Elites Tweet to get Feet off the Streets: Measuring Elite Reaction to Protest Using Social Media

Kevin Munger, Rich Bonneau, John T. Jost
Jonathan Nagler, Joshua Tucker*

January 26, 2015

Abstract

Social media use offers a useful avenue for analyzing strategic elite response to protest. We examine the frequency and content of Twitter usage among Venezuelan elites in the context of the 2014 protests. This analysis demonstrates that, compared to the opposition, the regime sent more messages during protests, and that the content of these messages addressed more topics. This observation supports theoretical predictions that elites strategically manipulate public information to promote or inhibit coordination among citizens. We also discuss the role of features unique to the medium of Twitter that elites use differentially to accomplish their respective goals. Our approach, of taking the strategic element of elite use of social media more seriously, contributes to the debate over whether social media hinders or promotes regime change.

*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 Invesetigators. km2713@nyu.edu

1 Introduction

Social media enables protesters to disseminate information and organize themselves efficiently and discreetly in the face of repressive regimes, as in the Arab Spring, Turkey in 2013 and Ukraine in 2014 (Bohdanova, 2014; Howard and Hussain, 2011, 2013; Metzger et al., 2014; Tucker et al., 2014). The widespread adoption of social media has changed the way that protests operate, and these changes leave a visible footprint: the use of Facebook and Twitter produces a huge and detailed collection of communication that was previously difficult or impossible to access.

Much of the interest in the subject of social media and protest has centered on ordinary citizens. Twitter makes it much easier for Western journalists and academics to learn about individuals engaged in protest on the ground, and the near-instantaneous spread of tactical information among protesters was simply impossible just ten years ago. However, the fact that social media has facilitated protest movements does not itself allow us to draw conclusions about whether it actually increases the chance of overthrowing repressive regimes. Regime elites can and do use social media to defuse protest movements.

We investigate elite Twitter use in response to protest. Regime elites oppose protests, whereas opposition elites support protests; we test specific predictions about how these behaviors manifest themselves on Twitter. In order to make the analysis of the corpus of tweets more tractable, we use a topic modeling algorithm that allows us to see how many and which topics each coalition emphasizes during the course of the protest. The crucial advantage of Twitter as a data source is that it provides a dense record of strategic elite communication from which we can extract trends, which then allow us to make inferences about the underlying strategy.

The Venezuelan protests of 2014 have been relatively understudied (perhaps because they were unsuccessful), and they offer an excellent opportunity to test our hypotheses. Although Venezuela edged toward authoritarianism in the 21st century, Venezuelan elites continue to have free speech. 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 Chávez is a close second, despite being dead for over a year.

To ensure a balanced sample of elites, we investigate the Twitter accounts of Venezuelan *diputados*, members of the unicameral National Assembly. 139 of 165 *diputados* had

Twitter accounts, and they tweeted over 100,000 times during the 5 months under study, giving us ample analytical leverage. Our findings provide support for our hypotheses: the Twitter use of the two coalitions differs in predictable and important ways that provide insight into their proximate and ultimate aims.

The paper proceeds as follows: Section 2 provides a background of the situation in Venezuela; Section 3 discusses the relevant literature; Section 4 outlines our hypotheses in detail; Section 5 presents the data; Section 6 explains the methodology; Section 7 gives results; Section 8 concludes.

2 Recent Venezuelan History

Venezuela has been a democracy since 1958.[1] The first 40 years 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. Though Chávez 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. After surviving an attempted coup and an oil worker strike in 2002 as well as a recall attempt in 2004, Chávez possessed the political capacity to favorably renegotiate contracts with international oil companies operating in Venezuela. This provided fiscal flexibility in the short term and, in concert with the oil boom that began soon thereafter, allowed the government to consolidate its rule by dramatically increasing social spending (Corrales and Penfold-Becerra, 2011).

