Elites Tweet to Get Feet Off the Streets: Measuring Regime Social Media Strategies During Protest*

KEVIN MUNGER, RICHARD BONNEAU, JONATHAN NAGLER AND JOSHUA A. TUCKER

As non-democratic regimes have adapted to the proliferation of social media, they have began actively engaging with Twitter to enhance regime resilience. Using data taken from the Twitter accounts of Venezuelan legislators during the 2014 anti-Maduro protests in Venezuela, we fit a topic model on the text of the tweets and analyze patterns in hashtag use by the two coalitions. We argue that the regime's best strategy in the face of an existential threat like the narrative developed by La Salida and promoted on Twitter was to advance many competing narratives that addressed issues unrelated to the opposition's criticism. Our results show that the two coalitions pursued different rhetorical strategies in keeping with our predictions about managing the conflict advanced by the protesters. This article extends the literature on social media use during protests by focusing on active engagement with social media on the part of the regime. This approach corroborates and expands on recent research on inferring regime strategies from propaganda and censorship.

Social media has changed the way that citizens communicate, both amongst themselves and with elites. In some contexts, most famously the Arab Spring in 2011, social media is reputed to have empowered citizens relative to established government regimes (Howard and Hussain 2011; Tufekci and Wilson 2012). Despite the initial enthusiasm caused by these revolutions, the effect of the diffusion of social media on this balance of power is far from obvious.

Technologically advanced and capable regimes in Russia and China have begun to actively and strategically engage in social media (Sanovich et al. 2018), rather than merely trying to ban its use, and it may be that by 2016, popular social media usage can actually enhances regime resilience.

The long-term use of social media by both pro- and anti-regime actors has received a good deal of scholarly attention, and its use by dissidents during acute protests has been even more studied. The fourth piece of the equation, the regime's strategic social media response to protests, however, has yet to be fully explored. In part, this is due to a lack of appropriate cases. There have been few large-scale protests against tech-savvy regimes during what Deibert et al. (2010) calls the "third generation" of social media use by regimes: active engagement. This engagement often takes the form of anonymous pro-government postings, but elite social media use offers an alternative avenue to flood and disrupt online forums. The largest and most protracted protests in the post-Arab Spring time period were in Turkey in 2013 and Ukraine in 2013 and 2014. Neither of these regimes had the capacity to use "third generation" strategies.[1]

This article fills this gap by analyzing the Venezuelan *La Salida* protests of 2014. This case is particularly appropriate to exploring elite social media response to protests. Social media (in this case, Twitter) is widely used by both citizens and elites: Venezuela is among the top five countries in terms of Twitter penetration (Schoonderwoerd 2013), and 2014 data suggests that 21 percent of internet users log on at least once a month (Friedman 2014). The "soft censorship" of mainstream media sources makes Twitter's speech-enabling function especially important, and, as is common in Latin America, Venezuelan politicians are more likely to speak their minds directly on Twitter rather than have a media team craft a sanitized presence. The *La Salida* ("the exit") protests of 2014 represented a serious effort to oust the regime of President Nicolás Maduro, and the protesters *en guarimba* (in a blockade) paralyzed parts of the capital of Caracas from February 12 to May 8.

We test a simple theoretical argument, inspired by descriptive accounts from qualitative research and by comparative statics from a related formal model, to analyze the strategic response of both pro- and anti-regime elites. Regime elites constitute the primary sample of interest, but any changes in their behavior can only be understood in contrast to that of opposition elites. We claim that the regime will try to distract attention away from the protesters' narrative, ignoring the specific grievances of the dissidents and instead bringing up other issues or engaging in general pro-regime cheerleading so that the protesters' narrative appears to be only one of several possible readings of current events. This strategy is similar to the one that King, Pan and Roberts (2017) find employed by the Chinese regime in response to online criticism.

This strategy will have several observable implications. First, we expect that the opposition will focus on the protests and try to draw more attention to the narrative *La Salida* is promoting. Practically, this will entail discussing a smaller number of different topics each day and using Twitter's "hashtag" function to mention people or events related to the protests in an effort to draw them into the conversation.[2] Conversely, the pro-Maduro coalition will bring up new issues and emphasize traditionally successful regime slogans, to give people the impression that *La Salida's* narrative is only one of many ongoing political discussions. The regime will use

---

[1] The Turkish regime merely tried to demonize Twitter use, and banned it entirely for several weeks (Tufekci 2014). Yanukovich's Ukrainian government badly underestimated the threat posed by social media, and their efforts to quickly crack down on its use without developing the proper infrastructure further incensed protesters (Leshchenko 2014).

[2] Note that the opposition politicians under study are more moderate, and not associated with the radical *La Salida* movement.

Twitter's "hashtag" function to insert explicitly crafted talking points into the public discussion, aiming to create conversations among their supporters.

We investigate the Twitter accounts of Venezuelan *diputados*, members of the unicameral Venezuelan National Assembly. The National Assembly is a powerful force in Venezuelan politics, as evidenced by the success of the opposition after winning a large majority in the 2015 parliamentary election. At the time of the 2014 protests, 121 of the 165 *diputados* had active Twitter accounts, and both regime and opposition *diputados* are well represented. These elites tweeted over 100,000 times during the five months under study, giving us ample analytical leverage.

We find evidence of the strategic use of Twitter by the two coalitions reflecting their proximate and ultimate goals, in accordance with the observable implications of our argument. By fitting a topic model on the text of the tweets, we find that opposition *diputados* focus their attention on the protesters and their demands, while regime *diputados* discuss a large number of different topics. Through analysis of the patterns of hashtag use, we find that the opposition mentions specific people and places, whereas the regime promotes new top-down narratives.

THE VENEZUELAN CONTEXT

Venezuela has been a democracy since 1958.[3] The first 40 years of Venezuelan democracy were dominated by the *punto fijo* system, with political power alternating between two clientelistic parties; 1998 saw the election of Hugo Chávez and a shift in Venezuelan economic and social policy.

Chávez won re-election in 2006 and 2012 with very high voter turnout, even as inflation and crime rates rose to high levels, as Chávez successfully but expensively expanded social services and consolidated support among poor Venezuelans (Corrales and Penfold-Becerra 2011). Chávez's tenure also saw the transition of Venezuela from a representative democracy to what Mainwaring (2012) calls "participatory competitive authoritarianism," a system which could be subsumed in the general category of a "hybrid" regime. Chávez's health declined rapidly in 2012, and he died in early 2013.

Vice president Nicolás Maduro won the 2013 special election against Henrique Capriles, a popular state governor and Chávez's opponent in 2012, but his margin of victory was only 1.5 percent, and there were claims of fraud and illegitimacy made by the losing party.

Under Maduro, continuing inflation and high violent crime rates led to rising discontent (Corrales 2013). The spark that ignited the 2014 protests was the January 6 robbery–murder of a former Miss Venezuela. On January 23, María Corina Machado and Leopoldo López called for *La Salida*, a movement to protest in the street with the aim of removing the Maduro from power. The plan was for the protest to begin around the country on February 12 (National Youth Day in Venezuela) (Diaz 2014). On February 5, the attempted rape of a university student on a campus in Táchira in the southeastern part of the country led to a small student protest that provoked a violent governmental response (Perez 2014). This was publicized by *La Salida*, which hoped to use student anger to advance their movement. Importantly, the moderate opposition elites, including all of the *diputados* in this study, were not involved in the planning or execution of these early protests; thus, the large-scale protests we study were similarly exogenous to both the pro- and anti-regime *diputados*.

---

[3] For a visual timeline of the events discussed here, see Appendix C.

As planned, protests began in Caracas and 38 other cities on February 12, and escalated quickly. López became sufficiently outspoken and high-profile that the regime jailed him on February 18. His arrest was precipitated by him tweeting directly at Maduro: ".@Nicolasmaduro Don't you have the guts to arrest me? Or are you waiting for orders from Havana? I tell you this: The truth is on our side" (authors' translation).[4] Twitter allowed López to spread his message directly and quickly to a wide mass of people—note the almost 45,000 retweets—with no risk that it would be distorted or misrepresented.

The use of Twitter and other social media by the opposition was especially important because of the media environment. Although Venezuela's constitution guarantees free speech, news outlets faced implicit threats, and engaged in significant self-censorship; Freedom House (2014) deemed Venezuelan media in 2014 "not free". The regime lacked the technical ability to curtail speech online, however, because the most commonly used social messaging platforms (Twitter and Facebook) were controlled by US multinational corporations (Pan 2017). Chávez used his Twitter account as an extension of his hours-long radio addresses, but this was not a major component of the regime's overall media strategy.

Personal Elite Communication on Twitter

During the height of the protests, opposition groups advocating *La Salida* in Caracas set up permanent blockades (*guarimbas*) designed to paralyze the city and heighten the economic crisis. In response, the regime allowed armed paramilitary groups called *colectivos* to attack, rob, and kill opposition protesters. The violence led to thousands of injuries and at least 40 deaths.

The first major sit-down between the regime and the moderate opposition took place on April 10, though no conclusion or agreement was announced. April 19–20 were a major revolutionary holiday (Beginning of the Independence Movement Day) and Easter, respectively, and this weekend represented the high water mark of the protests. The *guarimbas* were permanently dismantled on May 8.

The protests ended with a compelling display of strength by the government, but they revealed deep divisions within the country and among the opposition factions, neither of which was truly unified. Capriles and the moderate opposition that includes all of the opposition *diputados* continued to push for democratic reforms and non-violent methods of progress,

---

[4] The tweet begins with a period so that it is visible to everyone; at the time, tweets that began with @[username] were only displayed in the timeline of people who "followed" both parties.

whereas *La Salida* wanted nothing less than the removal of the ruling party, through constitutional means if possible but through violent street protests if necessary (Ciccariello-Maher 2014). The moderates tried to draw as much attention to the protests as possible, but whether this was out of opposition solidarity or in order to improve their bargaining power or electoral chances in the future remains an open question. In either case, *La Salida* was never the first choice of the opposition *diputados* under study.

