Drivers of Polarized Discussions on Twitter during Venezuela Political Crisis

SAMEERA HORAWALAVITHANA, Computer Science and Engineering, University of South Florida, USA
KIN WAI NG, Computer Science and Engineering, University of South Florida, USA
ADRIANA IAMNITCHI, Computer Science and Engineering, University of South Florida, USA

Social media activity is driven by real-world events (natural disasters, political unrest, etc.) and by processes within the platform itself (viral content, posts by influentials, etc). Understanding how these different factors affect social media conversations in polarized communities has practical implications, from identifying polarizing users to designing content promotion algorithms that alleviate polarization. Based on two datasets that record real-world events (ACLED and GDELT), we investigate how exogenous events drive related Twitter activity in the highly polarizing context of the Venezuela's political crisis from early 2019. Our findings show that users who tweet in different languages show different emotions and reactions to events; antagonistic communities react differently to different exogenous sources; influential users not only have an impact on the structure of the communities but also on polarization; and different events lead to different changes in network polarization.

ACM Reference Format:
Sameera Horawalavithana, Kin Wai NG, and Adriana Iamnitchi. 2021. Drivers of Polarized Discussions on Twitter during Venezuela Political Crisis. In *WebSci '21: The 13th International ACM Conference on Web Science, June 21–25, 2021, Southampton, UK*. ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Understanding if social media activity is responding to triggers from real-life events or is the result of in-platform discussions (such as information campaigns) has many applications. For example, detecting information campaigns that are often coordinated within a platform and are only loosely connected to real-life events [9] can have impacts from healthcare [18] to political unrest [1]. Another application could be developing simulators of social media activity that realistically incorporate reactions to events from the physical world. For example, seismic events were "seen" in Twitter activity [15], thus a realistic simulator of Twitter activity should react to relevant exogenous events.

This paper addresses the question of how external events correlate with related Twitter activity. To this end, we chose the Venezuelan political crisis of the early 2019 which involved mass protests and international political responses that were recorded in the news media. Because of the international attention, Twitter activity related to this episode spanned multiple languages. The politically polarized Venezuelan society [25] responded in different ways to the same event, depending on political ideology.

In this work, we ask the following questions: How does the Twitter community who engages with Venezuela-related topics during early 2019 react to real world events? Are there differences in their reactions if they are local (i.e., Spanish-tweeting) or international (English-tweeting)? How do the two ideologically opposing sides react to the same events? Are the influential users on Twitter increasing the polarization of the community? Which are the topics that affect polarization?

Via empirical analysis on three datasets (from Twitter, ACLED and GDELT) we show that users who tweet in different languages show different emotions and reactions to events; antagonistic communities react differently to different exogenous sources; influential users not only have an impact on the structure of the communities but also on polarization; and different events lead to different changes in network polarization.

2 RELATED WORK

Two main mechanisms were identified as the driving forces behind the information diffusion in social media discussions [29]. First, users share information because of the actions of their friends who share similar social identity, trust, etc. This is known as social influence. Second, exogenous factors that exist outside of the social media platforms can influence the online users to participate in social media discussions.

Previous studies analyzed the effect of internal factors driving the discussions on Twitter. Morales et al. [19] show that a minority of opinion leaders (e.g., political elites) tend to capitalize public attention in online discussions, and thus contribute to highly polarized conversations. Ozer et al. [22] studied the effect of bot accounts on online polarization over the US gun control debate. Their findings suggest that automated accounts contribute to more polarized online discussions related to US mass shooting events. Soares et al. [28] investigated the effect of opinion leaders and activists on shaping polarized discussions on social media within the Brazilian political context. The authors leverage in-degree centrality to identify influentials, and qualitatively found that politicians, journalists and media outlets are among the most popular accounts driving the debate.

Few studies show how Twitter users react to external events. It is widely known that Twitter reacts fast to offline events [16, 30] in discussions that are either responding to a message posted by an opinion leader (e.g., politician, media), or sharing information about offline events. For example, Myers et al. [21] discover that 29% of the URL mentions on Twitter are due to external events, while the rest of the URL mentions are driven internally. These offline events affect the diversity of actors participating in online discussions and fostering interactions between actors and communities of diverse backgrounds [4]. Freelon et al. [10] examined tweets regarding #BlackLivesMatter and found that the network composition was shaped at different phases of the movement by reactions from media and offline events.

3 DATA FROM THE POLITICAL CRISIS IN VENEZUELA

For the last two decades, Venezuela has experienced a pervasive sociopolitical fragmentation fueled by differences of interests, identities, and politics. In Venezuela, the political spectrum is for the most part divided into two ideologies: Chavism, embraced by those who support the political ideology of the late president Hugo Chavez, and Anti-Chavism, embraced by those who are strongly opposed to Chavez's legacy. Today, Chavism still maintains control of the Venezuelan political system, with Nicolas Maduro as the head of the state. However, failure to manage globalization, lack of investment in infrastructure, and a poor administration has put the country in the grip of a significant economic collapse. As a result, it has contributed to a significant rise in crime and violence, lack of essentials, shortages of medicines and food, and an unprecedented humanitarian crisis [7].

