Elites Tweet to get Feet off the Streets: Measuring Regime Response to Protest Using Social Media

Kevin Munger

August 22, 2014

Abstract

Social media use among elites offers a useful avenue for analyzing regime response to protest, especially in countries with some degree of freedom of speech. Here I examine the frequency and content of Twitter usage among Venezuelan elites in the context of the 2014 protests. This analysis demonstrates that the regime sends more signals during protests but that they become less focused compared with opposition elites, especially following acts of regime suppression of opposition-sponsored protests. This observation supports theoretical predictions that increasing public information can make coordination among protesters more difficult. The discussion of social media use by both pro- and anti-regime elites contributes to the debate over whether social media prevents or promotes regime change.

1 Introduction

From the Arab Spring (Howard and Hussain, 2011, 2013) to Turkey in 2013 (Metzger et al., 2014) and Ukraine in 2014 (Bohdanova, 2014; Tucker et al., 2014), social media has been analyzed as an essential tool that allows protesters to disseminate information and organize efficiently and discretely in the face of repressive regimes. Though social media use was widespread, whether it actually changed the protest/regime dynamic is an open question. Unambiguous is that social media provides a huge and detailed data set of protest and regime communication, making visible previously hidden or difficult to access data.

The Venezuelan protests of 2014 are an important and understudied case study of the role of social media in protest. Venezuela is a “hybrid regime” that took the unusual path from full democracy towards authoritarianism in the context of medium GDP per capita. Venezuela is also among the top five countries in terms of Twitter penetration (PeerReach, 2013), and its politicians are well-represented on Twitter. Henrique Capriles, the runner-up in the last two elections, is the most followed person in Venezuela, and Hugo Chàvez is a close second, despite being dead for over a year.

One of the central challenges of using social media to study politics is that the people using social media are not a representative sample of the population as a whole. I follow a growing trend in the literature (Barberá, 2013; Barbera et al., 2013; Cormack, 2013; Gulati and Williams, 2010) in focusing on the social media usage of a certain subsection of the elite; in this case, the members of the Venezuelan National Assembly (*diputados*). Although Venezuela has a strong presidential system, the National Assembly has played an important role in the current protests, voting in November 2013 to give President Nicolàs Maduro emergency powers to rule by decree. 139 of the 166 have Twitter accounts, providing ample analytical leverage.

This approach is well-suited to answering the question of how regimes respond to protests. Protesters and dissidents face a coordination problem in convincing enough people to engage in increasingly violent protest, and the regime's aim is to make this more difficult. The tactics they employ to do this include violent repression and the threat of economic/social sanctions. I focus on the role of information in regime strategy in preventing popular participation in protest. Though the use of propaganda to maintain power is as old as the concept of government itself, social media is distinct because it provides (potentially) diverse sources of propaganda, larger data sets from which to extract patterns, and faster turnaround between events and government response.

The corpus of tweets from the 139 *diputados* who have Twitter accounts (out of 166 total) are a large data set that allow us to compare the way that pro- and anti-regime *diputados* use information to respond to protests. The formal theory literature on coordination problems in the context of protest predict that the government will try to prevent coordination on rebellion and that the opposition will promote it. Twitter data allow us to test these theories.

This paper proceeds as follows: Section 2 contains background information on the situation in Venezuela; Section 3 is a review of the relevant literature; Section 4 presents my hypotheses; Section 5 outlines my data collection strategy; Section 6 contains my analytic strategy; Section 7 presents results; and Section 8 concludes.

2 Background

Venezuela has been a democracy since 1958. The first 40 years were dominated by the *punto fijo* system, with political power alternating between two clientelistic parties. 1999 saw the election of Hugo Chàvez and a shift in Venezuelan economic and social policy. Though Chàvez successfully instituted a new constitution that expanded the power of the executive and broke up the old political equilibrium, the country was in poor economic shape, and his initial economic reforms were modest. After surviving a 2002 attempted coup and both an oil worker strike and recall attempt in 2003, Chàvez had the political capacity to fully nationalize the oil industry. This provided fiscal flexibility in the short term and, in concert with the oil boom that began soon after, allowed the government to consolidate its rule by dramatically increasing social spending (Corrales and Penfold-Becerra, 2011).

The formula worked for Chàvez: he won re-election in 2006 and 2012 handily and with very high voter turnout. His health declined rapidly in 2012, and he died in early 2013. Though vice president Nicolàs Maduro's assumption of the presidency was uncontroversial, it was not obvious who would represent Chàvez's party (The United Socialist Party of Venezuela, PSUV) in the constitutionally-mandated election held 30 days after the death of a sitting president (Corrales, 2013).

Maduro won the 2013 special election against Henrique Capriles, a popular state governor and Chàvez's opponent in 2012, but his margin of victory was only 1.5% and there were claims of fraud and illegitimacy by the opposition. It was clear that Maduro was not Chàvez, lacking the latter's charisma and intellect, but Maduro also inherited a poor economic situation.

Under Maduro, continuing inflation and a violent crime rate among the worst in the world led to rising discontent. The sparks needed to ignite the protest that prompts this analysis were the January 6 robbery-murder of a former Miss Venezuela and the attempted rape, on February 5, of a university student on a campus in the southeast part of the country. The latter led to a small student protest against high crime rates that provoked a violent government response. (Perez, 2014). This inspired a much larger protest that eventually spread to Caracas. Opposition leaders, especially the radical faction headed by Marìa Corina Machado and Leopoldo Lòpez, used this momentum to advocate for regime change (known as *La Salida*), culminating in the arrest of Lòpez on February 19th.

These protests began to wind down in May 2014, but they thoroughly divided both the country and the opposition party. Capriles continued to push for democratic reforms and non-violent methods of progress, while *La Salida* wanted nothing less than the removal of the ruling party through violent street protests (Ciccariello-Maher, 2014). The primary complaints shared by both factions of the opposition concerned violence, inflation, corruption, shortages of basic goods, and government censorship.

As discussed below, most research on Twitter and protest has been in the context of significant state censorship, and Venezuela is no exception. Although free speech and freedom of the press are guaranteed in the Constitution, there have been significant restrictions put into place under Chàvez and Maduro. In keeping with Venezuela's status as a "hybrid" regime, with some characteristics of a democracy and others of an autocracy, this censorship has rarely been explicit. An illustrative example is that of RCTV: once Venezuela's largest private broadcasting company, it actively supported the coup attempt in 2002. In 2007, its license expired, but was not renewed (Corrales and Penfold-Becerra, 2011).

Direct broadcast of the protests has been banned from Venezuelan television, and CNN was kicked out of Venezuela in 2014. In early February, just as the protests were spreading nationwide, there were a number of reports that Twitter was not functioning and the Internet was being shut down. In April, Maduro retweeted a message from a follower that said, "If Twitter is being blocked, how are you reading this tweet?" (my translation). Maduro's tweet indicates that the regime was concerned that a significant portion of Venezuelans suspected that their Internet access was being curtailed. Internet censorship is not purely a recent phenomenon: websites tracking the true exchange rate between Venezuela's inflation-plagued bolivar and the US dollar have been blocked for years.

