Analyzing Social and Communication Network Structures of Social Bots and Humans

Tuja Khaund[1], Kiran Kumar Bandeli[1], Muhammad Nihal Hussain[1], Adewale Obadimu[1], Samer Al-khateeb[2], Nitin Agarwal[1]
[1]University of Arkansas at Little Rock
Little Rock, USA
{txkhaund, kxbandeli, mnhussain, amobadimu, nxagarwal}@ualr.edu
[2]Creighton University
Omaha, Nebraska, USA
sameral-khateeb1@creighton.edu

***Abstract*— Recently, several journalistic accounts have suggested that Twitter is becoming a bellwether for mis- and dis-information due to the pervasiveness of bots. These bots are either automated or semi-automated. Understanding the intent and usage of these bots has piqued the scientific curiosity among researchers. To that effect, in this study, we analyze the role of bots in two distinct categories of real-world events, i.e., natural disasters and sports. We collected over 1.2 million tweets that were generated by nearly 800,000 users for Hurricane Harvey, Hurricane Irma, Hurricane Maria, and Mexico Earthquake. We corroborate our analysis by examining bots that engaged with the 2018 Winter Olympics. We collected over 1.4 million tweets generated by nearly 700,000 users based on the hashtags #Olympics2018 and #PyeongChang2018. We examined the social and communication network of bots and humans for the aforementioned events. Our results show distinctive patterns in the network structures of bots when compared with that of humans. Content analysis of the tweets further revealed that bots used hashtags more uniformly than humans, across all the events.**

Keywords—Social bots, Twitter, social network, communication network, coordination, natural disasters, Hurricane Harvey, Hurricane Irma, Hurricane Maria, Mexico Earthquake, Olympics, PyeongChang

I. Introduction

A bot is a computer application that is designed to perform automated tasks over the Internet. The main idea behind creating a bot is to run simple tasks that are also structurally repetitive, at a rate much higher than humans [1]. A botnet refers to a collection of computer agents or bots that are programmed to act in a coordinated manner. Bots that mimic social behaviors of humans are referred to as social bots. Social bots could be of different types, viz., advertising bots, entertainment bots, spam bots, and influence bots. Social bots have various capabilities, e.g., learn a user's preferences over time and recommend/advertise products or services, share jokes or satirical remarks, post spam messages or phishing links, affect users' behaviors, among other similar tasks. With all these capabilities, social bots have inarguably played an active role in affecting public discourse in online spaces (e.g., social media and chat forums) [2]. In this research, we investigate the role of social bots during the 2017 natural disaster events (Hurricane Harvey, Hurricane Irma, Hurricane Maria, and Mexico Earthquake) and the 2018 Winter Olympics in PyeongChang, South Korea. Using state-of-the-art bot detection approaches, the objective of the study is to understand the differences in communication and coordination structures of bots and humans. More specifically, we seek answers to the following questions: (1) are there characteristic differences between the social and communication networks of bots and humans?; (2) if so, what are these differences?; (3) do bots and humans have specific Twitter usage and communication styles, i.e., using hashtags, etc.? To answer these questions, we examine social networks of bots and humans, coordination strategies used by bots, and analyze tweets shared by bots.

II. Literature Review

Authors in [3] define a "social bot" as a computer algorithm that automatically produces content and interacts with humans on social media, trying to emulate and possibly alter their behavior. According to a report by [4], social bots are currently being used to systematically exploit and alter the opinion of the general populace. There are recent claims that social bots played a crucial role in the United States 2016 Presidential election [5-7]. These bots can mess with the ecosystem of different social media channels such as Twitter or Facebook. They can create propaganda to promote a favorable image of a public figure/person, increase the followers of an account, or strategically retweet text created by a user [7].

Researchers in [8] identified predictive features such as invitation frequency, outgoing requests accepted, and network clustering coefficient and utilized network-based techniques, crowdsourcing strategies, feature-based supervised learning, and hybrid systems to detect bots to detect bots and developed a tool called BotOMeter, formerly known as BotOrNot [9].

An article from 2014 on detecting bot activity surrounding high impact events, took notice of the move towards link shortener IFTTT (if this, then that) [10] which later became a common practice for bots on Twitter. There are a number of case studies that focus on the dissemination of information on social networks. In one of the botnet case studies, Abokhodair et al. compared the expectations from botnets with the reality of the interactions observed with the botnet. Authors reported that the growth, behavior, and content of the botnet did not align with common conceptions of botnets [11]. For example, some of these botnets did not seem to possess a clear intent to mimic human behavior but rather flooding users' timeline with unrelated information. Researchers then applied these newfound characteristics discovered in the Syrian account botnet to identify more features that differentiate botnets from regular users.