SOCIAL MEDIA IN NON-DEMOCRATIC CONTEXTS

Social media has had a pronounced impact in non-democratic regimes like Venezuela in 2014. Much of the initial enthusiasm about social media surrounded the way it empowered dissidents. It offered citizens affordances to become informed and engage in protest (Tufekci and Wilson 2012; Christensen and Garfias 2015) and for those protests to be better organized (Earl et al. 2013; Agarwal et al. 2014; Bennett, Segerberg and Walker 2014).

However, dissidents are not the only actors who can use social media, and an acute political crisis is not the only context in which it can be used. Social media shapes the way that citizens communicate on a day-to-day basis, and tech-savvy regimes take an active role in trying to structure and engage with discussions online.

There is a healthy literature on this dynamic. Deibert et al. (2010) develop a three-generation framework for how regimes reduce the threat of social media. The first generation is to restrict access to the internet (Howard, Agarwal and Hussain 2011), and the second is to censor certain kinds of information posted online (King, Pan and Roberts 2013). These two types of censorship vary significantly in both the technical sophistication required to execute them and their observed effectiveness. A crude, blanket internet shutdown is within the power of many authoritarian regimes. However, employing such a strategy during a protest can backfire: Hassanpour (2014) finds that Mubarak's mass internet shutdown actually increased revolutionary mobilization during the 2011 Egyptian uprising. Targeted censorship of specific types of posts—those with the possibility of encouraging collective action—may be more effective, but they are beyond the reach of most national governments. As Pan (2017) argues, China is only able to execute this strategy because the Chinese internet market is dominated by domestic firms; Venezuela and other nations who primarily use US multinational firms are simply unable to dictate rules about content. First and second generation strategies for limiting the risk of the internet and social media are thus either ineffective or inaccessible to the Venezuelan regime.

This leaves the "third generation" strategies, which include information campaigns and actively paying supporters to post online. These strategies show how social media can be used as a tool for regime activism (Greitens 2013). For example, Oates (2013) details how post-Soviet regimes have taken advantage of weak civil societies to control the online discourse, providing a new, compelling narrative for people who felt directionless after the revelation of the emptiness of the Soviet project. Sanovich et al. (2018) demonstrate the dramatic changes in online engagement strategy in Russia since 2000. Recent evidence suggests that Russian social media activity has offsetting effects: areas with higher social media penetration were both more supportive of the government and more likely to experience a protest (Enikolopov, Makarin and Petrova 2015).

A comprehensive discussion of social media's potential to increase regime resilience using these "third generation" strategies is Gunitsky (2015). Social media can help stave off revolution by giving the regime information about specific grievances that they can address without relying on local bureaucrats with an incentive to falsify information (Chen, Pan and Xu 2015).

Social media may be especially useful for predicting collective action events and aggregating allegations of corruption (Qin, Strömberg and Wu 2017).

However, there is little research specifically testing the "third generation" strategies employed by non-democratic regimes on social media during times of crisis. Before social media, regimes that controlled the mass media could either suppress all mention of a protest, or craft their own preferred narrative about it. Explicit propaganda produced by the state was an effective tool for enabling violent repression and preventing dissent.

Consider the case of the Tiananmen Square protest in 1989. Early in the occupation of the Square by student protesters, the Chinese regime used the media to promote the narrative that these students were agents of the United States, aiming to undermine China. The students were unable to broadcast their true goal and grievances, and the regime's narrative was unchallenged throughout much of China. Once the regime found it necessary to begin shooting, they switched their media strategy, banning all mention of the protest and its repression.

Neither of these strategies is viable in a society—even one with control over mass media—in which social media use is widespread. Once a large protest has occupied space in a city, it will be able to broadcast its own narrative on social media, and the regime cannot physically remove or kill them without the whole country becoming aware.[5]

Instead, a powerful strategy for the regime is to compete with the protesters on social media. They cannot fully silence the protesters, but they can hope to drown them out and show citizens that the there are other political conversations going on, either by bringing up other issues or engaging in general pro-regime cheerleading (King, Pan and Roberts 2017). *La Salida*'s narrative focused on high crime and inflation rates. These are trends that individuals can notice, and efforts by the regime to deny them outright would ring false. However, by bringing up many different issues or posting classic regime slogans rather than engaging with the protesters on their own terms, the regime can create ongoing discussions about their efforts to help the poor, improve healthcare, and stand up to the United States. Citizens using social media will thus see a political climate with multiple issues on the agenda and groups of competing regime boosters and detractors, rather than perceiving a one-issue debate with two sides arguing with each other about the fundamental legitimacy of the regime.

This theory has much in common with the strategic predictions made in section 6 of Edmond (2013). In this model, the predicted amount of noise that a regime will insert into their communications will vary with the strength of the regime. Strong regimes (and most regimes are strong in that they are not vulnerable to being overthrown) will decrease noise, to accurately signal their strength. On the other hand, weak regimes (like Venezuela, facing an existential threat) will want to make their communications "noisier" (less focused), in order to make it harder for imperfectly coordinated protesters with incomplete information to infer that the regime is in fact weak.[6] We do not claim to formally test the model in the current article, but the confluence of predictions from the descriptive and formal literatures is encouraging.

The strategy for the moderate opposition, however, is straightforward: they largely agree with the protesters, but would prefer to see the removal of Maduro and his party through electoral success. They will thus support the protests and try to coordinate attention on them on Twitter, to increase their bargaining power and chances at victory in the next election. The opposition

---

[5] Perhaps for this reason, current Chinese media censorship strategy has been found to be explicitly designed to prevent mass public demonstrations (King, Pan and Roberts 2013).

[6] A regime's "strength," here, refers to their susceptibility to an organized revolution. So, the stronger a regime is, the more likely they are to survive such a revolution.

diputados are of interest primarily so that the change in their Twitter behavior can be contrasted with changes in the media strategy of the regime *diputados*.

HYPOTHESES

We test the theory of active, strategic use of social media during an acute crisis developed above in the case of the 2014 Venezuelan protests.

We expect the pro-regime elites to attempt to distract the public's attention from the protest's narrative concerning the regime's mishandling of the economy and failure to reign in violent crime. They will tweet about a variety of different subjects and engage in general pro-regime cheerleading, to make the opposition's narrative less central to the national discourse.

In contrast, we expect the opposition elites to focus on the protests and their specific message, at least as long as they believe it to be in their interest. We expect to see a contrast in the strategic reactions of the two coalitions of *diputados* to the onset of the protests as they pursue their parties' respective goals. Under normal conditions, politicians are expected to balance their personal interest against that of their party, but because these protests represented an existential threat to the government, the strategic incentives of each politician were closely aligned with that of their party.

The primary way we measure these strategies is by measuring what we call *focus*. Intuitively, this is a way of trying to get at how many different subjects are being addressed by each coalition's Tweets: *high focus* refers to a small number of subjects, whereas *low focus* refers to Tweets spread out across many different topics. Technically, we measure focus using a two-step process. We employ an unsupervised topic modeling technique that uses the co-occurrence of words to find topics in a given series of texts: Latent Dirichlet Allocation (LDA). LDA gives us an estimate of the topics discussed on Twitter every day by each coalition. We then operationalize *focus* through the topic diversity of the tweets from each coalition on each day. Explained in more detail below, topic diversity summarizes the amount of different information a given text contains. Higher topic diversity corresponds with less focus.

HYPOTHESIS 1: Compared to their pre-protest levels of focus, the opposition diputados' tweets will become more focused and their topic diversity scores will decrease, while the regime diputados' tweets will become less focused and their topic diversity scores will increase.

There are also important indicators of strategy that are unique to the medium of Twitter. The most relevant is the use of "hashtags" (#s), which structure discussion between people who might not know each other. There are two different strategies for using hashtags: include them in tweets to indicate the subject of your tweet, or invent specific, lengthy ones that themselves define subjects of discussion. The analysis of the use of hashtags does not easily generalize beyond the medium of Twitter, but we believe that Twitter plays a sufficiently large role in modern political communication to merit explicit study.

The former strategy has been shown to be useful in joining subjects together and connecting disparate networks (Bennett, Segerberg and Walker 2014), and is thus more useful for engaging with an existing narrative. There are two empirical implications of this strategy: using more short hashtags, and creating more tweets in which multiple hashtags co-occur. The more hashtags your tweet contains, the more (and potentially distinct) groups can be exposed to it. These implications are related because of Twitter's then 140-character limit: longer hashtags limit the number of hashtags that can be used in a single tweet. Long hashtags are useful,

TABLE 1   *Number of Tweets by Venezuelan Diputados*

Diputados	N	First Quartile	Median	Third Quartile	Mean	Total Tweets
Regime	65	109	299	794	663	43,093
Opposition	56	211	584	1231	1115	62,423

Note: Period of analysis: December 19, 2013 to May 29, 2014.

though, to define and attempt to create engagement with a specific narrative. They constrain both the topic and the tone of a discussion.

We expect the opposition *diputados* to use more short, "labeling" hashtags and to send more tweets containing multiple hashtags. In contrast, we expect the regime *diputados* to use more long,[7] "discourse-structuring" hashtags and to send fewer tweets containing multiple hashtags.

HYPOTHESIS 2:   During the protests, the regime diputados will use long hashtags more frequently than the opposition.

HYPOTHESIS 3:   During the protests, the regime diputados will send tweets containing multiple hashtags less frequently than the opposition.