3 Theories on the Role of Social Media in Politics

The potential importance of social media as a political technology has been discussed since its advent in the early 2000s. Analysis of social media has thus far been divided into two areas: political campaigns in fully consolidated democracies, and popular opposition and protest in non-democracies.

For the former, expectations of the impact of social media have been modest–politicians and campaigns use it as just another tool to communicate their message and position themselves. Some politicians have been quicker and savvier in incorporating social media into their campaigns, most famously Barack Obama in the 2008 US presidential election (Cogburn and Espinoza-Vasquez, 2011; Harfoush, 2009). Accessibility to the analyst means that social media is useful even if it is identical to other forms of communication, and there is a growing literature on Twitter use by US Members of Congress and the way it differs by gender (Cormack, 2013) and tweet content (Golbeck, Grimes, and Rogers, 2010); which covariates drive the difference in the number of followers (Gulati and Williams, 2010; Vaccari and Nielsen, 2013); and whether they tend to set the agenda or to respond to their followers (Barbera et al., 2013). Analysis of this kind has also been done in other fully democratic countries like South Korea (Hsu and Park, 2012), Germany (Jungherr, 2010) and Australia (Bertot, Jaeger, and Grimes, 2010).

There has been comparatively little research on this kind of activity in "hybrid" electoral regimes (Robertson, 2011) or dictatorships. Instead, theories of the political role of social media in non-democracies focus on the way in which it changes the dynamic between governing regimes and opposition groups. Spurred by the unexpected Arab Spring and the visible use of social media by protesters and revolutionaries, scholarship on the role of social media was initially enthusiastic (Gerbaudo, 2012; Khondker, 2011; Lotan et al., 2011). This enthusiasm stemmed to some extent from a failure to appreciate that, in equilibrium, the regime's adoption of new technologies makes their impact at best ambiguous. Authoritarian governments used social media to spread disinformation even before the Arab Spring (Esfandiari, 2010). Regimes have various strategies for neutering the revolutionary potential of social media, from broadly restricting access to the Internet (Howard, Agarwal, and Hussain, 2011) to elaborate censorship programs (King, Pan, and Roberts, 2013). Depending on the respective capacities of the two sides to take advantage of social media, it might even tip the balance of power towards the regime (Morozov, 2011; Rahimi, 2011).

Research on social media in democratic countries has centered on the way that elites use it in campaigns and to interact with constituents. On the other hand, the study of social media in non-democratic countries has been almost entirely about the ability of mass protesters to spread information or coordinated tactically and of the regime to prevent this. Elite usage of social media in non-democratic contexts is a lacuna in the literature which this paper intends to fill.

One reason for this gap may be captured by the following assumption: it is even more obvious that the government controls these Twitter accounts than they do other media, so people discount them entirely and they are not worthy of study. This assumption is mistaken. Even if regime social media is categorically similar to traditional media in its role as government propaganda, it still offers a large, accessible, and idiosyncratic (insofar as it varies from regime elite to regime elite) data set to analyze regime communication. The idiosyncrasies are important: most of the theoretical literature on authoritarian regime change involves some faction of the elites, either in cooperation with the masses or the military, overthrowing the government (Acemoglu and Robinson, 2005; Bueno de Mesquita et al., 2005), and it may be possible to track these divisions on social media, possibly even detecting signals that they don't intend to send.

4 Hypotheses

Expanding on similar work on endogenous information in global games by Morris and Shin (1998) and Angeletos, Hellwig, and Pavan (2006), the model presented in Edmond (2013) provides much of the theoretical underpinning for this paper. In modeling a competition between the regime and the citizens, the central assumption is one of endogenous information: the regime decides what information to share, with the goal of preventing the citizens from overthrowing them. The regime tries to disguise its true type by sending signals, via potentially heterogeneously biased media outlets, that inflate its true quality. Edwards calls this "signal jamming": each citizen observes an at least partially random draw from the distribution of signals sent by media outlets on some spectrum of pro- to anti-regime leanings. This approach does not minimize the number of citizens who oppose the regime, but is designed rather to keep that number below the threshold necessary for revolution. Each citizen cannot be sure of the degree of information manipulation in the signals they observe, and thus cannot be sure about the true quality of the regime. If an information technology simultaneously increases the amount of information transmitted and the ability of the regime to manipulate that information (e.g. the widespread use of radios by Nazi Germany (cf Adena et al., 2013)), this unambiguously increases the ability of the regime to convince citizens to coordinate on not revolting. On the other hand, with a technology like Twitter, where the regime cannot manipulate all information, the effects should be ambiguous. More signals increase the precision of the signal that each citizen receives, but does not improve their ability to infer the extent of the regime's manipulation.

The game modeled by Edwards is, crucially, one of coordination. For each citizen, being confident that the regime is worth overthrowing does not imply that she should take to the streets: they also have to believe that enough others citizens will join them that the revolution will be successful.[1] Signal jamming thus serves two purposes: it decreases the confidence of each citizen in *both* his knowledge the true type of the regime and in the opinions of other citizens. As such, the model predicts that regimes should respond to the creation of new forms of communication like social media by increasing the number of signals they transmit. An increase in the frequency of messages that citizens hear makes it harder for them to coordinate: if almost all of the information a citizen observes is pro-regime, it is unlikely that enough other citizens will see sufficient anti-regime information to infer that the regime is worth overthrowing.

Although Edmonds does not model a situation with an active and open opposition as is the case in Venezuela, he points to a similar model in de Mesquita (2010) in which a group of dissidents tries to convince citizens to coordinate on revolution by using violence as a credible signal of their willingness to revolt. This model implies that, in a country with largely free speech and widespread access to social media, the opposition should behave similarly to the government and flood the informational marketplace with anti-regime messages in support of the more violent tactics of the dissidents.

This theory gives rise to the first hypothesis I will test:

Hypothesis 1 *In response to large-scale protests, both the regime and opposition diputados will tweet more.*

In a normal, stable context, regime and opposition *diputados* should use social

[1]Note that this is not a prisoner's dilemma-style model in which each citizen's dominant strategy is to stay home even in the case of a successful revolution. It requires the assumption that citizens pay some cost, either in social sanctions or moral guilt, for not participating in a successful revolution.

media like any other politician–to promote party unity, to campaign for re-election or to interact with voters. Their social media use in this context will therefor be relatively *unfocused.* The change in the degree of *focus* exhibited by the Twitter messaging in response to the "shock" of the onset of the protests is central to my analysis. The model discussed above predicts that, as the threat of revolution or widespread protests becomes more salient, the regime *diputados* should strategically send multiple kinds of messages to make it harder for the citizens to coordinate and revolt; their messages should be less *focused* in that they address a number of issues. The opposition *diputados*, on the other hand, should be more *focused* precisely to encourage this kind of coordination; their messages should address only the most important issues. Note that these predictions pertain to the change in *focus*, not the absolute level. Because the protests dominated the national discussion, the government would have to actively try to avoid discussing them to become less *focused* on net, but their increase in focus will be lower than the attention-coordinating *focus* I expect the opposition to evince.

