BotSlayer: DIY Real-Time Influence Campaign Detection

Pik-Mai Hui, Kai-Cheng Yang, Christopher Torres-Lugo, Filippo Menczer
Center for Complex Networks and Systems Research
Luddy School of Informatics, Computing, and Engineering
Indiana University, Bloomington Bloomington, IN 47408
{huip, yangkc, torresch, fil}@iu.edu

Abstract

BotSlayer is an application that helps track and detect potential manipulation of information spreading on Twitter. It can be used by journalists, corporations, political candidates, and civil society organizations to discover online coordinated campaigns in real time. BotSlayer uses an anomaly detection algorithm to flag hashtags, links, accounts, phrases, and media that are trending and amplified in a coordinated fashion by likely bots. A Web dashboard lets users explore the tweets and accounts associated with suspicious campaigns, visualize their spread, and search related content on multiple search engines and social media platforms. BotSlayer is easily installed and configured in the cloud. It will aid in the study and early detection of social media manipulation phenomena.

Introduction

Social media can be manipulated through forms of abuse such as astroturf (Ratkiewicz et al. 2011), amplification of misinformation (Shao et al. 2018b), and trolling (Zannettou et al. 2019). These manipulations often involve social bots, inauthentic accounts controlled in part by software (Ferrara et al. 2016), which have been employed to influence the U.S. presidential election in 2016 (Bessi and Ferrara 2016; Shao et al. 2018a), the 2017 Catalan referendum in Spain (Stella, Ferrara, and De Domenico 2018), the French Presidential Election of 2017 (Ferrara 2017), and the 2018 U.S. midterm election (Deb et al. 2019; Yang, Hui, and Menczer 2019). The detection of such malicious influence campaigns is therefore of utmost importance to our society.

However, the collection of data from social media at scale requires computational resources and coding skills (Davis, Ciampaglia, and others 2016). In addition, the detection of online campaigns presents significant technical challenges (Varol et al. 2017). These difficulties have greatly limited the opportunities for journalists, corporations, political candidates, and civil society organizations to detect and study online manipulation campaigns.

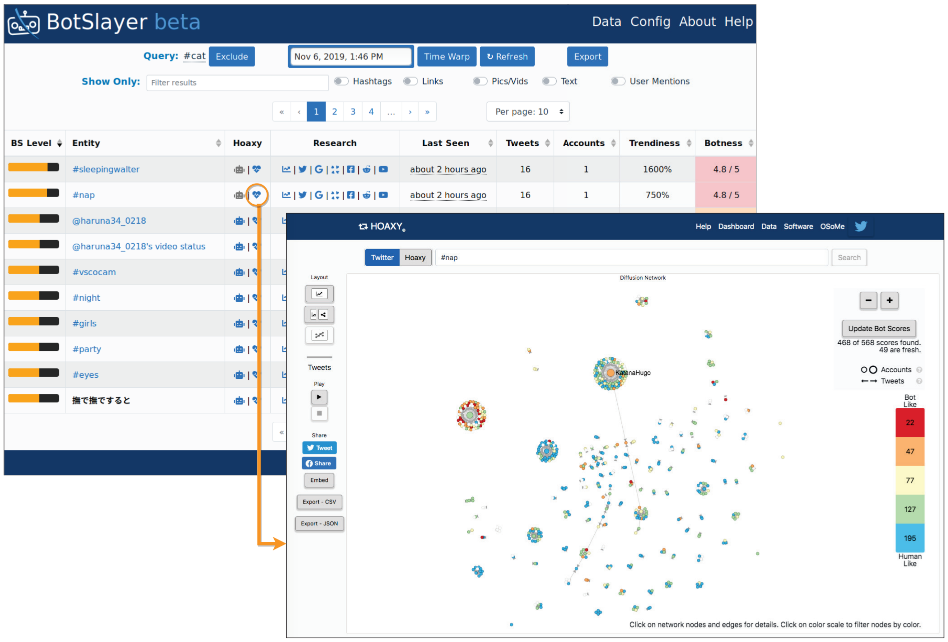
To fill this gap, here we present a free DIY toolkit called *BotSlayer* (URL omitted for anonymity). (osome.iuni.iu.edu/tools/botslayer). The software is available as open source (Hui et al. 2019). The tool is easy to setup and configure to collect public tweets matching a standing user query in real time. Its algorithms mine the data for potential amplification of information by likely coordinated bot accounts with a user-friendly web dashboard for in-depth investigation of suspicious content. In exchange for the free software, BotSlayer collects anonymous usage data for research, in a way that is compliant with Twitter's terms and guidelines.