The formula worked for Chávez: he won re-election in 2006 and 2012 handily and with very high voter turnout, even as inflation and crime rates rose to high levels. His health declined rapidly in 2012, and he died in early 2013. Vice president Nicolás Maduro's assumption of the presidency was controversial, and it was not obvious who would ultimately represent Chávez'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 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 a violent crime rate among the worst in

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

the world led to rising discontent. The sparks that ignited the protests prompting our 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 Táchira in the southeastern part of the country. The latter event led to a small student protest against high crime rates that provoked a violent governmental response (Perez, 2014). This inspired a much larger protest that eventually spread to Caracas. Radical opposition leaders, especially the faction headed by María Corina Machado and Leopoldo López, had been in the planning stages of a movement to advocate for regime change (known as *La Salida*), and they rode the wave of student protest to put their plans into action. Things escalated quickly, and López became sufficiently outspoken and high-profile that the regime jailed him on February 18th.

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. These *colectivos* often took the form of masked men on motorcycles armed with shotguns. The protesters *en guarimba* countered this tactic with thin tripwires at head level. 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 10th, though no conclusion or agreement was announced. April 19th and 20th were Independence Day and Easter, respectively, and this weekend represented the high water mark of the protests.

The protests began to wind down in May after 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, out of opposition solidarity or in order to improve their baragaining power or electoral chances in the future. In any case, *La Salida* was never the first choice of the opposition *diputados* under study.

3 Literature on the Role of Social Media in Politics

The importance of social media as a political technology has been discussed since its advent in the early 2000s. Analysis of social media has largely 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 to position themselves. In these contexts, social media has been seen as useful primarily because of its accessibility to the researcher, rather than a categorically different form of communication. There is a growing literature on Twitter use by US Members of Congress (Golbeck, Grimes, and Rogers, 2010) and the ways in which it differs by gender (Cormack, 2013); which factors determine a politicians' number of followers (Gulati and Williams, 2010; Vaccari and Nielsen, 2013); and whether politicians tend to set the agenda or respond to their followers (Barberá et al., 2013). Analyses of this latter kind have also been undertaken 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 elite social media usage in non-democracies. Work in this area instead centers on the way in which social media changes the dynamic between governing regimes and opposition groups. Spurred by the unexpected Arab Spring and the visible use of social media by protesters, scholarship on the role of social media was initially enthusiastic about its revolutionary potential (Gerbaudo, 2012; Khondker, 2011; Lotan et al., 2011). This enthusiasm stemmed perhaps from a failure to appreciate that the regime's adoption of new technologies makes the net impact of those technologies at best ambiguous. There are, however, several studies that treat authoritarian regimes as strategic actors who can react to the use of social media by their opponents (Esfandiari, 2010). Regimes have various strategies for counteracting 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 technological capacities of the two sides, social media might even tip the balance of power in favor of the regime (Morozov, 2011; Rahimi, 2011).

These two literatures each address important aspects of the role of social media in politics, but they are as yet disjointed. Elite usage of social media in non-democratic

contexts is a lacuna in the literature which this article intends to fill.

One possible reason that this area has attracted less research attention is that the questions that can be answered with non-democratic elite social media data are different than in other contexts, though no less interesting. Because it is explicitly propagandistic, the content of this communication cannot be taken at face value. But it is precisely this aspect that makes it especially useful for the quantitative analysis of regime strategy: the words may be cheap talk but the motivation is serious, as the fortunes of the regime may hang in the balance. Modern methods for textual analysis allow for the detection of signals that the elites could not have realized they were sending. Additionally, the personal nature of social media makes visible the messages directed to specific individuals that would previously have been unobservable.

A dramatic example of this came from radical mayor Leopoldo López. In early February, as the protests were gaining momentum, there were rumblings from the regime that if he did not moderate his rhetoric, he would be arrested. López took to Twitter and directed his response to 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"[2] (author's translation). López was arrested soon afterwards.

Twitter allowed López to spread his message directly and quickly to a wide mass of people, with no risk that it would be distorted or misrepresented. Before Twitter, with the regime controlling or exerting pressure on traditional media outlets, this would not

[2]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.

have been the case, even for a figure as influential as López. Twitter has even more transformative potential for marginal politicians who would not otherwise be able to make their personal views known.

