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

Kevin Munger, Richard Bonneau, John T. Jost
Jonathan Nagler, Joshua A. Tucker[*]

August 4, 2016

[*]The writing of this paper was supported by the Social Media and Political Participation Lab at NYU, which is funded in part by the INSPIRE program of the National Science Foundation (Award SES-1248077), and New York University's Dean Thomas Carew's Research Investment Fund. This paper was written in conjunction with NYU's Social Media and Political Participation (SMaPP) lab, of which Munger is a PhD student member and the remaining authors are Principal Investigators.
km2713@nyu.edu

Abstract

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 pursue different rhetorical strategies in keeping with our predictions about managing the conflict advanced by the protesters. This paper 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.

1 Introduction

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., 2015), 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 better 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 paper 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 (PeerReach, 2013), and 2014 data suggests that 21% 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 Nicolàs Maduro, and the protesters' *en guarimba* (blockade) paralyzed parts of the capital of Caracas from February 12th to May 8th.

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 protests, ignoring the specific grievances of the dissidents and instead trying to create a competing discourse that conforms to their preferred narrative.

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

[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).

related to the protests.[2] Conversely, the pro-Maduro coalition will bring up new issues unrelated to the protests, to give people the impression that *La Salida's* narrative is only one of many possible ways of viewing the political situation. The regime will use Twitter's "hashtag" function to insert explicitly crafted talking points into the public discussion.

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, 139 of the 165 *diputados* had Twitter accounts, and both regime and oppostion *diputados* are well-represented. They tweeted over 100,000 times during the 5 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 talks about specific people and places, whereas the regime promotes new top-down narratives.

2 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 representative democracy to what Mainwaring (2012) calls "participatory competitive authoritarianism", a system which could be subsumed in the general category of a

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

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

"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%, 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 12th (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*.

As planned, protests began in Caracas and 38 other cities on February 12th, 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"[4] (authors' translation). 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.

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

[4]The tweet begins with a period so that it is visible to everyone; normally, tweets that begin with @[username] are only visible to people who "follow" both parties.

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 8th.

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, 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.

3 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 (Christensen and Garfias, 2015; Tufekci and Wilson, 2012) and for those protests to be better organized (Agarwal et al., 2014; Bennett, Segerberg, and Walker, 2014; Earl et al., 2013).

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 approaches are fairly crude, and aim to hinder opposition groups and return to the pre-social media balance of power.

The "third generation" strategies, including information campaigns and actively paying supporters to post online, 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. (2015) demonstrate the dramatic changes in online engagement strategy in Russia since 2000. Depending on the technological and organizational capacities of the two sides, this regime activism might even tip the balance of power in favor of the regime (Morozov, 2011; Rahimi, 2011).

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). The use of social media by both central and local elites acts as a credible way for them to coordinate, lowering monitoring costs and lessening the chance of shirking on the part of local elites (Boix and Svolik, 2013).

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 US, 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 convince citizens that the there are other, more important things to pay attention to. The effects of infiltrating and diverting discussions on social media using paid pro-regime trolls should be amplified if their messages concord with those strategically sent out by elite actors. In this sense, the regime will engage in an attempt at agenda control: social media has democratized communication in non-democratic countries, and therefore it makes sense to ground authoritarian social media strategies in theoretical approaches usually applied to democratic contexts.

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 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. We do not claim to formally test the model in the current paper, but the confluence of predictions from the descriptive and formal literatures is encouraging.

A famous result in public opinion and mass politics is that citizens are more likely to think that issues that are commonly discussed in the media are important (Iyengar and Kinder, 1987; McCombs and Shaw, 1972). *La Salida*'s narrative was primarily

[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).

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; most citizens would evaluate the regime poorly if these were the most salient dimensions. However, by bringing up many different issues and not engaging with the protesters on their own terms, the regime will raise the salience of other issues. People who evaluate the regime primarily on their efforts to help the poor, improve healthcare, and stand up to the US are unlikely to view the regime in a negative light.

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 *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*.

4 Hypotheses

We test the theory of active, strategic use of social media during 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, to make the opposition's narrative less central to the national discourse. This will entail the use of a concerted, top-down media strategy.

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. First, 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* in two ways. The first is 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.*

We also operationalize *focus* by looking at which topics are most prominent during each day's worth of tweets from each coalition. During the protest, we expect the regime to make many different topics the main talking point for at least one day. For the opposition, however, we expect very few different topics to be the main talking point for at least one day. H_1 is concerned with the overall topic distribution, while H_2 tests the consistency of top topics.

Hypothesis 2 *During the protest, the opposition diputados will emphasize fewer distinct topics than will the regime diputados.*

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.[6] These implications are related because of Twitter's 140-character limit: longer hashtags limit the number of hashtags that can be used in a single tweet. Long hashtags are useful, 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 3 *During the protests, the regime diputados will use long hashtags more frequently than the opposition.*

Hypothesis 4 *During the protests, the regime diputados will send tweets containing multiple hashtags less frequently than the opposition.*

5 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.[8] 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.[9] We were able to locate accounts for 139 of the 166 *diputados* (84%): 63 of 66 for the opposition (95%, similar to US Members of Congress), and 76 of 99 (77%) for the regime.[10] For the subsample whose tweets were analyzed

[6] In fact, Twitter's website discourages abuse of this practice: "Don't #spam #with #hashtags. Don't over-tag a single Tweet. (Best practices recommend using no more than 2 hashtags per Tweet.)" (from Twitter.com as of February 2016, mentioned in Bennett, Segerberg, and Walker (2014).)

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

[8] In some cases, there were multiple accounts associated with a single politician–either a campaign account and a governing account or an official account and a personal account–but in all cases of multiple accounts, there was only one that was both active and which possessed a significant number of followers, which we then included in the study.

[9] It turns out that several Venezuelan politicians share names with professional baseball and soccer players.

[10] To check the validity of our selection, we had a research assistant recreate our analysis. There were only 2 discrepancies, the adjudication of which was obvious.

Table 1: **Number of Tweets by Venezuelan *Diputados***

Diputados	N	1st Quartile	Median	3rd Quartile	Mean	Total Tweets
Regime	65	109	308	799	664	43,174
Opposition	56	215	584	1234	1,117	62,534

Period of Analysis: December 19, 2013 - May 29, 2014

in this article–those who tweeted after December 18, 2013–there were 135 accounts, 65 regime and 56 opposition.

We used Twitter's REST API[11] via `tweepy`[12], in the Python programming language, to collect the most recent tweets for each account. Using the `/statuses/user_timelines` endpoint[13], Twitter's API allows fetching the latest 3,200 tweets for a given account.[14] 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 more than 3,200 tweets for some accounts. Many of the accounts had fewer than 3,200 tweets, so we have their entire history. Twitter's API also provides the metadata associated with each account, including their "biography," where they claim to be located, and the date they joined Twitter. We entered each *diputado's* party as an additional variable in the dataset.

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% 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 accounts; comparing the 1st quartiles, medians, and 3rd 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-February 11, before the protests; (2) February 12-May 8, protest blockades paralyze Caracas; (3) May 9-May 28, after the protests. The beginning and end of the main

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

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

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

[14]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.

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 and a summary of concurrent observed elite twitter behavior.

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 Maria Corina Machado being 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; note that in this figure, there is no noticeable change around the second vertical line, April 10, the televised sit-down between Maduro and Capriles.