IEEE/ACM ASONAM 2018, August 28-31, 2018, Barcelona, Spain
978-1-5386-6051-5/18/$31.00 © 2018 IEEE

III. METHODS

Google TAGS [12] was used to collect data from Twitter based on keywords and hashtags such as climatechange, #HurricanHarvey, #HurricaneIrma, #MexicoEarthquake, #hurricaneMaria, #PyeongChang2018, winterolympics2018. In our natural disasters dataset, we used four events, specifically Hurricane Harvey, Hurricane Irma, Hurricane Maria and Mexico Earthquake that resulted in over 1.2 million tweets generated by 776,702 Twitter accounts. The methodology is shown in Figure 1.

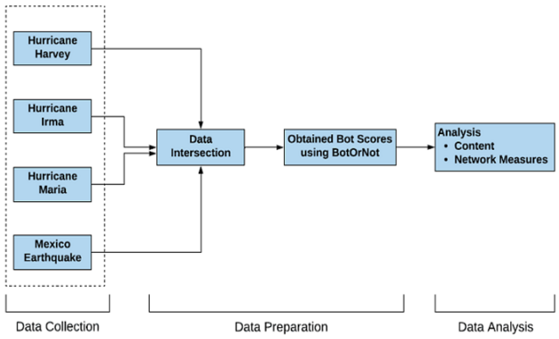

Fig. 1. The research methodology.

Similarly, for the Sport event dataset, we followed the same methodology and have collected over 1.4 million tweets generated by 660,897 Twitter users as shown in Table I and Table II.

TABLE I. 2017 NATURAL DISASTER EVENTS

Event Name	*Tweet Count*	*User Count*	*Time Period*
Hurricane Harvey	223,555	165,928	30/08/17 - 09/09/17
Hurricane Irma	590,748	355,582	05/09/17 - 28/09/17
Hurricane Maria	287,589	160,078	21/09/17 - 28/09/17
Mexico Earthquake	117,557	95,114	20/09/17 - 28/09/17

TABLE II. 2018 INTERNATIONAL SPORTING EVENT

Event Name	*Tweet Count*	*User Count*	*Time Period*
Winter Olympics	1,418,328	660,897	09/02/18 - 25/02/18

We identified common twitter accounts across the four events in Table I. For each user, we calculated their bot likelihood score using BotOrNot; the higher the score, the more likely the account is a bot. Similarly, we collected the bot likelihood scores of the users from Table II. For a robust analysis, top 100 bot and top 100 human accounts were selected for both the datasets (Tables I and II). Maltego [13] was used to collect friends' and followers' information of all the accounts to build a social network. For each account, their retweets and mentions were extracted from original datasets to build the communication network. We used a community detection algorithm proposed by Vincent et. al. [14], which optimizes modularity score, to identify clusters in the social and communication networks.

For content analysis, hashtags from each tweet were extracted to build a hashtag co-occurrence network. Two hashtags were connected if they were mentioned in the same tweet. The more they co-occur, the thicker the edge. Same clustering algorithm, mentioned above, was used to identify communities in the hashtag co-occurrence networks. Later, hashtags from each of these communities were analyzed to identify topics and information space of bots and humans. Significant differences were observed in their networks which will be explained in detail in the next section.

IV. RESULTS

A. Bots vs. Humans - Social and Communication Network Comparison for Natural Disaster Events

For the natural disaster events, the social network of bots (see Fig. 1) clearly reveal a strongly connected core network structure at the center and smaller botnets at the periphery as opposed to humans which consisted of one large component (see Fig. 2). However, more communities were detected in the human network (57 communities) as compared to the bot network (43 communities). Human communities were found to be smaller and denser than bots. Alternatively stated, while human communities are close-knit and more focused, bots tend to maintain a rather sparse connection with the members of the community.

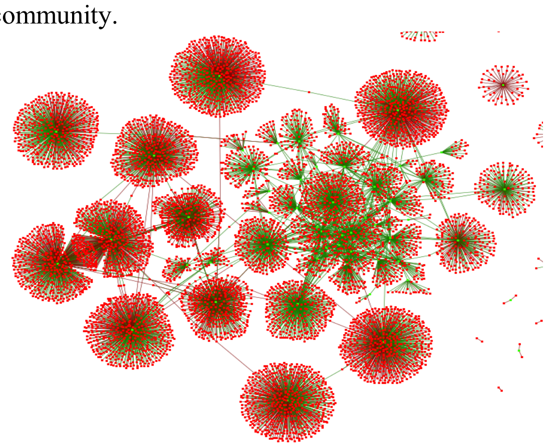

Fig. 2. Bot social network for the natural disaster events, green nodes indicate bots and red nodes indicate their friends and followers.