3.1 Political Crisis Events in Early 2019

The 2019 Venezuelan political crisis has its roots in the controversial re-election of Nicolas Maduro as the country's president on January 10th. This event marked the beginning of a presidential crisis driven by claims of illegitimacy and reports of coercion and fraud. During the following days, the opposition-controlled National Assembly widely denounced the re-election as fraudulent, and mandated an order of succession. On January 23, the opposition leader, Juan Guaidó, declared himself interim president of Venezuela in an effort to restore democracy and constitutional rights. The event erupted widespread protests to put pressure on Maduro's administration to resign from office, and it formed a coalition of countries in support of Guaidó. In response, Maduro's government ordered the armed forces into the streets to maintain social order and disperse mass protests. These intense and violent clashes between the military and opposition supporters continued during the first couple of weeks of February, and resulted in massive lootings, a large number of detentions, and dozens of injured.

In early February, Guaidó announced a plan to bring international humanitarian aid into Venezuela on February 23. Maduro rejected the international aid offers and ordered the closure of the Brazilian and Colombian borders to impede humanitarian aid delivery. A day before the international aid delivery, two dueling concerts took place simultaneously at the Colombia-Venezuela border. The "Aid Live" concert was organised with the purpose of raising money and support for the international humanitarian aid effort. In response, Maduro's government organised the "Hands Off Venezuela" concert with the goal of rejecting aid efforts by counteracting the rival concert. On February 23, the plan to bring the humanitarian aid into Venezuela was met with a violent standoff between military forces and those accompanying the aid and protesting against the regime. Clashes continued over the next couple of days, and eventually it was reported that none of the aid shipments were able to enter the country.

On March 8, the crisis intensified as a major nationwide blackout struck Venezuela, leaving most parts of the country in the dark over a period of five days. The blackout affected the daily routines of millions of Venezuelans as businesses shut down, public transportation was out of service, water supply was unavailable, and hospital operations were disrupted. This situation caused widespread protests all over the nation against the recurring blackouts that have aggravated the country's social crisis, and especially criticizing Maduro over his poor administration.

On March 25, Russian aircrafts were seen arriving at the Caracas airport guarded by the Venezuelan military. The Venezuelan government claimed that the visit was part of a supposedly joint military exercise and cooperation between the two nation allies. However, Guaidó and his supporters denounced the act as an illegal foreign military intervention seeking to threaten the national security of the country. Over the next couple of months, protests continued to run rampant across the country. Nonetheless, the opposition side started to lose momentum towards the end of the year, and attendance at public rallies eventually dropped. Figure 1 presents the timeline of these political events.

3.2 External Event Data

The events that unfolded during the 2019 Venezuelan political crisis have been analyzed and documented in various magazines, newspapers, and conflict databases. In this study, we rely on two such data sources, ACLED and GDELT, that collect information on the Venezuelan political conflict based on the reviews of human coders and machines.

Armed Conflict Location and Event Dataset (ACLED) [24] is collected primarily from local and regional news sources, Integrated Regional Information Network (IRIN), Relief Web, Factiva, and humanitarian agencies. The database is known to be highly curated, which often includes manually verified data [23]. ACLED typically reports violent (e.g., protests, riots, political repressions) and non-violent events (e.g., strategic developments) carried out by political agents, including government officials, rebels, and militias [23]. There are 1,485 ACLED events related to Venezuela's early 2019 conflicts in our dataset, grouped into six different event types (as shown in Figure 2a). "Protests" and "violence against civilians" events are the most popular. Due to its high accuracy in reporting real events, we used ACLED to drive the selection of similar topics in Twitter and GDELT.

In addition, we used a publicly available geopolitical event database, GDELT, to extract information as recorded in news articles [17]. The database consists of machine-coded events extracted from news reports on a variety of news sources. GDELT database is updated at every 15 minutes. We queried the database using the "Venezuela" search term for events recorded between December 24th, 2018 and April 1st, 2019. This query returned a total of 624,254 GDELT events in 20 event types, as shown in Figure 2b. Each GDELT event records the interaction between two actors defined by a CAMEO code book [26].

While GDELT relies on similar information sources as ACLED, there are differences regarding the events detected by each source [12]. For example, GDELT might pick events that could be missed by a human annotator. In addition, the GDELT database does not use background information to verify the accuracy of the automatically detected events. In this study, we selected GDELT events that are likely to be associated with event types also found in ACLED. In particular, we filtered our GDELT data based on seven CAMEO event types, namely Protest, Fight, Mass Violence, Coerce, Assault, Threaten, and Exhibit Force Posture.

3.3 Twitter Data

In the Venezuelan political crisis, Twitter has enabled the online public to stay informed on the latest real-world events, and allowed politicians to effectively mobilize protesters [20]. While Venezuelan authorities have regularly engaged in censoring news outlets and digital media platforms, Twitter has become an alternative platform to increase the awareness of on-the-ground conflict in both regional and international communities[1].

Table 1. Keywords used for Twitter data collection.