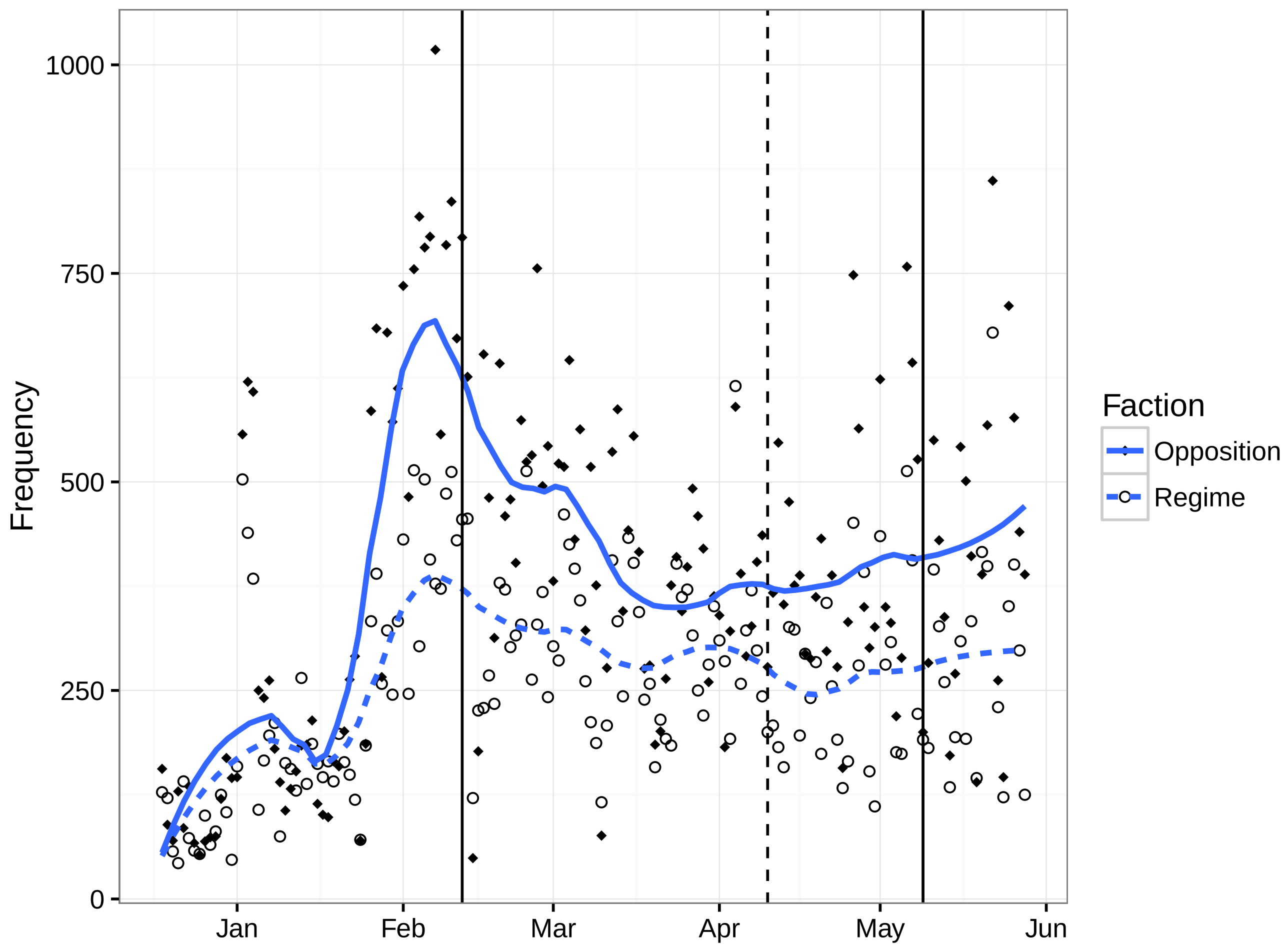
6 Analysis

To estimate the number of overall topics being discussed on a given day by each coalition, we employed Latent Dirichlet Allocation (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 (that is, the researcher does not decide which topics to study *ex ante*) and because it allows the entire corpus of information to be used.[15]

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,

[15]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 robust to using the CTM; see Appendix F.

Figure 1: Tweets per Day by Each Coalition



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. Trend lines for each coalition created with loess.

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, 2 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. Once aggregated into these documents, the terms comprising the tweets from that coalition-day 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 degree of coordination among the coalitions, however, this information is not germane.[16] The 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).[17]

To determine the number of topics k, we performed ten-fold cross-validation of both log-likelihood and perplexity analyses on the holdout sample.[18] Although the model fit improves monotonically in the number of topics, the gains from adding more topics

[16]This would be a problem if there were fewer accounts, or if one account were doing all the tweeting. Ideally, the tweets would be distributed among the accounts in a closer approximation to the normal distribution than the observed logarithmic distribution, but because the shape of the distribution is similar for the two coalitions, this weakness does not impair our comparative analysis.

[17]There are recent innovations in machine learning that attempt to improve on LDA performance modeling topics generated from short-length texts like tweets, either by using word dyad cooccurance (Cheng et al., 2014) or by pre-pooling the tweets by hashtags (Mehrotra et al., 2013); once these techniques are tested and fully validated, they could represent improvements on the naive tweet-pooling technique used in this manuscript.

[18]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.

diminish at around 50 topics (see Figure 1). 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.[19] 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.

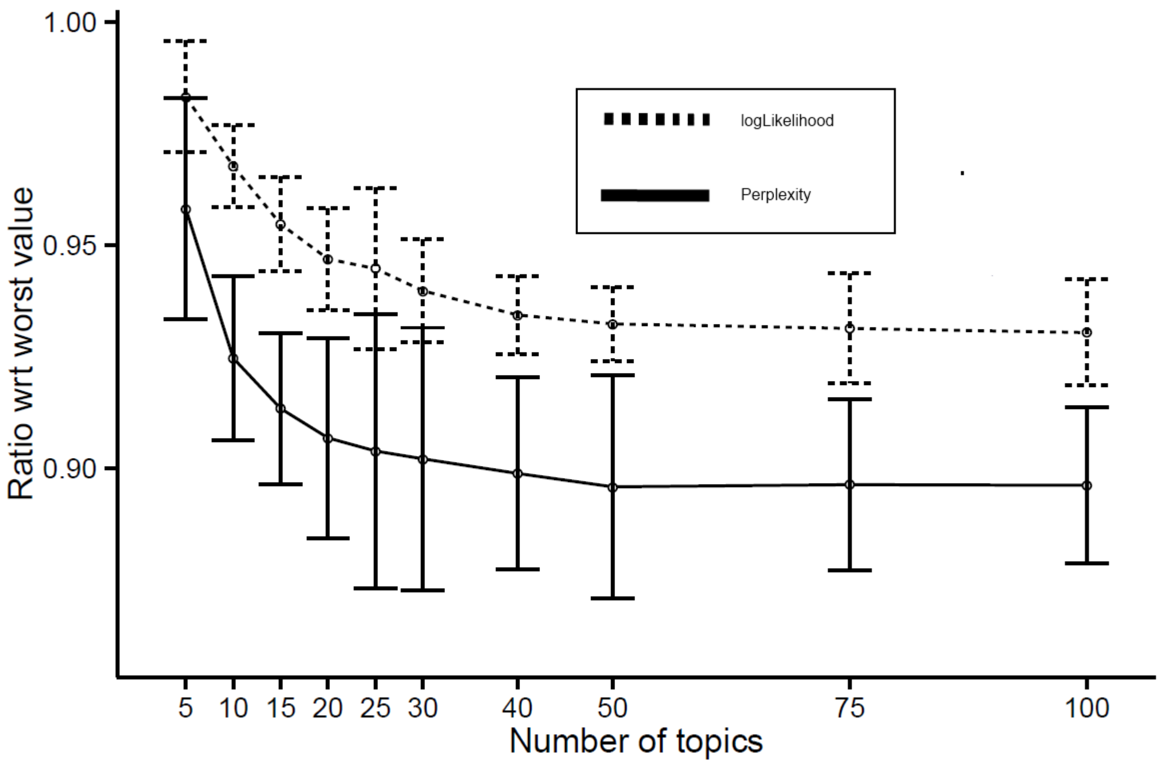
Following standards in the social science literature, we use the collapsed Gibbs sampler (Griffiths and Steyvers, 2004; Phan, Nguyen, and Horiguchi, 2008), a modification of the sampling method proposed in Blei, Ng, and Jordan (2003). Using the R package 'topicmodels' developed by Hornik and Grün (2011), 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 (collection of words) consisted of $N = 50,902$ terms.

In creating the topics, the algorithm estimates k=100 γ's 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 γ's.

To analyze how *focused* the coalitions are over time, we measured the Shannon Entropy (Shannon, 1948) 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 paper 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

[19] For α, we follow the standard established by Griffiths and Steyvers (2004) and set $\alpha = .5$, where $50/k = .5$.

Figure 2: Testing Different Numbers of Topics



The dotted line connects log-likelihood estimates for the model fitted with different number of topics, the dark line connects perplerxity estimates for the model fitted with different number of topics. Vertical bars are 95% confidence intervals. Models fit on the entire corpus of tweets.

imposition of arbitrary thresholds. The formula for Shannon Entropy is $-\sum_{i=1} p_i \log_2(p_i)$. In this case, because the γ's in each document must sum to 1, $p_i = \gamma_{wk}$ and the topic diversity score for each document is

$$\text{Topic Diversity} = -\sum_{k=1}^{100} \gamma_{w,k} \log_2(\gamma_{wk})$$

The possible topic diversity scores range from 0 (if the γ distribution is unitary) to $\log_2(k = 100)$ (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. Note that LDA identifies some "topics" that are not semantically meaningful and assigns them non-zero weights; as a result, all of our diversity scores are somewhat inflated, but this is not a threat to our analysis because of our difference-in-differences approach.