These expectations are captured in my second hypothesis:

Hypothesis 2 *In response to large-scale protests, the tweets of both opposition and government* diputados *will become more focused, but the change will be more pronounced for the opposition.*

5 Data

For each of the 166 *diputados* in Venezuela's unicameral legislature, I searched Twitter and Google to find an associated Twitter account. 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 there was usually only one that was both active and with a significant number of followers. If there was any ambiguity as to whether a Twitter account was a politician or just a citizen, I checked to see if the account was followed by one of the party elites from either side.[2] I was able to locate accounts for 139 of the 166 *diputados* (84%), though the sample is biased in favor of the opposition: I found 63 of 66 (95%, similar to US Members of Congress),

[2]Incidentally, it turns out that many Venezuelan politicians share names with professional baseball and soccer players.

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

Diputados	N	Min	1st Quartile	Median	3rd Quartile	Max	Mean	Tweets
Regime	76	1	128	993	2863	4525	1494	113584
Opposition	59	11	1302	3279	3484	4621	2570	151637
Total	135	1	413	1851	3331	4621	1951	264227

Period of Analysis: December 18, 2013 - May 29, 2014

but I found only 76 of 99 (77%) for the regime.[3] For the subsample whose tweets are analyzed in this paper–those who tweeted after December 18, 2013–there are 135 accounts, 76 regime and 59 opposition. Table 1 contains a summary of the number and distribution of the tweets collected.

Although the regime has a higher number of active accounts, the opposition produced roughly 40% more tweets during my observation. This isn't just driven by a few prolific opposition accounts; comparing the 1st quartiles, medians and 3rd quartiles of the regime and opposition indicates that the opposition is more active throughout the distribution. The largest difference is in the lower end of the distribution–there are 18 government accounts with fewer than 100 tweets compared to just 3 for the opposition.[4]

I used Twitter's REST API[5] via `tweepy`[6], in the Python programming language, to collect the most recent tweets for each account. Using the `/statuses/user_timelines` endpoint[7], Twitter's API allows fetching the latest 3,200 tweets for a given account. I did this on April 19th and then again on May 29th, at which point the protests had largely subsided. As a result, I have more than 3,200 tweets for some accounts. Many of the accounts have fewer than 3,200 tweets and so I have their entire history. Twitter's API also provides other bits of information associated with each account, including their "biography," where they claim to be located, and the date they joined Twitter. Additionally, I collected the party to which each *diputado* belongs.

[3]To check the validity of my selection, I had a research assistant recreate my analysis. There were only 2 discrepancies, the adjudication of which were obvious.

[4]Because the method I employ treats tweets from different members of the same coalition identically, there's no risk of bias from including these accounts.

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

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

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

6 Analysis

In order to track the topics being discussed on Twitter, I implemented Latent Dirichlet Allocation (Blei, Ng, and Jordan, 2003), an unsupervised, "bag-of-words" machine-learning algorithm used for topic modeling that is increasingly popular in the social sciences. As Barbera et al. (2013) point out, this approach is well-suited to analyzing the messaging by a group of elites over time because it is unsupervised (that is, the researcher cannot introduce bias by deciding which topics to study) and because it allows the entire corpus of information to be used.

LDA takes two objects as parameters: some number 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 which contains all the terms in the corpus; and K, the number of topics to be modeled. LDA functions by treating each document as a distribution over latent topics and each topic as a distribution over words. These are assumed to be Dirichlet-distributed, where the Dirichlet distribution is a multivariate extension of the beta distribution.

In my case, the "documents" consist of the text of the *diputados*' tweets. Taking the tweets from 2 months before the start of the protests as marked by Lòpez's fiery condemnation of the regime before he turned himself over for arrest on February 18, I divided each day's worth of tweets by each coalition into a separate document. There are 162 days included in my analysis, and thus 324 documents. Once aggregated into these documents, the terms compsing 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 ignoring which *diputado* produces each tweet. Since my aim is to measure the degree of coordination among the coalitions, this information is not relevant.[8] The technique also loses information by conflating all the tweets from each day. This is unavoidable–there isn't enough information in a single tweet to use them as the documents, and the machine learning literature indicates that aggregating tweets leads to better performance for LDA (Hong

[8]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 since the shape of the distribution is similar between the coalitions, this weakness does not impair my comparative analysis.

and Davison, 2010).[9]

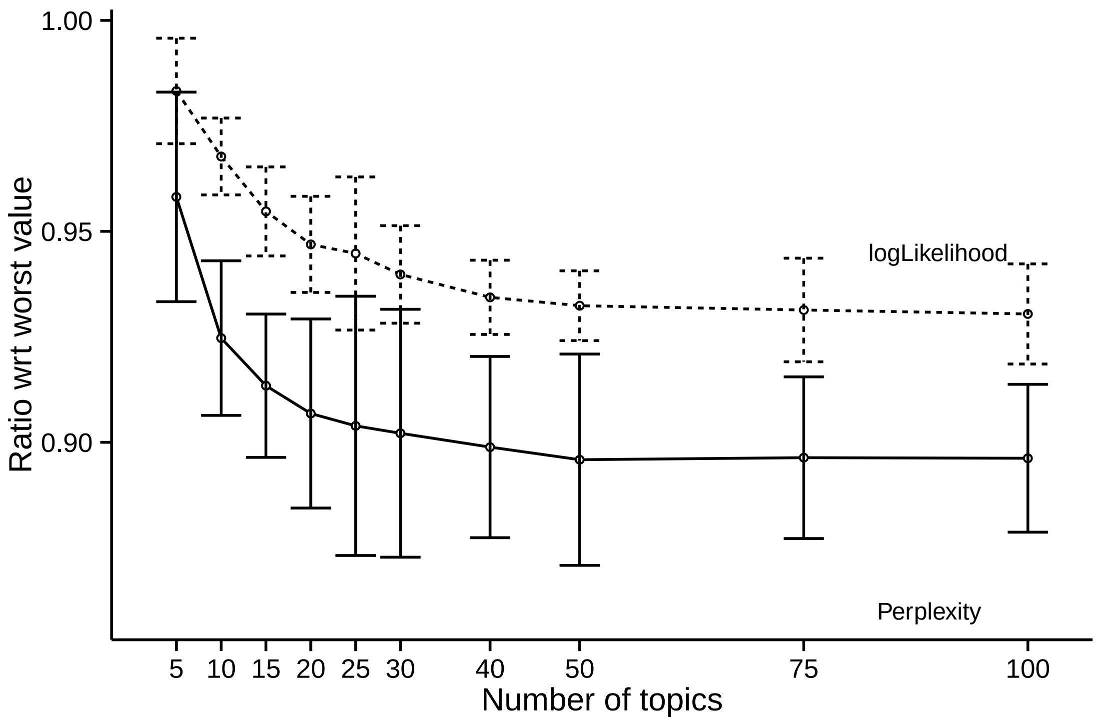
To determine the number of topics K, I performed ten-fold cross-validation of both log-likelihood and perplexity analyses on the holdout sample. This method works by repeatedly taking a subsection of the sample and generating predictions that are then tested on the remaining subsection. Though the model fit improves monotonically in the number of topics, the gains from adding more topics diminish at around 50 topics (see Figure 1). Although there exist standard rules for choosing K, like the conservative "one-standard-error" rule as outlined in Hastie et al. (2009), this choice is contingent on the question LDA is being used to answer. LDA has most commonly been used to identify specific topics, prioritizing the recognizability of the topics created; in this context, the main priority is to avoid overfitting the data via a choice of a conservative K. My aim is to study the change in focus over time, so this concern is less relevant and creating extra topics allows for greater variation in the quantity of interest, even if those topics are sparsely represented and hard to identify. As a result, I follow the guideline of doubling the number of topics that the conservative approach recommends, and select K=100.

