CORVINUS UNIVERSITY of BUDAPEST

Doctoral School of Sociology and Communication Science

Thesis summary

Bence Kollányi

The automatization of social media communication

Exploring the development of bot codes on GitHub and the use of open-source bots on Twitter

Supervisor:
Bence Ságvári, PhD

Budapest, 2022

Doctoral School of Sociology and Communication Science

Thesis summary

The automatization of social media communication

Exploring the development of bot codes on GitHub and the use of open-source bots on Twitter

Bence Kollányi

Table of Contents

1. Research Summary and Thesis Formulation

As social media platforms become more open to automation, allowing businesses and government institutions to use bots to mediate online transactions and services, bots are becoming an increasingly important component in the power relations of the online communication sphere. The production and deployment of social media bots signal the emergence of new political economies that redistribute agency around new technological actors. This has implications for marketing, political action and even private lives. Yet, we have very little systematic knowledge about how bots are produced and the role of sharing code online and using a collaborative platform.

Automated social media accounts are often portrayed in the literature as actors that endanger social media platforms by spamming users, distributing malicious code or using fake profiles to create an artificial grassroots movement that support certain political goals. Many of the early publications documented efforts to detect automated Twitter accounts to prevent commercial spam or the distribution of links pointing to malicious websites (Chu et al., 2010; Lee et al., 2011; Song et al., 2011). The literature has also addressed a new class of more sophisticated social bots that can be described as "software agents mimicking humans" and more difficult to detect (Ferrara et al., 2016). These social bots could still have various intentions, or more specifically, they could have been programmed or deployed with different motivations. Stieglitz et al. (2017) recognized these nuances and differentiated between bots designed for malicious, neutral, and benign purposes. The broader literature on automated social media accounts includes papers on spambots (Chen & Subramanian, 2018, Cresci et al., 2017), bots promoting academic papers (Haustein et al., 2014), newsbots and chatbots used by the media (Diakopoulos, 2019; Jones & Jones, 2019), bots used in crisis communication (Brachten et al, 2018; Hofeditz et al., 2019) and in politics (Bastos & Mercea, 2019; Caldarelli et al., 2020; Ferrara, 2017; Howard, Woolley & Calo, 2018; Wooley & Howard, 2016a, 2016b), as well as anti-harassment bots (Geiger, 2016).

At the beginning of my PhD research project I wrote a paper about how open source Twitter bots were designed and how the code was published on GitHub (Kollanyi, 2016). GitHub is the largest online repository for shared computer code and for open-source developers it is the de facto solution for collaborating and sharing their work (Gousios et al., 2014). Therefore, GitHub is a good place to explore open-source codes for Twitter bots and the social arena around bot development.

There is a growing body of literature based on GitHub data that examines how developers use the collaborative code repository. This growing interest is explained in part by the availability of data, the availability of big data technologies for collecting and distributing vast amounts of data, and in this particular case, the emergence of methods for overcoming the limitations on data collection imposed by GitHub as an organization.

The goal of my PhD research project is to understand the practice of writing and deploying bots on Twitter. To understand current practices of bot development and use, I propose to bring together (1) an examination of bot codes available on GitHub, the largest online code repository, and (2) data on how human users communicate with the automatized social media accounts on Twitter, a platform where these bots are deployed. Thus, my research method combines data about automated accounts deployed on a social media platform with the source code behind those same bots – this provides a unique lens on bots and provides important insights about how they work.

Combining API-based, data driven research with a classic social science approach helps to understand current practices of writing for and deploying bots on online social media platforms. As part of my PhD research, I examined bot repositories on GitHub and conducted a survey with the same bot developers to gain further insights into the nitty-gritty of writing bot code, including the most important challenges during the development phase and major barriers to deploying and operating bots on Twitter over an extensive period of time. The survey results also shed light on the motivations behind creating automated social media accounts. I also studied how programmers deploy their bots on Twitter and how other, mostly human, users react to the bot activity. The final part of my thesis contributes to a typology of open-source social media bots by systematically examining bot codes shared on GitHub and the activity of some of these bots on Twitter in tandem.

2. Research Methodology

I proposed a novel way to study Twitter bots by examining both the open-source bot codes on GitHub and the bots deployed on Twitter. Traditionally, research projects have examined bots as black boxes and attempted to "reverse engineer" the algorithms behind automated accounts. Instead, I tracked bots to their source code on GitHub and connect the algorithm (code) running behind the bot to data about bot activity on a social media platform. By connecting these two data sources, I gained unique access to the inner "mechanics" of certain bots on Twitter, and I can test my findings about how algorithms built into the bots work on a large "real world" dataset.

The data collection for my dissertation is a four-step, often iterative process that includes initial data collection from GitHub via the website's APIs, a survey of bot developers, Twitter data collection, and more specific data collection from GitHub. These steps are designed to build on each other and therefore can only be done in a specific order.