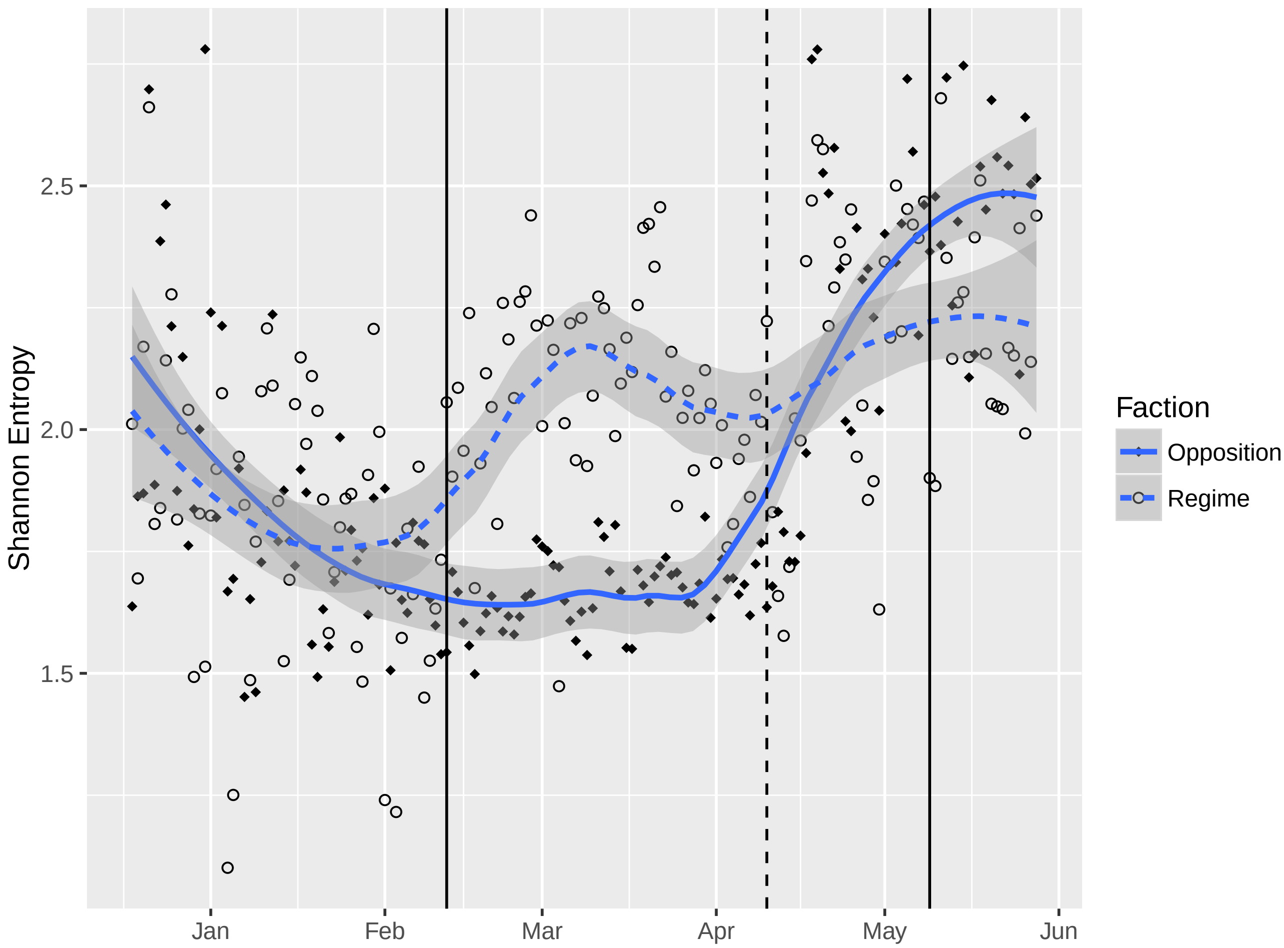
7 Results

We divide the time period under study into three periods:(1) December 19-February 11, before the protests; (2) February 12-May 8, protest blockades paralyze Caracas; (3) May 9-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.

7.1 Tweet Focus–topic diversity

The central prediction in H_1, implied by the argument that the regime cannot silence the protester's narrative but must instead try and distract from it and bring up other issues, 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 trend for both the regime and opposition *diputado* diversities are identical in the first period: each coalition is becoming more *focused* during the regional protests over crime and inflation. On the day of planned, country-wide protests that shut down Caracas, the government's trend abruptly changes: after nearly two months of indistinguishable levels of *focus*, the regime becomes significantly less *focused* than the opposition during the protests.

Figure 3: Focus–as Modeled by Topic Diversity–Over Time



Each point represents the Topic Diversity score for the opposition and regime tweets respectively, compuated 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 10th televised sit-down between Maduro and Capriles. Trend lines and 95% confidence intervals were created with loess.

The difference in topic diversity during the protests is significant at $p < .001$, using a two-tailed t-test. In fact, during the height of the protests, from February 12 to April 10, there was only a single day in which the topic diversity for the opposition was higher than for the regime. Additionally, the 95% confidence intervals generated by non-parametric local regression do not overlap until April 10th (see Figure 3). We thus find strong support for H_1.

After April 10, however, the opposition topic diversity measures begin to rise. This suggests that the messaging strategy of the opposition may have changed.[20]

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. The opposition *diputados* wanted to present a unified front against the regime and thus were careful to *focus* attention on the protest without going so far as to openly call for violent revolution. They may also have been using the pressure that the protests represented to improve their bargaining power with the regime with respect to policy. This explanation comports with the timing of the first major sit-down between the regime and the moderate opposition on April 10. Though these talks were widely seen as a failure in that they did not produce any concrete changes, they could have represented a public show of power to determine the relative strengths of the bargaining parties. If this was the case, then the talks may have convinced the moderate opposition that the government was willing and able to maintain the status quo.

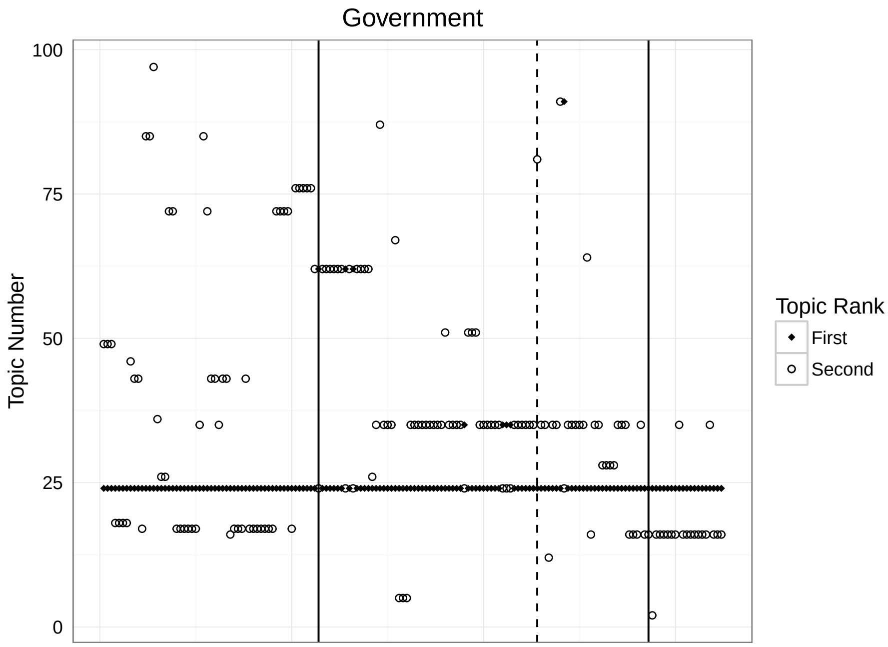
This unexpected development does not undermine the support for H_1; we had no expectations about how the coalitions would respond to the end of the protest. In fact, that we noticed the change in the behavior of the opposition illustrates a strength of the style of unsupervised text analysis of elite communication employed in this paper: it allows us to detect trends that take place at a scale that might elude a human analyst, and to at least make guesses as to the underlying strategy driving this change.