Following the standard in the social science literature, I 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), I implemented 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 aren't 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 100 γ parameters for each document. For each of the 324 documents w, the γ parameters are the probabilities that that document pertains to each topic. $\gamma_{w,k}$, then, is the probability that document w pertains to topic k; note that $\sum_{k=1}^{100} \gamma_{w,k} = 1$. There are 32,400 of these γ parameters.

[9] 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 focusing on word dyad co-occurance (Cheng et al., 2014) or by pre-pooling the tweets by hashtags (Mehrotra et al., 2013); once these techniques are tested and verified, they could represent improvements on the naive tweet-pooling used in this article.

Figure 1: Testing Different Numbers of Topics



Plotting the model fit for different numbers of topics. The dotted line connects log-likelihood estimates while the dark line connects perplexity estimates. Conservative approaches would call for 50 topics but I double that recommendation to allow for more variation in focus, the variable of interest.

To analyze how focused the coalitions are over time, I measure 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 (or Shannon Diversity) is well suited to measuring how focused these documents are. It efficiently captures information about the entirety of the distribution while avoiding imposing arbitrary thresholds.

The formula for Shannon Entropy is: $-\sum_{i=1} p_i \log_2(p_i)$, where each p_i is the number of individuals from species i divided by the total number of individuals in the ecosystem. Because the γ's in each document must sum to 1, a direct application of the formula generates a Shannon Entropy score for each document: $SE_w = -\sum_{k=1}^{100} \gamma_{w,k} \log_2(\gamma_{w,k})$. The possible SE scores range from 0 (if the γ distribution is unitary) to $\log_2(k = 100)$ (if the γ distribution is uniform). Generally, lower SE scores mean a more unequal distribution, and in the case being analyzed here, a more focused message.

7 Results

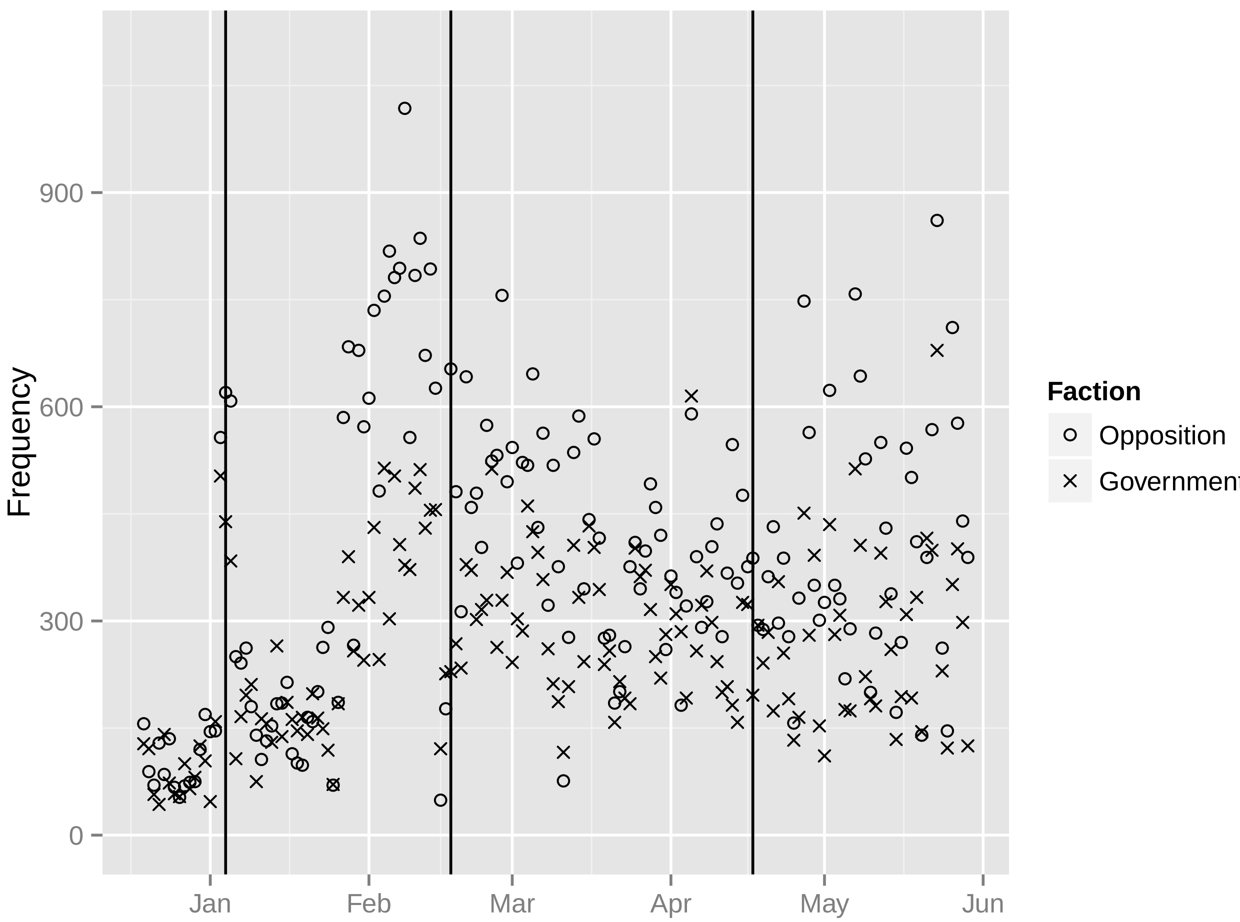
To test H_1, I need to demonstrate that Twitter activity among *diputados* increased in response to the protests. Figure 2 provides visual evidence that this is indeed the case. Though this effect is not causally identified, the protests were clearly the biggest events of the time, and the timing of the increased tweets is plausible.

Notice that before the protests, the tweet density was identical for the government and opposition. Both sides saw a flare-up around the time of the murder of Miss Venezuela on January 6th, which then subsided. The opposition began to tweet more often well in advance of the February 14 protest explosion, and sustained a higher level of tweeting throughout March and April. After Independence Day, the opposition had occasional spikes in activity and otherwise maintained their energetic Twitter usage.

The government also increased its rate of tweeting, but never reached the same heights. The government was generally less variable in the number of tweets it sent each day. This contrast in variability is consistent with the logic that the opposition was trying to bring attention to exogenous protest events whenever they happened while the government wanted to provide a steady stream of positive information.

This contrasts with the results from the Shannon Entropy scores, shown in Figure 3. Compare the evolution of the focus of the two coalitions over time. The SE scores track each other in the first period, but diverge afterwards. The government SE scores

Figure 2: Tweet Density by Each Coalition



The number of tweets sent by *dipudatos* from each faction per day. The vertical lines correspond to January 6 (the murder of former Miss Venezuela), February 18th (the arrest of Lòpez), and April 19th (Independence Day).

