Digital Weapons in Social Media Manipulation Campaigns

Felipe Cardoso, Luca Luceri, Silvia Giordano
University of Applied Sciences and Arts of Southern Switzerland (SUPSI), Manno, Switzerland
felipe.cardoso@supsi.ch, luca.luceri@supsi.ch, silvia.giordano@supsi.ch

Abstract
In the last few years, social media have been weaponized to spread false information and affect users’ opinions. Bots represent one of the main actors associated with manipulation campaigns and interference operations on social media. Uncovering the strategies behind bots’ activities is of paramount importance to develop computational tools for their detection. In this paper, we propose to explore bots’ behavior by inspecting the digital weapons (i.e., sharing activities) they utilize to spread content and interact with humans in the run-up to the 2018 US Midterm Election. We observe that bots strategically mimicked the human temporal activity (since several months before the election) and balanced their interaction among human and bot population. Human response to bots’ deceptive activity is alarming: One in three retweets performed by humans comes from a bot-generated content.

Introduction
False news, fake accounts, low-credibility sources, state-sponsored operators, and so on and so forth: The online ecosystem landscape appears loaded of threats and malicious entities disposed to undermine the integrity of social media discussion. Among those, bots (i.e., software-controlled accounts) and trolls (i.e., human operators often state-sponsored) have been recognized as the main responsible actors of manipulation and misinformation operations in diverse contexts (Chen et al. 2020; Yang, Torres-Lugo, and Menczer 2020; Nizzoli et al. 2020). The malicious activity of such actors received enormous resonance in the political domain, where the risk of manipulating public opinion creates concerns for the integrity of voting events (Chen et al. 2020; Stella, Cristoforetti, and De Domenico 2019; Badawy, Lerman, and Ferrara 2019). Hence, developing computational tools to detect orchestrated campaigns is of paramount importance for the health of online platforms and represents a challenging task (Ferrara et al. 2016b). In the last few years, researchers offered several approaches to identify bots (Varol et al. 2017; Yang et al. 2019; Cresci et al. 2019; Mazza et al. 2019), while solutions for unveiling the activity of trolls have been recently proposed (Luceri, Giordano, and Ferrara 2020; Addawood et al. 2019). However, social media abuse persists and the online ecosystem still presents a mix of organic and malicious users (Luceri et al. 2019b; Im et al. 2019).

It is, therefore, of fundamental importance to carry on research to uncover both the strategies and the driving forces behind the activity of malicious actors on social media with the objective of advancing their detection. Along this research direction, in this paper, we inspect the activity of bot accounts on Twitter during the year approaching the 2018 US Midterm Election. In particular, we aim to explore the strategies developed by automated accounts for spreading information and interacting with social media users, while resembling human appearance and avoiding detection. Specifically, we focus on the sharing activities that Twitter users can employ to create content (i.e., original tweets), re-share others’ tweets (i.e., retweets), and respond to others’ tweets (i.e., reply). We investigate how and to what extent bots use such digital weapons over time to identify cues that might enable an early detection of orchestrated campaigns.

Twitter Data and Accounts
To perform our analysis, we leverage the dataset published in (Wrubel, Littman, and Kerchner 2019), which gathers election-related messages shared on Twitter during the year (from January to December) of the Midterm Election (November 6, 2018). From the set of released tweet IDs, we collect 126M tweets by querying the Twitter API. Given the massive amount of accounts to be classified (bot vs. human), we consider to narrow the analysis to the tweets published by the set of ~1M accounts classified in (Luceri et al. 2019a; Deb et al. 2019) by means of Botometer (https://botometer.iuni.iu.edu/). In our dataset, we recog-

<table>
<thead>
<tr>
<th>Table 1: Account and tweet statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number (%)</td>
</tr>
<tr>
<td>accounts</td>
</tr>
<tr>
<td>tweets</td>
</tr>
<tr>
<td>original tweets</td>
</tr>
<tr>
<td>retweets</td>
</tr>
<tr>
<td>replies</td>
</tr>
</tbody>
</table>

Copyright © 2020, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.
nized 943k of these accounts, which were responsible for about 93M tweets, whose details are shown in Table 1.

**Results**

In this Section, we evaluate how bots strategically perform their activity over time by examining their strategies to spread information and interact with other users. In Table 1, we report the number (and percentage) of shared tweets by bot and human accounts. Humans represent the largest portion (80%) of the accounts and shared more content (about 57% of the tweets) with respect to the bot population (20% of the accounts and about 43% of the tweets). However, bots were more active than humans: On average, a bot shared three times the number of tweets shared by a human user (218 vs. 70 tweets/account), which is consistent with bots’ purpose of flooding online platforms with a high volume of tweet (Bessi and Ferrara 2016). In particular, retweeting is the most used sharing activity for both humans (74% of the time) and bots (81% of the time). As expected, bots heavily rely on the retweet action as it represents a simpler operation for automated accounts with respect to the creation of an original tweet or a reply, which requires the usage of more sophisticated techniques based on natural language processing and understanding (Ferrara 2019). This is also in line with previous findings (Luceri et al. 2019b), further confirming that retweets have been employed as bots’ favorite digital weapon to resonate messages and create an illusion of consensus within online communities (Ferrara et al. 2016a).

In Fig. 1a, we show the number of tweets shared weekly by humans and bots over the period under analysis. Interestingly, bots followed a similar temporal pattern of human users over the whole year, which might suggest that bots attempted to resemble human behavior by strategically mimicking their temporal activity (Luceri et al. 2019b). This is also true for the disentangled sharing activities, whose analogous plots are not shown in the interest of space. Indeed, the number of tweets shared by human and bot over time positively correlate ($\rho > 0.97$, p-value < 0.001) for each kind of activity (i.e., original tweets, retweets, replies). Although original tweets and replies require advanced AI techniques, the volume (and temporal patterns) of such activities performed by bots approaches human activities volume (see Table 1), especially if we consider that humans outnumber bots. This suggests that also more sophisticated bots operated along with less advanced *spamners*. Also, as expected, the content shared by both classes of account increases as the election approaches, and a peak of activity is noticeable during the election week. However, it is worth noting how bots started emulating human activity since the beginning of the year, further confirming how early detecting coordinated campaigns is a challenging task. Finally, in Fig. 1b, we show how bots and humans interact with each other in terms of retweets and replies. Node size is proportional to the percentage of accounts in each group for a given sharing activity, while edge size is proportional to the percentage of interactions between each group. We consider each group separately, thus, the edge size gives a measure of the group propensity to interact with the other groups. Interestingly, bots balanced their interactions with humans (49% of the time) and within the bot population (51% of the time), confirming bots’ strategy to amplify messages to elicit an illusion of public consensus. A similar but less balanced pattern can be appreciated for replies, which might suggest that bots mainly use reply to engage with humans. By considering humans interaction, it can be noticed that about one in three retweets performed by humans comes from a bot-generated content, which represents an alarming signal for the spread of misinformation and its impact on people’s beliefs.

**Conclusion**

Bots represent one of the most recognized threats for the integrity of social media. In this paper, we examined how bots strategically perform their online activity over time during the year approaching the 2018 US Midterm Election. We observed how bots reproduce humans’ temporal patterns for every sharing activity since the year before the election and we recognized their propensity of flooding online platforms with retweets. Also, we noticed the usage of increasingly advanced bot accounts capable of creating tweets and engaging with humans in conversations. Finally, our analysis presents an alarming finding: One in three posts re-shared by humans is an original content created by bots. In our future endeavors, we aim to extend this analysis to further characterize bots’ strategy over time.
References