7.2 Tweet Focus–Content of Top Topics

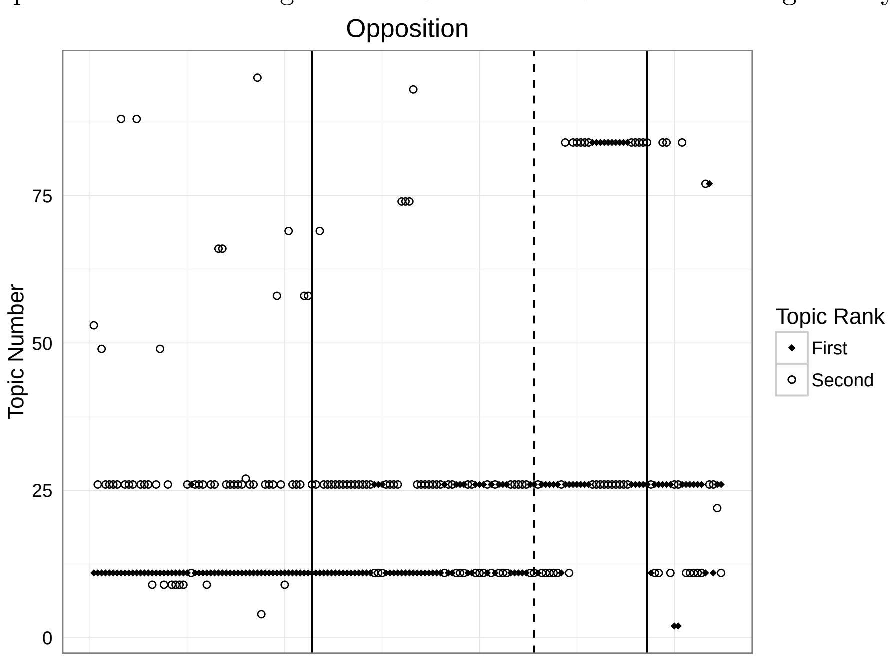
Using the topic diversity to operationalize *focus* captures information about the entire distribution of topics, but it is agnostic about which topics are the most important for

[20] 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. The placebo check in Appendix B, applying the same analysis as in Figure 3, demonstrates a lack of trend in the four months after the period studied here.

Figure 4: Top Topics Over Time



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. Cell entries indicate whether that topic number was the number 1 (diamond) or number 2 (circle) topic that day. The three most popular government topics, from top to bottom, are: 62, which paints López as a fascist; 35, which describes the opposition as terrorists; and 24, which equates the Maduro regime with Chávez and Venezuela more generally.



The three most popular opposition topics, from top to bottom are: 84, that the regime ruined the economy; 26, which promotes the student protest; and 11, encouraging people to work and vote with Capriles and his faction.

Table 2: **Top Terms for Relevant Topics**

Top Government Topics	
#	Terms
24	*nicolasmaduro*, Chavez, the people, PSUV, president, Venezuela, new, *dcabellor*, Maduro, homeland
35	Venezuela, *jmontillapsuv*, peace, terrorism, protest camp, Venezuelan, opposition, violent, rightist, violent protesters
62	peace, fascist, violent, violence, fascism, Venezuela, Leopoldo, march, rightist, López
Top Opposition Topics	
#	Terms
11	today, work, vote, the people, *hcapril*, unity, day, diputado, good, Venezuela
26	government, Venezuela, student, Maduro, protest, path, today, country, this way, diputado
84	*prmerojusticia*, *juliocmontoy*, *americodegrazi*, increase, law, *williamsdavil*, earn, day, economy, path

These topics are the most commonly discussed by each coaltion. The terms have been translated from Spanish by the lead author; terms in italics refer to specific Twitter accounts. The original Spanish terms can been found in Appendix A.

each day. Discussion could be evenly split between topics 1-50 on one day and evenly split between topics 51-100 the next and yield identical topic diversity scores. Both conceptions are essential to understanding the two coalitions' communication strategy, and Figure 4 provides a more detailed look at which topics are most central to each coalition's discussion on each day. The x-axis shows the timeline of the protests, and each day has two corresponding marks on the graph: the diamond indicates the topic that captured the highest percentage of the coalitions' tweets for that day, and the circle indicates the topic with the second highest percentage. We would find support for H_2 if, during the protest, there were fewer different topic numbers that made it to the top two topics for the opposition than for the regime.

The government did hold more consistently to their number one topic, but it also featured a wider range of second topics (25) in comparison with the opposition (18). During the protest, though, there were only 6 opposition top topics compared to 14 for the government. This comparison is not amenable to formal hypothesis testing, but it is consist with the expectation in H_2. Note that this is not simply an artefact of the government tweeting more; as noted above, the opposition actually produced 40% more total tweets, making this difference during the protest even more striking.

Table 2 lists the most common terms used in each of the most important topics.[21] Topic 24 cements the Chávez-Maduro connection and includes the President of the National Assembly (Cabello), the acronym of the regime's party (PSUV) and the words for "the people" and "homeland." Topic 62 is a clear indictment of López; five of the terms are designed to paint him as a "violent right-wing fascist." Topic 35 is a more general condemnation of the opposition's violent ("terrorist") tactics, including two variations on *guarimba*, the Venezuelan term for the permanent protest camps that also served as blockades designed to paralyze the government in Caracas.

For the opposition, Topic 11 seeks to connect "the people" with moderate opposition leader Henrique Capriles and asks them to "vote" and "work," and to do these things "today." Topic 26 describes the student protests as the way forward for the country and specifically addresses the government and governmental officials. Topic 84 is harder to interpret, but a majority of the terms have to do with government-mandated wage increases. Every year, the government raises mandatory salaries for the majority of workers on May 1, but in 2014 the nominal increase was only 30%, far less than the unofficial inflation rate of around 60%. Topic 84 also addresses *Primero Justicia*, a prominent opposition party headed by Capriles, and several affiliated *diputados*.

The specific topics reaching the top two are indicative of particular communication strategies they adopted at different points. The government switched from its overwhelming favorite topic (24: the regime is Venezuela) and adopted a new topic (62: López is a fascist) around the time of López's arrest. At the same time, its second topic (35: the opposition are terrorists) solidified during the main phase of the protest. For the opposition, however, the largest change is again seen in the wake of Easter/Beginning of the Independence Movement Day, as its previous top topic (11: work with Capriles) is abandoned when a new topic (84: the regime ruined the economy) comes to the top of the discussion. There are various other topics that are similarly event-specific[22], but our choice of a large value for k makes this kind of analysis, sometimes central to LDA models of political discussions over time, less essential.

[21] We over-partitioned the data by choosing to model 100 topics because we were primarily interested in the distribution over these topics rather than the topics themselves; these summaries are illustrative, but not central to our analysis.

[22] For example, the regime-sponsored topic around Christmas highlighted by the term "Chavidad," a portmanteau of Chávez and *Navidad*, the Spanish word for Christmas.

7.3 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 will be designed to connect with extant narratives and engage in conversations between different groups. In contrast, the regime will try to create their own narratives, limiting their ability to simultaneously engage with multiple ongoing conversations.

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 H_3.

To understand why this is important, consider the following example: the most frequent opposition hashtag to mention the president is simply "#Nicolàs", but for the regime it is "#RodillaEnTierraConNicolàsMaduro". 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. "#Nicolàs", however, could be used as part of a general discussion of Maduro's policies. Also, "#Nicolàs" is only 7 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 5(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 H_3, as the 95% 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 5(b) shows, both coalitions used a higher number of hashtags during the protest, and the increase is actually 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. #ChavezViveLaPatriaSigue was one of several discourse-structing hashtags that claimed that the spirit of Chàvez 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.

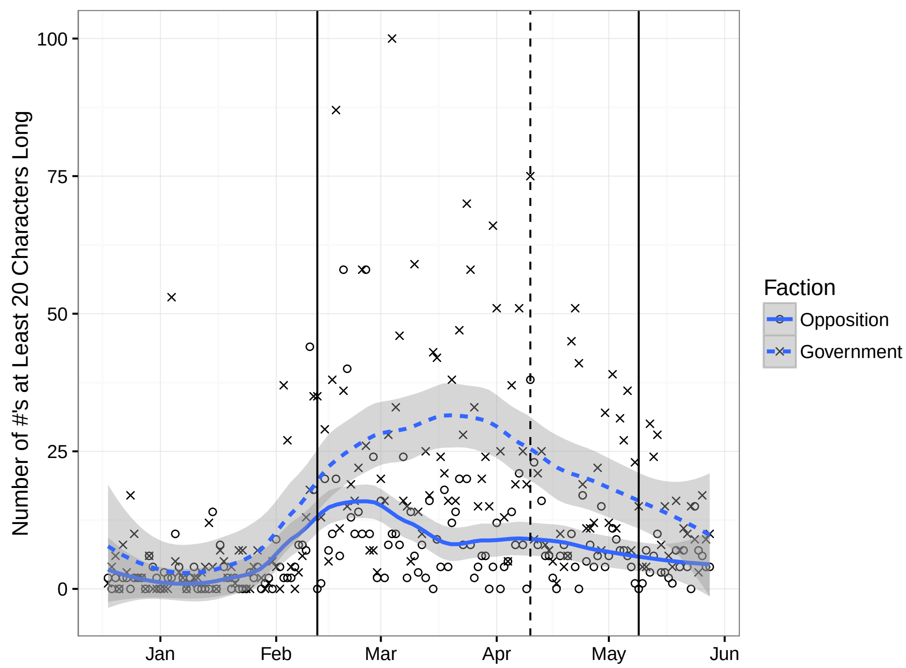
However, this strategy comes at a cost: the length of "discourse-structuring" hashtags makes them less useful than "labeling" hashtags for connecting different conversations by different groups. Following Bennett, Segerberg, and Walker (2014), H_4 predicts that the opposition will be more likely to send tweets containing multiple hashtags during the protest.