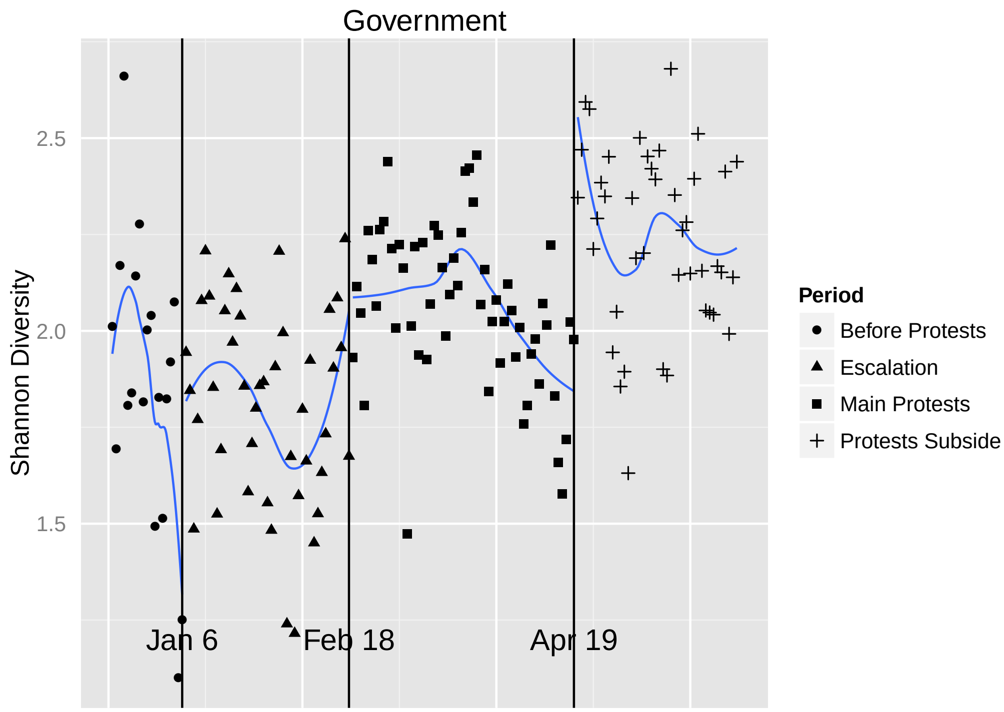
stay constant in the second period of heightened tensions, and then increase in the third period of full-blown protest. This is consistent with H_2 insofar as we expect that the focusing effects of the protests to be less pronounced for the government. However, that the degree of focus would actually *decrease* was unexpected.

The opposition, on the other hand, becomes more focused in the second and third periods. This is also consistent with H_2: the opposition message is focused on keeping citizens' attention on the crisis at hand, whatever the particular events of the day. What happens in the fourth period, after Independence Day and Easter, was unexpected: there's a large discontinuity in the opposition SE scores. This indicates that the messaging strategy of the opposition changed. Because there was no corresponding public event around which they might have coordinated, the discontinuity indicates a high degree of coordination among these *diputados*.

Though the data cannot conclusively explain why this happened, a reasonable interpretation is that the opposition "gave up" on the aggressive strategy of *La Salida*. Promoted and executed almost entirely by the radical faction of the opposition, *La Salida* was never the first choice of the moderate opposition elites, and almost all of the opposition *diputados* studied here were moderates. Indeed, the most radical *diputado*, Marìa Machado, was actually removed from her post. Although they were either strategically or ideologically opposed to the street protests and disruptions, these *diputados* wanted to present a unified front against the regime and thus were careful to focus attention on them 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 over policy. This explanation comports with the timing of the first major sit-down between the regime and the moderate opposition on April 10th. Though these talks were widely seen as a failure in that they did not lead to any concrete changes, they could have been a public show of power to determine the relative strengths of the bargaining parties. If that is the case, then the talks convinced the moderate opposition that the government was willing and able to continue along the status quo. The discontinuity in the fourth section of Figure 3 can be read as the opposition abandoning their strategy of supporting the radical wing of the opposition in favor of a return to their previous broad criticism of the regime.

The above analysis has emphasized the topic *distribution*, although the Shannon Entropy approach is agnostic about which topics are the biggest for 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 SE scores. Figure 4 provides a more detailed look at

Figure 3: Focus, as Modeled by Shannon Entropy, Over Time



The time under study divided into four periods to visualize the impact of real-world events on the degree of focus exhibited by the Twitter usage of the two factions. The government's focus remains constant or even increases gradually over time, while the opposition's focus decreases after the onset of the protests and then jumps upwards when they subside. The vertical lines correspond to the murder of Miss Venezuela, the arrest of Lòpez, and Independence Day.