#23Ene, #23Feb, 23 de Enero, 23 de Febrero, Aid Venezuela, #BravoPueblo, Caracas, Chacao, Maturin, Maracaibo, #Chavismo, #Chavistas, FANB, #FreeVenezuela, #FueraDictadura, Fuerza Venezuela, GNB, #GritemosConBrio, #HandsOffVenezuela, #GuaidoPresidente, #JGuaido, Juan Guaido, Maduro, #LasCallesSonDelChavismo, Leales siempre traidores nunca, Guaido, Libertad para Venezuela, Freedom for Venezuela, #VamosBien, #MaduroDictador, #MaduroUsurpador, Nicolas Maduro, #SOSVenezuela, #VenezolanosEnElMundo, Venezuela Aid Live, #WeAreMaduro, Yankee go Home, #FebreroRebelde, #NoMasDictadura, #AbajoCadenas Venezuela Crisis Humanitaria, Maduro Ilegitimo

The Twitter data was collected over a period of three months (December 24th to April 1st, 2019) using GNIP, a data collection API tool, and based on a list of keywords (shown in Table 1) relevant to the Venezuelan political crisis. We conducted a thorough exploration of our dataset corpus in order to identify the most representative topics originating from online social media discussions. We worked alongside three research collaborators who are subject-matter experts regarding the Venezuelan political context. The annotators are fluent in both English and Spanish, and also familiarized with particular jargon and specialized terms commonly used in Venezuela. They identified 10 top-level topic groups: "Guaidó", "Assembly", "Maduro", "protests", "arrests", "violence", "international", "military", "crisis", and "other". Because it is not feasible to manually label millions of messages, in order to automate the annotation process, we conducted a semi-supervised classification task consisting of two steps: (1) manually annotating an initial subset of messages, and (2) training a multilingual BERT model to classify each message with one or multiple topics.

The manual annotation process was conducted over a corpus of 11,218 messages, and consisted in a 8 to 1 ratio of single-annotator annotations to all-annotator annotations. These ten topics reported inter-annotator agreement scores of 0.64 for the weighted average Cohen's Kappa, and 0.7 for the Fleiss' Kappa measurement. After manual annotation, a BERT model was trained for topic annotation for multilingual text classification tasks [6]. The BERT model was trained on 10,097 distinct text documents and evaluated on a 10% test set (1,121 texts). Stratified sampling was used to ensure

[1]https://time.com/5571504/venezuela-internet-press-freedom/

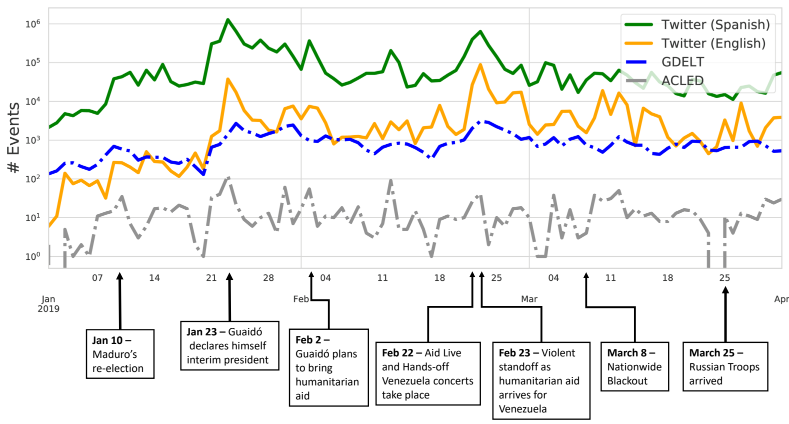
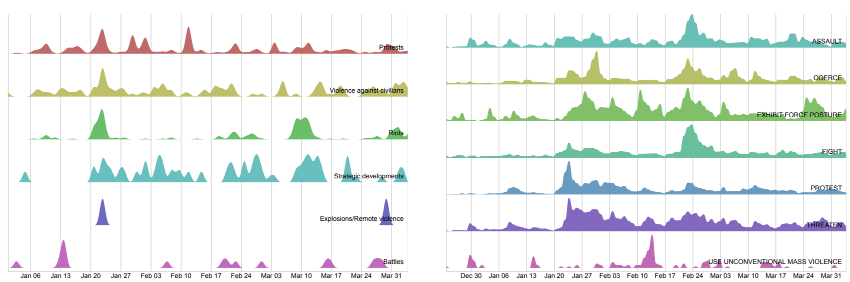


Fig. 1. Daily volume of Twitter messages, GDELT, and ACLED events. For Spanish messages, the correlation coefficient is 0.56 and 0.65 with GDELT and ACLED, respectively. For English messages, the correlation is 0.61 and 0.4 with GDELT and ACLED, respectively.



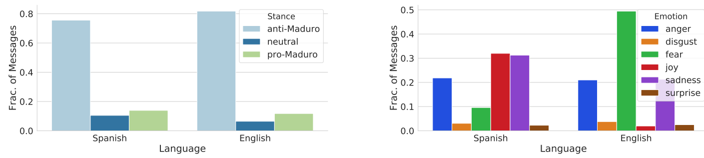
(a) ACLED (b) GDELT

Fig. 2. Timeseries of ACLED and GDELT events in various conflict event types

that the train and test sets have approximately the same percentage of samples of each topic class as in the original manually annotated corpus. The model obtained a precision of 67%, recall of 66%, and F1 score of 66%.

For this study, we focused on a subset of topics that directly relate to political conflict events and that are represented in ACLED and GDELT. This subset includes "protests", "arrests", "violence", and "military" topics. The resulting dataset consists of 723,883 seed messages including tweets, replies and quotes done by 125,778 users, and 8,278,527 retweets by 690,960 users. The majority of messages are in Spanish (86%) and English (6%). Each Twitter record in our dataset contains the following information: an assigned unique identifier, the unique (anonymized) ID of the user who posted it, the timestamp of the message, the respective content of the message and its type (whether a tweet, quote, reply or retweet), and the Twitter profile descriptions of each user.