In Figure 6, we find only limited 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 is only a brief period during the protests during which the opposition creates more co-occuring hashtags.

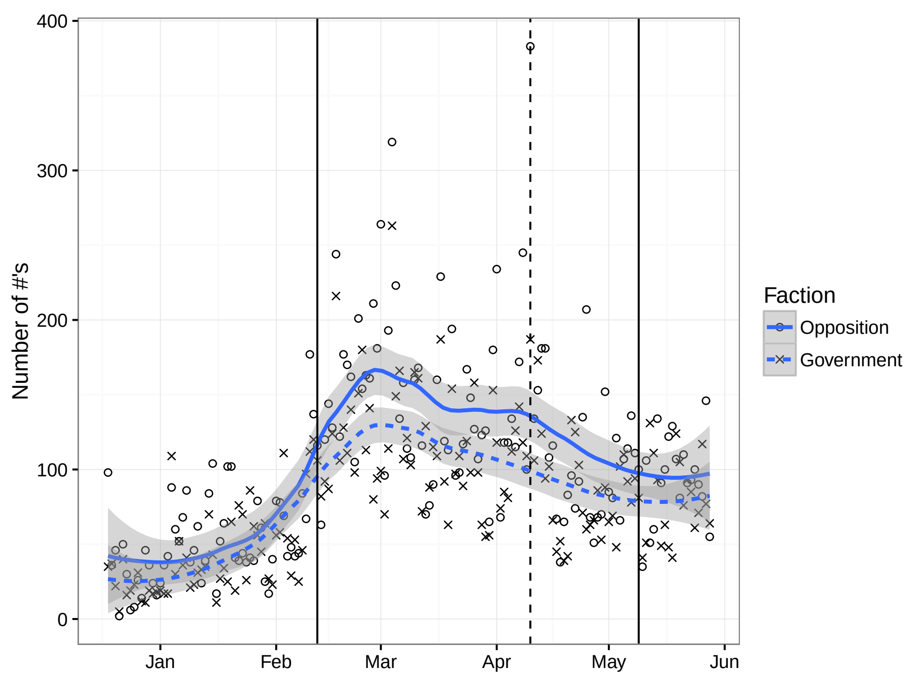
The reason for the weak results in this case may be that because all of the accounts we study are elites, they already have access to large follower networks that comprise all of the groups they might try to connect. Earlier results about hashtag co-occurrence during protest have focused on the tactics employed by normal citizens, and the logic does not appear to translate to this context (Bennett, Segerberg, and Walker, 2014).

Figure 5: Hashtag Use

(a) Number of Hashtags at Least 20 Characters Long



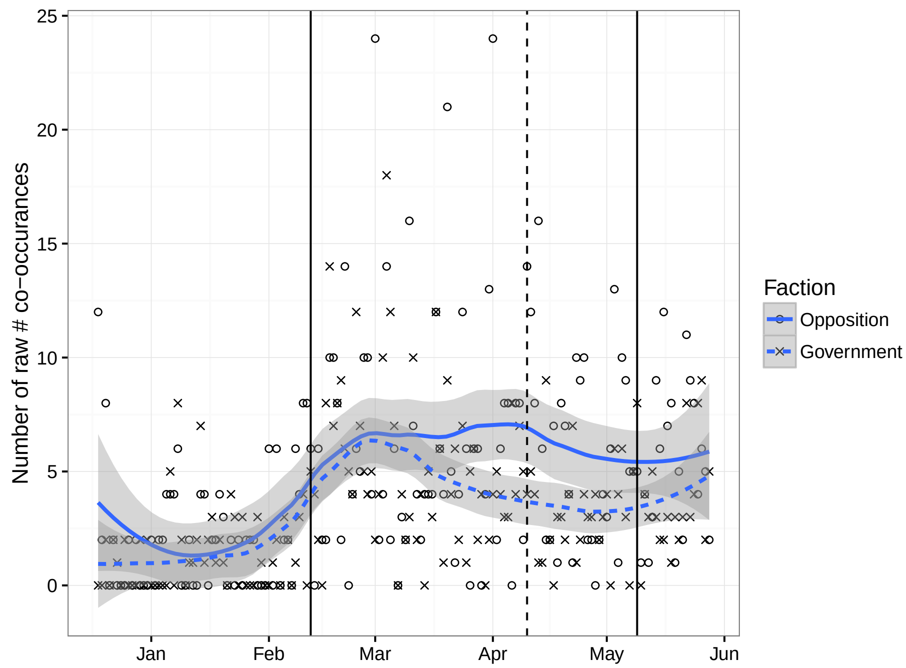
(b) Overall Hashtag Use



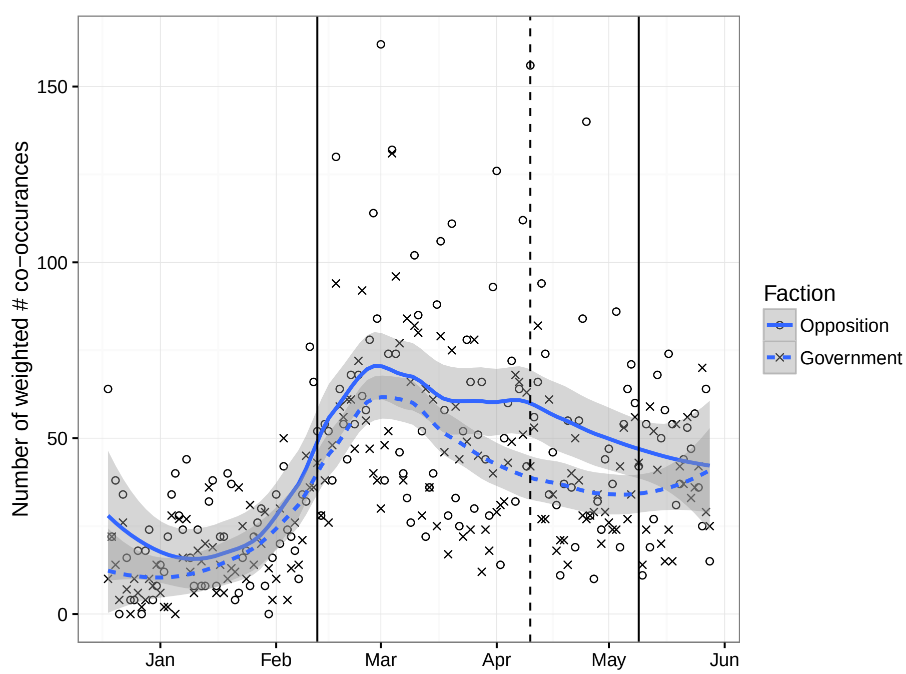
(a) plots the number of hashtags that were at least twenty characters long tweeted per day by the two colations. (b) expands this number to the total count of hashtags tweeted per day by the two coalitions. The solid vertical lines correspond to February 12 (the beginning of the protests in Caracas on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

Figure 6: Hashtag Co-occurances

(a) Raw Number of # Co-occurances



(b) Number of # Co-occurances Weighted by #'s per Tweet



Panel (a) shows the raw number of tweets containing multiple different hashtags tweeted per day by the two coalitions. Panel (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 10th televised sit-down between Maduro and Capriles.

Though hashtag use on Twitter is obviously endemic to the platform and has no direct parallels in other contexts, it does provide insight into the quantity of interest: the strategic use of social media by the regime to develop their own competing narratives. The fact that the opposition was no more likely to adopt this strategy means that it is not simply the case that elites in general use "discourse-structuring" hashtags in times of crisis; rather, elites use Twitter strategically to accomplish their respective goals.

7.4 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, and that they changed their daily topical focus more often than the opposition *diputados*. 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*.

8 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 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 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 the researcher to more easily observe and test these kinds of theories. While this paper 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 paper, but would be an excellent avenue for future research.

The quantitative textual analysis developed in this paper 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.

