



Selected Papers of Internet Research 16:
The 16th Annual Meeting of the
Association of Internet Researchers
Phoenix, AZ, USA / 21-24 October 2015

ARCHITECTURE FOR UNDERSTANDING THE AUTOMATED IMAGINARY: A WORKING QUALITATIVE METHODOLOGY FOR RESEARCH ON POLITICAL BOTS

Norah Abokhodair
University of Washington, Information School

Samuel Woolley
University of Washington, Communication Department

Philip Howard
University of Washington, Communication Department

David McDonald
University of Washington, Human Centered Design and Engineering Department

Introduction

Social media – especially social networking sites – have substantially transformed the ways in which people discuss current affairs and obtain political news and information. Because of increased affordances for building and maintaining social connections, young people are better able to cultivate a political identity and engage civically in both authoritarian and democratic regimes [1]. Activist causes and democratic movements have been born, organized and disseminated on sites including Facebook, Twitter, Weibo, and YouTube [2]. Like any technology, though, the interfaces, applications, and modes of communication on social networking sites are in constant negotiation, transformation, and repurposing – and by a wide variety of social and political actors.

On Social Bots

Bots, strings of automated code, are a complex and compelling case of this, as automated software agents are increasingly being developed and deployed on social networking sites for a diverse range of purposes. Security experts have found that bots generate over 10 percent of content on social media websites and 62 percent of all web traffic [3]. More recently, we have started seeing social bots emerging in contemporary

Suggested Citation (APA): Abokhodair, N., Woolley, S., Howard, P., & McDonald, D.. (2015, October 21-24). *Architecture For Understanding The Automated Imaginary: A Working Qualitative Methodology For Research On Political Bots*. Paper presented at Internet Research 16: The 16th Annual Meeting of the Association of Internet Researchers. Phoenix, AZ, USA: AoIR. Retrieved from <http://spir.aoir.org>.

social networking sites (SNS) such as Twitter, Facebook, and reddit. In 2012, Facebook estimated that 5-6% of all Facebook accounts were fake. This is a high percentage for Facebook, as it means that approximately 50 million user accounts are manufactured or not human [4]. A recent study of Twitter revealed that bots make up 32% of Twitter posts generated by the most active accounts [5]. Traditionally bots have been developed and deployed to automate benign tasks (e.g., fix syntax, fetch news headlines) and malicious tasks (e.g., spambots, bimbots, and denial of service (DoS) bots) on SNS platforms. These automated functions have been a part of the web almost since its inception. On the other hand, there has been a recent emergence of bots with a political agenda. These bots intervene in discourse about ongoing political events. The results have been interesting to observe. Social bots, software programs designed to appear as human users on social networking sites like Twitter, Instagram, and Facebook, are now not only deployed to peddle questionable pharmaceuticals or trick people into sharing their bank accounts. Another, more propagandistic, side of this automated technology exists.

Computational Propaganda

Regimes of all types are using bots across multiple SNS in efforts to manipulate public opinion, a phenomenon we identify and theorize as 'computational propaganda.' While there are similarities that unite this broader use of bots in the public political sphere, each case of political bot usage is also driven by the particular set of circumstances surrounding specific political events. During the on-going Syrian civil war, for instance, bot accounts have been spread on Twitter and Facebook by a group of programmers employed by The Syrian Intelligence [6]. Based on a recent study [7] which followed the botnet through the 35 weeks it was active, the botnet generated a massive amount of content with more than 3000 tweets a day both in English and Arabic. According to this research, the social bot had three main aims: first, flooding the Syrian revolution hashtags (for example #Syria, #Hama, #Daraa); second, overwhelming the pro-revolution discussion on Twitter and other social media portals; and third, misdirecting by trying to get the audience to attend to other content (for example, the protest in Bahrain and Libya).

The Big Picture

This paper is based on more globally focused study that plots out a working three part qualitative methodology for researching and understanding political bots. These methods are designed for usage by interdisciplinary teams of computer and social scientists who hope to use creative methods to work and communicate with diverse research and public communities. Part one focuses on the development of a codebook built to compile a comparative event dataset of all available instances of political bot usage worldwide. Part two aims at explaining approaches for interviewing the coders and trackers of political bots with an aim of understanding the developmental approach to building these automated scripts. Part three is a dialogue on the working development of computational and communication theory on political bots, an iterative

undertaking based upon interactions with the processes and findings of the first two methodological steps.