In order to identify message stances, we manually annotate an initial corpus of 11,218 messages with anti-Maduro, pro-Maduro and neutral positions. Particularly, each annotator was asked to classify messages as part of the anti-Maduro group if (1) they express disapproval of Nicolas Maduro, his administration, or his actions, or (2) if they express approval of Guaido, his administration, or his actions. The reverse is true for the pro-Maduro stance group. Moreover, messages



(a) (b)

Fig. 3. The number of Twitter messages grouped by different stances, and different emotions.

which do not take a stance in the political crisis were assigned a neutral stance. Lastly, we trained a multilingual BERT model to classify each message with one of three valid stance groups. The BERT model was trained on 10,097 unique text documents and evaluated on a 10% test set (1,121 texts). For this task the model obtained a precision of 81%, recall of 81%, and F1 score of 81%. Overall, we found that the vast majority of messages in both Spanish and English were classified as anti-Maduro (as shown in Figure 3a). This observation suggest that there is a significant community imbalance in Twitter with respect to online discussions about the Venezuelan political crisis. This is aligned with a recent survey, which showed that only a third of Venezuelans support the present government in power [25].

In addition, we analyzed emotions in the social media messages. We used SEL [27] and Empath [8] to count the words associated with different emotional categories in the Spanish, and English messages, respectively. SEL contains 2,036 Spanish words dictionary associated with sadness, joy, fear, anger, repulsion, and surprise emotion categories. Empath is a tool that uses deep learning and word embedding to build a semantically meaningful lexical categories. We selected six Empath categories (sadness, joy, fear, anger, repulsion, surprise) to cover the emotional words used in English messages. As shown in Figure 3b, both positive (e.g., joy) and negative (e.g., sadness) messages were more prevalent in the Spanish community. However, the messages from the English speaking community reveal fear.

4 INTERNAL DRIVERS

In order to identify the internal drivers of discussions in our Twitter dataset, we first ranked all users based on their spread score [2]. The spread score for user u is the product of the fraction of the number of tweets posted by u that get retweeted and the total number of retweets that user u gets for his tweets. Intuitively, the spread score captures the level of influence of a user: the higher the spread score, the more influential the user is.

We selected the top-200 influential users ranked by the spread score and analyzed their Twitter profile descriptions, account type (i.e., verified user or not), and most recent messages posted on the platform. Our main objective is to quantify the impact of these influential accounts on the online discussions during the Venezuelan crisis. To this end, we first classify each account into one of three classes: (1) political accounts, (2) media (e.g., news media outlets, journalists), and (3) other (those that are not mapped to any of the previous classes). This effort resulted in the identification of 48 political accounts (23 anti-Maduro and 25 pro-Maduro) and 60 media accounts. Second, we constructed the user-to-user interaction network for each of the two most active periods of the crisis: Guaidó president (January 23) and International aid (February 23). For this, we considered a five-day interval around each event—two days before the event, the day of the event, and two days after the event—and selected only the messages posted during this interval. Overall, we found that the "Guaidó president" episode generally sparked discussions related to *protests* and *violence* in Twitter (40% and

24%, respectively, out of the total activity). On the other hand, the *military* (51%) and *violence* (34%) topics were the most popular during the "International Aid" episode.

In order to investigate what is the role of these accounts in online discussions from a network perspective, we compute the clustering coefficient within each stance group and under certain what-if scenarios. Specifically, we look at how the structure of the networks changes when removing either political accounts or media accounts, especially in comparison with random node removal. Random nodes were selected based on their similarity to both political and media influential accounts with respect to their average number of activities. We ran the selection of random nodes for removal 10 times and report the average results.

Figure 4 presents the results of our experiments. We observe similar patterns in the two episodes. First, we found that the clustering coefficient in the anti-Maduro community decreases with the removal of media accounts. This suggests that media accounts tend to bind the anti-Maduro community to a greater degree than political accounts. On the other hand, the clustering coefficient of the pro-Maduro community decreases more when political accounts are removed from the network than when media accounts are removed, which suggests that political accounts tend to bind the pro-Maduro community more than the media outlets. Second, we show that these networks do not experience a significant change in clustering coefficient when a set of nodes are randomly removed. Thus, the changes on the network structure inflicted by the removal of political or media accounts are due to account status rather than to the activity of the account in the 5-day intervals of our investigation. This highlights the importance of these political and media accounts in their role as drivers of online discussions during the Venezuelan crisis.

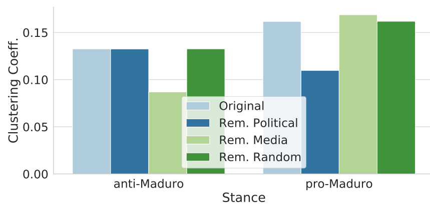
Another aspect to investigate is the impact of these political and media influential accounts on network polarization. We use the *Random Walk Controversy* (RWC) score [11], a network polarization measure based on random walks which ranges from 0 to 1. The polarization score is close to 1 when the probability of crossing sides is negligible (extreme polarization), and close to 0 when the probability of crossing sides is comparable to that of staying within the same community (absence of polarization). The measure is quantified as follows:

$$RWC = P_{XX}P_{YY} - P_{XY}PYX \tag{1}$$

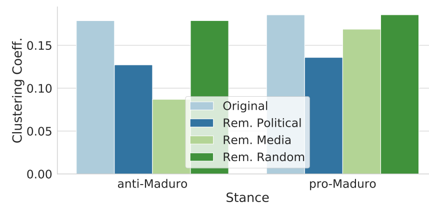
where $P_{AB}, A, B \in X, Y$, is the conditional probability $P_{AB} = P\{start\ in\ community\ A \mid end\ in\ community\ B\}$.