A: Top Terms for Relevant Topics

#	Terms
	Top Government Topics
#	Terms
24	nicolasmadur, chavez, puebl, psuv, president, venezuel, nuev, dcabellor, madur, patri
35	venezuel, jmontillapsuv, paz, terror, guarimb, venezolan, oposicion, violent, derech, guarimber
62	paz, fascist, violent, violenci, fascism, venezuel, leopold, march, derech, lopez
	Top Opposition Topics
#	Terms
11	hoy, trabaj, vot, puebl, hcapril, unid, dia, diput, buen, venezuel
26	gobiern, venezuel, estudiant, madur, protest, via, hoy, pais, asi, diput
84	prmerojustici, juliocmontoy, americodegrazi, aument, ley, williamsdavil, mer, dia, econom, via

Appendix

A Original Terms for Table 2

Table A contains the untranslated terms in Table 2.

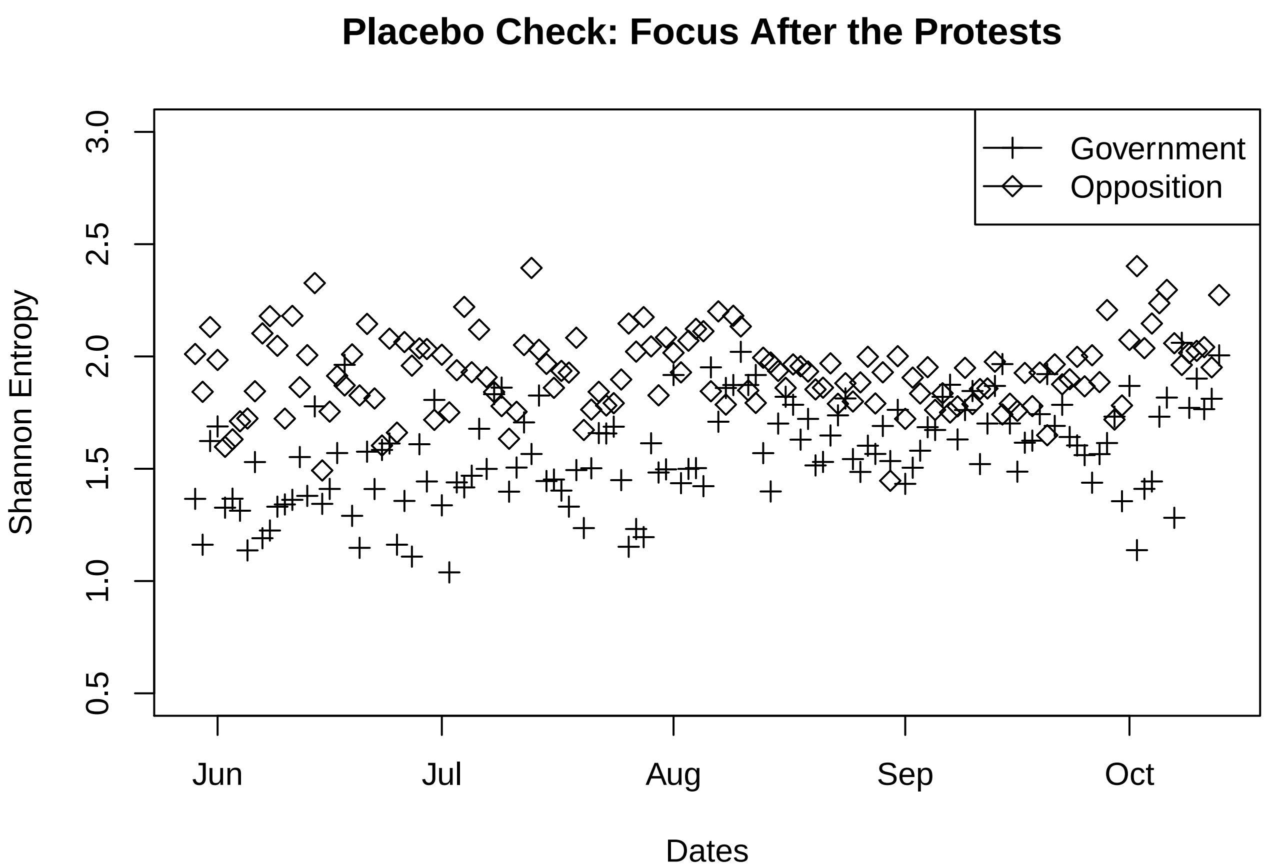
B Placebo Check for Focus

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.

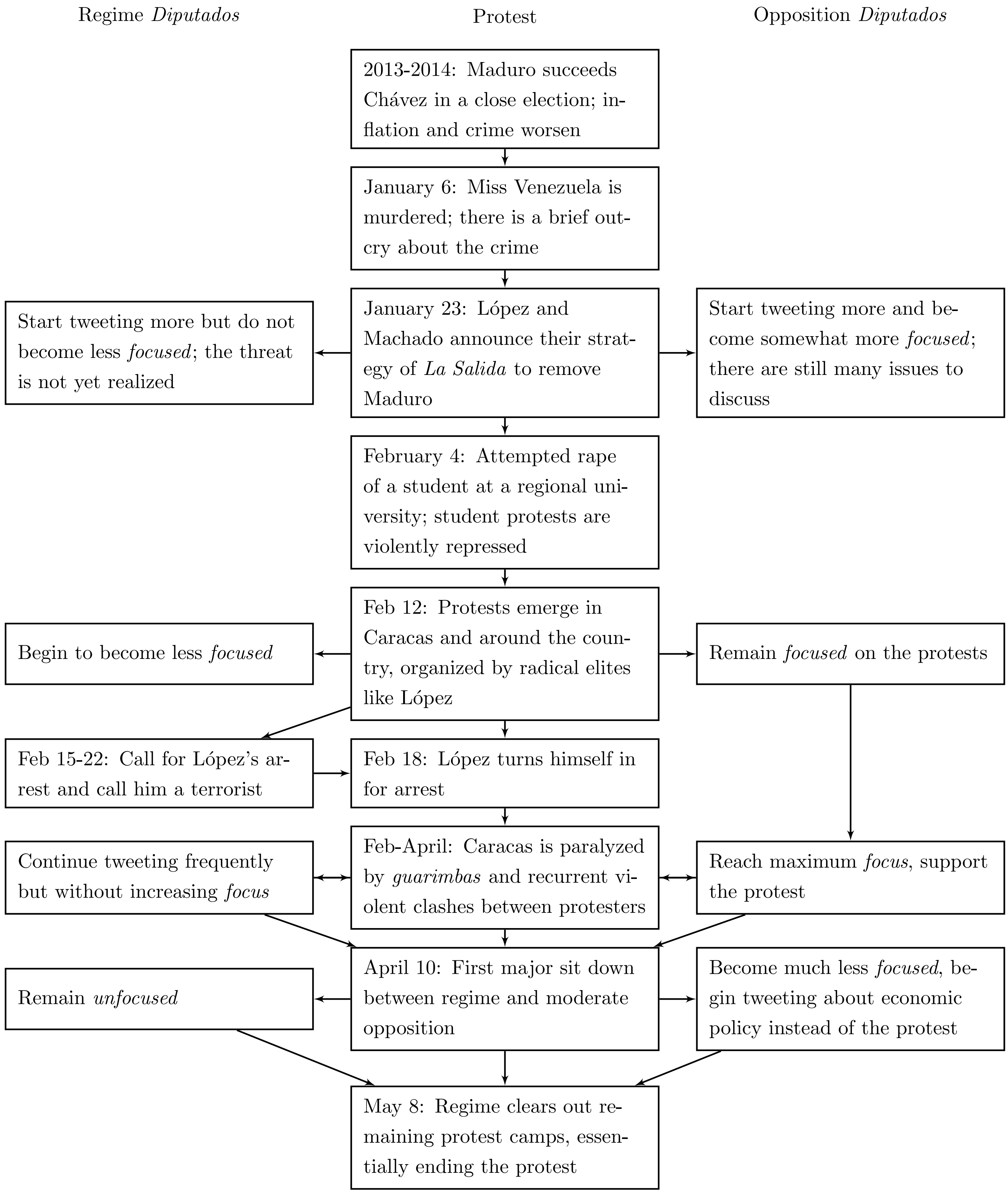
To that end, we collected the tweets from the same accounts from May 30th to October 13th, an arbitrary end date that was simply the present day. We collected 88,328 new tweets, 46,679 from the regime and 41,649 from the opposition. The ratio of regime to opposition is thus close to 1:1, compared to 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
Government
Opposition
Shannon Entropy
0.5
1.0
1.5
2.0
2.5
3.0
Jun
Jul
Aug
Sep
Oct
Dates