I began my data collection by exploring how to access data from GitHub, the largest online code repository. The website has an API that allows one to search for specific keywords in the name and description of the repositories.

The result of this data collection was a large database of metadata downloaded from the website about bot repositories and about all developers who had at least one bot repository available on the platform.

After the GitHub-based data collection, I contacted the developers to ask more questions about the motivation behind bot development, the skills required, interaction with other developers, and the biggest challenges in developing a bot and deploying it on Twitter. The GitHub dataset collected during the first phase of data collection allowed me to extract the contact information of all developers who set their email address public on the platform. I used a paid online survey tool called Survey Gizmo, which allowed me to contact all developers by sending emails in mass.

Although some of the bot repositories include either a Twitter handle or a link to Twitter in the description or readme.md file uploaded to the repository, this is not always the case. Therefore, I decided to include a question in the survey asking developers to provide a list of the Twitter handles of the bot they use on Twitter. This way, I could link data about the developers on GitHub (including their bot repositories) to the bots deployed on Twitter. I also had survey data available for the bot developers.

The following list provides an overview of the four major data sets collected during the data collection phase of my research project:

GitHub dataset: Metadata about 19K open-source bot repository;
Survey responses: Survey results from 860 bot developers;
Twitter dataset: Data and metadata about 321 open-source bot repositories with paired source code on GitHub (500K tweets and metadata, including user engagement);
Manual coding database: Data about 321 Twitter bots – manually labeled.

3. Research Questions

The following section provides a list of the five research questions that I addressed in my PhD research project. These research questions are centered around two main themes. The first three research questions address the following more general questions: who are the developers behind the open-source Twitter bots and how do they develop these bots. The second set of questions focuses on the Twitter bots themselves and attempts to answer questions about how they work and how they produce tweet after tweet.

3.1. Research questions on bot developers

These questions are aimed at the bot developers both professional developers and non-professionals who are able to write code.

RQ 1.1 What are the practices for code developing and sharing code for open-source Twitter bots? What are the most import reasons behind using GitHub as a tool[1] for developing and sharing code?

This question focuses on how developers use GitHub, such as how often they update their code, how much information they include for other developers, whether they receive support from other developers, etc. To answer these questions, GitHub provides access to almost all metadata about its repositories. However, for the question about the reasons for

[1] Some developers only host their code on GitHub and do not use the platform for aiding the code writing process by keeping track of changes or asking for contributions from other open-source developers. This practice is often called as code dumping. My study both includes developers who practice code dumping and developers who actively use the a wide range of functions of the platform.

using GitHub and the motivations for developing a Twitter bot, I relied on survey data.

RQ 1.2 How do developers acquire the skills needed to develop a Twitter bot? To what extent do these programming skills determine and facilitate the creation of Twitter bots?

I start from the assumption (based on my preliminary findings) that developing bots requires programming skills, and that bot codes are developed by a large and diverse group of developers, including programmers with computer science backgrounds, social scientists, journalists and artists. At the same time, the development of bots, like open-source software in general, is increasingly decentralized, with actors relying on reusable code that can be adapted to more specific needs, and sharing knowledge accordingly. This is the scene I would like to explore with my next research questions.

There are a handful of different sources of information available about creating Twitter bots, from blog posts to university courses to a look at the available bot codes on GitHub. What are the most important sources of information for bot development, and what does a bot developer do when an unexpected problem arises? To answer this research question, I rely mainly on survey data.

RQ 1.3 Is there a community of bot developers on GitHub? Or alternatively, is the code for various bots developed by lone developers?

Although GitHub provides a platform for collaboration on projects involving multiple developers, my previous research suggests that nearly 90 percent of the bot code available on GitHub was developed or at least published by only one developer.

The repositories on GitHub do not only keep track all the changes made by developers who have access to the code but also record who contributed to the project by either writing code or simply reporting issues.

All of this data is publicly available for the repositories I examined and provides important insights into how widespread bot development collaboration is and whether there is an active community of developers on GitHub focused on bots.

3.2. Research questions on Twitter bots

RQ 2.1 How do open-source bots generate, process and publish content on Twitter?

What can we learn about Twitter bots by examining the source codes on GitHub and the activity of deployed bots on Twitter in tandem? By combining these two data sources, GitHub and Twitter, in a novel way, I can describe how open-source Twitter bots work (e.g., I can rely on the Twitter bios and account descriptions and the code published on GitHub). All of the above data can be collected computationally through the various APIs of GitHub and Twitter.

Do Twitter bots generate their own content? If so, how do these accounts generate their content, what are the main sources of information used? Some Twitter bots do disclose information about the sources they rely on, either in the bot's bio on Twitter or in the tweets themselves. A good example for the latter is a link included in the tweet. However, the exact way these sources are accessed, processed, and how the bot generates content is often difficult or impossible to understand without looking into the code running behind the bot.

To answer this research question, I am also trying to quantify how much of the traffic generated by open-source bots on Twitter is original content, and how much of the content is simply retweets or quoting other Twitter accounts.

