Association for Information Systems

AIS Electronic Library (AISeL)

ICIS 2019 Proceedings

Cyber-security, Privacy and Ethics of IS

# Using Agent-Based Modelling to Address Malicious Behavior on Social Media

Agnieszka Onuchowska
*University of South Florida*, aonuchowska@mail.usf.edu

Donald J. Berndt
*University of South Florida*, dberndt@usf.edu

# Using Agent-Based Modelling to Address Malicious Behavior on Social Media

*Short Paper*

**Agnieszka Onuchowska**
University of South Florida
4202 E Fowler Ave, Tampa, FL 33620,
USA
aonuchowska@mail.usf.edu

**Donald J. Berndt**
University of South Florida
4202 E Fowler Ave, Tampa, FL 33620,
USA
dberndt@usf.edu

## Abstract

*In this study we create a platform for evaluating social media policies through simulation. We argue that social media policies need to be tested and refined before they can be successfully applied. We propose agent-based modelling (ABM) as a method for representing both malicious and legitimate social media agents, along with their key behaviors. Our two main research questions are as follows. 1. How do we build an agent-based model of a social media platform to address social media regulation? 2. How can an agent-based simulation approach be used to assess the effectiveness of social media policies? A preliminary agent-based model has been implemented (in Python), using the five human user types ('amplifier', 'broadcaster', 'commentator', 'influential user' and 'viewer') and two bot types ('simple' and 'sophisticated'). During the simulation, a social media network of 100 agents is created and the agents' behaviors are captured in this paper.*

**Keywords:** Twitter, Malicious Accounts, Agent-Based Modeling, Social Media Policy

## Introduction

The existence of malicious content on social media platforms has been an increasingly important issue over the past several years. Malicious actions on social media can be used to implement media manipulation campaigns (Bradshaw and Howard, 2018) and disseminate extremist propaganda (Ferrara, 2017). In the past, manipulation on social media has been used to influence presidential election outcomes (Ferrara, 2017, Howard et al., 2017), disseminate hate and harassment (Wolley and Howard, 2016) or propagate 'fake news' (Lazer et al., 2018). Social media platforms have been working intensively to reduce the number of malicious accounts and minimize adverse impacts of their actions. For example, Twitter strengthened their Automation Rules Policy in 2017 to ensure that automated tweets are filtered out (Twitter, 2017) whereas Facebook and Instagram blocked accounts which propagated inauthentic news prior to the US Midterm elections in November 2018 (Gleicher, 2018). Yet, it appears that social media companies have been unable to eradicate malicious actors from their platforms. As a result, malicious actions such as election manipulations, fake news propagation or privacy infringement of social media user accounts are still taking place (Yar, 2018). Malicious social media campaigns target local populations and global audiences, potentially harming both individual citizens as well as government structures (Bradshaw and Howard, 2017). Therefore, since 2016 at least 43 governments worldwide have been introducing regulations that address social media abuse (Bradshaw et al., 2018). In extreme cases, an absence of effective measures to prevent civil unrest fueled in part by the propagation of fake information related to terrorist attacks or ethnic violence caused some governments to temporarily block social media sites in their countries (Bogost, 2019).

Although some fear that social media policies can potentially curb important conversations, disallow free speech or oppress unpopular opinions (Arnold, 2018, Bradshaw et al., 2018), the introduction of such policies is likely to take place in the near future (Leetaru, 2019, Schwartz, 2019). Social media policies can help ensure responsible use of social media tools and mitigate the risks related to the use of social media platforms (Hrdinova et al., 2010). For example, Grinberg et al. (2019) suggest that social media outlets should actively seek ways to discourage social media users from following or sharing fake news content. To do that, social media platforms need to apply policies, which prevent unwanted content from flooding the network (Grinberg et al., 2019, Roberts, 2018).