C Timeline of Events



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 1915
#venezuela 1549
#tropa 446
#chavez 307
#caracas 164
#eeuu 121
#jpsuv 105
#cuba 103
#usa 90
#juventud 89
#tachira 79
#conelmazodando 79
#anzoategui 72
#rodillaentierraconnicolasmaduro 72
#a1aodetusiembracomandante 71
#barinas 69
#abrilvictoriosoenrevolucion 67
#dialogomaduroporlapaz 67
#envivo 58
#apure 57
#venezuelaserespeta 53
#bolivar 52
#maduroprotectordevenezuela 52
#a10mesesdetusiembracomandante 50
#estudiantes 49
#desconectatedelaguarimba 48
#maduroespueblo 48
#ahora 46
#merida 45
#hagamoslapaz 45
#deyarealcorazondelpueblo 45
#lafoto 44
#unasur 42
#madurohombredepaz 42
#maduroespuebloprospero 42
#vargas 41
#dialogosipactono 41
#rt 40
#13apatriadigna 40
#somosfanb 39
#chadertonvozdevzla 38
#apasofirmelarevolucion 38
#vtv 37
#a14mesesdetusiembracomandante 36
#encontactoconmaduronro6 35
#4frevolucionparasiempre 34
#zulia 33
#barriosunidosconmaduro 33
#maduroespueblochavista 33
#entrechavistasnosseguimos 32
#twitter 32
#lara 32

#desconectatedelodio 32
#conferenciadepaz 32
#chavezvivelapatriasigue 31
#pueblounidoconmaduro 31
#trujillo 30
#revolucionhaciendopatria 30
#mayoproductivoconmaduro 30
#vzlaunidacontraelfascismo 29
#gringosyfascistasrespeten 29
#vzlabajoataquemediatico 29
#encontactoconmaduro 29
#conmadurofuturodepaz 29
#colombia 28
#novolveran 28
#loschavistasnoperdemosbatallas 27
#conmadurotriunfalapaz 27
#construyendolapatriasegura 27
#mujerescontraelfascismo 26
#soyvenezuelanomasguarimba 26
#foto 25
#viveenpazsinguarimbas 25
#mayodelostrabajadores 25
#cojedes 24
#rusia 24
#juventudconmaduroalavictoria 24
#falcon 24
#conmadurocarnavalseguro 24
#daleunchancealapaz 24
#a13mesesdetusiembra 24
#circuito2 23
#oea 23
#encontactoconmaduronro7 23
#antv 22
#chvez 22
#video 22
#monagas 22
#a11mesesdetusiembracomandante 22
#encontactoconmaduronro4 22
#maduroespueblounido 22
#porelcaminodechavez 22
#opinin 21
#vzlaserespeta 21
#vzladisfrutaelcarnaval 21
#venezuelapueblodepaz 21
#gnbguardianadelapaz 21
#aunaodevictoriaconmaduro 20
#aragua 19
#pueblodepazalabolivar 19

D.2 Opposition Hashtags with Counts

#venezuela 642
#barcelona 486
#quenadatedetenga 303
#anzoategui 243
#caracas 212
#sucre 184
#guayana 177
#petare 177
#juntosgobernamos 166
#cantaura 166
#barinas 164
#mrida 157

#pjenvivo 156
#notitchira 147
#12f 129
#hayuncamino 129
#8d 126
#srafiscal 124
#ungobiernoparatodos 116
#unidad 109
#sosvenezuela 108
#anzotegui 102
#casaxcasa 102
#ff 100
#tchira 91
#lasalida 90
#lecheria 87
#futuroparatodos 80
#tachira 76
#nomasviolencia 76
#confesionesdemariosilva 76
#seguimosavanzando 74
#yaracuy 73
#freites 72
#nicols 71
#amnistaya 70
#futuroseguro 67
#an 65
#barquisimeto 64
#2m 64
#merida 63
#12m 62
#encuentroconmovimientosvecinales 60
#aumentochucuto 60
#vargas 59
#cojedes 58
#plc 58
#elsabado3porvzla 58
#quesigaelcambio 58
#puertoordaz 51
#sinluz 51
#toquedequeda 51
#amnistiaya 51
#valencia 50
#paquetazorojo 50
#zulia 49
#lacausar 48
#viernesamarillo 47
#vieneelcambio 47
#aragua 46
#miranda 45
#lara 45
#envivo 43
#noticias 42
#maracaibo 42
#resistenciavzla 42
#diputadosdelamudenpanama 42
#adelante 41
#video 40
#22f 40
#sosinternetve 40
#26a 40
#cuba 39
#victoriatotal 39
#carabobo 38
#chacao 38
#sancristobal 38
#14m 38
#oidoaltambor 38
#comitepoliticonacionalpj 38
#sucretienefuturo 36

#maturin 35
#leydesarme 35
#iribarren 34
#1m 33
#pablogana 32
#20f 32
#ucv 31
#medidahumanitariaparasimonovis 31
#8m 30
#mardosomostodos 30
#facebook 29
#titulares 29
#bna 29
#poltica 28
#envideo 28
#acn 28
#fb 28
#caprilestv 28
#sosvzla 28

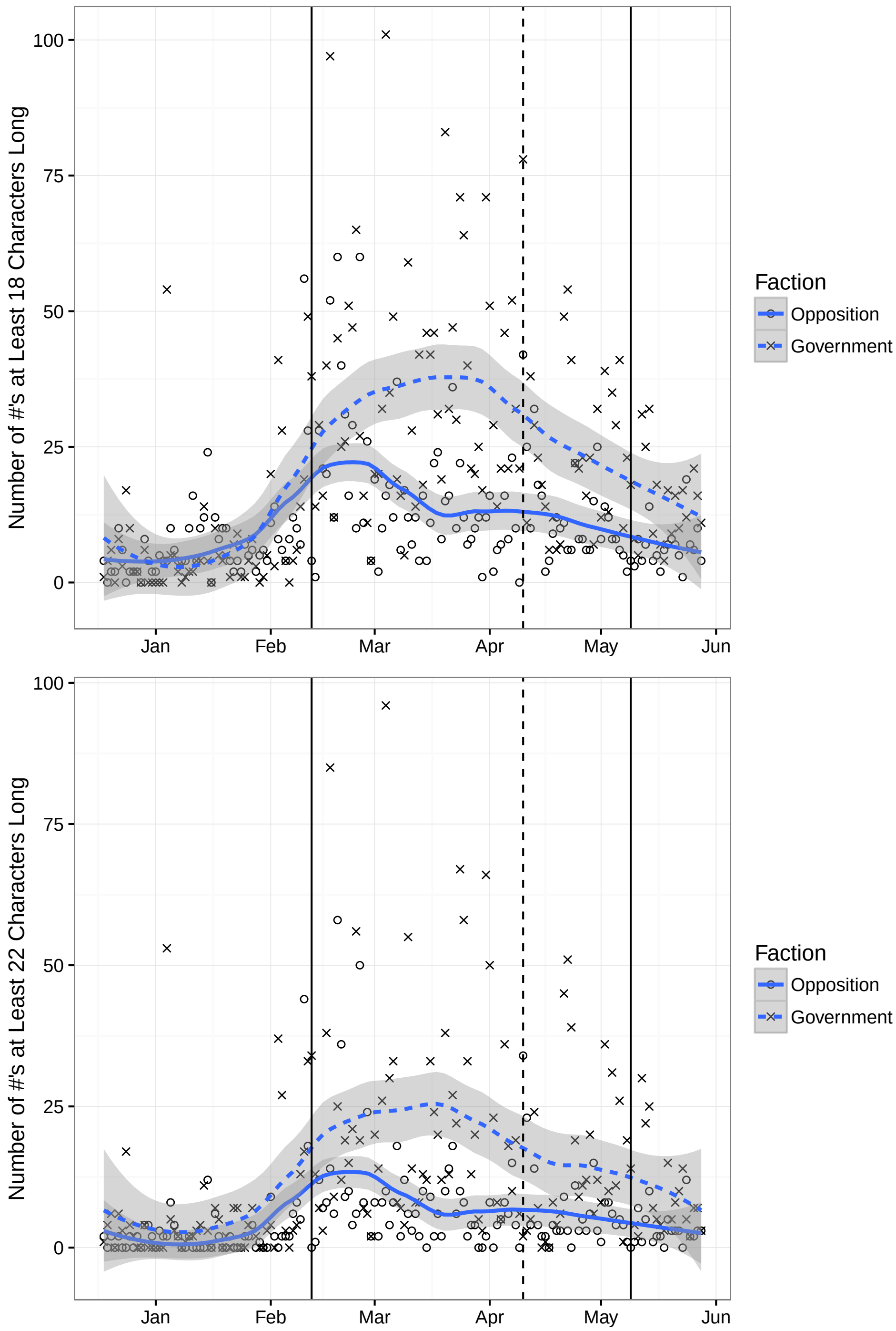
E Robustness Check–Hashtag Length

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 **??** 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

