. <https://policyreview.info/articles/news/personalisationalgorithms-and-elections-breaking-free-filter-bubble/1385>
- Ohme, J., Araujo, T., de Vreese, C. H., & Piotrowski, J. T. (2020). Mobile data donations: Assessing self-report accuracy and sample biases with the iOS Screen Time function. *Mobile Media & Communication*, 2050157920959106. <https://doi.org/10.1177/2050157920959106>
- Poell, T., Nieborg, D., & Van Dijck, J. (2019). Platformisation. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1425>
- Puschmann, C., & Ausserhofer, J. (2017). Social Data APIs. Origins, Types, Issues. In M. T. Schäfer & K. van Es (Eds.), *The Datafied Society. Studying Culture through Data* (pp. 147–154). Amsterdam University Press.
- Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014, May 22). *Auditing Algorithms: Research Methods for Detecting Discrimination on Internet Platforms*. Data and Discrimination: Converting Critical Concerns into Productive Inquiry, Seattle, Washington. <http://social.cs.uiuc.edu/papers/pdfs/ICA2014-Sandvig.pdf>
- 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*, 7(4), 532–540. <https://doi.org/10.1080/21670811.2019.1591927>