Our motivation for this study is to create a platform for evaluating social media policies through simulation. Based on our observations of past social media outlets' unsuccessful attempts to control malicious actors, we believe that new social media policies need to be tested and refined before they can be successfully applied in a live social network setting. Furthermore, the consequences of such policies should be better understood before implementation in live settings. We propose agent-based modelling (ABM) as a method for representing both malicious and legitimate social media agents, along with key behaviors. The goal is to define a network of Twitter actors, which consists of bot- and human-owned accounts. Our two main research questions are as follows. 1. How do we build an agent-based model of a social media platform to address social media regulation? 2. How can an agent-based simulation approach be used to assess the effectiveness of social media policies? In this paper, we show how agent-based models of a Twitter-like network can extend current research on Twitter and contribute to research on controlling malicious accounts in the more general social media environment. The design of the model is based on findings collected from the literature review and descriptive analyses we have recently conducted. The end goal of constructing an agent-based model of the social media environment is to assess the impacts of policies aimed at curbing malicious behavior on social media outlets through simulation.

This paper is structured as follows. First, we review past research that used agent-based modelling to investigate phenomena on the Twitter platform. We then focus on the review of past literature which classified the types of Twitter accounts and their behaviors. Third, we present an agent-based model and underlying assumptions tailored for policy evaluation. Finally, we discuss preliminary results and present next steps related to this research.

## Literature Review

### *Agent-Based Models of Twitter Network*

Agent-based models are often created with the goal of running granular or 'bottom-up' simulations, with systemwide behaviors emerging from the process. The agents in such models have unique characteristics and their behavior is simulated based on pre-defined rules and assumptions. As an output, macro-level patterns emerge from the agents' interactions (Groff, 2007). In particular, ABM seems like the right tool to be used for simulations of Twitter, which is a complex system defined by the interaction of many heterogeneous agents. Agent-based models have proven to be particularly useful in analyzing phenomena that are difficult to model using differential equation-based models (Rahmandad & Sterman, 2008). As far as Twitter is concerned, ABM is likely to be useful for investigating 'what-if' questions that project simulated future scenarios rather than modelling the past. Rahmandad & Sterman (2008) argue that agent-based modelling can be used to simulate targeted attacks or random failures and help test out network interactions using the creation or removal of links or nodes connecting network agents. Previous research on social media that uses agent-based modelling is quite diverse and explores many aspects related to social media networks. Table 1 serves as a summary of literature on agent-based simulations of social media networks.

| Year | Author | Agent-based Model Focus | Type of ABM Network Topology |
|---|---|---|---|
| 2018 | Fan et al. | Replication of diffusion patterns of emotion contagion on social media. | Directed network |
| 2016 | Charlton et al. | Relationship between Twitter users' sentiment levels | Not provided |
| 2015 | Sathanur et al. | Controlling of viral rumor activity spread | Stochastic block-model topology |

| | | | |
|---|---|---|---|
| 2015 | Plikynas et al. | Controlling of propagation of excitation on social media | Bi-modal model |
| 2015 | Attema et al. | Prediction of future volumes of tweets based on simulated past user behavior | Dynamic Random Graph Model |
| 2014 | Tang et al. | Social media policies for promoting information propagation | Zombie-city model |
| 2014 | Yang et al. | Population growth and message propagation among Twitter financial communities | Not defined |
| 2013 | Gatti et al. | Information spread across online social networks during 2012 US presidential race | Egocentric network |
| 2013 | Van Maanen et al. | Investigation of social influence on online social media | Not provided |
| 2011 | Liu & Chen | Rumor spreading on Twitter-like microblogging sites | Scale-free network |
| 2010 | Graham | Self-organization of communities through Twitter | Scale-free network |

**Table 1. Selected literature on agent-based simulations of social media networks**

For example, Sathanur et al. (2015) and Liu & Chen (2011) used agent-based simulation to investigate the propagation of rumors on social media outlets. Yang et al. (2014) and Gatti et al. (2013) introduced ABM to understand how information gets propagated on social media, whereas Tang et al. (2014), with the use of an agent-based model, tested social media policies for promoting information propagation on social media. Van Maanen and van der Vecht (2013) used agent-based simulation to investigate on online social influence. In turn, Fan et al. (2018) proposed an agent-based model to replicate emotion diffusion patterns and Plikynas et al. (2015) used ABM to simulate the control of excitation on social media. Graham (2010) applied ABM to investigate how communities self-organize on the Twitter network. Attema et al. (2015) used ABM to simulate a prediction of future volumes of tweets based on past user behavior. Charlton et al. (2016) used agent-based modelling to simulate the dynamics of sentiments and relations between users' sentiments on Twitter.