DATA COLLECTION

For each of the 99 regime and 66 opposition *diputados* in Venezuela's unicameral legislature, we performed a manual search for an associated Twitter account. If there was any ambiguity as to whether a Twitter account belonged to a politician or an ordinary citizen with the same name, we checked to see if the account was followed by one of the party elites from either side. We were able to locate accounts for 139 of the 166 *diputados* (84 percent): 63 of 66 for the opposition (95 percent, similar to US Members of Congress), and 76 of 99 (77 percent) for the regime.[8] For the subsample whose tweets were analyzed in this article—those who tweeted after December 19, 2013—there were 121 accounts, 65 regime, and 56 opposition.

We used Twitter's REST API[9] via tweepy,[10] in the Python programming language, to collect the most recent tweets for each account. Using the /statuses/user_timelines endpoint,[11] Twitter's API allows fetching the latest 3200 tweets for a given account.[12] We did this on April 19 and then again on May 29, at which point the protests had largely subsided. As a result, we obtained all of the tweets sent by these accounts during the time period under study. We entered each *diputado's* party as an additional variable in the data set.

Table 1 summarizes of the number and distribution of the tweets collected. Although the regime had a higher number of individuals with active accounts, the opposition produced roughly 40 percent more total tweets during our period of observation, and this difference only became pronounced once the protests started. The difference is not just driven by a few prolific opposition

---

[7] We discuss what constitutes "long" below, but in general these hashtags are 3+ word phrases.

[8] To check the validity of our selection, we had a research assistant recreate our analysis. There were only two discrepancies, the adjudication of which was obvious.

[9] https://dev.twitter.com/docs/api/1.1

[10] https://github.com/tweepy/tweepy

[11] https://dev.twitter.com/docs/api/1.1/get/statuses/user_timeline

[12] Note that this retroactive approach means that we may have missed any tweets that were deleted soon after being published. We are unaware of any scandals involving deleted tweets by Venezuelan diputados in this time period, so it is unlikely that any of them were particularly important.

accounts; comparing the first quartiles, medians, and third quartiles of the regime and opposition *diputados* indicates that the opposition *diputados* are more active throughout the distribution.

It is also important to note the distribution of tweets by each coalition over time. The timeline of our analysis can be divided into three periods: (1) December 19 to February 11, before the protests; (2) February 12 to May 8, protest blockades paralyze Caracas; (3) May 9 to May 28, after the protests. The beginning and end of the main protest period are noted by vertical lines in all of the figures on February 12 and May 8. For more details, see Appendix C, which contains an annotated visual timeline of the events.

Notice that before the protests, the tweet density was similar and low for both the government and opposition. Both sides saw a flare-up around the time of the murder of Miss Venezuela on January 6, which then subsided. On January 23, radical opposition leaders Leopoldo López and María Corina Machado began promoting *La Salida*, their plan to protest in the street and remove the regime from office. Both the regime and opposition *diputados* tweet most frequently in the time period between this announcement and the countrywide protests on National Youth Day, February 12. During the protest, both coalitions maintained a steady rate of tweeting. However, after the protests were effectively ended when the camps were bulldozed on May 8, there was a sharp decline in tweeting by both coalitions.

ANALYSIS

To estimate the number of overall topics being discussed on a given day by each coalition, we employed LDA (Blei, Ng and Jordan 2003) to generate topics and label each day's worth of tweets with a certain distribution over those topics. LDA is an unsupervised, "bag-of-words" machine-learning algorithm used for topic modeling that is increasingly popular in the social sciences. This approach is well suited to analyzing the communication of a group of elites over time because it is unsupervised (i.e., the researcher does not decide which topics to study *ex ante*) and because it allows the entire corpus of information to be used.[13]

The input to LDA is a corpus of documents composed of tweets aggregated as described below, each of which is a vector of N terms (within which order is irrelevant) taken from a vector of length V that contains all the terms in the corpus. LDA also requires the specification of two parameters: K, the number of topics to be modeled, and α, a concentration parameter that determines the shape of the probability distribution central to LDA. LDA analysis is conducted by treating each document as a probability distribution over latent topics and each topic as a distribution over words. That is, LDA assigns k weights to each word, one for each topic it is instructed to find and k weights to each document in the corpus.

In this case, the "documents" consist of the text of the *diputados*' tweets. The start date of our analysis is December 19, 2013, two months before the start of the height of the protests as marked by López's arrest, and the end is May 29, 2014, approximately one month after the protests subsided. We divided each day's worth of tweets by each coalition into a separate document. This decision, to treat each *diputado* within a coalition as interchangeable, is premised on the idea that each coalition is coordinating (either explicitly, or through cues from the leaders) on a communication strategy, discussed below. There are 162 days included in the analysis, and thus 324 documents. For a visual depiction of this tweeting pattern, see Figure 1. Once aggregated into these documents, the terms comprising the tweets from that coalition-day

---

[13] Another prominent topic model, the Correlated Topic Model (Blei and Lafferty 2007), functions similarly to LDA except that it allows for topics to be correlated within each document. We find that the inferences drawn from our main analysis in Figure 3 are descriptively supported by the correlated topic model results, but are not fully robust to this alternative model; see Appendix F.
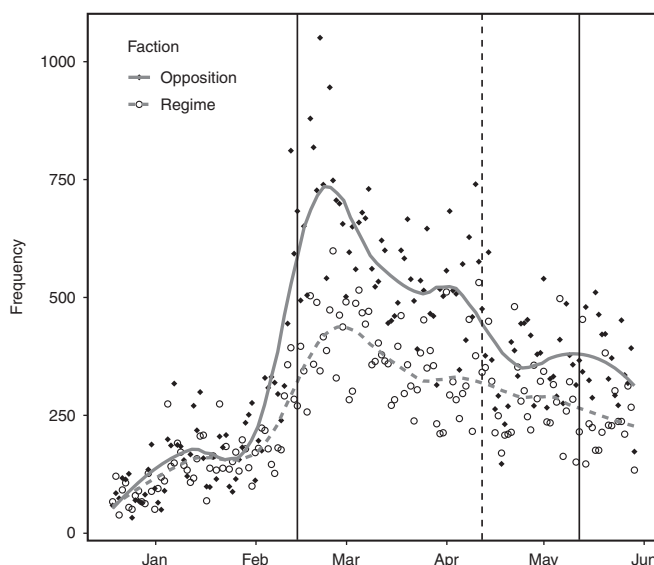
Fig. 1. Tweets per day by each coalition
Note: The number of tweets sent by *dipudatos* from each faction per day. The solid vertical lines correspond to February 12 (the beginning of the protests in Caracas on National Youth Day) and May 8 (protest camps are removed from Caracas). The dotted vertical line represents the April 10 televised sit-down between Maduro and Capriles. Trend lines for each coalition created with loess.

are treated identically: order ceases to matter, as does the number of tweets. For example, a dozen tweets that each say only "Venezuela" spread over a single day by a dozen different *diputados* from the same coalition has the same impact as a single tweet that says "Venezuela" a dozen times.

This approach does ignore potentially useful information by disregarding which individual *diputado* from a coalition produced each tweet. Given that our aim is to measure the different strategies between the coalitions, we are not concerned with which specific account sent each tweet. This assumption might not make sense if individuals within each coalition had substantially different incentives—for example, in the 2016 US election, some Republican candidates for Congress aligned themselves with the Trump presidential campaign, while others tried to distance themselves. However, in the relevant Venezuelan context, it would not make any sense for an individual PSUV *diputado* to distance herself from Maduro in the hope of holding office after the extralegal over-throw of the regime.[14] The tweet-pooling technique also loses information by conflating all the tweets from each day. This is unavoidable: there is not enough information in a single tweet to treat it as a document, and the machine-learning literature indicates that aggregating tweets leads to better performance for LDA (Hong and Davison 2010).

To determine the number of topics k, we performed tenfold cross-validation of both log-likelihood and perplexity analyses on the holdout sample.[15] Although the model fit

---

[14] There could be a technical problem if a small number of accounts were doing almost all of the tweeting. Indeed, the distribution of tweets is skewed, but because the shape of the distribution is similar for the two coalitions, this does not impair our comparative analysis.

[15] Perplexity is a measure of how well the LDA algorithm trained on the training sample is able to predict the holdout sample. Cross-validation works by dividing the sample up into several subsections (in this case, ten), iteratively using nine of the subsamples to predict the tenth, and seeing how accurate those predictions are.
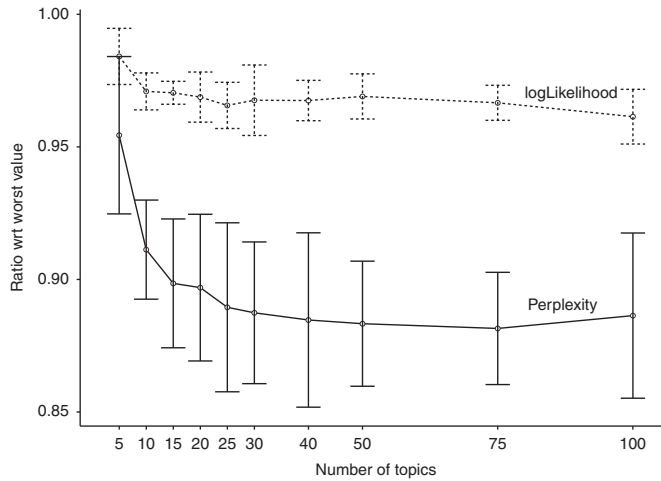
Fig. 2. Testing different numbers of topics
Note: The dotted line connects log-likelihood estimates for the model fitted with different number of topics, the dark line connects perplexity estimates for the model fitted with different number of topics. Vertical bars are 95 percent confidence intervals. Models fit on the entire corpus of tweets.

improves monotonically in the number of topics, the gains from adding more topics diminish at around 50 topics (see Figure 2). There exist standard rules for choosing k, such as the conservative "one-standard-error" rule outlined in Hastie et al. (2009), but this choice is contingent on the question LDA is being used to answer. LDA has been used most commonly to identify specific topics, prioritizing the recognizability of the topics created; in this context, the main priority is to avoid overfitting the data by choosing a conservative k. Our aim is to study the change in *focus* over time, so this concern is less relevant, and creating more topics allows for greater variation in the quantity of interest, even if those topics are sparsely represented and hard to identify. As a result, we follow a guideline of doubling the number of topics that the conservative approach recommends, and select $k = 100$.[16] Our results are not an artifact of this choice; the central finding of different degrees of *focus* in Figure 3 is statistically significant for $k = 30$ and $k = 50$, though the time period for which this is true is narrower than for $k = 100$.