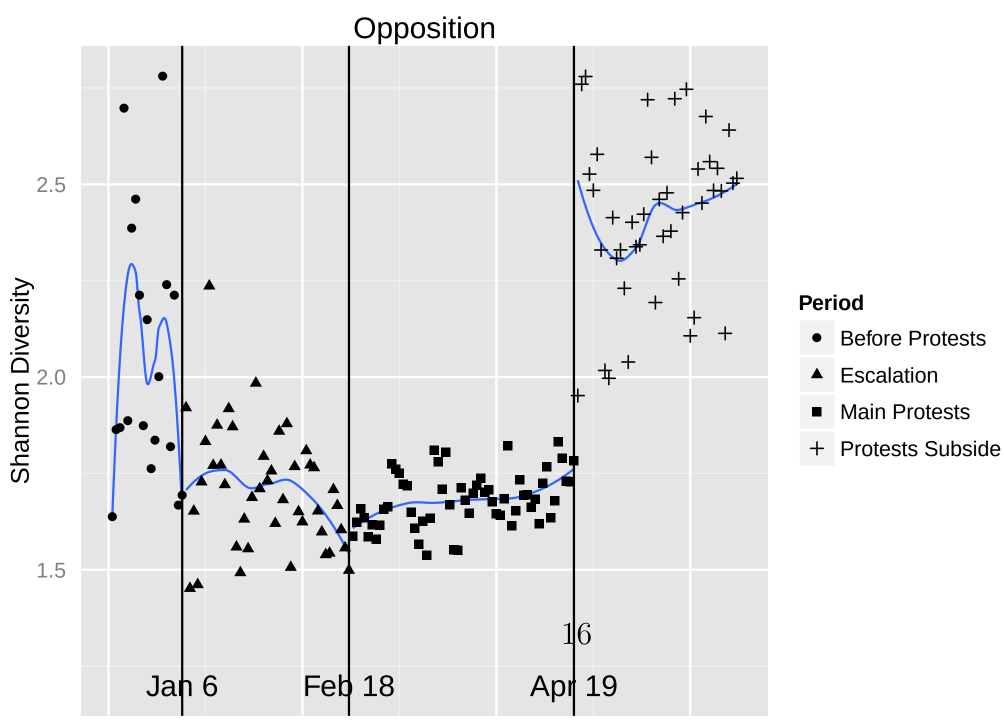
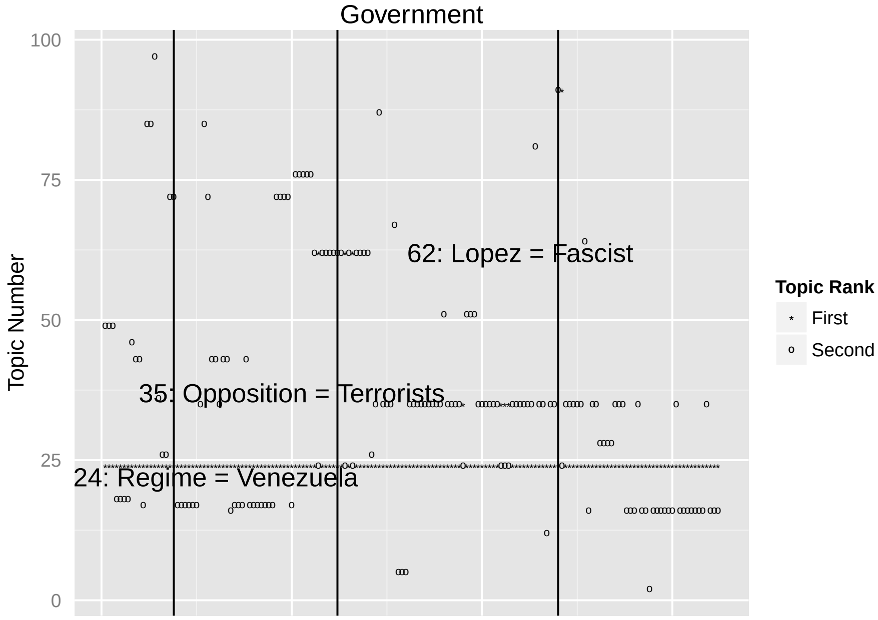
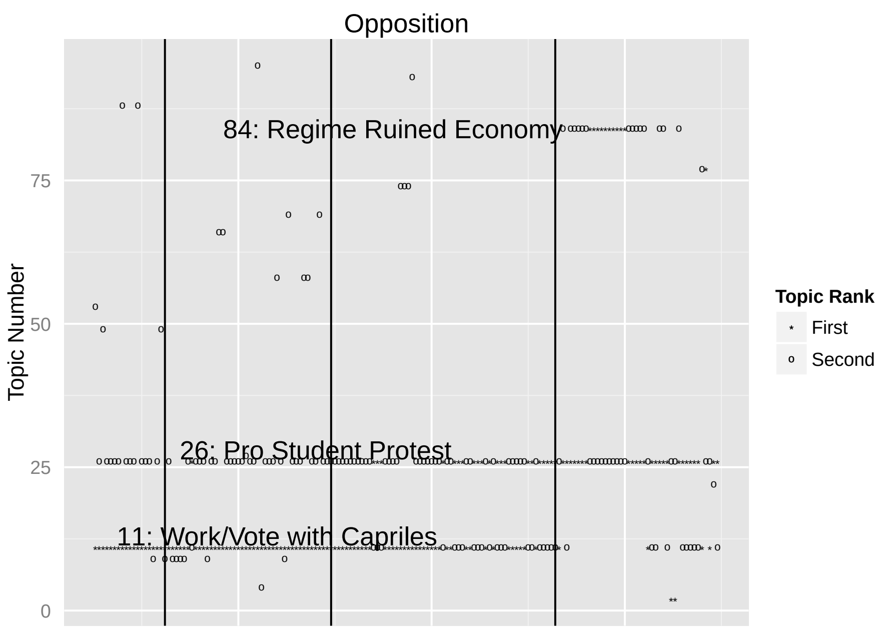


Figure 4: Top Topics Over Time



The first and second most popular topic each day for each faction. The three most popular government topics, from top to bottom, are: 62, which paints Lòpez as a fascist; 35, which describes the opposition as terrorists; and 24, which equates the regime with Chavez and Venezuela more generally. Though these topics dominate the discussion, the government emphasized a large number of different topics during the time period under study. The vertical lines correspond to January 6 (the murder of former Miss Venezuela), February 18th (the arrest of Lòpez), and April 19th (Independence Day).



The opposition's distribution of favorite topics is much tighter. 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**

#	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

which topics top each coalition's discussion on each day. In both graphs, it's clear that a small group of topics dominate each day's discussion, and that there is no overlap between the two coalitions' top topics. The government is more consistent with their number one topic, but features a wider range of second topics (25) than does the opposition (18).

The critical dates marked by the vertical lines provide less distinction than in the previous figures. The government switches from its overwhelming favorite topic (24: the regime is Venezuela) and adopts a new top topic (62: Lòpez is a fascist) around the time of Lòpez's arrest, and its second topic (35: the opposition are terrorists) solidifies during the main phase of the protest. For the opposition, however, the main effect is again seen in the wake of Easter/Independence 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 while its second topic (26: the student protest is good) remains important and becomes dominant by the end of the measurement period.

It is difficult to parse some of the terms in Table 2, and while direct translations can be difficult due to the stemming that is especially important for LDA with Spanish language terms, general summaries are possible.[10] Topic 24 cements the Chàvez-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 Lòpez–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*, a Venezuelan term for the permanent protest camps/blockades designed to paralyze the government in Caracas.

[10]Keep in mind that I over-partition the data by choosing to model 100 topics; these summaries are illustrative, but not the focus of my analysis.

For the opposition, Topic 11 tries to connect "the people" with moderate opposition leader Henrique Capriles and tells them to "vote" and "work," and to do these things "right now." Topic 26 talks about the student protests being the way forward for the country and specifically addresses the government and government officials. Topic 84 is harder to interpret, but a majority of terms have to do with government-mandated wage increases. Every year, the government raises mandatory salaries on May 1st, but in 2014 the increase was only 30% relative to an unofficial inflation rate of around 60%. Topic 84 also addresses *Primero Justicia*, a prominent opposition party headed by Capriles, and several affiliated *diputados*.

Table 2 gives the terms that comprise the two major topics address a specific event–84 for the opposition and 62 for the regime–and a comparison with Figure 4 shows that these topics become relevant precisely when they should, demonstrating the general validity of my method. There are various other topics that are similarly event-specific[11], but my over-partitioning of the data makes this kind of analysis, sometimes central to LDA models of political discussions over time, less appropriate.