## *Accounts' Behavior Classification*

Past literature provides valuable information on the classification of both legitimate and malicious actors on social media platforms based on the actors' behaviors, which can be used to create a simulated model. Table 2 summarizes the literature on the classification of Twitter accounts, which we relied on when defining an agent-based model of the Twitter network presented in Table 3. Below we provide examples of past research findings defining bot and human-owned account types, which later served as an input into the simulation model.

Varol et al. (2017) provided a distinction between (1) simple bots, which are likely to retweet other simple bots and (2) sophisticated bots, which usually retweet human accounts. Freitas et al. (2015) defined (1) high activity bots, which tweet between 1 and 60 minutes and (2) low activity bots, which generate tweets between 1 and 120 minutes. Abokhodair et al. (2015) defined five types of bots: peripheral bots followed by core bots and their subtypes: short-lived bots, long-lived bots and generator bots. Varol et al. (2014) provided a distinction between influential users and information consumers. Influential users' tweets are usually popular and receive high numbers of retweets whereas information consumers usually retweet others' tweets. Tinati et al. (2012) provided characteristics of five categories of human behavior on Twitter and grouped the users into the following categories: idea starters, amplifiers, curators, commentators, viewers.

| Year | Author | Identified Twitter Account Types |
|---|---|---|
| 2017 | Varol et al. | 'Simple bots,' 'sophisticated bots' |
| 2015 | Freitas et al. | 'High activity bots,' 'low activity bots' |
| 2015 | Abokhodair et al. | 'Peripheral bots,' 'core bots,' 'short-lived bots,' 'long-lived bots,' 'generator bots' |
| 2014 | Varol et al. | 'Influential users,' 'information consumers' |
| 2012 | Tinati et al. | 'Idea starters,' 'amplifiers,' 'curators,' 'commentators,' 'viewers' |

| | | |
|---|---|---|
| 2012 | Cha et al. | ‘Mass media,’ ‘grassroots,’ ‘evangelists’ |
| 2012 | Chu et al. | ‘Humans,’ ‘bots,’ ‘cyborgs’ |
| 2010 | Stringhini et al. | ‘Bragger,’ ‘whisperer’ |
| 2008 | Krishnamurthy et al. | ‘Broadcasters,’ ‘acquaintances,’ ‘miscreants,’ ‘evangelists’ |

**Table 2. Selected literature on Twitter account classification**

Cha et al. (2012) grouped Twitter users into three categories: (1) ‘mass media’ accounts that have a large number of followers but do not follow many accounts themselves, (2) ‘grassroots’ accounts represented by ordinary users and (3) ‘evangelists’ accounts represented by opinion leaders, celebrities or politician accounts. Chu et al. (2012), apart from identifying human and bot accounts, also provided a description of cyborg accounts (bot-assisted humans or human-assisted bots), which are an amalgamation of human and bot accounts. Stringhini et al. (2010) identified two categories of bots. The first category, called ‘bragger,’ posts spam tweets on their own Twitter pages, which are then visible on the followers’ feeds (only followers and not their contacts can see such spam tweets). The second category, ‘whisperer,’ sends direct spam messages to users, without posting the content on the Twitter webpage. Krishnamurthy et al. (2008) distinguished between the following types of human-owned accounts: (1) broadcasters run by media outlets such as newspapers or radio stations whose number of followers exceeds the number of followees, (2) acquaintances, whose number of followees and followers is usually similar; (3) miscreants represented by spammers or stalkers and (4) evangelists, who reach out to other users to collect new followers.

# Model Definition

After reviewing the past literature on Twitter actor classifications, we did not find a classification presented in a single paper which could serve as a basis for a simulated agent-based Twitter network containing a variety of legitimate and malicious actors. Past taxonomies of Twitter actors did not address the complexity of a Twitter network containing both malicious and legitimate entities. General findings are supported in past research and similar patterns emerge from the past taxonomies when one considers the behaviors of the identified actors. Therefore, when defining the ABM simulation, we decided to combine the findings from several past classifications and incorporate them in the simulated agent-based model. We follow the design science research methodology (Hevner et al., 2004) and propose an artefact in the form of a simulated agent-based model of a social media network.