Following standards in the social science literature, we use the collapsed Gibbs sampler (Griffiths and Steyvers 2004; Phan, Nguyen and Horiguchi 2008). Using the R package "topicmodels" developed by Hornik and Grün (2011), we ran LDA in a single chain for 1000 iterations. The text was pre-processed using "topicmodels" by removing numbers and punctuation, by converting all the text to lowercase, and by "stemming" the words so that different forms of the same word are not treated as entirely different words; stemming is especially important when dealing with Spanish objects that have four different endings depending on the number and gender of the subjects. After this pre-processing, the corpus consisted of $N = 50{,}902$ terms.

In creating the topics, the algorithm estimates $k = 100\,\gamma$ for each document. For each of the 324 documents w, γ_{wk} is the probability that document w pertains to topic k; note that $\sum_{k=1}^{100} \gamma_{wk} = 1$. There are 32,400 of these γs.

---

[16] For α, we follow the standard established by Griffiths and Steyvers (2004) and set $\alpha = 0.5$, where $50/k = 0.5$.
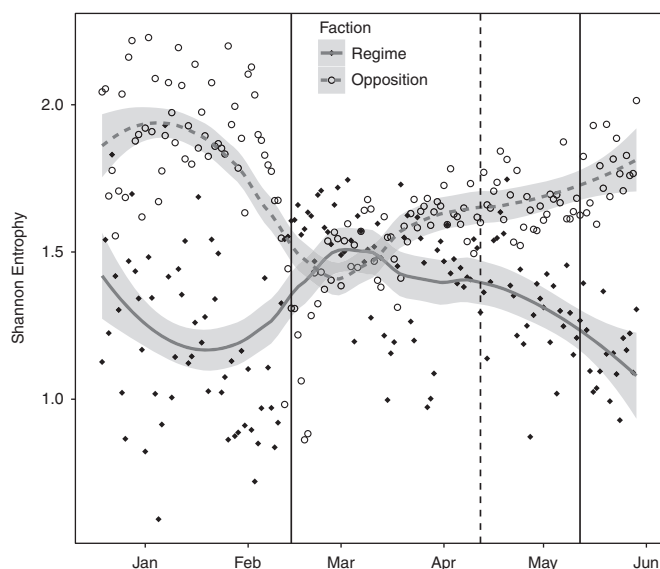
Fig. 3. Focus—as modeled by topic diversity—over time
Note: Each point represents the topic diversity score for the opposition and regime tweets respectively, computed based on their tweets as described in the text. The solid vertical lines correspond to February 12 (the beginning of the protests in Caracas on National Youth Day) and May 8 (protest camps are removed from Caracas). The dotted vertical line represents the April 10 televised sit-down between Maduro and Capriles. Trend lines and 95 percent confidence intervals were created with loess.

To analyze how *focused* the coalitions are over time, we measured the Shannon (1948) Entropy of the γ distribution of each document. Commonly used in the natural sciences to measure the diversity of an ecosystem by the relative counts of each species in that ecosystem, Shannon Entropy (what we refer to throughout the article as topic diversity) is well suited to measuring how *focused* these documents are. It efficiently captures information about the entirety of the distribution while avoiding the imposition of arbitrary thresholds. The formula for Shannon Entropy is $-\sum_{i=1} p_i \log_2(p_i)$. In this case, because the γs in each document must sum to 1, $p^i = \gamma^{wk}$ and the topic diversity score for each document is

$$\text{Topic Diversity} = -\sum_{1}^{k} \gamma_{w,k} \log_2(\gamma_{wk}).$$

The possible topic diversity scores range from 0 (if the γ distribution is unitary) to $\log_2(k)$ (if the γ distribution is uniform). Generally, lower topic diversity scores mean a less uniform distribution, and in the case being analyzed here, a more *focused* message. LDA identifies some "topics" that are not semantically meaningful and assigns them non-zero weights. In order to address this issue, we analyzed only the most meaningful topics: we calculated the "semantic coherence" of each topic, following Roberts, Stewart and Tingley (2014). There was a clear clustering between the topics with "high" and "low" values of semantic coherence, so we kept only the 43 in the former category.

RESULTS

We divide the time period under study into three periods:(1) December 19 to February 11, before the protests; (2) February 12 to May 8, protest blockades paralyze Caracas;

(3) May 9–28, after the protests. The beginning and end of the main protest period are noted by vertical lines in all of the figures on February 12 and May 8.

Tweet Focus: Topic Diversity

The central prediction in Hypothesis 1 is that the tweets sent by the regime will become less *focused* during the crisis. The results from the topic diversity scores are shown in Figure 3. Recall that lower topic diversity reflects more *focused* Twitter behavior (i.e., tweets are concentrated across a smaller number of topics). The absolute levels are quite different, as the opposition is far less *focused* than the regime. We did not explicitly predict this, but it concords with descriptive accounts of a generally well-coordinated regime and a comparatively disorganized opposition.

Importantly, the trends for the regime and opposition *diputado* diversities are similar in the first period: each coalition is becoming more *focused* during the regional protests over crime and inflation, though the opposition is consistently less focused.

Just before the planned, countrywide protests that shut down Caracas, the slope of the regime's *focus* abruptly changes: after nearly two months of becoming more *focused*, the regime becomes significantly less *focused* during the protests.[17] On the other hand, the opposition *dipudatos* continue to become more *focused* even after the protest begins.

The topic diversity during the beginning of the protests is significantly higher than during the pre-protest time period for the regime, and significantly lower for the opposition, both at $p < 0.01$, using a two-tailed *t*-test. However, the period for which this is the case is shorter than we had expected: beginning in mid-March, the regime begins to become more *focused*, reaching their pre-protest levels at around the April 10 sit-down with Maduro we had identified as a turning point for the protests. We thus find strong support for Hypothesis 1, with the caveat that the regime seems to have thought the threat of the protests had dissipated earlier than we anticipated.

The opposition, too, seems to have given up on their strategy of being *focused* on the protests earlier than anticipated. This suggests that the messaging strategy of the opposition may have changed. This was unexpected, and we cannot conclusively explain why this happened solely by looking at these data, a reasonable interpretation is that the moderate opposition "gave up" on supporting the aggressive strategy of *La Salida* to oust the Maduro government only a few weeks into the protest.

These results are supported by a brief inspection of the topics most discussed by each coalition. In addition to consistently drumming home the connection between Chávez, Maduro, President of the National Assembly (Cabello), the acronym of the regime's party (PSUV), and the words for "the people" and "homeland" in Topic 44, their two most common top topics refer to specific events that happened at times other than the protest. Table 2 lists the top terms for each topic, sorted by how specific they are to that topic. All ten terms are common in that topic, and the terms earlier in each list are least likely to be common in other topics. Topic 76 is centered around a Chávez remembrance hashtag and mainly took place in May; many of the terms ("I will do") are promises for what the regime will do in the future. Topic 77 is even more specific, referring to a meeting of the Community of Latin American and Caribbean States in

---

[17] This is a novel method, and it could be that these methods applied to any time period's worth of tweets would provide trends that could be interpreted as supportive of a given narrative. To test for this possibility, we re-ran our analysis on the four months after the period studied here. Appendix B presents the results of the "placebo check," and as we expected, does not show evidence of any noticeable trend. This finding supports the idea that the diputados under study were engaged in a specific communication strategy during the protests.

TABLE 2   *Top Terms for Relevant Topics*

	Top Government Topics
#	Terms
44	*dcabellor*, the people, PSUV, Chávez, Venezuela, president, peace, new, Maduro, *nicolasmaduro*
76	#TenMonthsWithoutYouChávez, Rendón, gold, protest camp, I will do, May, South American, Zerpa, Catia, *eduardopinata*
77	Magellan, against, January, #WeMakePeace, Caribbean, Havana, peacemaking, CELAC, birth, summit
	Top Opposition Topics
#	Terms
74	Biagio, Pilieri, paper, media, newspaper, safety, motherhood, violence, #WithoutPaperThereAreNoNewspapers, La Salida
94	today, peace, *hcapril*, protest, Venezuela, Maduro, march, the people, street, student
72	*pr1merojusticia*, path, today, diputado, dip, government, more, country, *juliocmontoy*

Note: The most common terms in the topics are the most commonly discussed by each coaltion. Each list of terms has been sorted so that the terms most specific to each topic are listed first. The terms have been translated from Spanish by the lead author; terms in italics refer to specific Twitter accounts. The original Spanish terms can be found in Appendix A.

Havana at the end of January which the regime touted as an opportunity to "make peace" with the opposition by heading off the planned protests on February 12.

Hashtags

Hashtags are a Twitter-specific linguistic tool used to coordinate discussion between strangers on a given subject. For example, the most common hashtag used by the regime elites in this sample was "#psuv," the acronym for the ruling party. By including "#psuv" in a tweet, they ensure that their tweet will be shown in a list of all tweets containing "#psuv" to anyone who searches for or clicks on that hashtag. Tweets can also contain multiple hashtags, both increasing their overall visibility and allowing them to comment on multiple discussions. However, Twitter's 140-character limit implies a tradeoff between characters used in hashtags and the amount of textual content in a tweet.

The final test of our theory of differential Twitter strategies comes from analyzing the use of hashtags. The opposition's use of hashtags is predicted to be designed to connect with the dominant extant narrative and encourage different groups to join in on the ongoing conversation. In contrast, the regime is predicted to try to create their own narratives, giving the impression of multiple parallel conversations but limiting their ability to simultaneously engage with disparate groups.