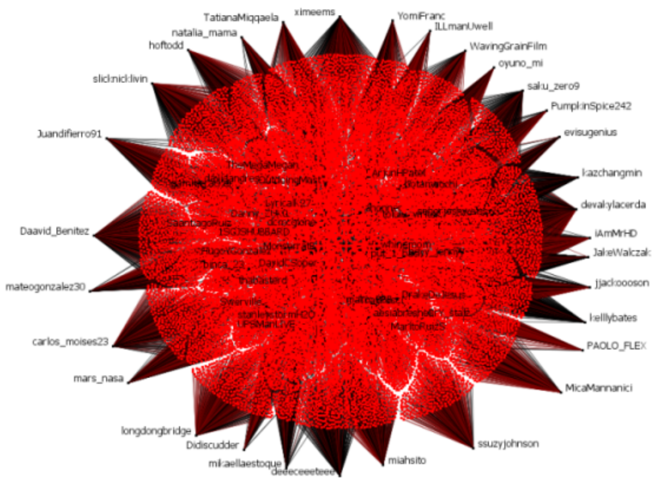

Fig. 3. Human social network for the natural disaster events, black nodes indicate humans and red nodes indicate their friends and followers.

Communication networks (retweets and mentions) of bots and humans are relatively sparse compared to their social networks. However clustering of communication networks resulted in similar findings as that of social networks, i.e., bots have fewer, sparser, and larger communities as compared to humans. Humans have more, denser, and smaller communities, as shown in Fig. 3 and Fig. 4.

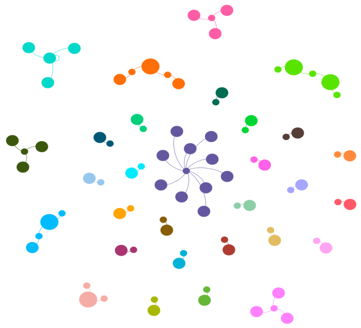

Fig. 4. Bot communication network for the disaster event clustered based on modularity where the colors indicate each cluster.

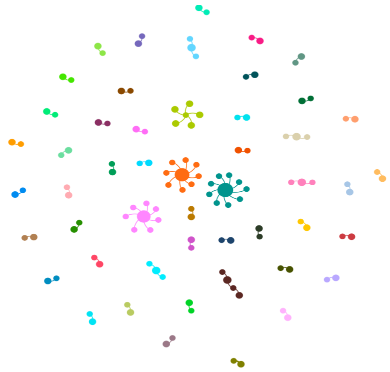

Fig. 5. Human communication network for the disaster event clustered based on modularity where the colors indicate each cluster.

B. Bots vs. Humans - Social and Communication Network Comparison for International Sporting Event

For the international sporting event data, the bot social network shows a distinct botnet in the center and a few isolated components on the periphery (see Fig. 5). Although the human social network consists of a few independent components, the rest of the network was interconnected as one component. Human social network has more communities and connections within the cluster are stronger as compared to the entire network. Due to space limitation, we were unable to include the human social network. Our analysis shows that the structure of the bot network and the human network in the natural disaster events are identical to the networks in the international sport event.

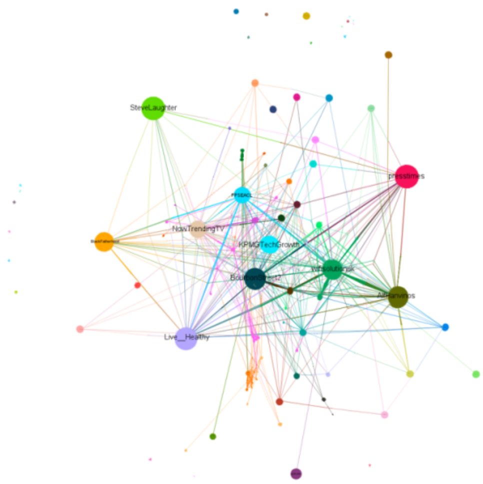

Fig. 6. Bot social network for the sporting event clustered based on modularity where the colors indicate each cluster.

Similar to the natural disaster events, bot and human communication network for sporting event was sparser than their respective social networks. Community extraction on these networks also resulted in similar findings, i.e., bots have fewer, sparser, and larger communities as compared to humans. Humans have more, denser, and smaller communities. Due to space limitations, only bot communication network is shown in Fig. 6.