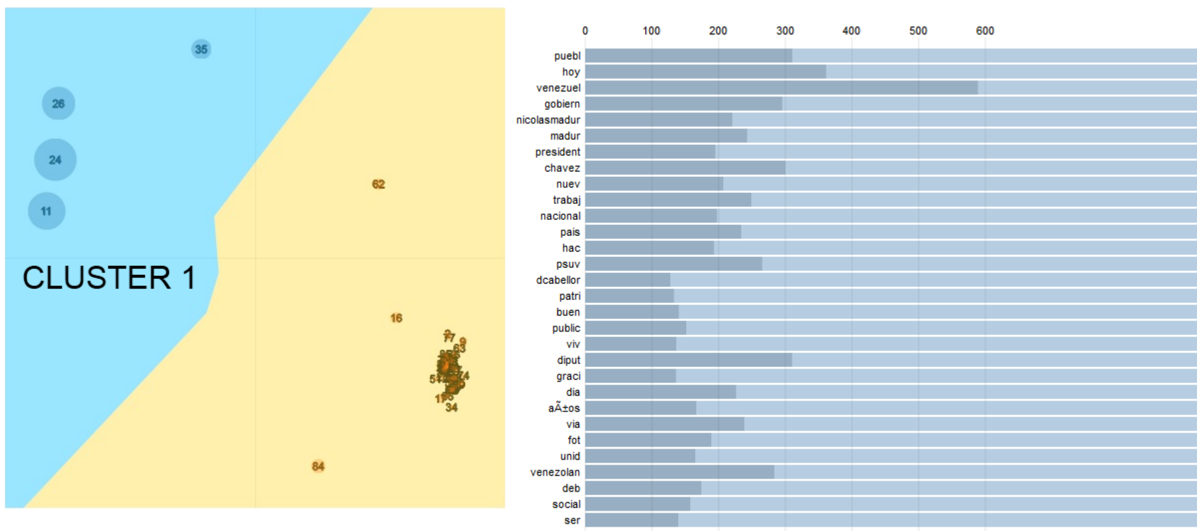
Because there is less of a dramatic shift between the first and second vertical lines in Figure 4 than in Figure 2, the increased Twitter usage predicted by H_1 does not appear to be driven by a change in the content of the messages. This supports the theoretical prediction that sheer quantity of information is an important variable in the coordination/disruption dynamic between the regime and the opposition.

The wider variety of top topics for the government than the opposition provides support for the differential increase in focus proposed in H_2. There is also evidence that this intentional, rather than because of lack of organization: the distinct shift to Topic 62 in the top panel of Figure 4 in response to an increasingly vocal and aggressive L'opez shows that the regime was capable of coordinating its *diputados* on certain topics if necessary. The switch to topic 84 after the third vertical bar in the bottom panel of Figure 4 provides evidence that the corresponding discontinuity in Figure 3 is part of a specific strategy on the part of the moderate opposition to cease to support *La Salida*.

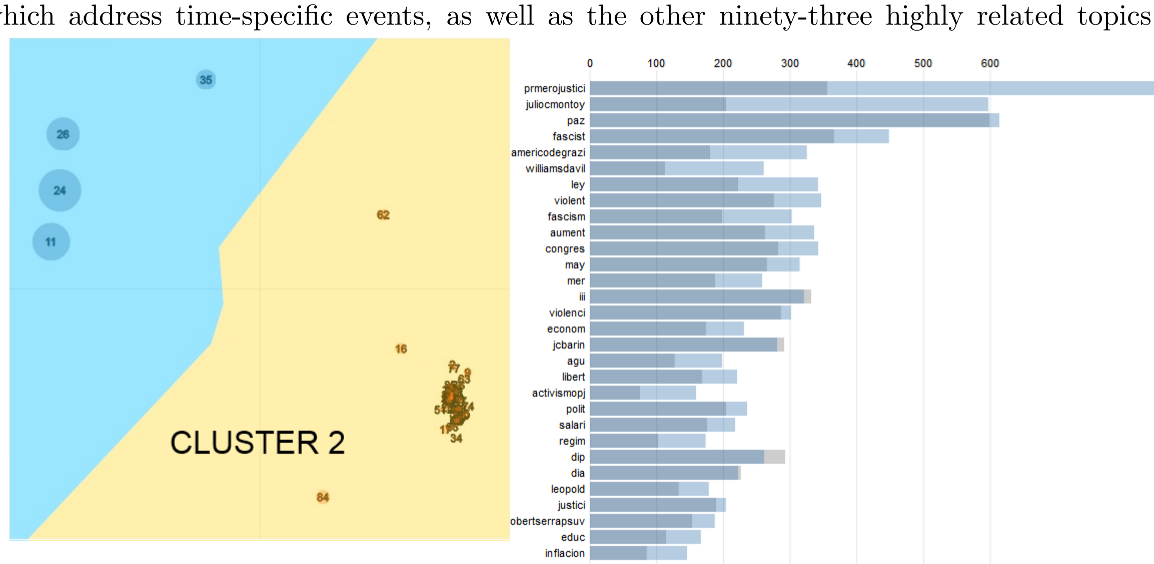
Figure 4 and Table 2 address only the most prominent topics. A better way to visualize the overall distribution of the topics can be seen in Figure 5, created with LDAvis. Developed by Sievert and Shirley (2014), LDAvis plots the topics created by LDA by creating circles proportional to each topic's term's frequencies over the entire corpus and positions them in two dimensions based on the overlap of terms in each topic.

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

Figure 5: Visualizing Topic Distribution with LDAVis



The bar graphs on the right compare how frequent each of the most relevant terms in each cluster is in the entire corpus (light line) to how frequent it is in that cluster (dark line). Above is the plot for Cluster 1, which contains 74.5 % of the corpus but only four of the topics. Below is Cluster 2, which contains three distinct topics (16, 62 and 84) which address time-specific events, as well as the other ninety-three highly related topics.



It also allows the user to cluster the topics, as seen in the two panels of Figure 5. Cluster 1 consists of the four main topics, which contain 74.5 % of the entire corpus, while cluster 2 captures the other ninety-six. Notice that topics 11, 24 and 26–the largest–are fairly close together, indicating that they are heavily composed of common terms. The major time-specific topics 84 and 62, however, are the farthest from other topics and thus contain the most distinct term frequencies. This independent visualization technique confirms the intuition in Figure 4. The majority of the conversation is dominated by the main topics from each side, which address roughly the same theme.

LDAvis offers another way to examine the terms that compose these important topics. The most relevant[12] terms in cluster 1 are too frequent in the total corpus for the light bars to be meaningful, but this bar chart does show both the most relevant terms in cluster 1 and their relative frequencies within that cluster. The most frequent term, for example, is "Venezuela," and many of the others are the sort of neutral, political terms that politicians would be expected to prioritize. There is a lack of partisan terms because of how the topics are aggregated–there are two topics from each coalition, and the politically-motivated terms thus correspond to no more than half the cluster. Cluster 2's bar graph is more interesting. The most prominent term, "prmerojustici," dominated topic 84, but is not itself dominated even by all of cluster 2. On the other hand, "paz," the top term from topic 62, is almost entirely contained in cluster 2. Generally, the most relevant terms in cluster 2 are nearly exclusive to it. LDAvis is optimized to be used interactively, though, and a fuller exploration of the information contained in Figure 5 is available in the Online Appendix.