RQ 2.2 What is the life cycle of an open-source bot, and how much traffic is generated by a bot on Twitter during that time? What are the challenges of running a Twitter bot for an extended period of time, and why do bots get banned or become inactive on Twitter?

The life cycle of a Twitter bot can be defined in several ways. We can look at the time between the first and last tweet generated by the account on Twitter or calculate the time spent developing and occasionally updating the bot's source code on GitHub. In most cases, these time periods overlap, but bot accounts are often suspended and sometime even redeployed with a different user handle by the developer. If the bot has a longer lifespan, code developers may need to address issues such as changes in how the Twitter APIs work or how the bot can access its sources (outside of Twitter). I am also trying to understand why some bots not working anymore. To figure out the main challenges, I rely both on studying dysfunctional or inactive bots on Twitter and on asking the developers themselves in the survey.

4. Results of the Thesis

My dissertation contributes to scientific knowledge about open source Twitter bots in two distinct ways. First, it provides a framework to study both the development and the use of open source bots on Twitter. Secondly, it explores the practices of developing open source bots and shed lights on the motivation behind developing Twitter bots, the long-term use of such bots including the challenges of running bots for an extensive period of time and the bots contribution to the Twitter eco-system.

I first investigated open-source bot development practices by examining metadata about bot codes shared on GitHub, the largest online code repository. After analyzing the available metadata about more than 19,000 Twitter bot repositories, I contacted developers who set their email address public on GitHub and asked about their motivations for developing a bot for Twitter and using the platform. In the same survey, I also asked questions about the challenges of developing and deploying a bot on Twitter and the ways in which they acquired the skills necessary to develop a bot. A large number of bot developers shared the Twitter username of their deployed bots - this gave me a unique opportunity to examine the source code of open-source Twitter bots along with the activity of deployed bots on Twitter.

Much of this thesis is concerned with developing a methodology for investigating the development of open-source software on GitHub. Working with digital platform data has its own limitations. In the case of GitHub repository data, the data contains very little information about the motivation behind a particular project, the challenges during development, or the communication related to the project outside of GitHub. On the other hand, answers to direct questions about the challenges during development based on surveys or interviews combine well with automatically collected metadata about the project.

The following section of the results chapter provides an overview of the most important findings in bullet-points style. In-line with the logic of thesis, the first part enlist the findings with regard to the development of the bots, mostly focusing on GitHub, and the second part sums up the findings about deploying and running the open-sources bots on Twitter.

A. Results about bot developers

A1. Based on an analysis of the source codes published on GitHub, it appears that Twitter bots can be programmed to perform a variety of functions and that is relatively easy to repurpose or adapt these bots to perform new tasks on Twitter. It can be argued that individuals and groups who do not have high-level programming skills or the means to pay for expensive IT services can easily use bots on Twitter, suggesting a democratization of these powerful social media tools. On the other hand, the survey results suggest that the vast majority of bot developers are programmers, as nearly 78 percent of bot developers have formal training in programming or computer science. This suggests that while developing a bot is not a super complex task, writing code and understanding other people's code still requires some level of programming knowledge.

A2. The results presented in my thesis suggest that the overwhelming majority of bot code on GitHub was developed by a single author in a relatively short period of time. More than 40% of bot repositories were one-day projects, or in some cases these bots were developed outside of GitHub and the author only uploaded the final code to the platform (using GitHub as a code dump site). This suggests that GitHub is not being used to bring together different skill sets, connect developers from different geographic locations, or help existing teams share tasks, compared to more complex software projects.

A3. Interestingly, projects developed by more than one developer are maintained or developed longer, and these projects receive more engagement on average, such as stars from other GitHub users. About 9 out of 10 bot repositories developed by one author were not forked on GitHub – meaning that these repositories were not copied to enable further work on the code by using the platform. On the other hand, half of the projects developed by multiple authors had at least one fork. Single-author projects received about 1 star on average, while multi-author repositories received 9 stars on average. Finally, single-author projects were maintained for an average 134 days, while repositories developed by more than one author were maintained for 428 days. This could be explained by both the complexity of the code (more complex codes can break more easily) and the increased attention paid to these repositories.

A4. In addition to programming classes in school and looking up bot code on GitHub, the typical developer consults multiple sources when developing a bot for Twitter. The main sources of information were Q&A (question and answer) sites (e.g., a solution to a

specific problem) and blog posts (e.g., a step-by-step description of how to develop a bot). The fact that they prefer to solve their problems individually rather than turning to the developer community on GitHub is partly related to the structure of GitHub. Although the platform is designed to encourage collaboration and teamwork, individual developers for one-person projects do not have good opportunities to get input or help from other developers. The survey of bot developer on GitHub also found that bot developers often do not use the platform to contribute to other people's repositories or reach out to developers. Using the site's version control system, for example, is much more important.

