Investigation of Emotion Exchange Motifs in Bot/Human Interactions during Riot Events

Ema Kušen
Vienna University of Economics and Business (WU)
Institute for Information Systems and New Media
ema.kusen@wu.ac.at

Mark Strembeck
Vienna University of Economics and Business (WU)
Secure Business Austria Research Center (SBA)
Complexity Science Hub Vienna (CSH)
mark.strembeck@wu.ac.at

***Abstract*—In this paper, we analyze a data-set consisting of 4.5 million tweets related to four highly emotional events and investigate the interaction patterns that emerge when social bots communicate with human users. In particular, we propose an emotion-annotated n-layer multiplex network model to study the involvement of bots in the exchange of emotional messages. To this end, we study four events: a) the riots that happened during the 2017 G20 summit in Hamburg, b) the 2017 Charlottesville riot, c) the 2017 Catalonia independence referendum riot, and d) the riots that happened after the Philadelphia Eagles won the 2018 Superbowl. In our analysis, we found that: 1) when identifying significant triadic patterns (motifs) in the the respective communication network, a number of star-like subgraphs emerge as representative and significant communication patterns, 2) bots tend to send messages of a higher emotional intensity, as compared to humans, 3) though bots predominantly send retweets, they also compose original messages that may potentially harm the reputation of the person they are targeted at, and 4) in contrast to previous findings, we found that during riot events up to 87.18% of the involved bots actively engage in a direct communication with human users.**

***Index Terms*—Emotions, Multilayer Network, Multiplex Network, Network Motif, Online Social Network, Social Bot, Twitter**

I. Introduction

In general, a complex network [1] represents various types of entities as vertices (e.g. computers or people) that are connected via edges. Depending on the network model, edges can have different semantic meanings, also known as edge types. For example, in an online social network (OSN) a person can be connected to friends, co-workers, and family members. Typically, different edge types are represented in the form of network layers, thereby generating a *multilayer network* [2]. A *multiplex network* is a special type of a multilayer network and assumes that the same set of vertices is connected via different types of edges [2].

In 2002, Milo et al. [3] proposed *network motifs* as statistically significant and over-represented subgraphs that serve as basic building blocks of complex networks. Since then, a large number of studies applied the concept of motifs to study the structural properties of predominantly biological networks, such as cellular networks and protein interaction networks [4], [5]. These studies have shown that motifs provide significant insights into how small subgraph interactions form the behavior on the macroscopic level of a complex network [6].

While the structural patterns emerging in biological networks have been a vivid area of research, the emergence of motifs in human communication networks such as OSNs, has generally been understudied so far.

In this paper, we focus on communication patterns that emerge from the Twitter interaction between bot accounts and human accounts during highly emotional events. In particular, this paper centers around four recent riot events, namely the riots that a) happened during the 2017 G20 summit in Hamburg, b) the 2017 Charlottesville riot, c) the 2017 Catalonia independence referendum riot, and d) the riots that happened after the Philadelphia Eagles won the 2018 Superbowl.

Previous studies have pointed to the negative influence that bot accounts may have on the way human users perceive an event or a particular person [7], or their tendency to manipulate opinion formation during elections [8] and distract human users from important news by sending spam messages [9]. In this paper, we show that bots also engage in a direct communication with human users and form statistically significant and representative subgraphs (motifs). In particular, the motifs we found reveal that bots tend to either flood humans with messages, begin a chain of messages, try to boost their own influence by mentioning themselves, or attract a substantial number of messages.

The remainder of this paper is organized as follows. In Section II we provide an overview of related work, followed by a brief description of the four riot events in Section III. We outline our research method in Section IV and present our results in Section V. Section VI provides a discussion on the most important findings. Section VII concludes the paper.

II. Related work

Existing studies on bot behavior have predominantly provided empirical evidence on the differences between human accounts and bot accounts by observing the social connectivity (e.g., number of followers), the content generation rate, as well as the topics of bot-generated OSN messages (e.g., news, opinion). For example, Riquelme et al. [10] found that, compared to humans, bots tend to be passive (i.e. they don't actively engage in discussions with particular human users).

In recent years, studies have reported on the potential influence of bots during important events, such as political elections or acts of war (see, e.g., [8], [11]–[16]). One such study is presented in [11] where Abokhodair et al. studied the role of bots in a Twitter discussion about the 2012 Syrian civil war. The findings suggest that human users author comparatively more tweets that express personal opinion, while bots tend to disseminate more informative tweets (e.g. news). Moreover, the study also suggests that bots did not aim to mimic the human-like behavior [11]. In contrast, other studies argued that bots are nowadays more sophisticated, can mimic human behavior, and are hard to distinguish from human users [17]–[19]. Some studies even found that sophisticated bots have the ability to deceive and influence people [20].

In this context, Everett et al. [18] have shown that, in general, messages that disagree with the public opinion increase the likelihood of deception. This observation has been empirically demonstrated in [12], where Dickerson et al. studied the case of the 2014 Indian elections. In their study, Dickerson et al. found that humans tend to disagree more with the general opinion (sentiment) when compared to bots. Moreover, the study has also found that humans express stronger positive sentiments than bots.

It has also been shown that bots are active during events where they might follow a strategic agenda [16]. This includes important events such as elections [8]. Gupta et al. [7] found that bots are prone to spreading misinformation due to their automatic retweeting of messages that have not gone through fact checking. For instance, Gupta et al. found that bots were responsible for spreading false accusations emerging after the 2013 Boston marathon bombing.

III. Events of study

Our analysis is based on the four recent riot events.

The 2017 G20 riots in Hamburg. The 2017 G20 summit took place on July 7-8, 2017 in Hamburg, Germany. About a week before the summit, minor clashes occurred between the protesters and the local police. The day before the summit (July 6th), 8,000 protesters gathered in a so-called "Welcome to Hell" march which escalated in violent confrontations between the protesters and the local police, leaving 14 injured demonstrators and 76 injured police officers. The first day of the G20 summit (July 7th) was met with further acts of civil unrest, with the protesters setting cars on fire, looting shops, and clashing with the local police. In the aftermath, 160 police officers were reported injured[1].