Figure 5 shows the polarization scores as reported by RWC for all our networks under different conditions and across the two episodes of interest. We compare these results against random node removals similar to our previous investigation. During the international aid episode, we observe that the removal of both political figures and media indeed had an impact on the overall polarization score of the network, where we see that it decreases (Figure 5b). However, we do not observe the same behavior in the Guaido interim president episode, where the impact on the polarization score is not significantly changed with the removal of both political figures and media. We note this is due to the influence of users who share both pro and anti Maduro messages. For example, there are more (18%) users who share both anti and pro Maduro messages during the Guaido interim president episode compared to the similar number (8%) of users during the international aid episode. In addition, we found that random node removals do not affect significantly the polarization score of the networks during these two episodes.

The next question we investigate is what is the impact of different topics on network polarization during different events. We again use RWC measurement to gauge polarization on user-to-user interaction networks under two different conditions: (1) by removing all messages associated to a particular topic, and (2) by only considering interactions within a single topic. Figure 6 shows the results of our experiments. First, we found that the removal of messages related to a single topic does not seem to have a significant impact on polarization during the international aid episode.

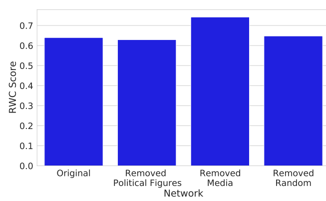


(a) Guaidó Interim President

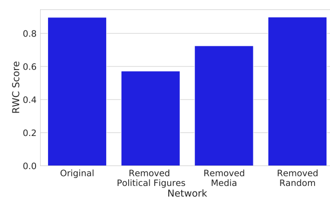


(b) International Aid

Fig. 4. Clustering coefficient for both anti-Maduro and pro-Maduro communities under certain what-if scenarios. The original networks were filtered by removing political and media accounts as well as random nodes with a similar activity rate than that of political figures and media.



(a) Guaidó Interim President



(b) International Aid

Fig. 5. Polarization scores (RWC) under varying what-if scenarios during Guaido interim president and international aid episodes. The original networks were filtered by removing political and media accounts as well as random nodes with a similar activity rate than that of political figures and media.

However, when removing certain topics in the Guaido interim president, we observed a slight effect on the overall polarization. In particular, removing *military*-related messages causes a drop in polarization, while removing either *arrests* or *protests* messages increases polarization. Second, we observed that when topics are considered independently from each other, the levels of polarization appear to vary widely during the two episodes of interest. We noticed that both *military* and *violence* are highly polarizing topics during both episodes, while *arrests* is the least polarizing. We found that on highly polarizing topics, the involvement of influential political and media accounts tends to be high. In particular, the percentage of identified political figures and media accounts, who were seen engaged with these topics, was approximately 80% and 90%, respectively. On the other hand, the involvement of influential users was much lower for less polarizing topics, especially from the political figures side. For instance, in the "International aid" episode, only 38% of political figures and 87% of media accounts were seen engaging with the *arrests* topic. Moreover, their involvement with the same topic was much lower during the "Guaido interim president" episode, where we observed only 25% of political figures and 65% of media accounts. This high/low engagement of influentials within particular topics seem to match the different levels of polarization observed in the networks.

5 EXTERNAL DRIVERS

In this section, we analyze the impact of exogenous events on polarized discussions on Twitter during the Venezuelan crisis. We calculate the correlation between the volume of anti-Maduro and pro-Maduro daily Twitter activities and the

(a) Guaidó Interim President

(b) International Aid

Fig. 6. Polarization scores (RWC) on networks where messages for a single topic are removed (Removed), and networks where messages for only one topic are considered (Topic). Percentage of messages within each topic during Guaido interim president episode: arrests (10%), military (25%), protests (41%), and violence (24%). Percentage of messages within each topic during international aid episode: arrests (5%), military (51%), protests (9%), and violence (35%).

volume of offline events as reported in the ACLED and GDELT databases, to gain insights on how opposing online communities react to real events as recorded and interpreted in these external datasets. We are also asking if users respond differently to the same events if they tweet in different languages.

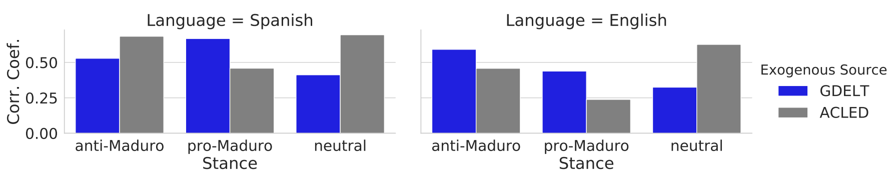


Fig. 7. Time series correlation between Twitter messages from either anti-Maduro, pro-Maduro, or neutral communities and the respective exogenous events. We grouped messages by Spanish and English languages.