BotSlayer Software Demo

A BotSlayer instance can be set up using an Amazon Machine Image (AMI) on Amazon Web Services (AWS). New AWS users can host the instance on a free-tier machine. We also provide instructions for an alternative setup using a Docker container. Pulling up the Web interface of a fresh instance of BotSlayer, the configuration page allows users to enter Twitter API keys (which must be obtained from the Twitter Developer platform), and a set of query terms for data collection. BotSlayer will then start collecting tweets matching the user query using the Twitter filter API.

The system extracts entities from each tweet, including hashtags, phrases, user mentions, images, videos, and links. The BotSlayer Web dashboard (Fig. 1) presents the most suspicious 1,000 entities that may be amplified through malicious coordination. Entities are presented in a table, ranked by their "BS" (BotSlayer) level — the higher, the more suspicious. This is a default setting; users can sort according to other criteria, such as the number of tweets containing an entity, its trendiness, the number of accounts, and their average bot score (Yang et al. 2020). The dashboard auto-refreshes to keep the information up-to-date.

The dashboard provides users with links to Hoaxy to investigate suspicious entities. Hoaxy is a system that visualizes diffusion networks — how an entity spreads from user to user — from the Twitter search API or user-provided data (Shao et al. 2018b). Fig. 1 shows the network visualization of one particular entity. Each node in the network represents a Twitter account, colored according to its bot score. Each edge in the network represents a retweet, quote, or mention by which the entity is transmitted. One can replay the spreading of the information and inspect the tweets and accounts involved. The BotSlayer dashboard has two Hoaxy



BotSlayer beta

Data Config About Help

Query: #cat Exclude | Nov 6, 2019, 1:46 PM | Time Warp | ↻ Refresh | Export

Show Only: Filter results ○ Hashtags ○ Links ○ Pics/Vids ○ Text ○ User Mentions

« ‹ 1 2 3 4 … › » Per page: 10

BS Level	Entity	Hoaxy	Research	Last Seen	Tweets	Accounts	Trendiness	Botness
	#sleepingwalter			about 2 hours ago	16	1	1600%	4.8 / 5
	#nap			about 2 hours ago	16	1	750%	4.8 / 5
	@haruna34_0218							
	@haruna34_0218's video status							
	#vscocam							
	#night							
	#girls							
	#party							
	#eyes							
	撫で撫ですると							

Figure 1: BotSlayer Web dashboard. Using action buttons in the Hoaxy column, one can visualize diffusion networks. The research buttons let users plot timelines and explore entities across search and social media platforms.

action buttons: the robot icon visualizes the data collected by BotSlayer that contains a matched query term and the selected entity; the heartbeat icon visualizes the tweets returned by a live search for the entity on Twitter.

For each entity, other functionalities provide timeline plots and search across multiple platforms for studying cross-domain suspicious activities. These include Twitter, Google, Facebook, Reddit, YouTube, and 4chan. If the entity is an image or a video, the Google search button performs a reverse search of the image or video thumbnail. If the entity is a link, searching on other social media may give content on other platforms that contains the same URL.

The dashboard shows information of real-time data by default. Through a "Time Warp" feature, it is possible to browse the past history of the entities tracked by the system. In this way, users can inspect snapshots of the BotSlayer state at previous points in time.