The 2017 "Unite the Right Rally" in Charlottesville. The 2017 Charlottesville (Virginia) riots, also called the "Unite the Right Rally", happened on August 11-12, 2017 as a response to the City Council's vote to rename two parks previously named after Confederate generals and remove the statue of the confederate general Robert E. Lee[2]. As the protesters marched the city they were met by the counter-protesters. Fights broke out between the two groups leaving 14 injured. On Saturday, August 12th, a man deliberately drove into a crowd with his car, killing one person and injuring 19[3].

The 2017 Catalonia independence referendum riots. The Catalonia independence referendum was held on October 1st, 2017 in Catalonia, a region in the north east of Spain. Though with a low turnout (43.32%), the referendum resulted in 92.1% of votes in favor of splitting Catalonia from Spain. During the referendum, the Spanish national police tried to prevent people from voting[4]. The clashes with the police resulted in about 900 people injured. As a response, about 15 thousand demonstrators gathered in Barcelona on October 3rd, 2017[5] as a sign of a protest against the police violence.

The 2018 Superbowl riots in Philadelphia. Superbowl LII was played on February 4, 2018. The Philadelphia Eagles beat the New England Patriots 41:33. On February 5, 2018, thousands of Eagles fans gathered on the streets of Philadelphia to celebrate the victory. However, the celebration evolved into a series of acts of vandalism and a riot, with people flipping over cars, attempting to tear down traffic lights and lamp post, and setting objects on fire[6]. The Philadelphia police called in additional support from the US National Guard. One police officer ended up injured while several rioters requested medical help.

IV. Method

Data extraction. We used Twitter's Search API[7] to extract publicly available data related to the four riot events (see Section III). Each of the extraction procedures started the day when the riot happened and stopped about a week later. We restricted our data extraction to tweets written in English language (Charlottesville, Catalonia, Hamburg, and Philadelphia riots) and German language (Hamburg riots). In total, we collected 4,519,152 unique tweets authored by 1,698,701 Twitter users (see Table I).

Emotion detection. Next, we pre-processed the data and identified the presence and the intensity of the eight basic emotions according to Plutchik's wheel of emotions [21] (anger, fear, sadness, disgust, joy, trust, anticipation, and surprise). To identify emotions, we applied our emotion detection algorithm (see [22]) which uses the NRC[8] lexicon, the AFINN lexicon of affect [23] (contains scores between [-5,5] to increase or decrease an intensity of a particular emotion-carrier word), as well as a set of heuristics that humans naturally use when

[1] Hamburg: https://www.theguardian.com/world/2017/jul/07/g20-protests-hamburg-altona-messehalle

[2] Charlottesville: https://edition.cnn.com/2017/08/14/us/charlottesville-rally-timeline-tick-tock

[3] Charlottesville: https://edition.cnn.com/2017/08/12/us/charlottesville-white-nationalists-rally

[4] Catalonia: https://www.nytimes.com/2017/10/01/world/europe/catalonia-independence-referendum.html

[5] Catalonia: https://www.theguardian.com/world/2017/oct/03/catalonia-holds-general-strike-protest-referendum-violence

[6] Philadelphia: http://www.bbc.com/news/world-us-canada-42943824

[7] For our data extraction, we used the following list of hashtags and search terms. **Hamburg**: #G20HH2017, #G20Hamburg, #G20HAM17, #G20HAM, "#G20 #Hamburg", "Hamburg riot", "Hamburg Unruhe"; **Charlottesville**: #Charlottesville, #UnitetheRight, "Charlottesville riot"; **Catalonia**: #Catalonia, #CatalanReferendum, #RepublicofCatalonia, "Catalonia violence", "Catalonia protest", "#1oct Catalonia"; **Philadelphia**: #PhillyBurning, #Phillyriot, "#superbowl #Philadelphia", "#Philadelphia #riot", and "Philadelphia riot".

[8] NRC: http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm

	Hamburg	Charlottesville	Catalonia	Philadelphia
Period	6.7-17.7 2017	10.8-16.8 2017	28.9-16.10 2017	4.2-10.2 2018
Tweets	653,568	2,202,682	1,640,829	22,073
Users	178,879	1,020,729	487,152	11,941
Bots	2,841	3,004	3,703	129
Vertices	982 (286)	5,550 (2,619)	1,071 (281)	6 (3)
Edges	1,100	3,335	5,091	3

TABLE I

BASIC INFORMATION ABOUT EACH DATA-SET: PERIOD OF DATA EXTRACTION, NUMBER OF UNIQUE TWEETS (AFTER THE DOUBLES REMOVAL), UNIQUE USERS, AND IDENTIFIED BOTS. THE NUMBERS OF VERTICES AND EDGES ARE BASED ON THE SUBSET OF BOT VERTICES THAT ARE ENGAGED IN A COMMUNICATION (VIA @SCREENNAME) WITH ANOTHER VERTEX (EITHER BOT OR HUMAN). THE NUMBERS IN BRACKETS SHOW THE NUMBER OF BOT VERTICES PARTICIPATING IN A DIRECT TWITTER COMMUNICATION (VIA @SCREENNAME).

assessing emotions in written texts. This particular set of heuristics was identified in previous scientific studies (see, e.g., [24]–[28]) and considers negation, misspellings, downtoners, boosters, amplifiers, maximizers, smileys, as well as common abbreviations.

Bot detection. For bot identification we used Botometer's Python API[9]. Botometer relies on a number of features associated with a Twitter account, such as the content of a tweet, friendship network, sentiments conveyed in tweets, temporal tweeting behavior, and user meta data [29]. A bot-score assigned by Botometer is the likelihood of a Twitter account to be controlled by a bot, ranging from 0 (definitely a human account) to 1 (definitely a bot account). In total, we processed 1,698,701 unique screen names. Following the procedure proposed in [30], we classified accounts as bots if the bot likelihood score assigned by Botometer was larger than 0.6. In total, we identified 9,548 bots that participated in the Twitter discourse about the riot events (see Table I).

