Studying the Role of Social Bots During Cyber Flash Mobs

Samer Al-khateeb[1(✉)], Madelyn Anderson[1], and Nitin Agarwal[2]

[1] Creighton University, Omaha, NE 68178, USA
{sameral-khateeb1,madelynanderson}@creighton.edu
[2] COSMOS Research Center, UA – Little Rock, Little Rock, AR 72204, USA
nxagarwal@ualr.edu

Abstract. A Cyber Flash Mob (CFM) is an event that is organized via social media, email, SMS, or other forms of digital communication technologies in which a group of people (who might have an agenda) get together online or offline to collectively conduct an act and then quickly disperse. In addition to the humans participating in these events, non-humans, i.e., artificial agents or social bots - which are computer software programmed to accomplish some tasks on your behalf such as tweeting, retweeting, and liking a tweet - also participate in a CFM. In this research, we study the shared orientations of the CFMs' participants and try to understand the role of social bots in disseminating CFMs' agendas by examining the communication network of these accounts, the toxicity of their posts, and the artifacts, e.g., the URLs they share. The goal is to understand how social bots help CFM organizers advertise, recruit, and share their products (e.g., videos, pictures) on various social media platforms.

Keywords: Cyber Flash Mobs · Social bots · Twitter · Toxicity · Social communication network · Human-machine teaming

1 Introduction

A Cyber Flash Mob (CFM) is an event that is organized via social media, email, SMS, or other forms of digital communication technologies in which a group of people (who might have an agenda) get together online or offline to collectively conduct an act and then quickly disperse [1]. To an outsider, such an event may seem arbitrary. However, a sophisticated amount of coordination is involved. In recent years, cyber flash mobs *"have taken a darker twist as criminals exploit the anonymity of crowds, using social networking to coordinate everything from robberies to fights to general chaos"* [2, 3]. More recently, the term "*mob*" has been increasingly used to remark an electronically orchestrated violence such as the recent attack on the State Capital in Washington by Pro-Trump protesters that lead to property damages, government disruption, and injuries or death for some of the protesters [4, 5]. In a recent incident, an army of small investors from all over the world used Reddit to coordinate "*flashmob investing*" [6] to create stock market frenzy causing GameStop's stock value to rise from $20 to $483 in less than a

R. Thomson et al. (Eds.): SBP-BRiMS 2021, LNCS 12720, pp. 164–173, 2021.
https://doi.org/10.1007/978-3-030-80387-2_16

month [7]. These events show that our systems (security, financial, etc.) are not equipped to handle such highly coordinated and flash actions, underscoring the importance of systematically studying such behaviors.

To study CFMs, it is essential to understand the *motivation* of the individuals that coordinate such events. Many researchers highlight the importance of studying the *shared orientations* such as language, location, religious or political views, among the group members as an indication of group organization. *Shared orientations* among individuals often form the basis for *motivation* resulting in *collective actions* [8, 9]. *Shared orientations* among individuals induce a sense of belongingness to the group, giving rise to the group's collective identity. Individuals may be connected along one or more social dimensions, resulting in multiple shared orientations and thus a stronger collective identity. Hence, in this study we analyze the *languages* and *locations* of the flash mobbers in an attempt to understand their motivation.

In addition to the human participants of such events, *automated actors* or *social bots* are also used during these events to advertise, recruit, and share the products of CFMs, e.g., videos and pictures on various social media platforms. One study estimates that 50% of Twitter accounts are automated or social bots [10] and around 16% of spammers on Twitter nowadays are social bots [11]. Social media moderators are aware of this problem and constantly suspend these accounts. However, there is a lack of systematic investigation of such a human-machine teaming that can affect cyber flash mobs. More specifically, what role do computer agents, or social bots play during CFMs? In this study, we try to answer this research question by examining the following: (1) *in general, where are citizens more interested in CFMs?* (2) *What is the role of social bots in disseminating cyber flash mobs' agenda?* More specifically, (a) *Who is more toxic, bots or humans?* (b) *What are the differences between bots and human's communication networks?* (c) *What resources (e.g., images, videos, URLs) are shared via Cyber Flash Mobs participants (bots and humans)?* Next, we provide a brief literature review of the topics related to this paper, then we discuss our methodology, results, and conclusion with possible future research directions.