Figure 7 helps make multiple observations. First, we found that the Spanish anti-Maduro community correlates more strongly with ACLED than with GDELT. This suggests that online discussions from this side tend to align very well with reports about protests and violent clashes happening within Venezuela as documented by ACLED. On the other hand, we observe that both the Spanish and English pro-Maduro communities seem to be better correlated with GDELT events than with ACLED. That is, rather than reacting to the on-the-ground events unfolding in the country, pro-Maduro messages tend to lean towards those events in GDELT that are extracted from news reports. One possible explanation for this observation is that pro-Maduro community might not engage often in discussions about the on-going conflicts/protests on the streets, as these demonstrations were strictly against the regime.

Moreover, we observe that the English-tweeting anti-Maduro community does not follow the same trend as its Spanish counterpart. Specifically, we observed a higher correlation between GDELT events and their Twitter online activities than what we see for ACLED. The English community is likely to reside outside Venezuela, and thus first-hand knowledge of the on-going real physical events might be limited. For this reason, their reactions tend to be better aligned with GDELT reports, as these events are made known in the news. Lastly, we noticed that in both languages the users who maintain a neutral tone correlate their messages more with ACLED. The messages from the neutral accounts might be more focused on timely reports about the situation within Venezuela without taking any sides.

To understand how information shared on Twitter could be related to news reports collected by GDELT, we analyzed the most popular URL domains promoted by both anti-Maduro and pro-Maduro communities. Table 2 presents the 20 most popular web domains in each group based on the number of shares. We found that 10 out of the top 20 domains shared in the pro-Maduro group overlap with those domains recorded in GDELT database while only 5 web domains shared by the anti-Maduro community overlap. This observation provides support towards the claim that news sources from GDELT seem to align more with the messages/sources shared by pro-Maduro community. Furthermore, we observe that a large portion of the URLs shared in pro-Maduro messages are from state-run sources controlled by Maduro's government (as shown by those web addresses under the .ve domain). Additionally, we noticed a high presence of state-sponsored Russian news organizations among these most popular domains (i.e., `actualidad.rt.com`, `rt.com`, and `mundo.sputniknews.com`). Russian media has been known for its extensive exploitation of the online environment through influence operations [3, 13]. However, the strong relationship between the Russian government and the Venezuelan regime has managed to establish these sources as trustworthy in the eyes of Maduro's supporters.

Table 2. Top-20 domains by number of mentions shared in both pro-Maduro and anti-Maduro tweets. We highlight in bold those domains which are also recorded in GDELT.

pro-Maduro		anti-Maduro	
Domain	# Messages	Domain	# Messages
caraotalibre.cf	1286	caraotalibre.cf	68381
elcooperante.com	990	**el-nacional.com**	17848
conelmazodando.com.ve	543	elcooperante.com	16878
minci.gob.ve	516	maduradas.com	9677
el-nacional.com	464	ntn24america.com	3796
maduradas.com	443	caraotalibre.ml	1649
konzapata.com	422	erpgkm.aws.ve.com	1590
telesurtv.net	363	**noticias.canalrcn.com**	1508
vtv.gob.ve	335	lapatilla.com	1388
ntn24america.com	224	caraotadigital.net	1343
actualidad.rt.com	219	dolartoday.com	1287
rt.com	196	elpitazo.net	1254
monagas.com.ve	182	donlengua.com	1100
mundo.sputniknews.com	174	caraotalibre.tk	1049
mindefensa.gob.ve	162	**notitarde.com**	1019
rnv.gob.ve	148	caraotadigital.cf	967
vicepresidencia.gob.ve	120	konzapata.com	966
radiomundial.com.ve	118	**noticias.caracoltv.com**	942
avn.info.ve	112	**elnuevoherald.com**	784
entornointeligente.com	110	epmundo.com	769

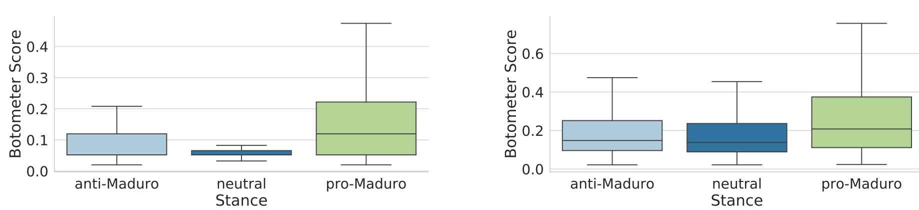
On the other side, the anti-Maduro community tends to share news from well-established national and international news media outlets, which are also recorded in GDELT (e.g., `noticias.canalrcn.com`, `noticias.caracoltv.com`, `el-nacional.com`, `elnuevoherald.com`). Many domains shared in this community are also from several independent media websites, whose popularity was built mainly through Twitter (e.g., `caraotadigital.net`, `caraotalibre.cf`, `lapatilla.com`, `elpitazo.net`, `dolartoday.com`). In Venezuela, internet censorship is a notorious problem. Over the years, the regime had frequently ordered local internet service providers to block access to websites whose viewpoints or reports conflict with those of the government [14]. As a result, many of these independent media sources tend to be volatile and short-lived as they often get attacked usually by means of DNS tampering. One such example is the