Reconstructing the bot-focused communication network. On Twitter, a user can directly communicate with another user by using the @ symbol followed by the receiver's screen name (@screenname). For our analysis, we first excluded all retweets (a retweet typically begins with the string "RT @+author's screenname"). By following the @-traces from the remaining tweets, we reconstructed a directed communication network for each of the four events (see Section III). Apart from the source's and the target's screenname, we also preserved the bot score for each of the vertices participating in the network and the corresponding dominant emotion communicated between two Twitter users as an edge attribute. Since this paper focuses on the emotion-exchange patterns that emerge in Twitter interactions which involve bot accounts, we subset the network with respect to bot vertices. The subset we analyze in-depth thus includes bots as well as those human vertices that are adjacent to bot vertices.

Modeling the interaction network as a multiplex network. For our analysis, we regard each emotion communicated between pairs of vertices as an own edge type. Thus, we model a multiplex network that consists of eight layers, each layer corresponding to one emotion in Plutchik's wheel of emotions (anger, fear, sadness, disgust, joy, trust, anticipation, and surprise).

[9]Botometer: https://botometer.iuni.iu.edu/

Motif detection. Next, we identified the resulting network motifs by applying the following steps (see also Algorithm 1):

1) apply the ESU [31] motif enumeration algorithm which enumerates all possible subgraphs of a particular size k (in this paper $k = 3$),
2) apply VF2 isomorphism testing [32] for each pair of subgraphs,
3) isomorphism classification,
4) construction of synthetically generated random networks (null models) which resemble the input network. For the purposes of this paper, we applied the stub-matching algorithm [33], [34]. We generated 300 null models[10] for each network,
5) enumerate and classify subgraphs in all null models, and
6) identify motifs by applying the Z-score measure and p-value, as well as construct a significance profile (SP) for each network [2].

In our analyses, we then enumerate all subgraphs in:

a) each of the eight emotion layers individually,
b) two aggregated valence layers for positive and negative emotions respectively (i.e. an aggregated layer consisting of the positive emotion layers joy, anticipation, trust, as well as an aggregated layer consisting of the negative emotion layers anger, fear, sadness, disgust),
c) a valence interlayer (i.e. the edges shared between the set of common nodes in the two aggregated valence layers for positive and negative emotions),
d) an aggregated network over all eight emotion layers.

The individual layers and the derived layers are sketched in Figure 1. Note that *surprise* (the yellow layer) is handled separately, because surprise might be both, negative or positive, depending on the context.

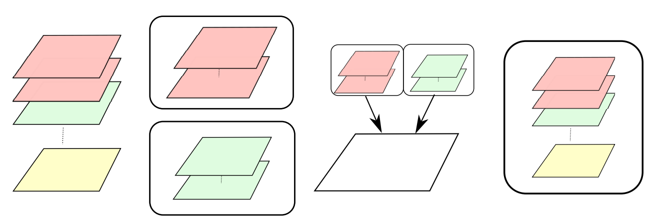


Fig. 1. Individual emotion-annotated layers and the corresponding derived layers used in our analyses (green = positive emotion layers, red = negative emotion layers, yellow = surprise).

[10]In order to determine if a network motif (i.e. a statistically significant and over-represented subgraph) has semantic meaning and results from the communication patterns in a real-world network, we have to make sure that it does not appear by chance. Thus, we generated null models and checked if the motifs found in the real-world network are also found in the corresponding null models. If the motifs appear significantly more often in the real-world network, they most likely result from the real-world communication patterns rather than from a chance process (see, e.g., [2], [3], [35]). Note that some papers argue that generating 200 null models is sufficient [35].

Algorithm 1: Motif detection.

```
Input: input_network;
Output: list_of_motifs = [];
Initialize: i = 0;
# ENUMERATE AND CLASSIFY SUBGRAPHS
def procedure: esu_vf2(list_layers)
foreach l in list_layers do
    subgraphs = esu(l)
    foreach s in subgraphs do
        subgraphs' = subgraphs \ s
        foreach s' in subgraphs' do
            if vf2(s, s') then
                assign_common_isomorphism_class
                subgraphs' = subgraphs' \ s'
                subgraphs = subgraphs \ s'
            end
        end
    end
end
end procedure
# GENERATE LAYERS AND INTER-LAYERS
detect layers in input_network
layer_negative.add_edges_from(layer_anger, layer_sadness, layer_disgust,
  layer_fear)
layer_positive.add_edges_from(layer_joy, layer_anticipation, layer_trust)
foreach i in range(length(V(input_network))) do
    if v_i ∈ V(layer_negative) & v_i ∈ V(layer_positive) then
        inter_layer.add_edges_from(layer_negative.edge_containing(v_i),
          layer_positive.edge_containing(v_i))
    end
end
list_layers = [layer_anger, layer_joy, ... , layer_surprise, layer_negative,
  layer_positive, interlayer, input_network]
esu_vf2(list_layers)
# GENERATE NULL MODELS
while i < 300 do
    foreach l ∈ list_layers do
        null[l] = matching(l.in_degree(), l.outdegree())
    end
    esu_vf2(null)
    i = i+1
end
```

In total, we generated 300 null models (synthetic random networks) corresponding to the eight emotion-annotated layers, the two valence aggregated layers, the valence interlayer, and the aggregated network for each of the four riot events. This resulted in 14,400 null models that went through the motif detection procedure. The emotion extraction as well as the motif identification procedures have been performed on a machine with Intel Xeon CPU E3-1240 v5 @ 3.5GHz (8 threads) and 32 GB RAM.