2 Literature Review

A flash mob is a phenomenon that has been studied in various disciplines such as communication studies [12], marketing [13], cultural studies [14], and other disciplines. However, there is a lack of a systematic and computational model of its formation and prediction of its occurrence or its success and failure. Our research is one step in this direction. On the other hand, social bots are a known problem that is facing social media sites. It has been studied by various researchers. A study that shares a similar methodology to our research is conducted by Khaund et al. [15]. They focused on the role and coordination strategies of Twitter social bots during four natural disasters that occurred in 2017. Here, we focus on the role of social bots during a CFM event and compare its behavior to humans' behaviors on Twitter. Online toxic content such as toxic posts or toxic comments is another problem that is facing social media sites. Using rude, disrespectful, hateful, and unreasonable language to provoke other users or make them leave a conversation are all examples of toxic behaviors. Toxicity analysis is different from

sentiment analysis as the latter usually gives a score ranking the text to be either positive, negative, or neutral [16]. In our study, we leverage Google Perspective API to assess a toxicity score for each post shared by bots and humans.

3 Methodology

In this section, we explain the methodology we followed to conduct our research. We first explain how we collected our data and provide information about it, then we explain our analysis.

3.1 Data Collection

For this research, we collected data using Twitter Restful API for the period July 3, 2018, to June 18, 2020. We used the keyword "flash mob" to pull data from Twitter using Twitter Archiving Google Sheets (available at https://tags.hawksey.info) then used Python with GSpread API (available at https://gspread.readthedocs.io) to upload the data to our MySQL database. This resulted in 766,111 records (including 108,372 Tweets[1], 581,484 Retweets[2], and 76,255 Mentions[3]) written by 508,029 Tweeters. We preprocessed the data, e.g., reformatted the date and time columns to match the MySQL required date format. Then we extracted all the hashtags and URLs included in the records. This resulted in 29,931 unique hashtags and 50,634 unique URLs. We used Google Perspective API (available at https://www.perspectiveapi.com/) to obtain the toxicity score (i.e., how toxic a text is) of the 766,111 records we collected. Finally, we used Botometer Pro API (available at https://botometer.osome.iu.edu/api) to obtain the bot score (i.e., how likely a Twitter account is a bot or a human) of all the Tweeters.

3.2 Data Analysis

In this subsection we highlight the analysis we conducted then we explain our results in the following subsection. **First**, we analyzed the diversity of the data by analyzing the user's languages and locations of the records we collected. This analysis should help in shedding a light on the ethnographical nature of CFMs. **Second**, we ran a Python script that utilizes the *Botometer Pro API* on our 508,029 unique Twitter accounts to find the likelihood of an account being bot or human. This API calculates various scores, so we selected the "universal" score because it is language-independent, and we have multi-language records. The score is returned in the range of 0.0 to 1.0 representing the likelihood of an account being bot or not. The closer the score is to 1.0, the more likely that account is a social bot. We multiplied the returned values by 100 to better visualize the likelihood of an account being a bot. Since the returned values are continuous and we wanted to have two distinct classes to categorize our Tweeters, i.e., human or bot we considered all the users who have a $\leq$ 10% bot score to be most likely human

[1] These are the records that do not start with RT and do not contain @.

[2] These are the records that start with RT.