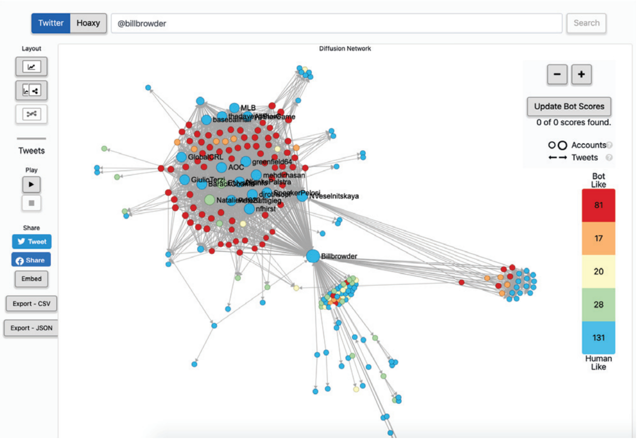


Figure 2: A coordinated campaign detected by BotSlayer

Applications and Impact

Almost 300 users world-wide have installed BotSlayer as of the writing of this paper. They include journalists, academic researchers in various fields, and numerous civil society organizations. One user reported that BotSlayer was able to identify a large number of coordinated accounts spreading ISIS propaganda right after the death of Abu Bakr al-Baghdadi. The majority of these accounts were suspended by Twitter almost immediately. Another user leverages the tool to maintain a list of likely bots posting large volumes of partisan content. Our research lab has uncovered a few instances of coordination using the tool. For example, Fig. 2 shows a diffusion network involving a Russian botnet that promoted a YouTube video. The video is a false news report attacking American activist Bill Browder.

BotSlayer is under active development. Future improvements will focus on better algorithms for the detection of coordinated manipulation even among accounts that are not automated. We hope that BotSlayer will enable new multidisciplinary research on manipulation of online social media by citizens and experts world-wide.

Acknowledgments. We thank Alessandro Flammini, Simon Rosenberg, Nic Dias, and Cameron Hickey for helpful discussions. We gratefully acknowledge support from Democracy Fund, Craig Newmark Philanthropies, The Knight Foundation, DARPA, and the IU Network Science

Institute.

References

Bessi, A., and Ferrara, E. 2016. Social bots distort the 2016 us presidential election online discussion. *First Monday* 21(11).

Davis, C. A.; Ciampaglia, G. L.; et al. 2016. OSoMe: The IUNI Observatory on Social Media. *PeerJ Computer Science* 2:e87.

Deb, A.; Luceri, L.; Badaway, A.; and Ferrara, E. 2019. Perils and Challenges of Social Media and Election Manipulation Analysis: The 2018 US Midterms. In *Companion Proc. WWW Conference*.

Ferrara, E.; Varol, O.; Davis, C.; Menczer, F.; and Flammini, A. 2016. The rise of social bots. *Comm. ACM* 59(7):96–104.

Ferrara, E. 2017. Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election. *First Monday* 22(8).

Hui, P.-M.; Yang, K.-C.; Torres-Lugo, C.; Monroe, Z.; McCarty, M.; Serrette, B. D.; Pentchev, V.; and Menczer, F. 2019. Botslayer: real-time detection of bot amplification on twitter.

Ratkiewicz, J.; Conover, M.; Meiss, M.; Gonçalves, B.; Flammini, A.; and Menczer, F. 2011. Detecting and tracking political abuse in social media. In *Proc. 5th ICWSM*.

Shao, C.; Ciampaglia, G. L.; Varol, O.; Yang, K.-C.; Flammini, A.; and Menczer, F. 2018a. The spread of low-credibility content by social bots. *Nature Communications* 9:4787.

Shao, C.; Hui, P.-M.; Wang, L.; Jiang, X.; Flammini, A.; Menczer, F.; and Ciampaglia, G. L. 2018b. Anatomy of an online misinformation network. *PLoS ONE* 13(4):e0196087.

Stella, M.; Ferrara, E.; and De Domenico, M. 2018. Bots increase exposure to negative and inflammatory content in online social systems. *PNAS* 115(49):12435–12440.

Varol, O.; Ferrara, E.; Menczer, F.; and Flammini, A. 2017. Early detection of promoted campaigns on social media. *EPJ Data Science* 6(13).

Yang, K.-C.; Varol, O.; Hui, P.-M.; and Menczer, F. 2020. Scalable and generalizable social bot detection through data selection. In *Proc. 34th AAAI Conf. on Artificial Intelligence (AAAI)*.

Yang, K.-C.; Hui, P.-M.; and Menczer, F. 2019. Bot electioneering volume: Visualizing social bot activity during elections. In *Companion Proc. WWW Conference*, 214–217.

Zannettou, S.; Caulfield, T.; De Cristofaro, E.; Sirivianos, M.; Stringhini, G.; and Blackburn, J. 2019. Disinformation warfare: Understanding state-sponsored trolls on Twitter and their influence on the web. In *Companion Proc. WWW Conference*, 218–226.