V. Results

A. Overall emotionality during riots

As riots are highly emotional events, we first analyzed the overall intensities of different emotions during the four events. As shown in Figure 2, two emotions are particularly dominant across all four riot events, namely *anger* and *fear*. We also tested for the effects of retweets (copies of a tweet disseminated by Twitter users) in our data-set and found that retweets consistently amplified *fear* across all four riot events.

Some examples of messages expressing fear and anger from our data-set:

> "*RT @screenname: Remember this? I sure as hell do. Where are the police and their riot gear now?!? #Charlottesville*",
>
> "*Now a helicopter crashed near the white supremacy Nazi march in #Charlottesville... It's complete chaos out there today.*",
>
> "*RT @screenname: #Tarragona #1Oct #Catalonia One of the injured was attacked again on his way to an ambulance (by Spanish riot cops)!*"
>
> "*RT @screenname: #NoG20 #Hamburg: Cops lost control over parts of the city.*"
>
> "*RT @screenname: Hamburg: ANTIFA tears apart a sidewalk - so they can throw it at police. #G20HAM #G20Summit.*"

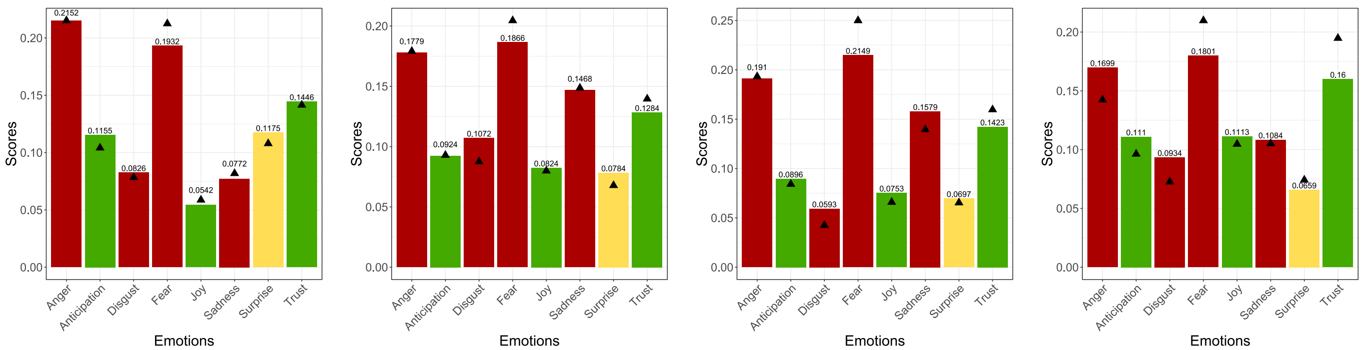
Next, we examine to which extent bots contribute to the overall perceived emotionality of the four events. We conducted Welch's two sample t-test with a 0.95 confidence level to examine whether there is any statistically significant difference in the intensities of emotions communicated by bot and human accounts. The results presented in Table II show that bots tend to disseminate messages that are emotionally more intense as compared to human accounts. This is particularly evident during the Charlottesville and Philadelphia riots.

In terms of content generation, a Twitter user may author an original tweet or disseminate copies of existing tweets (retweets). In our analysis, we found that most of the bot-generated content is attributed to retweets (Hamburg 57%, Charlottesville 65.98%, Catalonia 70.72%), with the exception of tweets related to the Philadelphia riots (38.89% retweets in related bot-generated messages). Bot-generated tweets related to the Philadelphia riots contained a substantial number of bot-authored spam messages (27%), such as: "*8 positive rules of Life <3 #PositiveVibes #Philadelphia*" or "*I am looking for a man who wants to see a responsible and serious woman beside him. Free registration #SuperBowlChamps*", news tweets (18%), as well as rumors (6%), such as "*I'm embarrassed to see a video of a man in Philly eating horse defecation (yes, for real) and a crowd of people cheer.*".

Furthermore, it is worth mentioning that in the four events we studied, similar content-generation patterns hold for human accounts who also predominantly disseminated retweets (Hamburg 79%, Charlottesville 87.53%, Catalonia 90.79%, Philadelphia 74.36%).

Next, we examine the temporal flow of aggregated positive and aggregated negative emotions communicated by bot and human accounts. Figure 3 shows the negative and positive emotions communicated by human accounts via red and green solid lines respectively, while emotions communicated by bots are depicted via red dots (negative emotions) and green triangles (positive emotions). As shown in Figure 3, negative emotion peaks are predominantly attributed to bot accounts, suggesting that bots tend to amplify the perceived negative emotionality of an event.

While previous studies suggested that bots do not tend to engage in direct interactions with the human users (see, e.g., [10]), our data-sets show that such interactions indeed happen in certain events. In particular, the Charlottesville data-set exhibits a substantial number of bots participating in a direct



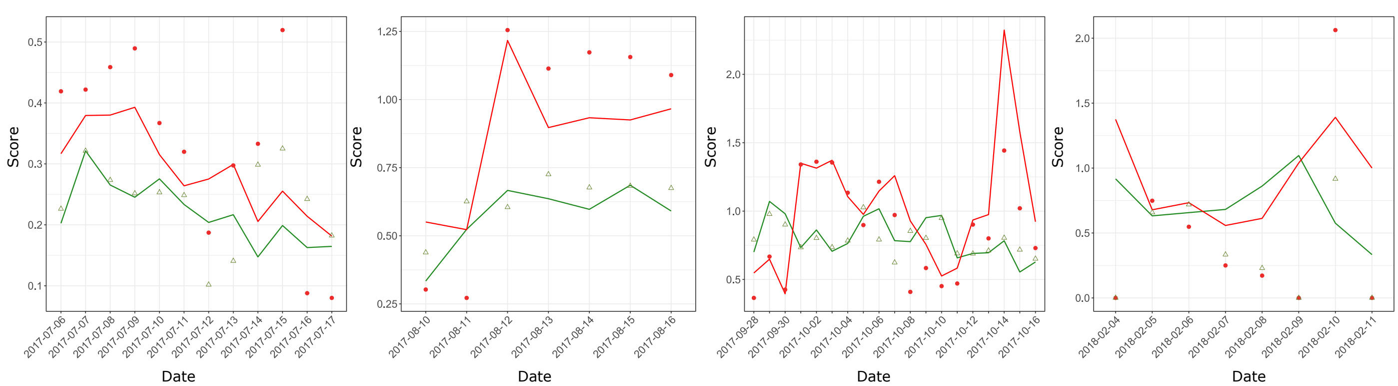
a) Hamburg b) Charlottesville c) Catalonia d) Philadelphia