[3] These are the records that do not start with RT but contain @.

accounts and all the users who have a ≥90% bot score to be most likely social bots [15]. This method should help in accounting for the Botometer Pro API misclassification and making sure that these accounts have distinct features to be considered as human or bot accounts. **Third**, we used the *Google Perspective API* to calculate the toxicity score of each record (tweets, mentions, or retweets) in our database. The score is returned in the range of 0.0 to 1.0 representing the likelihood of a text being toxic or not. The closer the score is to 1.0, the more likely that text is toxic. We multiplied the returned values by 100 to better visualize the likelihood of toxicity. Then, we calculated the user toxicity score by taking the average toxicity score of all the records of a Tweeter. A threshold of 0.5 (50%) was used to determine if a user is toxic or not. **Fourth**, we extracted all the users mentioned or retweeted from each tweet by using a Python script to create two communication networks, namely the *bot's communication network* and the *human communication network*. This resulted in a *bot's communication network* that consists of 1,861 nodes and 1,675 edges. This network is divided into 407 connected components, which consist of 2 Isolates (single nodes), 229 Dyads (two nodes connected), 81 Triads (three nodes connected), and 95 larger than Triads (more than three nodes connected) (See Fig. 1). We filtered the network and removed all the small-connected components (contain less than 100 nodes), as they did not contribute to our analysis and instead focused on the two largest-connected components which have 495 nodes and 658 edges (See Fig. 2). On the other hand, the *human communication network* consists of 59,530 and 70,418 edges. This network is divided into 6,287 connected components which consist of 46 Isolates (single nodes), 3,719 Dyads (two nodes connected), 1,254 Triads (three nodes connected), and 1,268 larger than Triads (more than three nodes connected) (See Fig. 3). We filtered the network and removed all the small-connected components (contain less than 100 nodes) as they did not contribute to our analysis and instead focused on the largest-connected component which contains 39,182 nodes and 54,520 edges (See Fig. 4). **Fifth**, we applied the *Newman Clustering algorithm* to auto locate communities in the filtered version of both networks. The *bot's communication network* has a *global clustering coefficien*t (*transitivity*) of 0.0 and a *Newman modularity* of 0.679, while the *human communication network* has a *global clustering coefficient* (*transitivity*) of 0.069 and a *Newman modularity* of 0.667. **Sixth**, we analyzed the number of retweets and mentions shared by bots and humans. Then, we calculated the *Retweets to Mentions Ratio* to find differences and similarities of these accounts' behaviors. The *Retweets to Mentions Ratio* give us an idea about how many retweets a bot or a human account posted for each mention. **Finally**, we extracted all the URLs included in the collected records using another Python script. We investigated the popularity of the twelve most known social media sites. This analysis can help us in identifying potential platforms for CFM activities and the difference in humans and bots sharing activities. Finally, we investigated the URL shortening services used by both account types. We used the list of URL shortening services mentioned in [17] to guide us. This analysis also shed light on the difference between humans and bots' behaviors when it comes to shortening URLs and including them in a tweet.

4 Results and Analysis

In this section, we try to answer the research questions listed in our *introduction* using the analysis mentioned in the *data analysis* subsection. We use the **first** analysis to answer the first research question (**RQ1: In general, where are citizens more interested in CFMs?**) We found that 239,985 (31.3%) records were shared with Tweeter's language information (49,216 tweets, 34,950 mentions, and 155,819 retweets) and only 186,553 (36.7%) users shared their language (1,503 bots and 36,872 humans). There are 47 different languages of users who share these records as shown in Fig. 5. The top 10 languages are English, Italian, Portuguese, Spanish, Indonesian, French, Thai, Japanese, Korean, German.

We also found 941 (0.123%) records were shared with geolocation information (517 tweets and 424 mentions) and only 532 (0.105%) users shared their geo-location (0 bots and 88 humans). Figure 6 shows a map of the users (humans) who shared their location. We note that most of the users who shared their geolocations are located in Europe and the USA with few in other parts of the world. This also aligns with our findings of the user's languages, as six out of the top 10 languages mentioned above are languages usually spoken in Europe and the USA. This analysis also shows that social bots share their location and language less than humans.