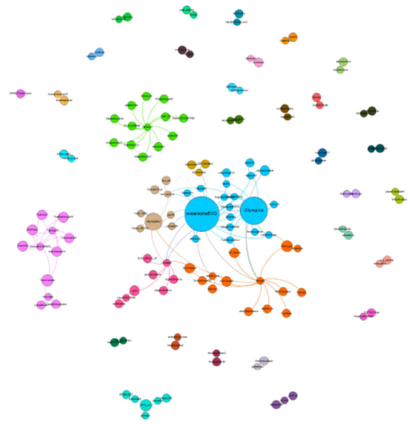

Fig. 7. Bot communication network for the sporting event clustered based on modularity where the colors indicate each cluster.

C. Bots vs. Humans - Hashtag Co-occurrence Network Comparison

For natural disaster events, we found 11 communities in the bots' hashtag co-occurrence network (Fig. 7) and 9 communities for humans' hashtag co-occurrence network (Fig. 8). Upon analyzing these communities, we observed that human communities were more focused as compared to bots. These communities were split based on the areas that were affected, damage and aftermath of these events, and humanitarian relief operations. Whereas, in the bot network, there were only a few communities that represented individual natural disasters and rest of the communities served as bridge between them.

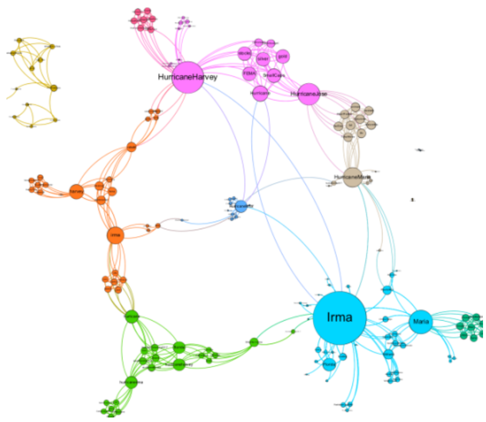

Fig. 8. Bot Hashtag Co-occurrence Network for the natural disaster events clustered based on modularity where the colors indicate each cluster.

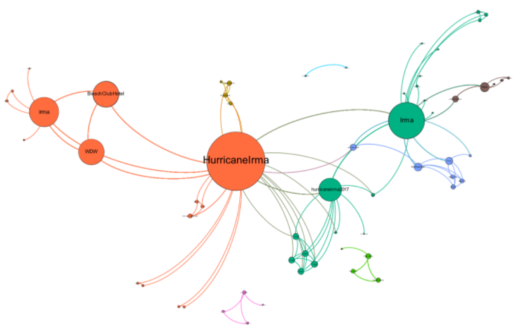

Fig. 9. Human hashtag co-occurrence network for the natural disaster events clustered based on modularity where the colors indicate each cluster.

Similar behavior was observed in the hashtag co-occurrence network for the sporting event (viz., Winter Olympics). There were 12 communities in the hashtag co-occurrence network for bots (Fig. 9). Each of these communities discussed the event in general and contained hashtags related to the opening and closing ceremonies. These communities were primarily separated based on language (Arabic, French, Japanese, Korean, and Russian). Human hashtag co-occurrence network (Fig. 10) had 9 communities and each community discussed specific sport category from the Olympics (men's ice hockey final, women's ice hockey final, halfpipe, snowboard and skeleton racing, speed skating, and ski jump and figure skating).

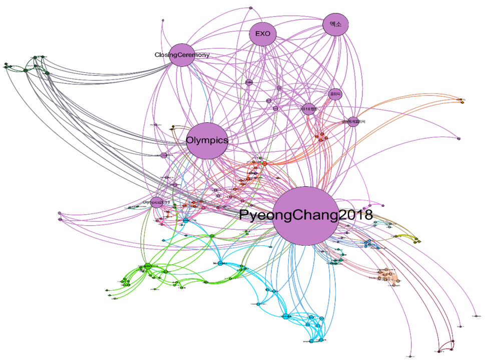

Fig. 10. Bot hashtag co-occurrence network for the sporting event clustered based on modularity where the colors indicate each cluster.

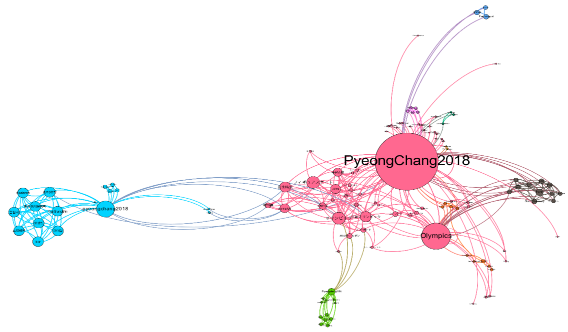

Fig. 11. Human hashtag no-occurrence network for the sporting event clustered based on modularity where the colors indicate each cluster.