The most straightforward way to examine hashtag use is to look at a summary of their frequencies; Appendix D presents the 100 most common hashtags for each coalition. Visual inspection suggests that the regime tweets more very long hashtags, and more often, supporting Hypothesis 2.

To understand why this is important, consider the following example: the most frequent opposition hashtag to mention the president is simply "#Nicolás," but for the regime it is "#RodillaEnTierraConNicolásMaduro." The latter is clearly intended as support for the president, in keeping with an explicitly designed narrative: "rodilla en tierra" is a term commonly used by the regime to connote dedication and resolve. We call this a "discourse-structuring" hashtag. "#Nicolás," however, could be used as part of a general discussion of Maduro's policies. Also,

"#Nicolás" is only seven characters longer (compared to 31), so a tweet containing it has more room to include other hashtags or more discussion. We call this a "labeling" hashtag.

Figure 4(a) demonstrates that the regime was more likely to use discourse-structuring hashtags, but only once the protest started; this is evidence in support of Hypothesis 2, as the 95 percent confidence intervals created by non-parametric regression do not overlap for the duration of the protest. For each coalition-day, it shows the number of hashtags that were at least 20 characters long. This threshold is strongly indicative of a hashtag that consists of at least three substantive words, but see Appendix E for evidence that our results are robust to different choices for this cut-off. This effect is not being driven by a general increase in the use of hashtags by the regime. Indeed, as Figure 4(b) shows, both coalitions used a higher number of hashtags during the protest, and the increase is actually much more pronounced for the opposition.

Before the protest, the two coalitions were equally (un)likely to use a discourse-structuring hashtag, but after the protests began, the regime started employing these lengthy hastags to develop alternative narratives to compete with the narrative of *La Salida*.

To illustrate, the regime *diputados* used the following hashtags at different times to encourage their followers to talk about specific topics. #DesconectateDeLaGuarimba explicitly encouraged people to stop participating in and paying attention to the protest camps in Caracas. #ChávezViveLaPatriaSigue was one of several discourse-structuring hashtags that claimed that the spirit of Chávez lived on and that patriots should support the ruling party. #GringosyFascistasRespeten emphasized the popular regime narrative calling their opponents facist US sympathizers who need to respect Venezuelan sovereignty. Relatedly, #VzlaBajoAtaqueMediatico claimed that the international press was being unfairly supportive of the opposition as part of an effort to unseat the anti-US regime. While each of these acknowledged that the protests were going on, each was an attempt to develop alternative narratives and to generate discussion of the situation that was unrelated to the demands and complaints of the protesters.

However, this strategy comes at a cost: the length of "discourse-structuring" hashtags makes them less useful than "labeling" hashtags for connecting different conversations by different groups. Following Bennett, Segerberg and Walker (2014), Hypothesis 3 predicts that the opposition will be more likely to send tweets containing multiple hashtags during the protest.

In Figure 5, we find consistent support for this hypothesis. Panel (a) plots the total number of tweets sent by each coalition containing more than one hashtag, while panel (b) weights each occurrence of multiple hashtags by the number of hashtags in that tweet. In both cases, there were few tweets containing multiple hashtags before the protest started, and a much larger increase in this technique for the opposition than for the regime during the high point of the protests.

Regime Social Media Strategy During Crisis

We have found support for our claim that regimes employ "third generation" social media strategies during times of acute political crisis. In particular, in a context in which they could not silence or sufficiently discredit *La Salida's* narrative, the regime *diputados* attempted to distract from the protesters by discussing many different issues and developing competing narratives and talking points. We have shown that the regime tended to discuss more distinct topics per day during the protest than before or after the protest, and the opposite is the case for the opposition. We have also shown that the regime used more "discourse-structuring" hashtags during the protest, each of which was designed to develop a conversation unrelated to the talking points of *La Salida*.
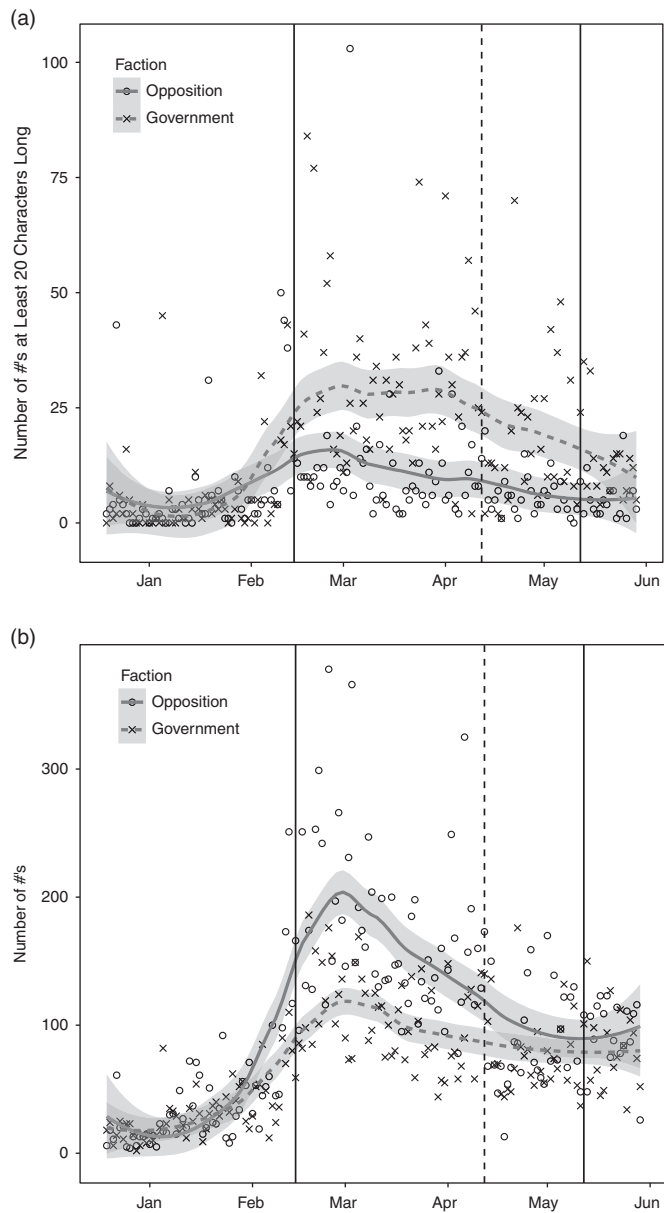
Fig. 4. Hashtag use

Note: *(a)* Plots the number of hashtags that were at least 20 characters long tweeted per day by the two colations. *(b)* expands this number to the total count of hashtags tweeted per day by the two coalitions. The solid vertical lines correspond to February 12 (the beginning of the protests in Caracas on National Youth Day) and May 8 (protest camps are removed from Caracas). The dotted vertical line represents the April 10 televised sit-down between Maduro and Capriles.

CONCLUSION

The stated goal of the *La Salida* movement, to remove the regime from power before the next election, was not accomplished during the summer of 2014. Although underlying macroeconomic

(a)

(b)
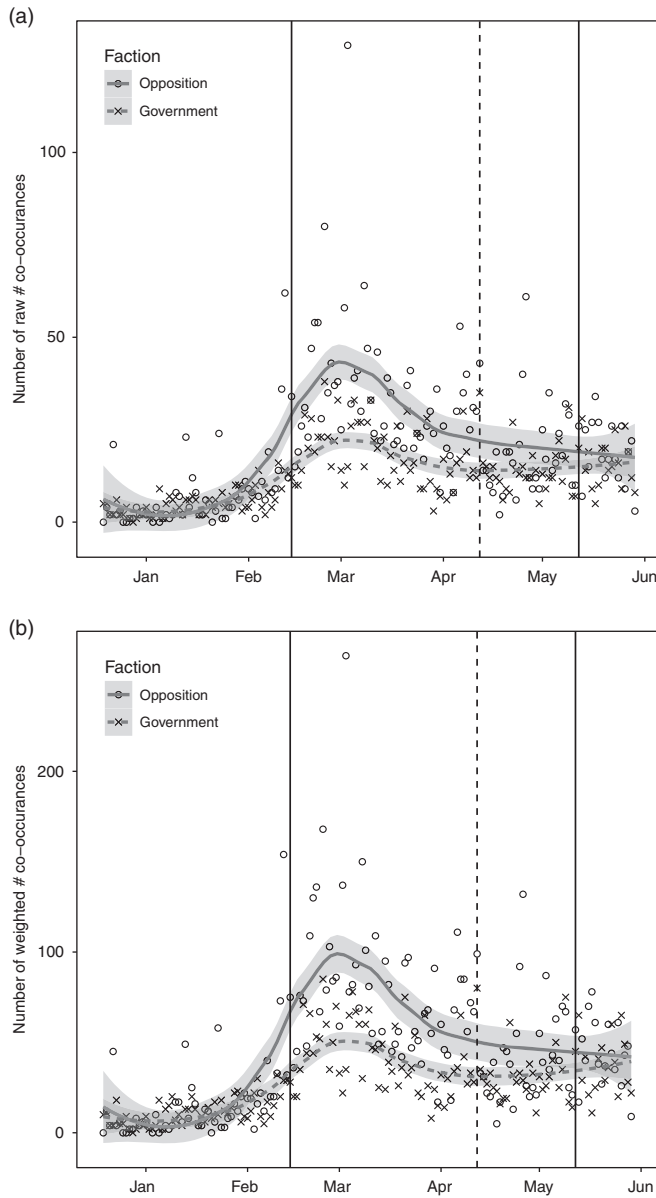
Fig. 5. Hashtag co-occurrences

Note: (a) Shows the raw number of tweets containing multiple different hashtags tweeted per day by the two coalitions. *(b)* Weights this number by the number of hashtags each of those tweets contained. The solid vertical lines correspond to February 12 (the beginning of the protests in Caracas on National Youth Day) and May 8 (protest camps are removed from Caracas). The dotted vertical line represents the April 10 televised sit-down between Maduro and Capriles.

and political concerns persisted, the regime effectively demonstrated their repressive capacity. Our findings show that part of their strategy involved actively engaging with Twitter to try and change the agenda away from the one driven by *La Salida*. The moderate opposition *diputados*, on the other hand, used Twitter to draw more attention to this established narrative, but only as long as it

was in their interest—they seemed to have abruptly changed after the April 10 sit-down with Maduro. The moderate opposition was apprently not permanently committed to the strategy of *La Salida*.