Fig. 2. Presence and intensity of emotions (sorted alphabetically) during the four riot events. Positive emotions (anticipation, joy, trust) are shown in green, negative (anger, disgust, fear, sadness) in red, and surprise in yellow. The effects of retweets are shown via a black arrow-head.

	Hamburg	Charlottesville	Catalonia	Philadelphia
Anger	μ_B=**0.63**, μ_H=0.48, t=9.98*	μ_B=**1.34**, μ_H=1.1, t=12.39*	μ_B=1.44, μ_H=1.48, t=2.91*	μ_B=**1.34**, μ_H=1.1, t=12.39*
Fear	μ_B=**0.62**, μ_H=0.48, t=8.44*	μ_B=**1.38**, μ_H=1.27, t=5.78*	μ_B=1.74, μ_H=1.95, t=10.74*	μ_B=**1.38**, μ_H=1.26, t=5.78*
Sadness	μ_B=0.21, μ_H=0.22	μ_B=**1.1**, μ_H=0.92, t=9.36*	μ_B=1.12, μ_H=1.09	μ_B=**1.1**, μ_H=0.91, t=9.36*
Disgust	μ_B=0.16, μ_H=0.18	μ_B=**0.72**, μ_H=0.55, t=12.39*	μ_B=0.29, μ_H=0.32, t=4.26*	μ_B=**0.72**, μ_H=0.55, t=12.39*
Joy	μ_B=0.17, μ_H=0.18	μ_B=0.51, μ_H=0.49	μ_B=**0.53**, μ_H=0.49, t=3.87*	μ_B=0.51, μ_H=0.5
Trust	μ_B=0.37, μ_H=0.35	μ_B=**0.86**, μ_H=0.81, t=4.84*	μ_B=1.1, μ_H=1.19, t=7.83*	μ_B=**0.86**, μ_H=0.81, t=4.84*
Anticipation	μ_B=0.24, μ_H=0.25	μ_B=**0.59**, μ_H=0.56, t=3.14*	μ_B=**0.64**, μ_H=0.62, t=2.86*	μ_B=**0.59**, μ_H=0.56, t=3.14*
Surprise	μ_B=0.23, μ_H=0.26, t=2.57*	μ_B=**0.5**, μ_H=0.42, t=8.63*	μ_B=0.56, μ_H=0.55	μ_B=**0.5**, μ_H=0.42, t=8.63*

TABLE II

EMOTION INTENSITIES COMMUNICATED BY BOT AND HUMAN ACCOUNTS (μ_B REPRESENTS A MEAN EMOTION SCORE COMMUNICATED BY BOT ACCOUNTS, μ_H REPRESENTS A MEAN EMOTION SCORE COMMUNICATED BY HUMAN ACCOUNTS). RESULTS OF THE T-TEST ARE SHOWN FOR THE CONFIDENCE LEVEL *0.95.



a) Hamburg b) Charlottesville c) Catalonia d) Philadelphia

Fig. 3. Temporal flow of positive (green) and negative (red) emotions during the four riot events (emotions spread by human accounts are presented via solid lines, negative emotions spread by bots via red dots, and positive emotions spread by bots via green triangles).

message exchange with human users (87.18%), while the remaining data-sets count between 2.33%-10.07% of bots that engaged in a direct communication with human users. Given such a high involvement of bots in a direct communication with human accounts, the question remains which structural patterns emerge due to the bots' engagement in direct message exchanges.

B. Emotion exchange motifs

For motif identification, we reconstructed a directed communication network by following the @-traces in the tweets (excluding retweets) for each event. To examine the role of bots in the corresponding communication network, we extracted a subnetwork which contained all participating bot vertices as well as their adjacent human accounts. As summarized in Table I, the bot-focused communication network reconstructed from the Philadelphia riot data-set resulted in a network consisting of only three edges. For the remainder of our analysis, we therefore focus on the remaining three data-sets.

As shown in Figure 4, the emotions exchanged between bot and human accounts are comparable across the three events. In particular, we observe a dominance of anger, fear, and anticipation across all three riot events. To measure the similarity between the networks with respect to the communicated emotions, we use Kendall's τ rank correlation

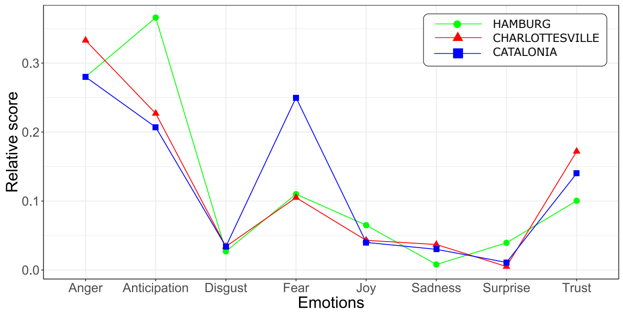


Fig. 4. Emotions communicated in the bot-focused communication networks

coefficient. We found that the communication network reconstructed from the Hamburg data-set strongly correlates with the one reconstructed from Charlottesville and Catalonia data-sets ($\tau_{H-CHA} = 0.71$, $\tau_{H-CAT} = 0.71$), while the correlation drops to a moderate $\tau_{CAT-CHA} = 0.57$ between the communication networks resulting from the Catalonia and Charlottesville events.