## *Twitter Agents and Relations Between Them*

Since we were not able to identify one single classification in the past literature which would thoroughly define our agent-based model, we decided to combine past findings on the types of Twitter agents. Following Bessi & Ferrara (2016) we assumed that 15% of all accounts are represented by bots and we assigned human accounts to the remaining 85% of the agents. In order to define bot-type agents in the agent-based model, we followed the Varol et al. (2017) classification and generated simple bot agents that retweet other simple bot agents as well as sophisticated bot agents, which focus on retweeting human agents. We followed the findings of Chu et al. (2012) and assumed that 60% of the defined bot accounts have fewer followers than followees. Among the remaining 40% of bots, we assumed that half of the bot population has a balanced number of followers and followees, whereas the other half has more followers than followees. As far as legitimate accounts are concerned, we defined the agents following the findings published by Krishnamurthy et al. (2008) and Tinati et al. (2012). We defined five types of legitimate agents: (1) ‘influential users,’ (2) ‘broadcasters,’ (3) ‘amplifiers,’ (4) ‘commentators’ and (5) ‘viewers’ (40% of the whole simulated population (Echeverria & Zhou, 2017)). Table 3 presents the sets of characteristics of the agents, which were populated in the simulated agent-based model.

| Agent | % | Characteristics | Tweets per day | Retweets per day |
|---|---|---|---|---|
| Simple bot (SIM) | 9 | No. of followees > no. of followers; posts more than 1600 tweets/ week; retweets other simple bot agents (Varol et al., 2017); over 50% of populated content is related to retweeting activities (Bessi & Ferrara, 2016) | 5-50 | 15-150 |

| | | | | |
|---|---|---|---|---|
| | | Agents with similar behavior patterns: 'high activity bot' (Freitas et al., 2015), 'bot' (Chu et al., 2012), 'short-lived bot' (Abokhodair et al., 2015) | | |
| Sophisticated bot<br>(SMT) | 6 | Posts on average 4 tweets a day and less than 70 tweets/week, usually not more than 2 tweets a day (Lee et al, 2011); high retweet levels of human-generated content (Varol et al., 2017)<br>Agents with similar behavior patterns: 'low activity bot' (Freitas et al., 2015), 'long-lived bot' (Abokhodair et al., 2015) | 0-5 | 1-10 |
| Influential user<br>(INF) | 0.05 | Generates popular tweets and receives high numbers of retweets (Varol et al., 2014)<br>Agents with similar behavior patterns: 'idea starter' (Tinati et al., 2012), 'evangelist' (Cha et al., 2012) | 0-5 | 0-3 |
| Broadcaster<br>(BRD) | 0.95 | Accounts run by newspapers, radio stations, etc.; no. of followees < no. of followers (Krishnamurthy et al., 2008)<br>Agents with similar behavior patterns: 'mass media user' (Cha et al., 2012) | 5-15 | 1-10 |
| Amplifier<br>(AMP) | 14 | Shares others' ideas; more likely to retweet others' ideas than post own tweets (Tinati et al., 2012)<br>Agents with similar behavior patterns: 'information consumer' (Varol et al., 2014), 'curator' (Tinati et al., 2012) | 0-5 | 2-20 |
| Commentator<br>(COM) | 30 | Ordinary human-owned accounts; number of followees = number of followers (Krishnamurthy et al., 2008); usually has 100 - 3000 followers and followees; most active users tweet 10 -200 times/ week; usually retweet others' tweets (Xu & Yang, 2012)<br>Agents with similar behavior patterns: 'acquaintance' (Krishnamurthy et al., 2008) | 0-2 | 0-5 |
| Viewer<br>(USR) | 40 | Takes passive interest in conversations on Twitter (Tinati et al., 2012); does not post any tweets or retweets (Echeverria & Zhou, 2017)<br>Agents with similar behavior patterns: 'grassroots user' (Cha et al., 2012) | 0 | 0 |

**Table 3. Defined sets of characteristics for simulated network agents**

## Results

The agent-based model of the Twitter network comprises the following parts: (1) the agent ecosystem and (2) behavioral rules taken from past research for each agent as defined in Table 3, as well as (3) the communication mechanism (such as tweets, likes and retweets). A preliminary agent-based model has been implemented (in Python), using the five human user types ('amplifier', 'broadcaster', 'commentator', 'influential user' and 'viewer') and two bots ('simple' and 'sophisticated'). The percentages in Table 3 are used as a guide for populating the entire agent ecosystem, while the tweet/retweet rates are parameterized along with other factors that affect individual agent behaviors.