This change in strategy could not have been detected without the use of social media data. The underlying theoretical argument about the powerful trying to limit the scope of conflict and the less powerful trying to expand it dates back at least to Schattschneider (1960). Social media allows researchers to more easily observe and test these kinds of theories. While this article has set out to establish that social media is used by authoritarian regimes as a part of a strategy of regime resilience, and thus that social media use by elites in these countries is intrinsically important, it also affords opportunities to learn about strategic goals in otherwise opaque political contexts.

We cannot measure whether or not the Twitter strategies employed by the regime were decisive or even effective; that question is outside the scope of this article, but would be an excellent avenue for future research.

The quantitative textual analysis developed in this article is a novel implementation of well-established machine-learning techniques. Although looking at the topic diversity scores and distribution of top topics is less qualitative than traditional methods, it has the advantages of being objective and easily replicable. Moreover, what this approach allows us to study—the amount of information being transmitted—is essentially impossible for humans to measure. Machine learning allows researchers to extract patterns from elite communication, even if the literal content of that communication is propagandistic.

Future research should take even more advantage of idiosyncratic social media use by elites to try to better understand the unobservable inner workings of non-democratic regimes. An excellent example of this approach is Malesky and Schuler (2010), who use the behavior of Vietnamese legislators in questioning sessions to draw inferences about their incentives and that of the regime more generally. That institution is peculiar to Vietnam, but social media is not, and social media thus offers a similar avenue for analysis in many contexts.

REFERENCES

Agarwal, Sheetal D., W. Lance Bennett, Courtney N. Johnson, and Shawn Walker. 2014. 'A Model of Crowd Enabled Organization: Theory and Methods for Understanding the Role of Twitter in the Occupy Protests'. *International Journal of Communication* 8:646–72.

Bennett, W. Lance, Alexandra Segerberg, and Shawn Walker. 2014. 'Organization in the Crowd: Peer Production in Large-Scale Networked Protests'. *Information, Communication & Society* 17(2):232–60.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. 'Latent Dirichlet Allocation'. *The Journal of Machine Learning Research* 3:993–1022.

Blei, David M., and John D. Lafferty. 2007. 'A Correlated Topic Model of Science'. *The Annals of Applied Statistics* 1:17–35.

Chen, Jidong, Jennifer Pan, and Yiqing Xu. 2015. 'Sources of Authoritarian Responsiveness: A Field Experiment in China'. *American Journal of Political Science* 60(2):383–400.

Christensen, Darin, and Francisco Garfias. 2015. 'Can You Hear Me Now?: How Communication Technology Affects Protest and Repression'. Available at https://ssrn.com/abstract=2529769, accessed 11 August 2015.

Ciccariello-Maher, Georeg. 2014. 'LaSalida? Venezuela at a Crossroads'. The Nation. Available at https://www.thenation.com/article/lasalida-venezuela-crossroads/, accessed 1 September 2015.

Corrales, Javier. 2013. 'Chavismo After Chavez'. Foreign Affairs. Available at https://www.foreignaffairs.com/articles/venezuela/2013-01-04/chavismo-after-ch-vez, accessed 1 September 2015.

Corrales, Javier, and Michael Penfold-Becerra. 2011. *Dragon in the Tropics: Hugo Chavez and the Political Economy of Revolution in Venezuela*. Washington, DC: Brookings Institution Press.

Deibert, Ronald, John Palfrey, Rafal Rohozinski, Jonathan Zittrain, and Miklos Haraszti. 2010. *Access Controlled: The Shaping of Power, Rights, and Rule in Cyberspace*. Cambridge: MIT Press.

Diaz, Sara Carolina. 2014. 'Sector de la oposicin convoca a marcha para el 12 de febrero'. El Universal. Available at http://www.eluniversal.com/nacional-y-politica/140202/sector-de-la-oposicion-convoca-a-marcha-para-el-12-de-febrero, accessed 1 September 2015.

Earl, Jennifer, Heather McKee Hurwitz, Analicia Mejia Mesinas, Margaret Tolan, and Ashley Arlotti. 2013. 'This Protest will be Tweeted: Twitter and Protest Policing During the Pittsburgh G20'. *Information, Communication & Society* 16(4):459–78.

Edmond, Chris. 2013. 'Information Manipulation, Coordination, and Regime Change'. *The Review of Economic Studies* 80(4):1422–58.

Enikolopov, Ruben, Alexey Makarin, and Maria Petrova. 2015. 'Social Media and Protest Participation: Evidence from Russia'. Available at https://ssrn.com/abstract=2777540, accessed 31 May 2016.

Freedom House. 2014. *Freedom in the World 2014: The Annual Survey of Political Rights and Civil Liberties*. New York: Rowman & Littlefield.

Friedman, Uri. 2014. 'Why Venezuela's Revolution Will Be Tweeted'. The Atlantic. Available at https://www.theatlantic.com/international/archive/2014/02/why-venezuelas-revolution-will-be-tweeted/283904/, accessed 11 August 2015.

Greitens, Sheena Chestnut. 2013. 'Authoritarianism Online: What Can We Learn from Internet Data in Nondemocracies?'. *PS: Political Science & Politics* 46(2):262–70.

Griffiths, Thomas L., and Mark Steyvers. 2004. 'Finding Scientific Topics'. *Proceedings of the National Academy of Sciences of the United States of America* 101(Suppl 1):5228–35.

Gunitsky, Seva. 2015. 'Corrupting the Cyber-Commons: Social Media as a Tool of Autocratic Stability'. Available at https://doi.org/10.1017/S153792714003120, accessed 31 May 2016.

Hassanpour, Navid. 2014. 'Media Disruption and Revolutionary Unrest: Evidence From Mubarak's Quasi-Experiment'. *Political Communication* 31(1):1–24.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning*, vol. 2. New York, NY: Springer.

Hong, Liangjie, and Brian D. Davison. 2010. 'Empirical Study of Topic Modeling in Twitter'. Proceedings of the First Workshop on Social Media Analytics, Association for Computing Machinery, Washington, DC, 80–88.

Hornik, Kurt, and Bettina Grün. 2011. 'topicmodels: An R Package for Fitting Topic Models'. *Journal of Statistical Software* 40(13):1–30.

Howard, Philip N., and Muzammil M. Hussain. 2011. 'The Role of Digital Media'. *Journal of Democracy* 22(3):35–48.

Howard, Philip N., Sheetal D. Agarwal, and Muzammil M. Hussain. 2011. 'When Do States Disconnect Their Digital Networks? Regime Responses to the Political Uses of Social Media'. *The Communication Review* 14(3):216–32.

King, Gary, Jennifer Pan, and Margaret E. Roberts. 2013. 'How Censorship in China Allows Government Criticism But Silences Collective Expression'. *American Political Science Review* 107(2):326–43.

King, Gary, Jennifer Pan, and Margaret E. Roberts. 2017. 'How the Chinese Government Fabricates Social Media Posts for Strategic Distraction, Not Engaged Argument'. *American Political Science Review* 111(3):484–501.

Leshchenko, Sergii. 2014. 'The Maidan and Beyond: The Media's Role'. *Journal of Democracy* 25(3):52–57.

Mainwaring, Scott. 2012. 'From Representative Democracy to Participatory Competitive Authoritarianism: Hugo Chavez and Venezuelan Politics'. *Perspectives on Politics* 10(4):955–67.

Malesky, Edmund, and Paul Schuler. 2010. 'Nodding or Needling: Analyzing Delegate Responsiveness in an Authoritarian Parliament'. *American Political Science Review* 104(3):482–502.

Oates, Sarah. 2013. *Revolution Stalled: The Political Limits of the Internet in the Post-Soviet Sphere*. Oxford: Oxford University Press.

Pan, Jennifer. 2017. 'How Market Dynamics of Domestic and Foreign Social Media Firms Shape Strategies of Internet Censorship'. *Problems of Post-Communism* 64(3–4): 167–88.

Perez, Valentina. 2014. 'The Grim Reality of Venezuelan Protests'. *Harvard Political Review*. Available at http://harvardpolitics.com/world/grim-reality-venezuela-protests/, accessed 1 September 2015.

Phan, Xuan-Hieu, Le-Minh Nguyen, and Susumu Horiguchi. 2008. 'Learning to Classify Short and Sparse Text & Web With Hidden Topics From Large-Scale Data Collections'. Proceedings of the 17th International Conference on World Wide Web, Association for Computing Machinery, Beijing, China, 91–100.

Qin, Bei, David Strömberg, and Yanhui Wu. 2017. 'Why Does China Allow Freer Social Media? Protests Versus Surveillance and Propaganda'. *Journal of Economic Perspectives* 31(1):117–40.

Roberts, Margaret E., Brandon M. Stewart, and Dustin Tingley. 2014. 'stm: R Package for Structural Topic Models'. *R Package Version 0.6 1*. Available at https://cran.r-project.org/web/packages/stm/vignettes/stmVignette.pdf, accessed 1 June 2016.

Sanovich, Sergey, Denis Stukal, and Joshua A. Tucker. 2018. 'Turning the Virtual Tables: Government Strategies for Addressing Online Opposition with an Application to Russia'. *Comparative Politics* (forthcoming).