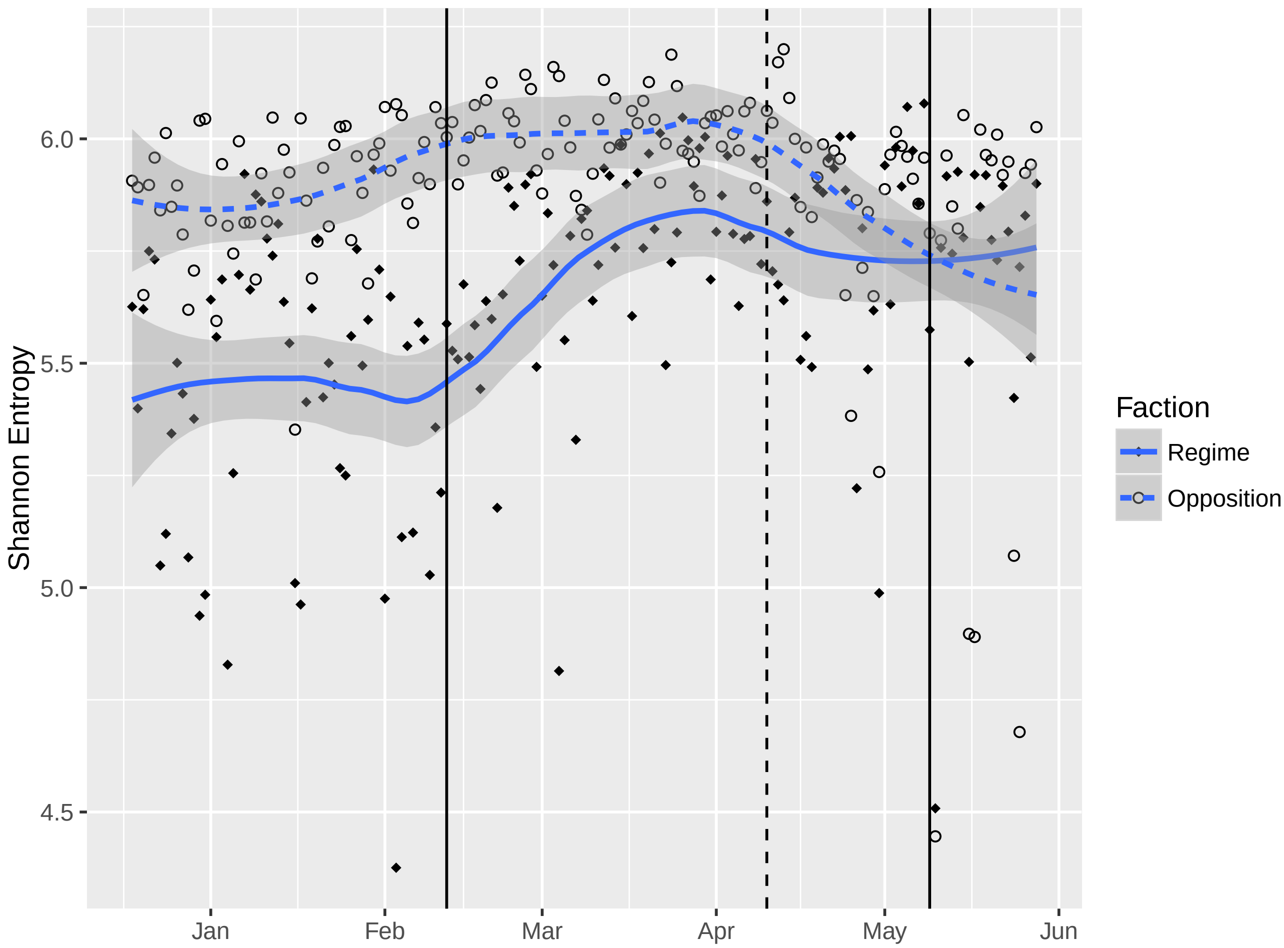
Analysis using an alternative topic model, called the Correlated Topic Model, leads to similar inferences about the changing communication strategy of the regime *diputados*. The absolute levels of *focus* are different from the LDA model; notably, the opposition is much less *focused* than the regime before the protest. However, the opposition's level of *focus* remains mostly constant while the regime becomes more *focused* during the protest.

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

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: 27.

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): 232–260.

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* pp. 17–35.

Boix, Carles, and Milan W Svolik. 2013. "The foundations of limited authoritarian government: Institutions, commitment, and power-sharing in dictatorships." *The Journal of Politics* 75 (02): 300–316.

Chen, Jidong, Jennifer Pan, and Yiqing Xu. 2015. "Sources of authoritarian responsiveness: A field experiment in China." *American Journal of Political Science* .

Cheng, Xueqi, Yanyan Lan, Jiafeng Guo, and Xiaohui Yan. 2014. "BTM: Topic Modeling over Short Texts." *IEEE Transactions on Knowledge and Data Engineering* p. 1.

Christensen, Darin, and Francisco Garfias. 2015. "Can You Hear Me Now?: How Communication Technology Affects Protest and Repression." *Available at SSRN 2529769* .

Ciccariello-Maher, Georeg. 2014. "LaSalida? Venezuela at a Crossroads." *The Nation* .

Corrales, Javier. 2013. "Chavismo After Chavez." *Foreign Affairs* .

Corrales, Javier, and Michael Penfold-Becerra. 2011. *Dragon in the tropics: Hugo Chavez and the political economy of revolution in Venezuela.* 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.* MIT Press.

Diaz, Sara Carolina. 2014. “Sector de la oposicin convoca a marcha para el 12 de febrero.” *El Universal* .

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): 459–478.

Edmond, Chris. 2013. “Information manipulation, coordination, and regime change.” *The Review of Economic Studies* 80 (4): 1422–1458.

Friedman, Uri. 2014. “Why Venezuela’s Revolution Will Be Tweeted.” *The Atlantic* .

Greitens, Sheena Chestnut. 2013. “Authoritarianism Online: What Can We Learn from Internet Data in Nondemocracies?” *PS: Political Science & Politics* 46 (02): 262–270.

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): 5228–5235.

Gunitsky, Seva. 2015. “Corrupting the Cyber-Commons: Social Media as a Tool of Autocratic Stability.” doi:10.1017/S153792714003120.

Hastie, Trevor, Robert Tibshirani, Jerome Friedman, T Hastie, J Friedman, and R Tibshirani. 2009. *The elements of statistical learning.* Vol. 2 Springer.

Hong, Liangjie, and Brian D Davison. 2010. Empirical study of topic modeling in twitter. In *Proceedings of the First Workshop on Social Media Analytics.* ACM pp. 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): 216–232.

Iyengar, Shanto, and Donald R Kinder. 1987. "News that matters: Agenda-setting and priming in a television age." *News that Matters: Agenda-Setting and Priming in a Television Age* .

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 (02): 326–343.

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

Mainwaring, Scott. 2012. "From representative democracy to participatory competitive authoritarianism: Hugo Chavez and Venezuelan politics." *Perspectives on Politics* 10 (04): 955–967.

Malesky, Edmund, and Paul Schuler. 2010. "Nodding or needling: analyzing delegate responsiveness in an authoritarian parliament." *American Political Science Review* 104 (03): 482–502.

McCombs, Maxwell E, and Donald L Shaw. 1972. "The agenda-setting function of mass media." *Public Opinion Quarterly* 36 (2): 176–187.

Mehrotra, Rishabh, Scott Sanner, Wray Buntine, and Lexing Xie. 2013. Improving lda topic models for microblogs via tweet pooling and automatic labeling. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. ACM pp. 889–892.

Morozov, Evgeny. 2011. "Whither Internet Control?" *Journal of Democracy* 22 (2): 62–74.

Oates, Sarah. 2013. *Revolution stalled: The political limits of the Internet in the post-Soviet sphere*. Oxford University Press.

Perez, Valentina. 2014. "The Grim Reality of Venezuelan Protests." *Harvard Political Review* .

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. In *Proceedings of the 17th international conference on World Wide Web*. ACM pp. 91–100.

Rahimi, Babak. 2011. "The agonistic social media: cyberspace in the formation of dissent and consolidation of state power in postelection Iran." *The Communication Review* 14 (3): 158–178.

Sanovich, Sergey, Denis Stukal, Duncan Penfold-Brown, and Joshua Tucker. 2015. "Turning the Virtual Tables: Government Strategies for Addressing Online Opposition with an Application to Russia.".

Schattschneider, Elmer E. 1960. "The Semi-Sovereign People: A Realist's View of Democracy in America.".

Shannon, Claude. 1948. "A Mathematical Theory of Communication." *The Bell System Technical Journal* pp. 379–423.

Tufekci, Zeynep. 2014. "Social movements and governments in the digital age: Evaluating a complex landscape." *Journal of International Affairs* 68 (1): 1.

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): 363–379.