To describe the characteristics of the generated groups of agents, we used the evaluation metrics ('tweet frequency' and 'total retweet count') presented in Table 4. In the course of this research we are yet to implement and assess 'Klout score' and 'account reputation metric.'

| Metric | Source | Formula |
|---|---|---|
| Tweet frequency | Dickerson et al., 2014 | The average number of tweets generated by an account on a daily basis |
| Total retweet count | Cha et al., 2010 | Total number of retweets all tweets received (Cha et al., 2010) |

| | | |
|---|---|---|
| Klout score | Zhang et al., 2016 | How frequently is the account retweeted |
| Account reputation | Chu et al., 2012 | Follower count/ (follower count + following count) |

**Table 4. Description of proposed Twitter account evaluation metrics**

For illustrative purposes, a small simulation with 100 agents is run for 100 ticks (simulated days) to generate some descriptive statistics. The tweet and retweet rates for the different kinds of agents are shown in Figure 1 (with the associated data in Table 4). (Please note that the 'simple bot' is not pictured in Figure 1 since the tweet/retweet rates are much higher (see Table 4)).Since the ecosystem is small (100 agents), there is only a single agent representing an 'influential user' (INF) and 'broadcaster' (BRD). These are the rare celebrity and professional media types. Other agents presented in the simulated network are as follows: 'amplifier' (AMP), consisting of 14 agents, 'commentator' (COM) that consists of 30 agents, 'simple bot' (SIM) represented by 9 agents and 'sophisticated bots' (SMT) represented by 6 agents. The base User (USR) type is a variant of the Commentator (COM) agent class and consists of 39 agents. The tweet and retweet rates are different for individual agents since a random process generates daily actions on a tick-by-tick basis.

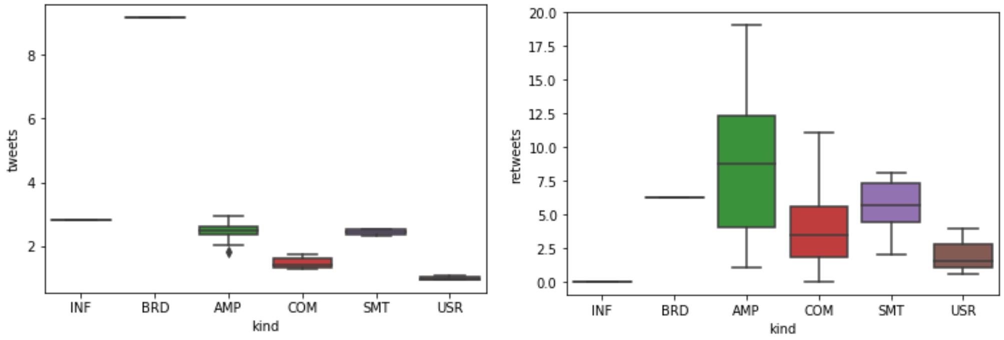

**Figure 1. Tweet (left) and retweet (right) rates (per day) for the agent ecosystem, except Simple Bot with much higher rates.**

| Kind of Agent | Number of Agents in the model | Tweets per Day: Mean | Minimum | Maximum | Retweets per Day: Mean | Maximum | Minimum |
|---|---|---|---|---|---|---|---|
| AMP | 14 | 2.46 | 1.84 | 2.96 | 8.47 | 19.04 | 1.08 |
| BRD | 1 | 9.16 | 9.16 | 9.16 | 6.26 | 6.26 | 6.26 |
| COM | 30 | 1.47 | 1.28 | 1.76 | 4.10 | 11.04 | 0.00 |
| INF | 1 | 2.82 | 2.82 | 2.82 | 0.00 | 0.00 | 0.00 |
| SIM | 9 | 25.53 | 25.26 | 27.64 | 15.29 | 45.70 | 0.00 |
| SMT | 6 | 2.44 | 2.34 | 2.54 | 5.56 | 8.12 | 1.98 |
| USR | 39 | 1.01 | 0.96 | 1.10 | 2.04 | 3.94 | 0.60 |

**Table 5. Tweet and retweet rates by agent types.**

During the simulation, a social media network is created and captured (using the NetworkX Python package for graphs). The final network for this 100-agent simulation and the zoom-in on the 'influential user' (INF) agent and its' network is shown in Figure 2. The different types of agents are named with varying three-character metasyntactic variables (such as "Foo"), along with an integer. The edges are the following relationships, which determine the message propagation pathways.