We used the rest of the analysis to answer the second research question (**RQ2: What is the role of social bots in disseminating cyber flash mobs' agenda?**). More specifically, to answer **a) Who is more toxic, bots or humans?** Using the **second** and **third** analysis, we found that 44,642 Twitter accounts were humans, 1,647 were social bots, and 387,359 accounts were "unknown" because Botometer Pro API did not return their bot score, which could be due to the account being set to private (most likely human in this case) or suspended (most likely a bot that is caught by Twitter algorithm). We found 20,710 (4.1%) toxic users, i.e., users who have a $\geq$ 50% toxicity score. We also found that social bots tend to post less toxic tweets than humans, i.e., the average user toxicity score for social bots is 15.63% while the average user toxicity score for humans is 17.58%. This could be due to the fact that humans tend to express their feelings about a specific CFM which might lead to more toxic posts (e.g., if people are against the CFM), while bots tend to share a URL of the event or prompt participation in a CFM.

We used the **fourth**, **fifth**, and **sixth** analysis to answer, **b) What are the differences between bots and human's communication networks?** We found that the human communication network is more fractured but has stronger within-community cohesion than the bot communication network. There are 6,287 connected components in human communication network (the mean is 38.07) compared to 407 connected components in bot communication network (the mean is 12.18). The *human communication network* has a higher value of transitivity which indicates that the human communities are more cohesive than the *bot communities*. This means the humans are more communicative about the cyber flash mob, while the bots work as disseminators of information about CFMs.

Finally, we found that both bot and human networks have more retweets than mentions. However, the *retweet to mention ratio* is much higher for the bots than the humans, meaning bots have a higher tendency to retweet than humans (see Table 1).

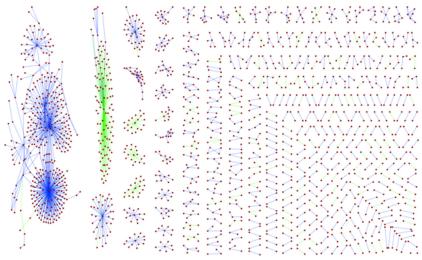

Fig. 1. Bot communication network. Green edges are mentions and blue edges are retweets (Color figure online)

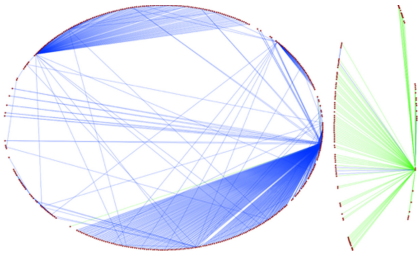

Fig. 2. Filtered bot communication network. Central nodes are high in betweenness centrality (bridges). Green edges are mentions and blue edges are retweets (Color figure online)

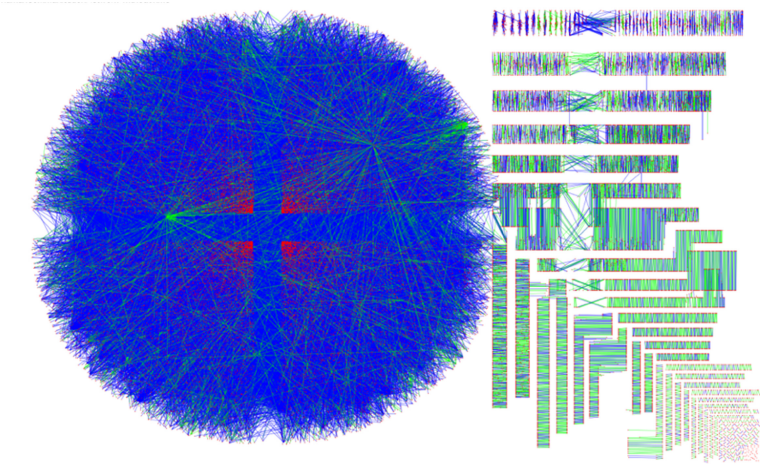

Fig. 3. Human communication network. Green edges are mentions and blue edges are retweets (Color figure online)

We used the **seventh** analysis to answer, **c) What resources (e.g., images, videos, URLs) are shared via Cyber Flash Mobs participants (bots and humans)?** Humans share more URLs than bots regardless of the site type. While humans share more Twitter URLs (e.g., status, picture of the CFM), bots share a disproportionately large number of YouTube URLs. We also found Twitter, YouTube, Facebook, Instagram, and Blogs to be the top 5 social media sites shared by these two account types. Other social media sites seem to be of less interest to both account types. Several researchers note that cyber commentaries on YouTube and Facebook pages are usually written by supporters and/or participants of CFMs, while blogs and online news sources attract "*more heterogeneous readership and offer a glimpse of strong criticism that is also directed at flash mobs*" [18].