A5. The survey results also suggest that bot developers use GitHub to gain exposure and build their careers. Especially for young developers, recognition and career building are among the top motivations for using the platform. Aside from having a good idea and implementing it, creating a bot seems to be a learning project for many developers. When I asked developers about the most important reason for creating a bot, the majority of survey respondents indicated that learning how the APIs work was a very important or important reason for starting a bot project. Interestingly, self-expression was much less important, and the vast majority rejected the idea that they were developing the bot to support a political cause. From background conversations with developers, it was clear that developers often post bot codes to their GitHub profiles to show that they have personal projects and are interested in programming outside of work.

B. Twitter-based results

Using the Twitter usernames provided by survey participants and the bot repositories owned by developers on GitHub, I was able to compile a list of 321 paired bot repositories. By querying each Twitter account via Twitter's REST API, I was able to download and analyze nearly 500,000 tweets. I first compared this Twitter dataset with the data collected from GitHub. The registration time for bot accounts on Twitter and the estimated distribution of traffic generated by such accounts coincide with the creation of bot repositories on GitHub.

B1. The changes in the number of bot repositories on GitHub and the bot traffic studied on Twitter both suggest that there has been a dynamic expansion of bot activity over the past 6 years. This also suggests that the survey respondents represent the entirety of open-source bot developer quite well.

B2. By modelling the temporal distribution of tweets, I was able to provide

estimations about the volume of traffic generated by a sample of open-source bots. This is a methodological novelty and can be used in similar settings for analyzing the contribution from high-frequency tweeting accounts.

B3. The vast majority of content posted to Twitter by the open-source bots studied is original content as opposed to retweets or quotes of existing content. This is one of the most interesting findings of the study of bot activity on Twitter. Although the review of bot repositories revealed bot accounts that either only amplify other accounts (retweeting them automatically) or curate content posted on Twitter (e.g., only retweet content that contains certain keywords and has certain popularity level), most accounts would fall into to the generative and transmitter bots categories. While generative bots generate their content based on code (e.g., images rendered based on an algorithm), transmitters use widely available web APIs to access content from other platforms, and then post the information (or content) to Twitter.

5. Authors's relevant publications

5.1. Peer-reviewed journal articles

Marchal, Nahema, Lisa-Maria Neudert, Bence Kollanyi, and Philip N. Howard. 'Investigating Visual Content Shared over Twitter during the 2019 EU Parliamentary Election Campaign'. *Media and Communication* 9, no. 1 (3 February 2021): 158–70. https://doi.org/10.17645/mac.v9i1.3421.

Bradshaw, Samantha, Philip N. Howard, Bence Kollanyi, and Lisa-Maria Neudert. 'Sourcing and Automation of Political News and Information over Social Media in the United States, 2016-2018'. *Political Communication* 37, no. 2 (3 March 2020): 173–93. https://doi.org/10.1080/10584609.2019.1663322.

Neudert, Lisa-Maria, Philip Howard, and Bence Kollanyi. 'Sourcing and Automation of Political News and Information During Three European Elections'. *Social Media + Society* 5, no. 3 (July 2019): 205630511986314. https://doi.org/10.1177/2056305119863147.

Kollanyi, Bence. 'Where Do Bots Come from? An Analysis of Bot Codes Shared on GitHub'. *International Journal of Communication* 10 (2016): 4932–51.

5.2. Conference presentations and publications

Machado, C., Kira, B., Narayanan, V., Kollanyi, B., & Howard, P. (2019, May). A Study of Misinformation in WhatsApp groups with a focus on the Brazilian Presidential Elections. In Companion Proceedings of The 2019 World Wide Web Conference (WWW '19), Ling Liu and Ryen White (Eds.). ACM, New York, NY, USA.

Gorwa, R., Kollanyi, B., & Howard, P. (2018, May). A critical analysis of bot detection methodologies In: Methodological Challenges to Studying Misinformation and Disinformation in Data-Driven Politics: Fake News, Bots, and Digital Campaigns, Panel Discussion at the ICA 2018, Prague.

Kollanyi, B. (2016, June). How to write a Twitter bot? Bot Codes Shared on GitHub. Algorithms, Automation and Politics, Preconference of the ICA 2016, Fukuoka.

6. Selected references

Assenmacher, Dennis, Lena Clever, Lena Frischlich, Thorsten Quandt, Heike Trautmann, and Christian Grimme. 'Demystifying Social Bots: On the Intelligence of Automated Social Media Actors'. *Social Media + Society* 6, no. 3 (July 2020): 205630512093926. https://doi.org/10.1177/2056305120939264.

Bastos, Marco T., & Dan Mercea. 'The Brexit Botnet and User-Generated Hyperpartisan News'. *Social Science Computer Review* 37, no. 1 (February 2019): 38–54. https://doi.org/10.1177/0894439317734157.