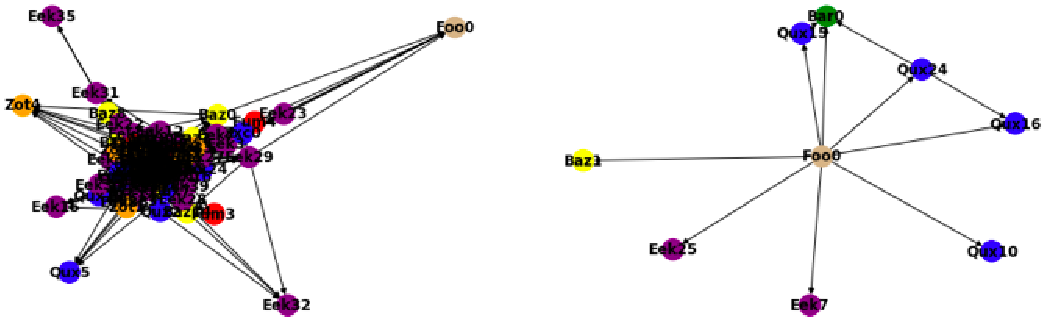

**Figure 2. Social media network depicting the simulation with 100 agents (left), a zoom-in on Fooo, an 'influential user' agent with a circle of followers (right).**

## Discussion and Next Steps

We see this article as another step in our comprehensive and on-going stream of research on the implications of malicious accounts on social media outlets. As a next step, we plan to use the agent-based model defined in this paper to test the implications of selected past policy implementations on simulated malicious behaviors. Using an agent-based model, we plan to simulate scenarios which show how the introduction of policies on social media outlets can help reduce malicious behavior. The extension of this paper will focus on the following research question: Can we adequately evaluate policy alternatives using agent-based modelling and simulation? Future research will apply social media policies on simulated scenarios based on documented past examples of disruptive behavior on social media. Next, the consequences of the introduction of the defined policies on the behavior of simulated agents will be analyzed.

We plan to support our future research using the diffusion of innovations theory, which serves as a base for the network interventions process (Valente, 2012). When testing malicious accounts' policy scenarios on a simulated agent-based model of a social network, we plan to design and apply network interventions (Valente, 2012), which are based on the diffusion of innovations theory (Rogers, 2003). Valente (2012) argues that network interventions' goal is to "use social networks or social network data to generate social influence, accelerate behavior change, improve performance, and/or achieve desirable outcomes among individuals, communities, organizations, or populations." Aral (2012) recommended the application of network interventions in a social media context in order to contain negative user behaviors on social media outlets or to support positive behaviors among social media users. Lazer et al. (2018) defined two types of interventions, which can be used to target fake news propagation. First, changes in the structure of the social network are proposed to prevent a user from being exposed to fake news. Second, intervention proposes that users should be empowered to critically evaluate the news on social media in terms of their authenticity. Harris et al. (2014) proposed node removal in the Twitter network to address malicious consequences of astroturfing as a part of network interventions strategy. The types of interventions listed above will be studied using the enhanced agent-based models and simulation.

## References

Aral, S. 2012. “Social Science: Poked to Vote,” *Nature*, 489(7415), p. 212.

Arnold, A. 2018. “Do We Really Need To Start Regulating Social Media?”, *Forbes*, August 24 (https://www.forbes.com/sites/andrewarnold/2018/07/30/do-we-really-need-to-start-regulating-social-media/#27f9b68a193d, accessed August 21, 2019).

Attema, T., van Maanen, P. P., & Meeuwissen, E. 2015. “Development and Evaluation of Multi-Agent Models Predicting Twitter Trends in Multiple Domains,” in *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pp. 1133-1140.