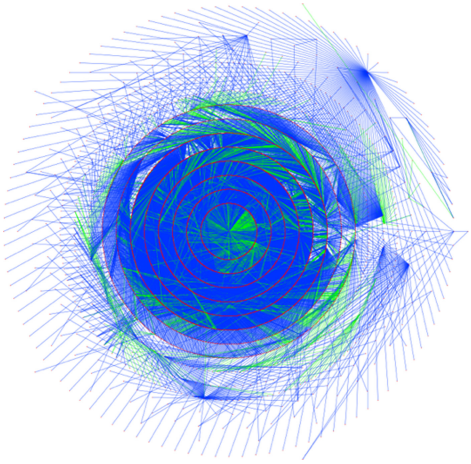

Fig. 4. Filtered human communication network. Central nodes are high in betweenness centrality (bridges). Green edges are mentions and blue edges are retweets (Color figure online)

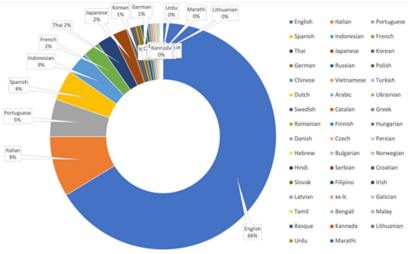

Fig. 5. Distribution of the tweeters' languages

Fig. 6. Tweeters' self-disclosed location

Table 1. The retweets and mentions count for both human and bot communication network. Bots have a higher retweet to mention ratio.

	Retweets	Mentions	(Retweet/Mentions) ratio
Bots	1,345	330	≈4.1
Humans	52,192	18,226	≈2.7

Figure 7 shows these sites and the number of URLs shared by both account types. We also found that humans tend to shorten URLs more than bots, probably due to the number of characters limit imposed by Twitter and so humans can squeeze in more content along with the URL (See Fig. 8).

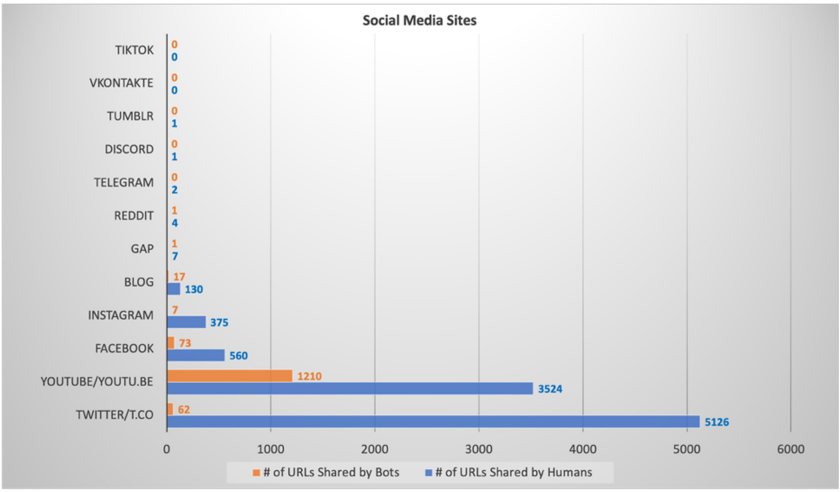

Fig. 7. Top social media sites shared by social bots (orange bars) and humans (blue bars). (Color figure online)