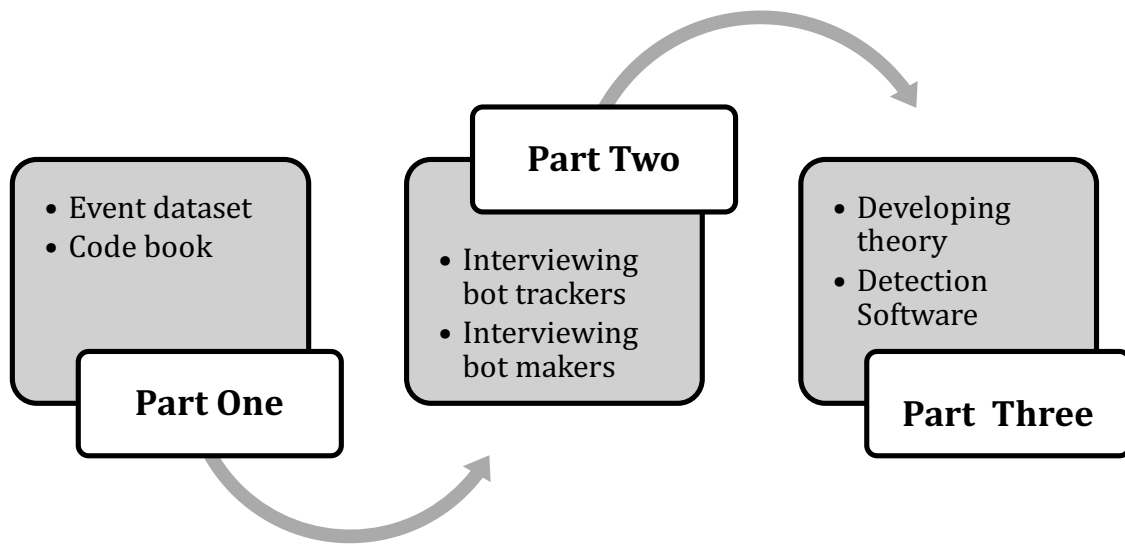


Figure 1: Three Part Qualitative Methodology

Event Dataset & Code Book

In this submission, our aim is to discuss the first part of this project and to provide a guide to developing coding manuals from scratch. Since the inception of this project in August 2014 the team was involved in the construction of an original event data set (with more than 68 cases) that is more comprehensive than any previously available data sets on political bots. Following the classical methodology of the study of unusual phenomena in technology diffusion and usability, we have conducted our sampling of bot use by means of news reports about them [8]. The media-based approach to data collection is especially valuable when the phenomenon at hand is particularly new. The main goal of the event dataset is to find themes and similarities across all cases to help us in the project final step. From the dataset we sampled several cases in order to develop a codebook using qualitative content analysis. This step was necessary to aid the coding of all the available cases of political bots. In the codebook we included seven overarching categories: Botnet Name, Botnet description, The Domain Impacted, The Target country, The Target Organization, The Suspected Deployer, and The Depolyer Goal. These categories were developed based on an iterative exploration of samples of the events, assumptions about the way a political botnet might work, and what we have observed in the prior literature on political bots.

Each case was coded three times by three different coders resulting in a total of 174 coded cases. Finally, we conducted an inter-coder reliability test to finalize the codes and improve the code book. Our goal is to make the code book available on the project

website¹ for other researchers, with hopes that they might benefit from the tool. We have also included a political bot reporting tool on the website in an effort to crowd-source more cases worldwide. As mentioned earlier, our goal is to find themes across these cases to inform theory and to build a stronger detection software.

Conclusion

We are currently concluding the first part of the project and finding many important lessons to report for researchers interested in similar work. Our proposed methodology makes use of triangulation of evidence from online and offline sources through the analysis of data emerging from the qualitative coding stage (part one) with the data we gain from interviewing botnet makers and botnet trackers (part 2) in order to piece together parts of the narrative of a contemporary ICT phenomenon. The results of this work will contribute to the literature on internet research and the use of automated scripts in many ways. One way is through the development of a comprehensive analysis of the current ways political bots are being utilized and how they influence public opinion. We believe that this work will help local and foreign policy makers better understand the relationship between technology diffusion and political processes.

References

[1] W. Lance Bennett and Alexandra Segerberg, *The Logic of Connective Action: Digital Media and the Personalization of Contentious Politics*, 2013; Philip N. Howard, *The Digital Origins of Dictatorship and Democracy: Information Technology and Political Islam* (New York, NY: Oxford University Press, 2010).

[2] F. Edwards, Philip N. Howard, and Mary Joyce, "Digital Activism and Non-Violent Conflict" (Digital Activism Research Project, 2013); Jennifer Earl and Katrina Kimport, *Digitally Enabled Social Change* (The MIT Press, 2011).

[3] Yuval Rosenberg, "62 Percent of All Web Traffic Comes from Bots," *The Week*, December 16, 2013, <http://theweek.com/article/index/254183/62-percent-of-all-web-traffic-comes-from-bots>.

[4] Protalinski, E. (2012). Facebook: 5-6% of accounts are fake | ZDNet.ZDNet. Retrieved May 27, 2014, from <http://www.zdnet.com/blog/facebook/facebook-5-6-of-accounts-are-fake/10167>

[5] Cheng, A., & Evans, M. (2009, August 1). Twitter Statistics - An In-depth Report At Most Active Twitter Users. *Social media monitoring and analytics solutions*. Retrieved June 3, 2014, from <https://www.sysomos.com/insidetwitter/mostactiveusers/>

¹ Politicalbots.org

[6] York, J. C. (2011, April). Syria's Twitter spambots. (T. Guardian, Producer)
Retrieved June 2014, from The Guardian:
<http://www.theguardian.com/commentisfree/2011/apr/21/syria-twitter-spambots-pro-revolution>

[7] Abokhodair, N., Yoo, D., McDonald, D. (2015). Dissecting a Social Botnet: Growth, Content and Influence on Twitter. Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW 2015). ACM

[8] Earl, J., Martin, A., McCarthy, J.D., Soule, S.A. (2004). The Use of Newspaper Data in the Study of Collective Action. Annual Review of Sociology. 30, 65-80.