Bessi, A., & Ferrara, E. 2016. “Social bots distort the 2016 US Presidential election online discussion,” *First Monday*, 21 (11) (doi: 10.5210/fm.v21i11.7090).

Bogost, I. 2019. “When a Country Bans Social Media,” *The Atlantic*, April 22 (https://www.theatlantic.com/technology/archive/2019/04/sri-lanka-social-media-ban-tk/587728/, accessed August 21, 2019).

Bradshaw, P. & Howard, P. 2017. “Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation,” *Oxford, Oxford Internet Institute*, (https://www. oii. ox. ac. uk/blog/troops-trolls-and-troublemakers-a-global-inventory-of-organized-social-media-manipulation/, accessed August 21, 2019).

Bradshaw, S., & Howard, P. 2018. “Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation,” *The Computational Propaganda Project, University of Oxford* (http://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2018/07/ct2018.pdf, accessed August 21, 2019).

Bradshaw, S., Neudert, L. M., & Howard, P. N. 2018. “Government Responses to Malicious Use of Social Media,” *NATO StratCom COE* (https://www.stratcomcoe.org/government-responses-malicious-use-social-media, accessed August 21, 2019).

Cha, M., Benevenuto, F., Haddadi, H., & Gummadi, K. 2012. “The World of Connections and Information Flow in Twitter,” *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 42(4), pp. 991-998.

Charlton, N., Singleton, C., & Greetham, D. V. 2016. “In the Mood: The Dynamics of Collective Sentiments on Twitter,” *Royal Society Open Science*, 3(6), 160162.

Chu, Z., Gianvecchio, S., Wang, H., & Jajodia, S. 2012. “Detecting Automation of Twitter Accounts: Are You a Human, Bot, or Cyborg?,” *IEEE Transactions on Dependable and Secure Computing*, 9(6), pp. 811-824.

Echeverria, J., & Zhou, S. 2017. “The ‘Star Wars’ Botnet with> 350k Twitter Bots,” arXiv preprint arXiv:1701.02405.

Fan, R., Xu, K., & Zhao, J. 2018. “An Agent-Based Model for Emotion Contagion and Competition in Online Social Media,” *Physica A: Statistical Mechanics and its Applications*, 495, pp. 245-259.

Ferrara, E. 2017. “Contagion Dynamics of Extremist Propaganda in Social Networks,” *Information Sciences*, 418, pp. 1-12.

Ferrara, E. 2017. “Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election,” *First Monday*, 22(8) (10.5210/fm.v22i8.8005).

Gatti, M. A. D. C., Appel, A. P., dos Santos, C. N., Pinhanez, C. S., Cavalin, P. R., & Neto, S. B. 2013. “A Simulation-Based Approach to Analyze the Information Diffusion in Microblogging Online Social Network,” in *Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World,* pp. 1685-1696.

Gleicher, N. 2018. “Election Update,” *Facebook Newsroom*, November 5, (https://newsroom.fb.com/news/2018/11/election-update/, accessed August 21, 2019).

Graham, K. C. 2010. “Complexity Science and Social Media: Network Modeling in Following “Tweets”,” in *2010 IEEE Systems and Information Engineering Design Symposium,* pp. 141-146.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. 2019. “Fake News on Twitter During the 2016 US Presidential Election,” *Science*, 363(6425), pp. 374-378.

Groff, E. R. 2007. “Simulation for Theory Testing and Experimentation: An Example Using Routine Activity Theory and Street Robbery,” *Journal of Quantitative Criminology*, 23(2), pp. 75-103.

Harris, J. K., Moreland-Russell, S., Choucair, B., Mansour, R., Staub, M., & Simmons, K. 2014. “Tweeting For and Against Public Health Policy: Response to the Chicago Department of Public Health's

Electronic Cigarette Twitter Campaign," *Journal of Medical Internet Research*, 16(10): e238. (doi:10.2196/jmir.3622).

Hevner A. R., March, S. T., Park, J., & Ram, S. 2004. "Design Science in Information Systems Research," *MIS Quarterly*, 28(1), pp. 75-105.