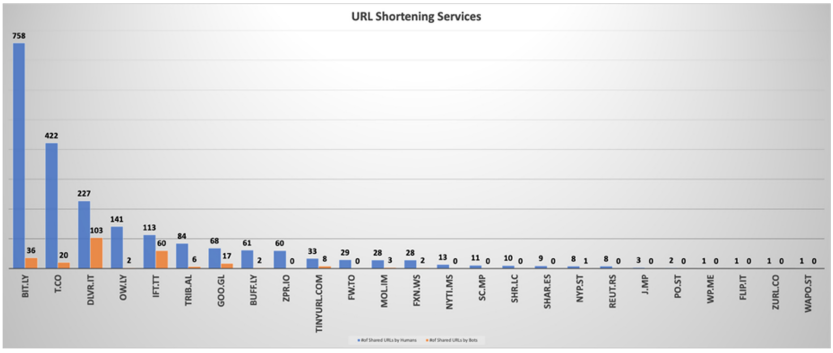

Fig. 8. Top 25 URL shortening services along with the number of URLs shortened and shared by humans (blue bars) and bots (orange bars). (Color figure online)

5 Conclusion and Future Research Directions

In this study, we examined the shared orientations of the CFMs' participants. We focused on the role of social bots and humans in disseminating CFM agenda by examining the communication networks of these two types of accounts, the toxicity of their posts, and the artifacts shared by such accounts, i.e., the URLs in an attempt to understand how social bots help CFM organizers advertise, recruit, and share the products of the CFMs (e.g., videos, pictures) on various social media platforms. We found various differences between the two types of networks highlighted in our analysis subsection. This research sets the foundation for a deeper examination of the role of human-machine teaming on CFM events. The CFM process includes the *planning phase*, the *recruitments phase*, the *execution phase*, and the *replaying and republishing of the products of the flash mob*

phase such as a retweet on Twitter, a Facebook post, or a YouTube Video [19]. In many of the scholarly articles we reviewed about this topic, the organizers of the CFMs cared so much about posting the event's video online and mostly on YouTube (i.e., the *replaying and republishing* phase) - some of them even considered recording the cyber flash mob and posting it on YouTube more important than the cyber flash mob itself because it gives people the re-view capability which is not possible without recording and posting online [20]. As *Lissa Soep* has called it, the "*digital afterlife*" [21] has a significant "*potential for a reaction and recontextualization by other users*" [21]. Hence, one future research direction is to study CFMs on YouTube. There are many CFM videos uploaded to YouTube. Using Google Advanced search and setting words to *flash mob* and the site to *YouTube* returns more than 808,000 hits (as of March 15, 2021). These events can be identified and studied to understand the viewers' behavior toward the various CFM forms.

Acknowledgments. This work is funded in part by the Center for Undergraduate Research and Scholarship (CURAS) at Creighton University, U.S. National Science Foundation (OIA-1946391, OIA-1920920, IIS-1636933, ACI-1429160, and IIS-1110868), U.S. Office of Naval Research (N00014-10-1-0091, N00014-14-1-0489, N00014-15-P-1187, N00014-16-1-2016, N00014-16-1-2412, N00014-17-1-2675, N00014-17-1-2605, N68335-19-C-0359, N00014-19-1-2336, N68335-20-C-0540, N00014-21-1-2121), U.S. Air Force Research Lab, U.S. Army Research Office (W911NF-20-1-0262, W911NF-16-1-0189), U.S. Defense Advanced Research Projects Agency (W31P4Q-17-C-0059), Arkansas Research Alliance, the Jerry L. Maulden/Entergy Endowment, and the Australian Department of Defense Strategic Policy Grants Program (SPGP) (award number: 2020-106-094) at the University of Arkansas at Little Rock. Any opinions, findings, and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding organizations. The researchers gratefully acknowledge the generous support.