Bessi, Alessandro, & Emilio Ferrara. 'Social Bots Distort the 2016 US Presidential Election Online Discussion'. *First Monday* 21, no. 11–7 (2016). https://doi.org/10.5210/fm.v21i11.7090.

Brachten, Florian, Milad Mirbabaie, Stefan Stieglitz, Olivia Berger, Sarah Bludau, & Kristina Schrickel. 'Threat or Opportunity? - Examining Social Bots in Social Media Crisis Communication'. In *Australasian Conference on Information Systems 2018*. University of Technology, Sydney, 2018. https://doi.org/10.5130/acis2018.bo.

Bright, Jonathan, Scott Hale, Bharath Ganesh, Andrew Bulovsky, Helen Margetts, & Phil Howard. 'Does Campaigning on Social Media Make a Difference? Evidence From Candidate Use of Twitter During the 2015 and 2017 U.K. Elections'. *Communication Research* 47, no. 7 (October 2020): 988–1009. https://doi.org/10.1177/0093650219872394.

Broniatowski, David A., Amelia M. Jamison, SiHua Qi, Lulwah AlKulaib, Tao Chen, Adrian Benton, Sandra C. Quinn, and Mark Dredze. 'Weaponized Health Communication: Twitter Bots and Russian Trolls Amplify the Vaccine Debate'. *American Journal of Public Health* 108, no. 10 (October 2018): 1378–84. https://doi.org/10.2105/AJPH.2018.304567.

Bruns, Axel, & Jean Burgess. 'Twitter Hashtags from Ad Hoc to Calculated Publics'. *Hashtag Publics: The Power and Politics of Discursive Networks*, 2015, 13–28.

Caldarelli, Guido, Rocco De Nicola, Fabio Del Vigna, Marinella Petrocchi, & Fabio Saracco. 'The Role of Bot Squads in the Political Propaganda on Twitter'.

Communications Physics 3, no. 1 (December 2020): 81. https://doi.org/10.1038/s42005-020-0340-4.

Chavoshi, Nikan, Hossein Hamooni, & Abdullah Mueen. 'DeBot: Twitter Bot Detection via Warped Correlation'. In *2016 IEEE 16th International Conference on Data Mining (ICDM)*, 817–22. Barcelona, Spain: IEEE, 2016. https://doi.org/10.1109/ICDM.2016.0096.

Chen, Zhouhan, & Devika Subramanian. 'An Unsupervised Approach to Detect Spam Campaigns That Use Botnets on Twitter'. *ArXiv:1804.05232 [Cs]*, 14 April 2018. http://arxiv.org/abs/1804.05232.

Chen, Zhouhan, Rima S. Tanash, Richard Stoll, & Devika Subramanian. 'Hunting Malicious Bots on Twitter: An Unsupervised Approach'. In *Social Informatics*, edited by Giovanni Luca Ciampaglia, Afra Mashhadi, & Taha Yasseri, 10540:501–10. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2017. https://doi.org/10.1007/978-3-319-67256-4_40.

Chu, Zi, Steven Gianvecchio, Haining Wang, & Sushil Jajodia. 'Who Is Tweeting on Twitter: Human, Bot, or Cyborg?' In *Proceedings of the 26th Annual Computer Security Applications Conference on - ACSAC '10*, 21. Austin, Texas: ACM Press, 2010. https://doi.org/10.1145/1920261.1920265.

Chu, Zi, Steven Gianvecchio, Haining Wang, & Sushil Jajodia. 'Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?' *IEEE Transactions on Dependable and Secure Computing* 9, no. 6 (November 2012): 811–24. https://doi.org/10.1109/TDSC.2012.75.

Cresci, Stefano. 'A Decade of Social Bot Detection'. *Communications of the ACM 63,* no. 10 (23 September 2020): 72–83. https://doi.org/10.1145/3409116.

Cresci, Stefano, Roberto Di Pietro, Marinella Petrocchi, Angelo Spognardi, & Maurizio Tesconi. 'The Paradigm-Shift of Social Spambots: Evidence, Theories, and Tools for the Arms Race'. *Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion*, 2017, 963–72. https://doi.org/10.1145/3041021.3055135.

Dabbish, Laura, Colleen Stuart, Jason Tsay, & Jim Herbsleb. 'Social Coding in GitHub: Transparency and Collaboration in an Open Software Repository'. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work - CSCW '12*, 1277. Seattle, Washington, USA: ACM Press, 2012. https://doi.org/10.1145/2145204.2145396.

Diakopoulos, Nicholas. *Automating the News: How Algorithms Are Rewriting the Media*. Cambridge, Massachusetts: Harvard University Press, 2019.

Edwards, Chad, Autumn Edwards, Patric R. Spence, & Ashleigh K. Shelton. 'Is That a Bot Running the Social Media Feed? Testing the Differences in Perceptions of Communication Quality for a Human Agent and a Bot Agent on Twitter'. *Computers in Human Behavior* 33 (April 2014): 372–76. https://doi.org/10.1016/j.chb.2013.08.013.