Howard, P. N., Bradshaw, S., Kollanyi, B., & Bolsolver, G. 2017. "Junk News and Bots during the French Presidential Election: What Are French Voters Sharing Over Twitter In Round Two?," Technical Report, Data Memo 2017.4. Project on Computational Propaganda, Oxford, UK.

Howard, P. N., & Woolley, S. C. 2016. "Political Communication, Computational Propaganda, and Autonomous Agents-Introduction," *International Journal of Communication*, 10(2016), pp. 4882-4890.

Hrdinová, J., Helbig, N., & Peters, C. S. 2010. *Designing Social Media Policy for Government: Eight Essential Elements*, Albany, NY: Center for Technology in Government, University at Albany

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Schudson, M. 2018. "The Science of Fake News," *Science*, 359(6380), pp. 1094-1096

Leetaru, K. 2019. "History Tells Us Social Media Regulation Is Inevitable," *Forbes*, April 22 (https://www.forbes.com/sites/kalevleetaru/2019/04/22/history-tells-us-social-media-regulation-is-inevitable/#5c37fe9521be, accessed August 21, 2019).

Lee, K., Eoff, B. D., & Caverlee, J. 2011. "Seven Months with the Devils: A Long-Term Study of Content Polluters on Twitter," In *International AAAI Conference on Weblogs and Social Media*, pp. 185-192.

Liu, D., & Chen, X. 2011. "Rumor Propagation in Online Social Networks Like twitter--A Simulation Study," in *2011 Third International Conference on Multimedia Information Networking and Security,* pp. 278-282.

Plikynas, D., Raudys, A., & Raudys, S. 2015. "Agent-Based Modelling of Excitation Propagation in Social Media Groups," *Journal of Experimental & Theoretical Artificial Intelligence*, 27(4), pp. 373-388.

Rahmandad, H., & Sterman, J. 2008. "Heterogeneity and Network Structure in the Dynamics of Diffusion: Comparing Agent-Based and Differential Equation Models," *Management Science*, 54(5), pp. 998-1014.

Roberts, M. E. 2018. *Censored: Distraction and Diversion Inside China's Great Firewall*, Princeton University Press.

Rogers, E. M. 2003. *Diffusion of Innovations*, New York: Free Press.

Sathanur, A. V., Sui, M., & Jandhyala, V. 2015. "Assessing Strategies for Controlling Viral Rumor Propagation on Social Media-a Simulation Approach," in *2015 IEEE International Symposium on Technologies for Homeland Security (HST),* pp. 1-6.

Schwartz, M. S. 2019. "U.K. Regulators Propose Broad Social Media Regulations To Counter 'Online Harms',", April 8, (https://www.npr.org/2019/04/08/711091689/u-k-regulators-propose-broad-social-media-regulations-to-counter-online-harms, accessed August 21, 2019).

Tang, M., Mao, X., Yang, S., & Zhu, H. 2014. "Policy Evaluation and Analysis of Choosing Whom to Tweet Information on Social Media," in *2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC),* pp. 45-50.

Twitter. 2017. "Automation Rules," *Twitter*, December 6 (https://help.twitter.com/en/rules-and-policies/twitter-automation, accessed August 21, 2019).

Valente, T. W. 2012. "Network Interventions," *Science*, 337(6090), pp. 49-53.

Van Maanen, P. P., & van der Vecht, B. 2013. "An Agent-Based Approach to Modeling Online Social Influence," in *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* pp. 600-607.

Xu, Z. and Yang, Q., 2012. "Analyzing User Retweet Behavior on Twitter," In *Proceedings of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pp. 46-50.

Yang, S. Y., Liu, A., & Mo, S. Y. K. 2014. "Twitter Financial Community Modeling Using Agent Based Simulation," in *2014 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr),* pp. 63-70.

Yar, M. 2018. "A Failure to Regulate? The Demands and Dilemmas of Tackling Illegal Content and Behaviour on *Social Media," International Journal of Cybersecurity Intelligence & Cybercrime, 1(1), pp. 5-20.*

Zhang, J., Zhang, R., Zhang, Y., & Yan, G. 2016. "The Rise of Social Botnets: Attacks and Countermeasures," *IEEE Transactions on Dependable and Secure Computing*, pp. 1068-1082.