After finding that there is a positive correlation regarding the emotions communicated during the three riot events, we next examine whether the corresponding communication networks are also similar on a structural level.

To this end, we enumerated all 3-node subgraphs and analyzed them for their significance in the input network with respect to the null models. As a result, we identified in total 38 different isomorphic classes that are considered motifs (25 in the Charlottesville data-set, 18 in the Catalonia data-set, and 7 in the Hamburg data-set). Of the 38 classes, ten appear in at least two data-sets and are thus considered common motifs in the bot-focused communication networks (Table III).

When comparing multiple networks with respect to their motifs, a *motif significance profile* (SP) is used to normalize the Z-scores of a set of motifs $m_1, m_2, ..., m_k$ [2]:

$$SP(m_i) = \frac{Z(m_i)}{\sqrt{\sum_{i=1}^{n} Z(m_i)^2}}.$$

The significance profile of the motifs identified in the three bot-focused communication networks is shown in Figure 5 (the layers in which a particular motif has been identified are color-coded and label-annotated).

When disregarding the edge density in the common motifs, the motifs can be classified as either in-star, out-star, or chain triadic patterns. While other studies pointed to the presence of additional triadic patterns, such as transitive triads [36] or triads with reciprocal edges, no such common patterns have been found in our analysis.

The common motifs we found predominantly take over a configuration of various star-like configurations (see Table III). In particular, these are triadic in-stars (IDs C1-C3, C5-C7) and out-stars (IDs C4 and C10), whereby the out-star motifs are denser than the in-star motifs (on average 4.83 in-star edges and 6.5 out-star edges). Moreover, we identified two chain triadic motifs (the motifs with the IDs C8 and C9 in Table III).

When comparing the role of bots in the motifs, we observed a predominantly higher out-degree (*message sending* behavior) of the bot nodes (average out-degree 3, while human nodes exhibit an average out-degree of 1.9).

With respect to the different emotion layers, the motifs were especially identified on the valence-interlayer, signaling that when communicating with a bot, characteristic patterns emerge as accounts exchange disparate emotions belonging to both positive and negative valence.

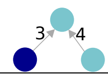
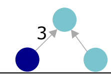
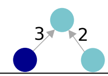
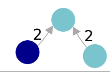
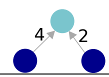
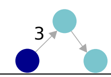
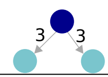
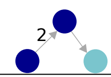
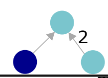
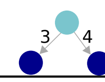
ID	Common motifs	ID	Common motifs
C1		C6	
C2		C7	
C3		C8	
C4		C9	
C5		C10	

TABLE III
COMMON MOTIFS IDENTIFIED IN BOT-FOCUSED COMMUNICATION NETWORKS. DARK BLUE NODES DEPICT THE POSITION OF A BOT IN A MOTIF ($> 50\%$). ALL MOTIFS IDENTIFIED FOR $p < 0.05$.

In addition to the common motifs, Table IV shows characteristic event-specific motifs, i.e. motifs that emerge only in one specific riot event. The role of bots is again depicted in dark blue, showing that bots regularly form self-loops (mention themselves). Such motifs (ID S1) were found on the trust layer ($SP = 0.00004$) and interlayer ($SP = 0.00005$) in the Charlottesville data-set. Motif S2 shows a bot engaging in a reciprocal message exchange with a human user (found in the Catalonia data-set, aggregated layer ($SP = 0.013$)). Moreover, bots may form transitive triads (motif ID S3 found in the interlayer ($SP = 0.003$) of the Hamburg data-set and ID S4 found in the Charlottesville data-set on the aggregated layer ($SP = 0.001$) and negative layer ($SP = 0.0001$)).

VI. Discussion

By studying bot behavior during four riot events, we found that bots considerably contribute to the overall perceived emotionality of an event. In particular, while studying the temporal evolution of emotions, we found that bots especially amplify negative emotions. While such an augmentation of negativity can result from retweeting messages that convey negative emotions, bots in our data-sets actually authored and disseminated original content. Though a substantial portion of such content is spam, we found empirical evidence of bots also spreading rumors. This observation is in line with the existing

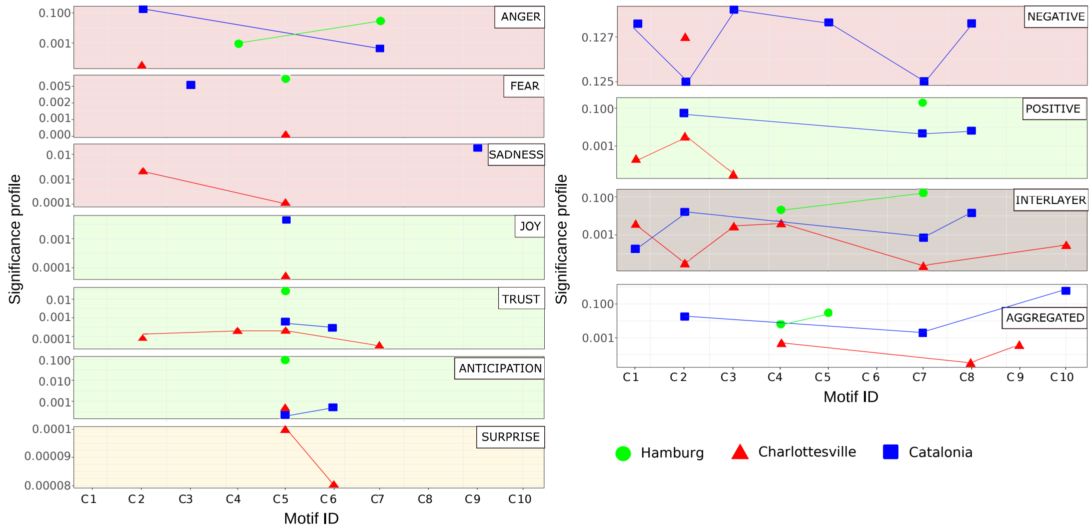


Fig. 5. Motif significance profile for the communication networks).