Table 3. Top-10 domains promoted by Bot accounts in both pro-Maduro and anti-Maduro retweet messages. We identify Bot accounts using Botometer score (> 0.5). We highlighted the government domains in bold.

pro-Maduro Bots	anti-Maduro Bots
mindefensa.gob.ve	ntn24america.com
minci.gob.ve	maduradas.com
vtv.gob.ve	erpgkm.awsve.com
vicepresidencia.gob.ve	dolartoday.com
telesurtv.net	caraotalibre.cf
j.mp	lapatilla.com
vivoplay.net	elpitazo.net
maduradas.com	vivoplay.net
psuv.org.ve	el-nacional.com
rnv.gob.ve	caraotadigital.cf

`caraotalibre` web domain. As shown in Table 2, this domain was the most shared in each of the communities. Despite its neutrality in reporting news from each side, `caraotalibre` alongside many other independent sources tend to also report coverage of the on-going nationwide unrest. Hence, efforts to limit access to this kind of information could have led to the emergence of multiple `caraotalibre` websites under different free top-level domains (i.e., .cf, .tk, and .ml). At the time of this study, the `caraotalibre` website and its various domains were no longer accessible.

The next issue we investigate is the promotion of domains by bots. We used Botometer [5] to identify potential bots who might also be promoting these popular domains. Figures 8a and 8b show a higher presence of bot-like behavior in the pro-Maduro community working towards pushing and spreading these popular websites. More precisely, Table 3 shows the top 10 domains promoted by the identified bot accounts in each community. We found that pro-Maduro bots tend to heavily promote Venezuela state-controlled websites, whereas anti-Maduro bots are more likely to promote information from independent news sources (e.g., `dolartoday`, `caraotalibre`, `lapatilla`).



(a) Tweeters (b) Retweeters

Fig. 8. Botometer score of users who cited URLs in tweets and retweets.

6 SUMMARY AND DISCUSSIONS

This paper addresses the following question: how do external events, as seen from the lens of physical conflicts and news, drive related Twitter activity? Understanding this question is useful for a number of applications. One such example is the problem that motivated our analysis in the first place, which is how to build a simulator of social media activity that reacts realistically to real-world events. The Venezuelan context of early 2019 is particularly relevant because the social crisis in the country is well documented and thus news articles are available and accessible in public datasets. We selected two publicly accessible datasets that record real-wold events: ACLED [24], which records conflicts of different severity from peaceful protests to military interventions, and GDELT [17], a Google-managed database of news reports across the world. On one hand, ACLED gives the bare facts of conflicts in the physical world; on the other hand, GDELT provides a record of the interpretation of these events as published by journalists and interpreted for the public at large. Thus, our question: how do journalists and in-the-street conflicts affect the conversations in social media, and in particular on Twitter?

Our quantitative investigation leads to the following observations. First, users who tweet in different languages show different emotions to events: the locals (if we consider the Spanish-tweeting users overwhelmingly Venezuelans or South Americans) react with more joy to events to which the English-tweeting community reacts with fear. This could be explained by the subjective reactions of Venezuelans who see the hope of change where the outsiders see political instability. In support of this hypothesis is also the observation that both pro-Maduro and anti-Maduro English speaking communities correlate to news with less discrimination than the Spanish-tweeting community, where the pro-Maduro group's activity correlates with the volume of news while the anti-Maduro group's tweets correlates better with the volume of reported conflicts. What this means for a simulator of social activity is that focusing on language-driven subcommunities has to be nuanced.

Second, in heavily polarized communities such as the Venezuelan Twitter community, the two sides show other differences as well. The pro-Maduro community tweets in sync with the government-controlled news as recorded in GDELT and shares messages from Russian-sponsored channels (such as Russia Today and Sputnik). On the other hand, the anti-Maduro community is more receptive to anti-government street protests and thus correlates better with ACLED-recorded events than with GDELT records. In a country where the opposition is censored in media, the anti government users share URLs from independent news media, some of which have a volatile presence online, which challenges efforts to reproduce results. Moreover, the pro-Maduro activity includes users with more bot-like characteristics than the anti-Maduro side. These distinctions in behavior are again relevant for simulations at meso-level granularity, where the volume of Twitter activity on each side is an important objective.

Third, we show that different types of influential users contribute differently to different topics and to the polarization of the user interaction network. Specifically, we observe that media outlets contribute to a stronger connection within the anti-Maduro community. On the other hand, political figures bind the pro-Maduro network more than media does. And finally, we see that different events lead to different changes in network polarization. These observations can help social media platforms fine-tune their content promotion algorithm to alleviate or at least not exacerbate polarization.

While these observations are drawn from a specific politico-cultural context, many of these observations will hold for other contexts and other communities. We mainly highlighted different reactions to real-life events as reflected in Twitter activity by different groupings of users: English vs. Spanish-tweeters, pro vs anti-government, media vs politician influencers vs the rest. Our observations can be used as guidelines to what parameters to choose for social media simulations or even what to look for when characterizing the social media activity of a community in times of political crisis.

REFERENCES

[1] Max Aliapoulios, Emmi Bevensee, Jeremy Blackburn, Emiliano De Cristofaro, Gianluca Stringhini, and Savvas Zannettou. [n.d.]. An Early Look at the Parler Online Social Network. *arXiv preprint arXiv:2101.03820* ([n. d.]).