Efthimion, Phillip George, Scott Payne, & Nicholas Proferes. 'Supervised Machine Learning Bot Detection Techniques to Identify Social Twitter Bots' 1, no. 2 (2018): 71.

Efthimion, Phillip George, Scott Payne, & Nicholas Proferes. 'Supervised Machine Learning Bot Detection Techniques to Identify Social Twitter Bots' 1, no. 2 (2018): 71.

Ferrara, Emilio, Onur Varol, Clayton Davis, Filippo Menczer, & Alessandro Flammini. 'The Rise of Social Bots'. *Communications of the ACM* 59, no. 7 (24 June 2016): 96–104. https://doi.org/10.1145/2818717.

Ferrara, Emilio. 'Disinformation and Social Bot Operations in the Run up to the 2017 French Presidential Election', 2017, 33.

Floridi, Luciano. *Information: A Very Short Introduction*. OUP Oxford, 2010.

Forsgren, N., G., Cecarelli, D., Fordi, V., Gennarell, Y., Huang, & Zimmerman, T. (2020). 2020 State of the Octoverse: Empowering healthy communities. https://octoverse.github.com/.

Gaffney, Devin, & Cornelius Puschmann. 'Data Collection on Twitter'. *Twitter and Society* 55 (2014): 67.

Geiger, R. Stuart. 'Bot-Based Collective Blocklists in Twitter: The Counterpublic Moderation of Harassment in a Networked Public Space'. *Information, Communication & Society* 19, no. 6 (2 June 2016): 787–803. https://doi.org/10.1080/1369118X.2016.1153700.

Gilani, Zafar, Reza Farahbakhsh, Gareth Tyson, Liang Wang, & Jon Crowcroft. 'Of Bots and Humans (on Twitter)'. In *Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017*, 349–54. Sydney Australia: ACM, 2017. https://doi.org/10.1145/3110025.3110090.

Gilani, Zafar, Reza Farahbakhsh, Gareth Tyson, and Jon Crowcroft. 'A Large-Scale Behavioural Analysis of Bots and Humans on Twitter'. *ACM Transactions on the Web* 13, no. 1 (20 February 2019): 1–23. https://doi.org/10.1145/3298789.

Gorwa, Robert, & Douglas Guilbeault. 'Unpacking the Social Media Bot: A Typology to Guide Research and Policy'. *Policy & Internet* 12, no. 2 (June 2020): 225–48. https://doi.org/10.1002/poi3.184.

Gousios, Georgios, Bogdan Vasilescu, Alexander Serebrenik, & Andy Zaidman. 'Lean GHTorrent: GitHub Data on Demand'. In *Proceedings of the 11th Working Conference on Mining Software Repositories - MSR 2014*, 384–87. Hyderabad, India: ACM Press, 2014. https://doi.org/10.1145/2597073.2597126.

Haustein, Stefanie, Timothy D Bowman, Benoît Macaluso, Cassidy R Sugimoto, & Vincent Larivière. 'Measuring Twitter Activity of ArXiv E-Prints and Published Papers', 3, 2014. https://doi.org/10.6084/m9.figshare.1041514.

Hofeditz, Lennart, Christian Ehnis, Deborah Bunker, Florian Brachten, & Stefan Stieglitz. 'Meaningful Use of Social Bots? Possible Applications in Crisis Communication during Disasters', 17. Stockholm & Uppsala, Sweden, 2019.

Howard, Philip N., Samuel Woolley, & Ryan Calo. 'Algorithms, Bots, and Political Communication in the US 2016 Election: The Challenge of Automated Political Communication for Election Law and Administration'. *Journal of Information*

Technology & Politics 15, no. 2 (3 April 2018): 81–93. https://doi.org/10.1080/19331681.2018.1448735.

Hwang, Tim, Ian Pearce, & Max Nanis. 'Socialbots: Voices from the Fronts'. *Interactions* 19, no. 2 (March 2012): 38–45. https://doi.org/10.1145/2090150.2090161.

Jensen, Jens F. 'Interactivity: Tracking a New Concept in Media and Communication Studies' 12 (1998): 20.

Jones, Bronwyn, & Rhianne Jones. 'Public Service Chatbots: Automating Conversation with BBC News'. *Digital Journalism* 7, no. 8 (2019): 1032–53. https://doi.org/10.1080/21670811.2019.1609371.

Kalliamvakou, Eirini, Kelly Blincoe, Leif Singer, Daniel M German, & Daniela Damian. 'The Promises and Perils of Mining GitHub (Extended Version)', 39, 2014.

Kollanyi, Bence. 'Where Do Bots Come from? An Analysis of Bot Codes Shared on GitHub'. *International Journal of Communication* 10 (2016): 4932–51.

Kudugunta, Sneha, & Emilio Ferrara. 'Deep Neural Networks for Bot Detection'. *Information Sciences* 467 (October 2018): 312–22. https://doi.org/10.1016/j.ins.2018.08.019.