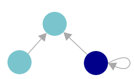
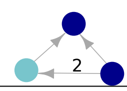
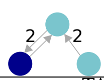
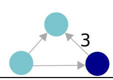
ID	Specific motifs	ID	Specific motifs
S1		S3	
S2		S4	

TABLE IV
EVENT-SPECIFIC MOTIFS IDENTIFIED IN BOT-FOCUSED COMMUNICATION NETWORKS. DARK BLUE NODES DEPICT THE POSITION OF A BOT IN A MOTIF ($> 50\%$). ALL MOTIFS IDENTIFIED FOR $p < 0.05$.

literature. For example, [7] suggested that bots are prone to spreading misinformation due to their inability to assess the credibility of the corresponding sources. For example, we found messages such as "*Philadelphia police looking for help identifying rioters. I know two of them! Guy on right, 17 is NAME*" or "*@screenname NAME's son marches with the protesters! #G20 #g20ham17*"[11], which may potentially harm the reputation of the publicly named person.

In our data-sets we further found that on occasion bots actively engage in a one-to-one message exchange with other Twitter users. In our Charlottesville data-set, the number of such proactive bots is as high as 87.18%, revealing that bots in the data-set predominantly interacted with human users.

Upon dissecting the bot-focused communication network into triadic patterns, we identified various motifs (statistically significant and representative patterns). In particular, the motifs we found reveal the tendency of bots to either flood human users with messages (motif IDs C1-C4, C6-C8, S4), mention themselves (ID S1), or participate in a messaging chain (ID C8, C9), thus potentially flooding human users with irrelevant spam, influencing human users by occupying their attention, or spreading rumors.

[11]Please note that we anonymized the tweets by replacing the actual names mentioned in the tweets with NAME.

VII. Conclusion

As online social networks (OSNs) have become a vital part of our society, it also becomes more important to understand OSN-based communication. Aside from numerous positive aspects, OSNs also have potentially negative effects on their users, ranging from manipulation and voter's opinion swaying during elections to malicious spreading of misinformation (see, e.g., [17], [37]). Among the culprits responsible for these negative effects are automated accounts (social bots). Previous studies have predominantly focused on various aspects distinguishing bots from human users in order to successfully detect bots, while other studies provided empirical evidence of the negative influence of such automated accounts.

In contrast to related studies which suggested that bots tend to be passive, we found that in certain events bots engage in a direct communication with human accounts. Our analysis is based on four highly emotional events, namely the 2017 G20 riot in Hamburg, the 2017 Charlottesville riot, the 2017 Catalonia referendum riot, and the 2018 Superbowl riot in Philadelphia.

To date, there is a lack of studies that investigate structural patterns emerging from the direct communication of bots and human users. Because emotions are one of the most prominent drivers of human interactions [38], we focused our analysis on the structural patterns that emerge as bots communicate emotions with humans. In particular, we proposed a multiplex network consisting of eight layers, corresponding to Plutchik's eight basic emotions. In our motif detection procedure, we investigated eight individual emotion layers (anger, disgust, fear, sadness, joy, trust, anticipation, surprise), two aggregated valence-based layers (positive and negative), one interlayer network, as well as an overall aggregated network.

Our findings suggest that as bots communicate emotions to human users, specific significant triadic subgraphs (motifs) emerge. The motifs we found include predominantly star-like structures, with a presence of self-loops. Interestingly, we found that motifs especially emerge on the interlayer between the positive and negative valence layer networks, meaning that significant triadic patterns are formed when bots and humans exchange mixed emotions.

In our future work, we plan to further study the emergence of motifs in OSNs and focus specifically on the motifs in emotion-annotated multilayer networks.

References

[1] M. Newman, *Networks: An Introduction*. Oxford University Press, 2010.

[2] K. Erciyes, *Complex Networks: An Algorithmic Perspective*. CRC Press, Inc., 2014.

[3] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, "Network motifs: Simple building blocks of complex networks," *Science*, vol. 298, no. 5594, pp. 824–827, 2002.

[4] E. Yeger-Lotem, S. Sattath, N. Kashtan, S. Itzkovitz, R. Milo, R. Y. Pinter, U. Alon, and H. Margalit, "Network motifs in integrated cellular networks of transcription–regulation and protein–protein interaction," *Proc. of the National Academy of Sciences*, vol. 101, no. 16, pp. 5934–5939, 2004.

[5] L. A., P. S.P., I. R., and M. A., "Functions of bifans in context of multiple regulatory motifs in signaling networks," *Biophysical Journal*, vol. 94, no. 7, pp. 2566–2579, 2008.

[6] N. T. L. Tran, L. DeLuccia, A. F. McDonald, and C.-H. Huang, "Cross-disciplinary detection and analysis of network motifs," *Bioinformatics and Biology Insights*, vol. 9, p. BBI.S23619, 2015.

[7] A. Gupta, H. Lamba, and P. Kumaraguru, "$1.00 per RT #BostonMarathon #PrayForBoston: Analyzing fake content on Twitter," *APWG eCrime Researchers Summit*, pp. 1–12, 2013.

[8] S. M. Mohammad, X. Zhu, S. Kiritchenko, and J. Martin, "Sentiment, emotion, purpose, and style in electoral tweets," *Information Processing and Management*, vol. 51, no. 4, pp. 480 – 499, 2015.

[9] H. Stefanie, B. T. D., H. Kim, T. Andrew, S. C. R., and L. Vincent, "Tweets as impact indicators: Examining the implications of automated bot accounts on Twitter," *Journal of the Association for Information Science and Technology*, vol. 67, no. 1, pp. 232–238, 2016.

[10] F. Riquelme and P. Gonzlez-Cantergiani, "Measuring user influence on Twitter: A survey," *Information Processing and Management*, vol. 52, no. 5, pp. 949 – 975, 2016.