V. Discussions and Conclusions

In this paper, we have analyzed the role social bots during two different events, natural disaster events and international sporting event. We studied their different networks, alongside humans, to identify different communication and coordination structures and analyzed the content shared by them. Our research successfully identified distinct network characteristics between bots and humans across different events, such as natural disasters and international sporting event. The stark differences observed between bots' and humans' social network and communication network were consistent across the various events that were studied. Communication networks were sparser than social networks. Bots' social networks had a core and periphery structure unlike humans' social network. In both communication and social network, bots had fewer, sparser, and larger communities, while humans had more, denser, and smaller communities.

Furthermore, for the natural disaster event, bots demonstrated stronger connections among its neighboring nodes in the social network as compared to their communication network. This implies, bots tend to engage better with their friends and followers for information dissemination. However, for the 2018 Winter Olympics, bots demonstrated similar amount of engagement in both their social and communication network. Analysis of the hashtags revealed that, humans used more specific hashtags (niche communities were found), while bots used more generic hashtags. Bots' usage of hashtags was more uniform, while humans' usage of hashtag was more skewed towards the event-specific hashtags.

The research presented here aims to study the behavior of bots in a social space and analyze their communication and network coordination strategies. This research did not develop new or enhance existing bot detection approaches. However, the findings suggest possible future research in incorporating characteristic network structural differences in bot classification approaches. Furthermore, as bots become more sophisticated, we would examine their network coordination and communication evolution trajectory to help develop proactive behavioral models.

Acknowledgments

This research is funded in part by the U.S. National Science, U.S. Office of Naval, U.S. Air Force Research Lab, U.S. Army Research, U.S. Defense Advanced Research Projects Agency and the Jerry L. Maulden/Entergy Endowment at the University of Arkansas at Little Rock.

References

[1] O. Gayer, "What is an Internet Bot | How Bots Can Hurt Your Business," Incapsula Blog, 02-Feb-2016. [Online]. Available: https://www.incapsula.com/blog/understanding-bots-and-your-business.html. [Accessed: 20-Mar-2018].
[2] @DFRLab, "Le Pen's (Small) Online Army," DFRLab, 06-Mar-2017. .
[3] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The rise of social bots," Commun. ACM, vol. 59, no. 7, pp. 96–104, 2016.
[4] E. Ferrara, "Disinformation and social bot operations in the run up to the 2017 French presidential election," First Monday, vol. 22, no. 8, Jul. 2017.
[5] N. Agarwal, S. Al-khateeb, R. Galeano, and R. Goolsby, "Examining the use of botnets and their evolution in propaganda dissemination," J. Def. Strateg. Commun., vol. 2, no. 2, pp. 87–112, 2017.
[6] J. Ratkiewicz, M. Conover, M. R. Meiss, B. Gonc¸alves, A. Flammini, and F. Menczer, "Detecting and Tracking Political Abuse in Social Media," in Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, 2011, pp. 297–304.
[7] S. Al-khateeb and N. Agarwal, "Examining botnet behaviors for propaganda dissemination: A case study of isil's beheading videos-based propaganda," in Data Mining Workshop (ICDMW), 2015 IEEE International Conference on, 2015, pp. 51–57.
[8] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini, "Online Human-Bot Interactions: Detection, Estimation, and Characterization," ArXiv170303107 Cs, Mar. 2017.
[9] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer, "BotOrNot: A System to Evaluate Social Bots," ArXiv160200975 Cs, pp. 273–274, 2016.
[10] S. Mittal and P. Kumaraguru, "Broker Bots: Analyzing automated activity during High Impact Events on Twitter," ArXiv14064286 Phys., Jun. 2014.
[11] N. Abokhodair, D. Yoo, and D. W. McDonald, "Dissecting a social botnet: Growth, content and influence in Twitter," in Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, 2015, pp. 839–851.
[12] M. Hawksey, "TAGS v6.0ns," Google Docs. [Online]. Available: https://docs.google.com/spreadsheets/d/1EqFm184RiXsAA0TQkOyWQDsr4eZ0XRuSFryIDun_AA4/edit?pli=1&usp=embed_facebook&usp=embed_facebook. [Accessed: 20-Mar-2018].
[13] "Paterva Home." [Online]. Available: https://www.paterva.com/web7/. [Accessed: 26-Apr-2018].
[14] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," J. Stat. Mech. Theory Exp., vol. 2008, no. 10, p. P10008, 2008.