References

1. Al-khateeb, S., Agarwal, N.: Cyber flash mobs: a multidisciplinary review. Soc. Netw. Anal. Mining (2021)
2. Tucker, E., Watkins, T.: More flash mobs gather with criminal intent. NBC News, 09 August 2011. https://www.nbcnews.com/id/wbna44077826. Accessed 7 Jan 2021
3. Steinblatt, H.: E-incitement: a framework for regulating the incitement of criminal flash mobs. Fordham Intell. Prop. Media Ent. LJ **22**, 753 (2011)
4. Staff, W.P.: Woman dies after shooting in U.S. Capitol; D.C. National Guard activated after mob breaches building. Washington Post, 06 January 2021
5. Barry, D., McIntire, M., Rosenberg, M.: 'Our president wants us here': the mob that stormed the capitol. The New York Times, 09 January 2021
6. Pratley, N.: The Reddit flash mob won't be able to work the GameStop magic on silver. The Guardian, 01 February 2021. http://www.theguardian.com/business/nils-pratley-on-finance/2021/feb/01/reddits-flash-mob-wont-be-able-to-work-the-gamestop-magic-on-silver. Accessed 5 Feb 2021
7. Brignall, M.: How GameStop traders fired the first shots in millennials' war on Wall Street. The Guardian, 30 January 2021. http://www.theguardian.com/business/2021/jan/30/how-gamestop-traders-fired-the-first-shots-in-millenials-war-on-wall-street. Accessed 5 Feb 2021

8. Melucci, A.: Nomads of the Present: Social Movements and Individual Needs in Contemporary Society. Temple University Press, Philadelphia (1989)
9. Maheu, L.: Social Movements and Social Classes: The Future of Collective Action, vol. 46. SAGE Publications Ltd., London (1995)
10. Koh, Y.: Only 11% of new twitter users in 2012 are still tweeting. Wall Street Journal, 21 March 2014
11. Grier, C., Thomas, K., Paxson, V., Zhang, M.: @spam: the underground on 140 characters or less. In: Proceedings of the 17th ACM Conference on Computer and Communications Security, pp. 27–37. New York, NY, USA, October 2010. https://doi.org/10.1145/1866307.1866311
12. Nicholson, J.A.: Flash! Mobs in the age of mobile connectivity. Fibreculture J. (6), 1–15 (2005). http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.566.6906, http://www.softhook.com/coverage/Fibreculture.pdf
13. Barnes, N.G.: Mob it and sell it: creating marketing opportunity through the replication of flash mobs. Market. Manag. J. **16**(1), 174–180 (2006)
14. Do Vale, S.: Trash mob: Zombie walks and the positivity of monsters in western popular culture. In: The Domination of Fear, pp. 191–202. Brill Rodopi, Amsterdam (2010)
15. Khaund, T., Al-Khateeb, S., Tokdemir, S., Agarwal, N.: Analyzing social bots and their coordination during natural disasters. In: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, pp. 207–212 (2018)
16. Obadimu, A., Mead, E., Hussain, M.N., Agarwal, N.: Identifying toxicity within youtube video comment. In: International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation, pp. 214–223 (2019)
17. Yang, K.-C. et al.: The COVID-19 infodemic: twitter versus facebook. arXiv preprint arXiv:2012.09353 (2020)
18. Molnár, V.: Reframing public space through digital mobilization: flash mobs and contemporary urban youth culture. Space Cult. **17**(1), 43–58 (2014). https://doi.org/10.1177/1206331212452368
19. Zellner, A., Sloan, C., Koehler, M.: The educational affordances of the flash mob: from mobs to smart mobs. In: Koehler, M., Mishra, P. (eds.) Proceedings of SITE 2011–Society for Information Technology & Teacher Education International Conference, Nashville, Tennessee, USA, pp. 3042–3047. Association for the Advancement of Computing in Education (AACE) (2011). https://www.learntechlib.org/primary/p/36779/. Accessed 11 June 2021
20. Venning, W.: Funk up your Japanese with a flashmob!. Thomson (ed.) pp. 60–69 (2016)
21. Soep, E.: The digital afterlife of youth-made media: implications for media literacy education. Comun. Media Educ. Res. J. **20**(1) , 93–100 (2012)