[11] N. Abokhodair, D. Yoo, and D. W. McDonald, "Dissecting a social botnet: Growth, content and influence in Twitter," in *Proc.of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 2015, pp. 839–851.

[12] J. P. Dickerson, V. Kagan, and V. S. Subrahmanian, "Using sentiment to detect bots on Twitter: Are humans more opinionated than bots?" in *Proc. of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2014, pp. 620–627.

[13] E. Ferrara, "Disinformation and social bot operations in the run up to the 2017 French presidential election," *First Monday*, vol. 22, no. 8, pp. 1–33, 2017.

[14] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, A. Flammini, and F. Menczer, "Detecting and tracking political abuse in social media," in *Proc. 5th International AAAI Conference on Weblogs and Social Media*, 2011, pp. 297–304.

[15] N. Kollanyi, P. Howard, and S. Woolley, "Bots and automation over Twitter during the U.S. election," *Data Memo. Oxford, UK: Project on Computational Propaganda*, vol. 2016, no. 4, pp. 1–5, 2016.

[16] E. Kušen and M. Strembeck, "Why so emotional? An analysis of emotional bot-generated content on Twitter," in *Proc. of the 3rd International Conference on Complexity, Future Information Systems and Risk (COMPLEXIS 2018)*, 2018, pp. 13–22.

[17] E. Ferrara, O. Varol, C. Davis, F. Menczer, and A. Flammini, "The Rise of Social Bots," *Communications of the ACM (CACM)*, vol. 59, no. 7, 2016.

[18] R. M. Everett, J. R. C. Nurse, and A. Erola, "The anatomy of online deception: What makes automated text convincing?" in *Proc. of the 31st Annual ACM Symposium on Applied Computing*, 2016, pp. 1115–1120.

[19] U. Yaqub, S. A. Chun, V. Atluri, and J. Vaidya, "Analysis of political discourse on twitter in the context of the 2016 US presidential elections," *Government Information Quarterly*, vol. 34, no. 4, pp. 613 – 626, 2017.

[20] J. Xie, C. Zhang, M. Wu, and Y. Huang, "Influence inflation in online social networks," in *Proc. of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE Press, 2014, pp. 435–442.

[21] R. Plutchik, "The nature of emotions," *American Scientist*, vol. 89, no. 4, 2001.

[22] E. Kušen, G. Cascavilla, K. Figl, M. Conti, and M. Strembeck, "Identifying Emotions in Social Media: Comparison of Word-emotion Lexicons," in *Proc. of the 4th International Symposium on Social Networks Analysis, Management and Security (SNAMS)*. IEEE, August 2017.

[23] L. K. Hansen, A. Arvidsson, F. A. Nielsen, E. Colleoni, and M. Etter, "Good friends, bad news - affect and virality in twitter," in *Future Information Technology*, J. J. Park, L. T. Yang, and C. Lee, Eds. Springer Berlin Heidelberg, 2011, pp. 34–43.

[24] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Computational Linguistics*, vol. 37, no. 2, pp. 267–307, June 2011.

[25] C. Hutto and E. Gilbert, "VADER: A parsimonious rule-based model for sentiment analysis of social media text," in *International AAAI Conference on Web and Social Media*, 2014.

[26] F. Benamara, C. Cesarano, A. Picariello, D. Reforgiato, and V. S. Subrahmanian, "Sentiment analysis: Adjectives and adverbs are better than adjectives alone," in *Proc. of the International Conference on Weblogs and Social Media (ICWSM)*, 2007.

[27] M. Wiegand, A. Balahur, B. Roth, D. Klakow, and A. Montoyo, "A survey on the role of negation in sentiment analysis," in *Proc. of the Workshop on Negation and Speculation in Natural Language Processing*, 2010, pp. 60–68.

[28] R. Meo and E. Sulis, "Processing affect in social media: A comparison of methods to distinguish emotions in Tweets," *ACM Transactions on Internet Technologies*, 2017.

[29] C. A. Davis, O. Varol, E. Ferrara, A. Flammini, and F. Menczer, "Botornot: A system to evaluate social bots," in *Proc. of the 25th International Conference Companion on World Wide Web*, 2016, pp. 273–274.

[30] O. Varol, E. Ferrara, C. Davis, F. Menczer, and A. Flammini, "Online human-bot interactions: Detection, estimation, and characterization," in *In International AAAI Conference on Web and Social Media*, 2017, p. 280289.

[31] S. Wernicke, "Efficient detection of network motifs," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 3, no. 4, pp. 347–359, Oct. 2006.

[32] L. P. Cordella, P. Foggia, C. Sansone, and M. Vento, "A (sub)graph isomorphism algorithm for matching large graphs," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 10, pp. 1367–1372, Oct 2004.

[33] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Random graphs with arbitrary degree distributions and their applications," *Physics Review E*, vol. 64, no. 2, p. 026118, Jul. 2001.

[34] E. A. Bender and E. Canfield, "The asymptotic number of labeled graphs with given degree sequences," *Journal of Combinatorial Theory, Series A*, vol. 24, no. 3, pp. 296 – 307, 1978.

[35] W. E. Schlauch, K. A. Zweig, G. Theory, and N. Analysis, "Influence of the null-model on motif detection," in *Proc. of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2015, pp. 514–519.

[36] A. Topirceanu, A. Duma, and M. Udrescu, "Uncovering the fingerprint of online social networks using a network motif based approach," *Computer Communications*, vol. 73, pp. 167 – 175, 2016.

[37] V. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini, and F. Menczer, "The DARPA Twitter Bot Challenge," *IEEE Computer*, vol. 49, no. 6, 2016.

[38] E. Kušen, M. Strembeck, G. Cascavilla, and M. Conti, "On the Influence of Emotional Valence Shifts on the Spread of Information in Social Networks," in *Proc. of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, July/August 2017.