Lazer, D., A. Pentland, L. Adamic, S. Aral, A.-L. Barabasi, D. Brewer, N. Christakis, et al. 'SOCIAL SCIENCE: Computational Social Science'. *Science* 323, no. 5915 (6 February 2009): 721–23. https://doi.org/10.1126/science.1167742.

Lazer, David M. J., Alex Pentland, Duncan J. Watts, Sinan Aral, Susan Athey, Noshir Contractor, Deen Freelon, et al. 'Computational Social Science: Obstacles and Opportunities'. *Science* 369, no. 6507 (28 August 2020): 1060–62. https://doi.org/10.1126/science.aaz8170.

Lee, Kyumin, James Caverlee, & Steve Webb. 'The Social Honeypot Project: Protecting Online Communities from Spammers'. In *Proceedings of the 19th International Conference on World Wide Web - WWW '10*, 1139. Raleigh, North Carolina, USA: ACM Press, 2010. https://doi.org/10.1145/1772690.1772843.

Lee, Kyumin, Brian David Eoff, & James Caverlee. 'Seven Months with the Devils: A Long-Term Study of Content Polluters on Twitter'. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 8, 2011.

Lima, Antonio, Luca Rossi, & Mirco Musolesi. 'Coding Together at Scale: GitHub as a Collaborative Social Network'. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media*, 10, 2014.

Lin, Po-Ching, & Po-Min Huang. (2013, January). A study of effective features for detecting long-surviving Twitter spam accounts. In Proceedings of the 15th International Conference on Advanced Communication Technology (pp. 841–846). Pyeonchang, Korea: IEEE.

Lokot, Tetyana, & Nicholas Diakopoulos. 'News Bots: Automating News and Information Dissemination on Twitter'. *Digital Journalism* 4, no. 6 (17 August 2016): 682–99. https://doi.org/10.1080/21670811.2015.1081822.

Luceri, Luca, Ashok Deb, Silvia Giordano, and Emilio Ferrara. 'Evolution of Bot and Human Behavior during Elections'. *First Monday*, 31 August 2019. https://doi.org/10.5210/fm.v24i9.10213.

Lumezanu, Cristian, Nick Feamster, & Hans Klein. '#bias: Measuring the Tweeting Behavior of Propagandists'. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media*, 8. Dublin, Ireland, 2012.

Manovich, Lev. *The Language of New Media*. MIT press, 2002.

Marres, Noortje. 'Foreword by Noortje Marres'. In Snee, Helene, Christine Hine, Yvette Morey, Steven Roberts, and Hayley Watson, eds. *Digital Methods for Social Science*. London: Palgrave Macmillan UK, 2016. https://doi.org/10.1057/9781137453662.

Maus, Gregory, & Onur Varol. 'A Typology of Socialbots', 2017, 8.

Metaxas, Panagiotis Takis, & Eni Mustafaraj. 'From Obscurity to Prominence in Minutes: Political Speech and Real-Time Search', 7. Raleigh, NC, USA, 2010.

Minnich, Amanda, Nikan Chavoshi, Danai Koutra, & Abdullah Mueen. 'BotWalk: Efficient Adaptive Exploration of Twitter Bot Networks'. In *Proceedings of the 2017*

IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, 467–74. Sydney Australia: ACM, 2017. https://doi.org/10.1145/3110025.3110163.

Murthy, Dhiraj. *Social Communication in the Twitter Age (Digital Media and Society)*. Cambridge: Polity Press, 2013.

Mustafaraj, Eni, & Panagiotis Takis Metaxas. 'What Edited Retweets Reveal about Online Political Discourse'. In *Conference: Analyzing Microtext, Papers from the 2011 AAAI Workshop*, 6. San Francisco, California, USA, 2011.

Oentaryo, Richard J., Arinto Murdopo, Philips K. Prasetyo, & Ee-Peng Lim. 'On Profiling Bots in Social Media'. In *Social Informatics*, edited by Emma Spiro and Yong-Yeol Ahn, 10046:92–109. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2016. https://doi.org/10.1007/978-3-319-47880-7_6.

Puschmann, Cornelius, & Jean Burgess. 'The Politics of Twitter Data'. *SSRN Electronic Journal*, 2013. https://doi.org/10.2139/ssrn.2206225.

Ratkiewicz, J, M D Conover, M Meiss, B Goncalves, A Flammini, & F Menczer. 'Detecting and Tracking Political Abuse in Social Media'. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 8, 2011.

Rebillard, Franck, & Annelise Touboul. 'Promises Unfulfilled? "Journalism 2.0", User Participation and Editorial Policy on Newspaper Websites'. *Media, Culture & Society* 32, no. 2 (March 2010): 323–34. https://doi.org/10.1177/0163443709356142.

Rogers, Richard. *Digital Methods*. Cambridge, Massachusetts: The MIT Press, 2013.