Schattschneider, Elmer E. 1960. *The Semi-Sovereign People: A Realist's View of Democracy in America*. New York: Wadsworth Publishing.

Schoonderwoerd, Nico. 2013. '4 Ways How Twitter Can Keep Growing-PeerReach Blog', Available at http://peerreach.com/2013/11/4-ways-howtwitter-can-keep-growing/, accessed 13 April 2015.

Shannon, Claude. 1948. 'A Mathematical Theory of Communication'. *The Bell System Technical Journal* 27:379–423.

Tufekci, Zeynep. 2014. 'Social Movements and Governments in the Digital Age: Evaluating a Complex Landscape'. *Journal of International Affairs* 68(1):1–18.

Tufekci, Zeynep, and Christopher Wilson. 2012. 'Social Media and the Decision to Participate in Political Protest: Observations from Tahrir Square'. *Journal of Communication* 62(2):363–79.

Elites Tweet to get Feet off the Streets:
Measuring Regime Social Media Strategies During Protest
Online Appendix

Kevin Munger, Richard Bonneau,

Jonathan Nagler, Joshua A. Tucker[*]

December 26, 2017

Table 1: **Top Terms for Relevant Topics**

	Top Government Topics
#	Terms
44	nicolasmadur, venezuel, puebl, psuv, paz, president, nuev, chávez, dcabellor, madur
76	suramerican, eduardopinata, zerp, campament, cati, háro, may, rendón, oro, amesesdetusiembracomandant
77	celac, ener, pacificación, cumbr, contrab, caribe, natalici, haban, hagamoslapaz, magallan
	Top Opposition Topics
#	Terms
74	papel, periód, violenci, lasal, segur, matern, sinpapelnohayperiód, medi, pilieri, biagi
94	estudiant, venezuel, protest, puebl, paz, march, call, maduro, hcapril, hoy
72	prmerojustici, gobiern, diput, hoy, más, juliocmontoy, país, dip, vía

Appendix

A   Original Terms for Table 2

Table A contains the untranslated terms in Table 2.

B   Placebo Check for Focus

The results of the primary analysis in Table 2 depict trends in Twitter usage. The claim is that these trends correspond to important real-world events, so it's important that a similar (or similarly distinctive) graph does not obtain for data collected from a less tumultuous time period.

To that end, we collected the tweets from the same accounts from May 30th to October 13th, an arbitrary end date that was simply the present day. We collected 88,328 new tweets, 46,679 from the regime and 41,649 from the opposition. The ratio of regime to opposition is thus close to 1:1, compared to 2:3 in the original analysis. We ran the same LDA model as before, and used the same measure of topic diversity to track *focus*. The results are depicted in Figure B below.

Very little changes during the four months, providing supporting evidence that the dynamism observed during the crisis is unique and important. Though the government and opposition evince similarly stable trends in Figure B, the opposition tends to have slightly higher topic diversity scores and thus be less *focused*, further marking their high *focus* during the protests as significant.

Figure B

Placebo Check: Focus After the Protests

days_p

C   Timeline of Events

Regime *Diputados*   Protest   Opposition *Diputados*

	2013-2014: Maduro succeeds Chávez in a close election; inflation and crime worsen	
	January 6: Miss Venezuela is murdered; there is a brief outcry about the crime	
Start tweeting more but do not become less *focused*; the threat is not yet realized	January 23: López and Machado announce their strategy of *La Salida* to remove Maduro	Start tweeting more and become somewhat more *focused*; there are still many issues to discuss
	February 4: Attempted rape of a student at a regional university; student protests are violently repressed	
Begin to become less *focused*	Feb 12: Protests emerge in Caracas and around the country, organized by radical elites like López	Remain *focused* on the protests
Feb 15-22: Call for López's arrest and call him a terrorist	Feb 18: López turns himself in for arrest	
Continue tweeting frequently but without increasing *focus*	Feb-April: Caracas is paralyzed by *guarimbas* and recurrent violent clashes between protesters	Reach maximum *focus*, support the protest
Remain *unfocused*	April 10: First major sit down between regime and moderate opposition	Become much less *focused*, begin tweeting about economic policy instead of the protest
	May 8: Regime clears out remaining protest camps, essentially ending the protest	

D   Table of Raw Hashtag Frequencies

These are the raw counts of the hashtags used by each coalition. Inspection of the two lists provides evidence for our claim that short hashtags are used to denote specific places, people or events, while long hashtags contain more crafted talking points or narratives.

D.1   Government Hashtags with Counts

#psuv 1618

#venezuela 1079

#tropa 381

#chavez 172

#caracas 109

#eeuu 100

#cuba 93

#rodillaentierraconnicolasmaduro 85

#conelmazodando 73

#venezuelaserespeta 69

#jpsuv 64

#usa 64

#dialogomaduroporlapaz 64

#a1a 62

#abrilvictoriosoenrevolucion 62

#envivo 60

#madurohombredepaz 58

#tachira 56

#juventud 52

#conferenciadepaz 52

#desconectatedelaguarimba 48

#barinas 47

#ahora 46

#apure 45

#maduroprotectordevenezuela 45

#maduroespueblo 45

#anzoategui 44

#lafoto 44

43

#apasofirmelarevolucion 43

#maduroespuebloprospero 43

#mujerescontraelfascismo 42

#deyarealcorazondelpueblo 42

#dialogosipactono 41

#13apatriadigna 38

#a10mesesdetusiembracomandante 37

#hagamoslapaz 37

#15a 37

#ch 36

#a14mesesdetusiembracomandante 36

#bolivar 34

#4frevolucionparasiempre 34

#encontactoconmaduronro6 34

#chavezvivelapatriasigue 33

#estudiantes 33

#vzlabajoataquemediatico 33

#maduroespueblochavista 33

#rt 32

#chadertonvozdevzla 32

#vargas 31

#unasur 31

#fueracnndevenezuela 30

5

#loschavistasnoperdemosbatallas 30
#desconectatedelodio 30
#encontactoconmaduro 30
#barriosunidosconmaduro 30
#conmadurofuturodepaz 30
#mayoproductivoconmaduro 30
#rusia 29
#merida 29
#construyendolapatriasegura 27
#somosfanb 26
#conmadurotriunfalapaz 26
#vzlaunidacontraelfascismo 25
#gringosyfascistasrespeten 25
#conmadurocarnavalseguro 25
#auna 25
#mayodelostrabajadores 25
#foto 24
#opini 24
#di 24
#lara 24
#derechaasesinayfascista 24
#vtv 23
#circuito2 23
#envideo 23
#novolveran 23
#venezuelapueblodepaz 23

#pueblounidoconmaduro 23
#revolucionhaciendopatria 23
#encontactoconmaduronro7 23
#video 22
#zulia 22
#vzlaserespeta 22
#soyvenezuelanomasguarimba 22
#encontactoconmaduronro4 22
#antv 21
#twitter 21
#colombia 21
#a11mesesdetusiembracomandante 21
#a13mesesdetusiembra 21
#porelcaminodechavez 21
#ciudadccs 20
#entrechavistasnosseguimos 20
#trujillo 20
#juventudconmaduroalavictoria 20
#oea 20
#chavezcorazondelapatria 20
#gnbguardianadelapaz 20
#viveenpazsinguarimbas 20

D.2   Opposition Hashtags with Counts

#pjenvivo 905
#venezuela 687
#m 344
#caracas 232
#sosvenezuela 227
#nomasviolencia 188

#merida 176
#guayana 176
#12f 174
#yaracuy 143
#anzoategui 130
#lasalida 127

#tachira 102
#t 102
#12m 99
#22m 94
#sinpapelnohayperi 83
#protestaporlosproblemasdelagente 83
#barquisimeto 81
79
#notit 79
#envivo 78
#lecheria 77
#22f 76
#convencionpj 76
#nom 75
#zulia 74
#viernesamarillo 74
#valencia 70
#resistenciavzla 70
#maracaibo 66
#elquesecansapierde 65
#cuba 64
#ucv 64
#carabobo 62
#la 61
#casaxcasa 60
#aumentochucuto 60
#8m 58
#nohay 57
#lacausar 57
#16m 57
#lucharhastavencer 55
#2m 55
#10m 55
#puertoordaz 54

#asambleaspopulares 53
#lara 50
#an 49
#altamira 49
#paquetazorojo 49
#amnistiaya 49
#noticias 48
#barcelona 48
#sosvzla 47
#chacao 45
#toquedequeda 45
#5m 44
#facebook 43
#oea 43
#14m 43
#rt 42
#primerojusticiatv 42
#26f 41
#sossaludve 41
#barinas 40
#sancristobal 40
#26a 40
#cojedes 39
#maturin 39
#18f 39
#4m 39
#vargas 38
#margarita 38
#sinopuedesrenuncia 37
#prayforvenezuela 36
#27m 36
#ciudadbolivar 36
#libertad 36
#yabasta 36
#sinpapelnohayperiodico 35

#mujeresporlavida 35

#miranda 34

#estudiantes 34

#estarinformadoesunderecho 34

#resistencia 34

#noalracionamiento 34

#protestasenvenezuela 33

#6m 33

#tuvozenlaoea 33

#nicol 33

#1m 33

#20m 33

#24m 33

#porestoprotestamos 33

#opini 32

#27f 32

#lacallevence 31

#360ucv 31

#anzo 30

E   Robustness Check–Hashtag Length

The analysis of the strategic use of "long" hashtags sets the cut-off point at twenty characters; our findings are robust to different cut-off points. Below in Figure E are two versions of Figue 5(b); the top one has a cut-off of eighteen characters, and the bottom one has a cut-off of twenty-two characters. In neither case is the result substantially different from that shown in the body of the paper, although the time period in which there is a significant difference between the two strategies is longer for the case of twenty-two characters.