Figures 2, 3, 4 and 5 provide different ways of visualizing the same data, and each provides some degree of support for my hypotheses. Figure 2 shows that the tweet frequency for both coalitions was highest as the protests were gaining momentum and higher for the duration of the protests than before they started, strong evidence for H_1. Figure 3 shows that this increase actually coincides with a decrease in the Shannon Diversity and thus increase in focus for the opposition while barely changing the focus for the government, in line with H_2. The lack of a change in the top topics during this time period documented in Figure 4 indicates that this change in focus was a strategic choice in and of itself. Figure 5 demonstrates that most of the changes in topics take

[12]This is a technical term proposed by Sievert and Shirley (2014) to choose which terms to display to summarize an LDA topic; it combines straightforward term frequency, which is how the terms in Table 2 were selected, with "lift," the ratio of a term's probability within a topic to its probability throughout the corpus.

place at the margins and that the bulk of communication for the entire corpus can be modeled with just a few topics. These latter two Figures support the intuition behind H_2 that political communication in the context of a coordination problem is more about focus than it is about content.

8 Conclusion

The Venezuelan protests of 2014 were a failure. The regime maintained its grip on power; though the underlying macroeconomic and social issues persist, the regime demonstrated their repressive capacity to the opposition. Part of their strategy was to target specific aspects of the opposition with messages on Twitter while not focusing too heavily on the protests. The opposition tried to focus attention on the protests, but their rapid reversal to a more specific economic criticism after Easter weekend represented a tipping point for their cause; the precise timing and degree of coordination could not have been observed without analyzing Twitter usage. This strategy by the opposition might have been one of necessity rather than their first choice; the *diputados* represented the moderate faction of the opposition and may have been balancing their desire to provide a unified front with skepticism of Lòpez and Machado's radical "*La Salida*" to oust Maduro.

Analysis of the kind presented in this paper could not have been done without social media data. New information technology always increases the amount of information in society (at least in "hybrid" regimes like Venezuela in 2014) and, relative to traditional media, makes it more difficult for the regime to control the information that is being spread. Theory predicts that in this context, the regime's best response is to counterbalance the high number of messages from the opposition with a high number of messages themselves. Using Twitter allowed me to gather data in a well-defined sphere of communication and provide evidence in support of these theoretical predictions.

Although much of the discourse on regime response to protest focuses on either physical repression and intimidation or on fraud, the theoretical literature on global games emphasizes the role of information and how it contributes to citizens solving the coordination problem that is revolution. I have demonstrated that elites try to manipulate that information to advance their own ends. Future research in this area should treat these elites as individual actors, to see how the variable incentives they face change their strategies.

References

Acemoglu, Daron, and James A Robinson. 2005. *Economic origins of dictatorship and democracy.* Cambridge University Press.

Adena, Maja, Ruben Enikolopov, Maria Petrova, Veronica Santarosa, and Ekaterina Zhuravskaya. 2013. "Radio and the rise of Nazis in pre-war Germany." *Available at SSRN 2242446* .

Angeletos, George-Marios, Christian Hellwig, and Alessandro Pavan. 2006. "Signaling in a global game: Coordination and policy traps." *Journal of Political Economy* 114 (3): 452–484.

Barberá, Pablo. 2013. "Birds of the Same Feather Tweet Together. Bayesian Ideal Point Estimation Using Twitter Data." *Proceedings of the Social Media and Political Participation, Florence, Italy* pp. 10–11.

Barbera, Pablo, Richard Bonneau, John T Jost, Jonathan Nagler, and Joshua Tucker. 2013. "Is There Anybody Out There? The Effects of Legislators' Communication with their Constituents.".

Bertot, John C, Paul T Jaeger, and Justin M Grimes. 2010. "Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies." *Government Information Quarterly* 27 (3): 264–271.

Blei, David M, Andrew Y Ng, and Michael I Jordan. 2003. "Latent dirichlet allocation." *the Journal of machine Learning research* 3: 993–1022.

Bohdanova, Tetyana. 2014. "Unexpected revolution: the role of social media in Ukraines Euromaidan uprising." *European View* pp. 1–10.

Bueno de Mesquita, Bruce, Alastair Smith, Randolph Siverson, and James D. Morrow. 2005. *The logic of political survival.* MIT press.

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.

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

Cogburn, Derrick L, and Fatima K Espinoza-Vasquez. 2011. "From networked nominee to networked nation: Examining the impact of Web 2.0 and social media on political participation and civic engagement in the 2008 Obama campaign." *Journal of Political Marketing* 10 (1-2): 189–213.

Cormack, Lindsey. 2013. "Gender and Vote Revelation Strategy in Congress." *Unpublished Manuscript* .

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.

de Mesquita, Ethan Bueno. 2010. "Regime change and revolutionary entrepreneurs." *American Political Science Review* 104 (03): 446–466.

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

Esfandiari, Golnaz. 2010. "The Twitter Devolution." *Foreign Policy* 7: 2010.

Gerbaudo, Paolo. 2012. *Tweets and the streets: Social media and contemporary activism*. Pluto Press.

Golbeck, Jennifer, Justin M Grimes, and Anthony Rogers. 2010. "Twitter use by the US Congress." *Journal of the American Society for Information Science and Technology* 61 (8): 1612–1621.

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.

Gulati, Jeff, and Christine B Williams. 2010. "Communicating with constituents in 140 characters or less: Twitter and the diffusion of technology innovation in the United States Congress." *Available at SSRN 1628247* .

Harfoush, Rahaf. 2009. *Yes We Did! An inside look at how social media built the Obama brand*. New Riders.

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, and Muzammil M Hussain. 2013. *Democracy's Fourth Wave?: Digital Media and the Arab Spring.* Oxford University Press.

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.

Hsu, Chien-leng, and Han Woo Park. 2012. "Mapping online social networks of Korean politicians." *Government Information Quarterly* 29 (2): 169–181.

Jungherr, Andreas. 2010. "Twitter in politics: lessons learned during the german superwahljahr 2009." *CHI 2010* pp. 10–15.

Khondker, Habibul Haque. 2011. "Role of the new media in the Arab Spring." *Globalizations* 8 (5): 675–679.

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.

Lotan, Gilad, Erhardt Graeff, Mike Ananny, Devin Gaffney, Ian Pearce et al. 2011. "The Arab Spring— the revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions." *International Journal of Communication* 5: 31.

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.

Metzger, M. M., D. Penfold-Brown, P. Barbera, R. Bonneau, J. Jost, J. Nagler, and J. Tucker. 2014. "Dynamics of influence in online protest networks: Evidence from the 2013 Turkish protests." *Paper presented at the annual meeting of the Midwest Political Science Association* .

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

Morris, Stephen, and Hyun Song Shin. 1998. "Unique equilibrium in a model of self-fulfilling currency attacks." *American Economic Review* pp. 587–597.

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.

Robertson, Graeme. 2011. "The politics of protest in hybrid regimes." *Managing dissent in post-communist Russia* .

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

Sievert, Carson, and Kenneth E Shirley. 2014. "LDAvis: A method for visualizing and interpreting topics.".

Tucker, Josh, Megan Metzger, Duncan Penfold-Brown, Richard Bonneau, John Jost, and Johnathan Nagler. 2014. "Protest in the Age of Social Media: Ukraines Euromaidan." *Carnegie Corporation* .

Vaccari, Cristian, and Rasmus Kleis Nielsen. 2013. "What Drives Politicians' Online Popularity? An Analysis of the 2010 US Midterm Elections." *Journal of Information Technology & Politics* 10 (2): 208–222.