Rogers. 'Political Research in the Digital Age'. *International Public Policy Review* 8, no. 1 (2014): 73-87.

Schultz, Tanjev. 'Interactive Options in Online Journalism: A Content Analysis of 100 U.S. Newspapers'. *Journal of Computer-Mediated Communication* 5, no. 1 (23 June 2006): 0–0. https://doi.org/10.1111/j.1083-6101.1999.tb00331.x.

Song, Jonghyuk, Sangho Lee, & Jong Kim. 'Spam Filtering in Twitter Using Sender-Receiver Relationship'. In *Recent Advances in Intrusion Detection*, edited by Robin Sommer, Davide Balzarotti, & Gregor Maier, 6961:301–17. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011. https://doi.org/10.1007/978-3-642-23644-0_16.

Spyridou, Paschalia-Lia, & Andreas Veglis. 'Exploring Structural Interactivity in Online Newspapers: A Look at the Greek Web Landscape'. *First Monday*, 2008.

Stieglitz, Stefan, Florian Brachten, Björn Ross, & Anna-Katharina Jung. 'Do Social Bots Dream of Electric Sheep? A Categorisation of Social Media Bot Accounts', 2017, 11.

Storey, Margaret-Anne, Leif Singer, Brendan Cleary, Fernando Figueira Filho, & Alexey Zagalsky. 'The (R) Evolution of Social Media in Software Engineering'. In *Future of Software Engineering Proceedings*, 100–116. Hyderabad India: ACM, 2014. https://doi.org/10.1145/2593882.2593887.

Stukal, Denis, Sergey Sanovich, Richard Bonneau, & Joshua A. Tucker. 'Detecting Bots on Russian Political Twitter'. *Big Data* 5, no. 4 (December 2017): 310–24. https://doi.org/10.1089/big.2017.0038.

Suchacka, Grażyna, & Jacek Iwański. 'Identifying Legitimate Web Users and Bots with Different Traffic Profiles — an Information Bottleneck Approach'. *Knowledge-Based Systems* 197 (June 2020): 105875. https://doi.org/10.1016/j.knosys.2020.105875.

Sundar, S. Shyam. 'Theorizing Interactivity's Effects'. *The Information Society* 20, no. 5 (November 2004): 385–89. https://doi.org/10.1080/01972240490508072.

Takhteyev, Yuri, & Andrew Hilts. 'Investigating the Geography of Open Source Software through GitHub', 2010, 10.

Tao, Y., Dang, Y., Xie, T., Zhang, D., & Kim, S. (2012, November). How do software engineers understand code changes? An exploratory study in industry. In Proceedings of the ACM SIGSOFT 20th International Symposium on the Foundations of Software Engineering (pp. 1-11).

Tarte, S., Willcox, P., Glaser, H., & De Roure, D. (2015, June). Archetypal narratives in social machines: approaching sociality through prosopography. In Proceedings of the ACM web science conference (pp. 1-10).

Toffler, Alvin. *The Third Wave*. A Bantam Book. New York Toronto London Sydney Auckland: Bantam Books, 1990.

Tromble, Rebekah. 'Thanks for (Actually) Responding! How Citizen Demand Shapes Politicians' Interactive Practices on Twitter'. *New Media & Society* 20, no. 2 (February 2018): 676–97. https://doi.org/10.1177/1461444816669158.

Varol, Onur, Emilio Ferrara, Clayton A Davis, Filippo Menczer, & Alessandro Flammini. 'Online Human-Bot Interactions: Detection, Estimation, and Characterization'. In *Proceedings of the Eleventh International AAAI Conference on Web and Social Media*, 10, 2017. https://arxiv.org/abs/1703.03107.

Wang, Alex Hai. 'Don't Follow Me: Spam Detection in Twitter'. In *2010 International Conference on Security and Cryptography (SECRYPT)*, 10. Athens, 2010.

Woolley, Samuel C., & Philip N. Howard. (2016). Social media, revolution, and the rise of the political bot. *Routledge handbook of media, conflict, and security. New York, NY: Routledge*, 282-292.

Woolley, Samuel C. 'Automating Power: Social Bot Interference in Global Politics'. *First Monday*, 21, no. 4 (2016). https://doi.org/10.5210/fm.v21i4.6161.

Wu, Yu, Jessica Kropczynski, Patrick C. Shih, & John M. Carroll. 'Exploring the Ecosystem of Software Developers on GitHub and Other Platforms'. In *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW Companion '14*, 265–68. Baltimore, Maryland, USA: ACM Press, 2014. https://doi.org/10.1145/2556420.2556483.

Zagalsky, Alexey, Joseph Feliciano, Margaret-Anne Storey, Yiyun Zhao, & Weiliang Wang. 'The Emergence of GitHub as a Collaborative Platform for Education'. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 1906–17. Vancouver BC Canada: ACM, 2015. https://doi.org/10.1145/2675133.2675284.