F   Robustness Check–Correlated Topic Model

Analysis using an alternative topic model, called the Correlated Topic Model, leads to similar inferences about the changing communication strategy of the regime *diputados*.

We fit the Correlated Topic Model using the stm **R** package developed by **?**, which reduces to the CTM in the absence of covariates. We used the same preprocessing steps as in the LDA model in the body of the text: we stemmed the text and removed stopwords and numbers. As with the LDA model, we calculated the semantic coherence of each topic and kept only topics above the semantic coherence threshold of -150 selected for the LDA model; unlike with the LDA model, there was no obvious clustering between "high" and "low" levels of semantic coherence. The CTM model tended to produce topics with higher semantic coherence overall, so this process retained 68 topics compared to only 43 for LDA.

Just as in the LDA model, the regime becomes less *focused* during the protest than before it; indeed, this trend is stronger than in the LDA results. However, in contrast to the main results, the opposition also appears to be getting less *focused* in the build-up to the protest. The opposition trend levels off during the protest while the regime continues to decrease its focus, comporting with the findings from the LDA model, but this contrast is less pronounced.

E: Robustness Check–Hashtag Length, 18 and 22 Characters
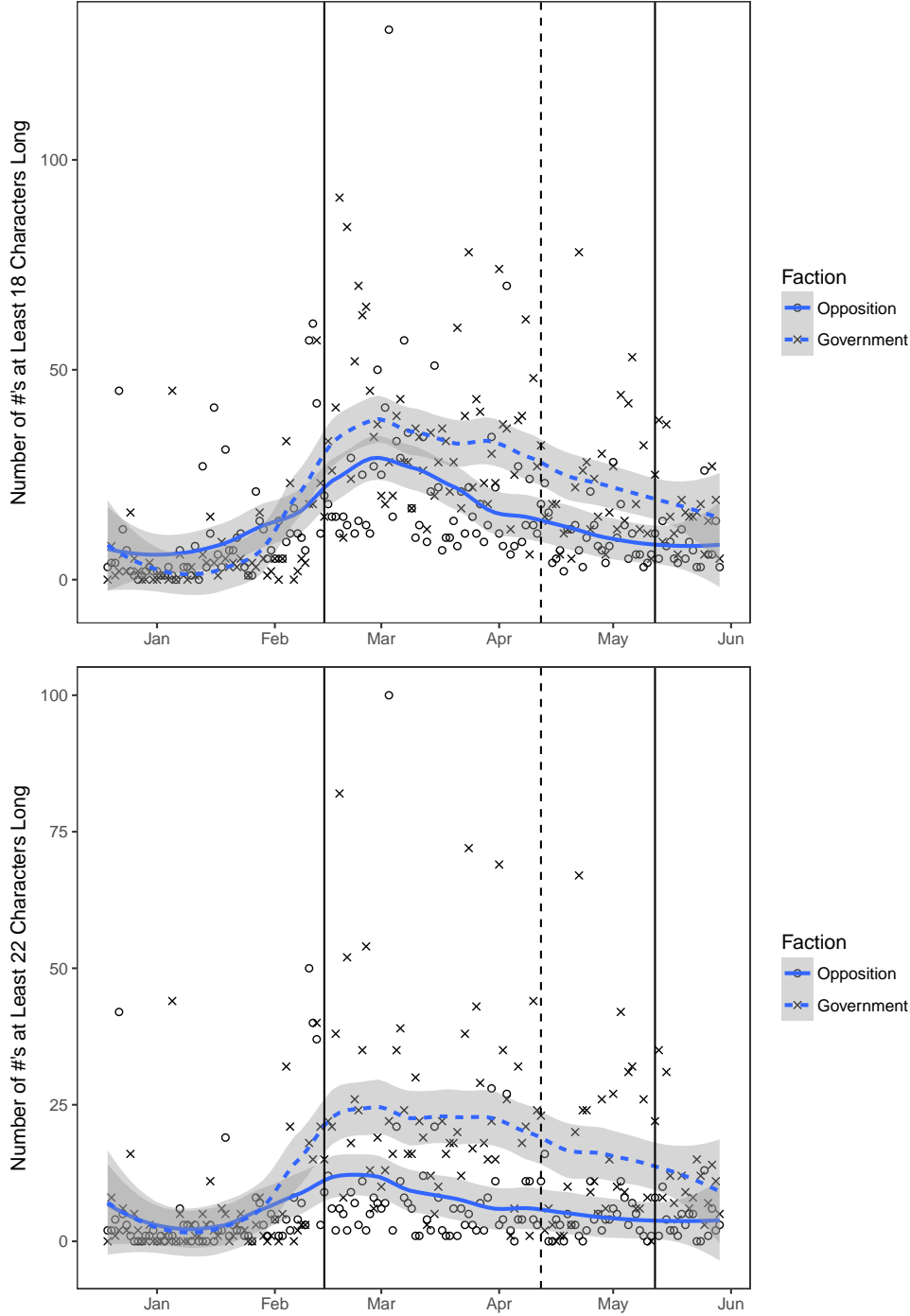
The number of tweets sent by *dipudatos* from each faction per day. The solid vertical lines correspond to February 12 (the beginning of the protests in Caracas on National Youth Day) and May 8 (protest camps are removed from Caracas). The dotted vertical line represents the April 10th televised sit-down between Maduro and Capriles. The top panel shows the number of hashtags used by each coalition on a given day, that were at least 18 characters long; the bottom panel, 22 characters long.

F: Robustness Check–Correlated Topic Model
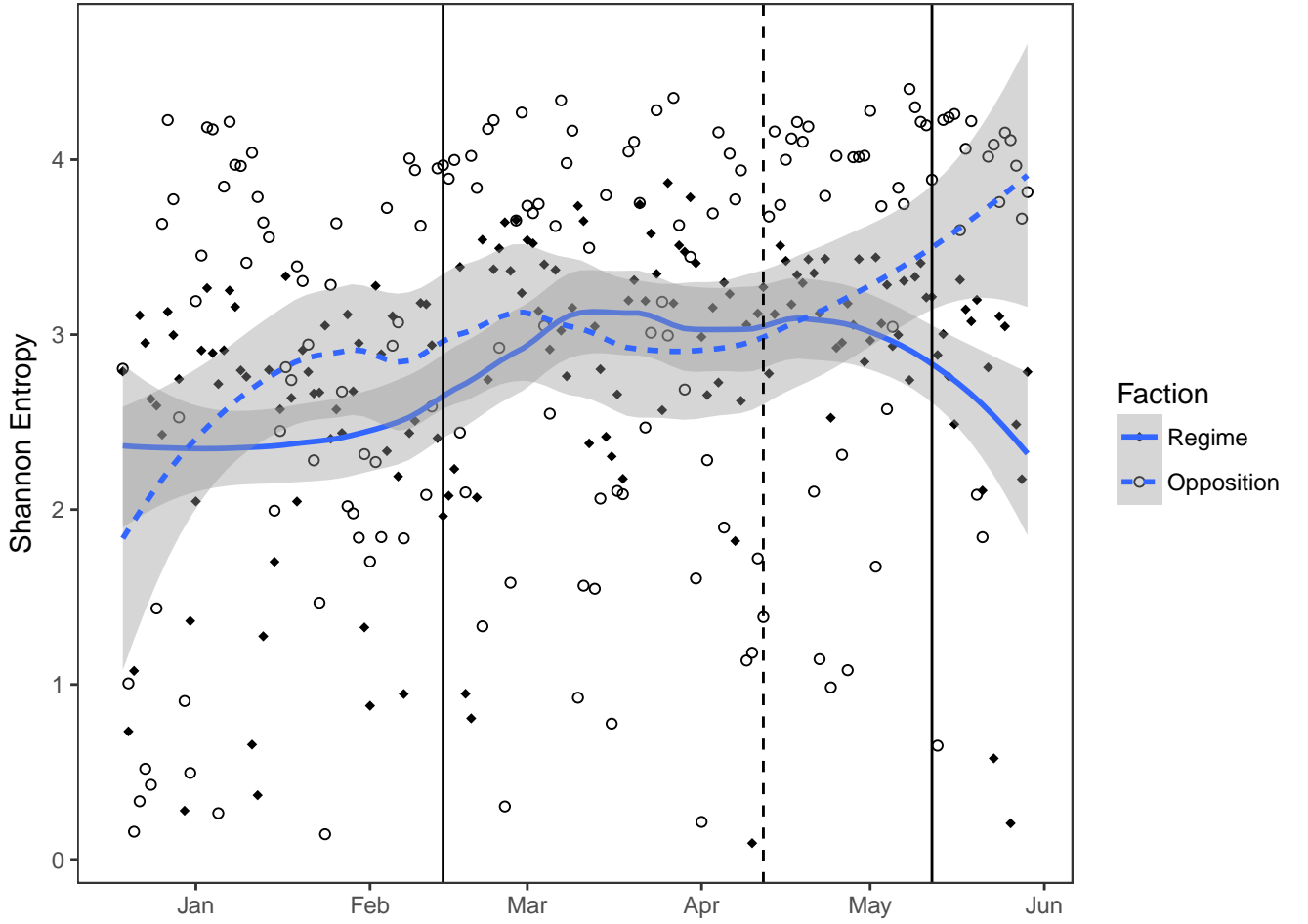
Each point represents the Topic Diversity score for the opposition and regime tweets respectively, compuated based on their tweets using the Correlated Topic Model. The solid vertical lines correspond to February 12 (the beginning of the protests in Caracas on National Youth Day) and May 8 (protest camps are removed from Caracas). The dotted vertical line represents the April 10th televised sit-down between Maduro and Capriles. Trend lines and 95% confidence intervals were created with loess.

References

Roberts, Margaret E, Brandon M Stewart, Dustin Tingley, Edoardo M Airoldi et al. 2013. "The structural topic model and applied social science." *Advances in Neural Information Processing Systems Workshop on Topic Models: Computation, Application, and Evaluation* .