[2] Zeynep Zengin Alp and Şule Gündüz Öğüdücü. 2018. Identifying topical influencers on twitter based on user behavior and network topology. *Knowledge-Based Systems* 141 (2018), 211–221.

[3] Nazim Choudhury, Kin Wai Ng, and Adriana Iamnitchi. 2020. Strategic Information Operation in YouTube: The Case of the White Helmets. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*. Springer, 318–328.

[4] Te-Lin Doreen Chung, Olivia Johnson, Adrienne Hall-Phillips, and Kyuree Kim. 2020. The effects of offline events on online connective actions: An examination of# BoycottNFL using social network analysis. *Computers in Human Behavior* (2020), 106623.

[5] Clayton Allen Davis, Onur Varol, Emilio Ferrara, Alessandro Flammini, and Filippo Menczer. 2016. Botornot: A system to evaluate social bots. In *Proceedings of the 25th international conference companion on world wide web*. 273–274.

[6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*.

[7] The Economist. 2017. How Chávez and Maduro have impoverished Venezuela. (2017).

[8] Ethan Fast, Binbin Chen, and Michael S Bernstein. 2016. Empath: Understanding topic signals in large-scale text. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 4647–4657.

[9] Emilio Ferrara, Onur Varol, Filippo Menczer, and Alessandro Flammini. 2016. Detection of promoted social media campaigns. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 10.

[10] Deen Freelon, Charlton D McIlwain, and Meredith Clark. 2016. Beyond the hashtags:# Ferguson,# Blacklivesmatter, and the online struggle for offline justice. *Center for Media & Social Impact, American University, Forthcoming* (2016).

[11] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2018. Quantifying controversy on social media. *ACM Transactions on Social Computing* 1, 1 (2018), 1–27.

[12] Jesse Hammond and Nils B Weidmann. 2014. Using machine-coded event data for the micro-level study of political violence. *Research & Politics* 1, 2 (2014), 2053168014539924.

[13] Sameera Horawalavithana, Kin Wai Ng, and Adriana Iamnitchi. 2020. Twitter Is the Megaphone of Cross-platform Messaging on the White Helmets. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation*. Springer, 235–244.

[14] Freedom House. 2017. Freedom on the net: Venezuela. (2017).

[15] Joo-Young Jung and Munehito Moro. 2014. Multi-level functionality of social media in the aftermath of the Great East Japan Earthquake. *Disasters* 38, s2 (2014), s123–s143.

[16] Janani Kalyanam, Mauricio Quezada, Barbara Poblete, and Gert Lanckriet. 2016. Prediction and Characterization of High-Activity Events in Social Media Triggered by Real-World News. *PLOS ONE* 11, 12 (12 2016), 1–13. https://doi.org/10.1371/journal.pone.0166694

[17] Kalev Leetaru and Philip A Schrodt. 2013. Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA annual convention*, Vol. 2. Citeseer, 1–49.

[18] Tanushree Mitra, Scott Counts, and James Pennebaker. 2016. Understanding anti-vaccination attitudes in social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 10.

[19] Alfredo Jose Morales, Javier Borondo, Juan Carlos Losada, and Rosa M Benito. 2015. Measuring political polarization: Twitter shows the two sides of Venezuela. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 25, 3 (2015), 033114.

[20] Davide Morselli, Stefano Passini, and Craig McGarty. 2020. Sos Venezuela: an analysis of the anti-Maduro protest movements using Twitter. *Social Movement Studies* (2020), 1–22.

[21] Seth A Myers, Chenguang Zhu, and Jure Leskovec. 2012. Information diffusion and external influence in networks. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. 33–41.

[22] Mert Ozer, Mehmet Yigit Yildirim, and Hasan Davulcu. 2019. Measuring the polarization effects of bot accounts in the us gun control debate on social media. (2019).

[23] Clionadh Raleigh and Caitriona Dowd. 2015. Armed conflict location and event data project (ACLED) codebook. *Find this resource* (2015).

[24] Clionadh Raleigh, Andrew Linke, Håvard Hegre, and Joakim Karlsen. 2010. Introducing ACLED: an armed conflict location and event dataset: special data feature. *Journal of peace research* 47, 5 (2010), 651–660.

[25] Pew Research. 2019. Venezuelans have little trust in national government, say economy is in poor shape. (2019).

[26] Philip A Schrodt. 2012. CAMEO Conflict and Mediation Event Observations Event and Actor Codebook Version 1.1 b3. *Retrieved November* 11 (2012), 2016.

[27] G Sidorov, S Miranda-Jiménez, F Viveros-Jiménez, A Gelbukh, N Castro-Sánchez, F Velásquez, I Dıaz-Rangel, S Suárez-Guerra, A Trevino, and J Gordon. 2012. Empirical study of opinion mining in Spanish tweets. MICAI 2012. *Lect Notes Comput Sci* 7629 (2012), 1–14.

[28] Felipe Bonow Soares, Raquel Recuero, and Gabriela Zago. 2018. Influencers in polarized political networks on Twitter. In *Proceedings of the 9th international conference on social media and society*. 168–177.

[29] Shaomei Wu. 2013. The dynamics of information diffusion on on-line social networks. (2013).

[30] Sifan Xu and Alvin Zhou. 2020. Hashtag homophily in twitter network: Examining a controversial cause-related marketing campaign. *Computers in Human Behavior* 102 (2020), 87–96.