4 Hypotheses

The opposition and regime *diputados* had opposing goals with respect to the protest movement, and a number of strategies they might employ. Regime response to social media use in protest varies widely, from Egypt shutting down internet access entirely to US elites largely ignoring Occupy Wall Street. In order to make sense of the trends we observe, we have to compare regime and opposition strategies. There are two stark possibilities: the two coalitions behave symmetrically or asymmetrically. We had reason to predict the latter.

Hypothesis 1 *The response of regime and opposition diputados will be quantitatively different.*

Our motivation for H_1 comes from the literature on the role of information in global games (Angeletos, Hellwig, and Pavan, 2006; Morris and Shin, 1998). The mechanism by which elite communication tends to influence support for the protest is through the amount and type of information that citizens acquire. The opposition strategy is thus intuitive: they want to produce as many tweets as they can, and they want them to be focused on the protest. We expect the regime to behave differently.

The game modeled in Edmond (2013) guides our expectation of the regime response. In the model, the regime has to choose an information strategy to prevent the citizens from overthrowing them. If everyone knew the truth about the regime, they would favor overthrow, so the regime inflates its true quality. However, if this inflation is predictable, the citizens will be able to effectively discount it. The regime's optimal strategy is thus to introduce a randomly varied amount of noise into the signals they send, resulting in a more diverse messaging strategy. Edmond calls this "signal-jamming" because it maximizes the confusion about the true state of the regime experienced by each citizen.

This is a reasonable goal because revolution, both in theory and in practice, is a coordination problem. For each citizen, being confident that the regime is worth overthrowing does not imply that she should take to the streets: she also has to believe that enough other citizens will join her that the revolution will be successful. Signal-jamming thus decreases the confidence of each citizen in his knowledge *both* of the true nature of the regime and in the opinions of other citizens. There will be some people who are utterly convinced that the regime should be overthrown, but this strategy sets them aside: the regime's goal is to prevent those who are not already convinced from becoming convinced and passing the threshold necessary for revolution.

The first specific test of H_1 addresses the frequency of tweets in response to protest. The opposition should tweet as frequently as they can. The regime response should be more nuanced: they want to engage in signal-jamming by producing enough tweets, but if they react too strongly, they will appear to be too concerned about the threat posed by the protests. The Edmond model makes no specific prediction about message frequency, but we predict that in a context like Twitter, where the number of messages can vary dramatically, a dramatic change in frequency is itself a strong signal.

Hypothesis 2 *Both coalitions will tweet more in response to the protests, but the increase will be greater for the opposition.*

The protests will also change the content of the elite tweets. The opposition should rapidly focus much or all of their attention on the protests. The regime, though, should avoid doing this because it gives the protests too much credibility. Instead, they should signal-jam by tweeting about various non-trivial political issues, reducing the protests to one issue among many and preventing the citizens from coordinating. We operationalize these measures in two ways, the details of which are discussed below.

Hypothesis 3 *During the protests, the opposition will become more focused, but the regime will not.*

Hypothesis 4 *During the protests, there will be more different topics that dominate a day's discussion for the opposition than for the regime.*

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 latter is more common among elites who aim to guide online discourse, and we expect them to be especially common among the regime:

Hypothesis 5 *During the protests, the regime will use lengthy discourse-structuring hashtags more frequently than the opposition.*

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

5 Data

For each of the 99 regime and 66 opposition *diputados* in Venezuela's unicameral legislature, we 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 which possessed a significant number of followers. 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.[3] 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.[4] For the subsample whose tweets are analyzed in this paper–those who tweeted after December 18, 2013–there are 135 accounts, 65 regime and 56 opposition. Table 1 contains a summary of the number and distribution of the tweets collected.

Although the regime had a higher number of 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 indicates that the opposition is more active throughout the distribution.

We used Twitter's REST API[5] via `tweepy`[6], in the Python programming language,

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

[4]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.

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

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

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. We did this on April 19th and then again on May 29th, 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 other bits of information 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.

6 Analysis

To explore the topics being discussed on Twitter, we employed 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 Barberá et al. (2013) point out, 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.

The input to LDA is 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. It also requires the specification of two parameters: K, the number of topics to be modeled, and α, a concentration parameter which determines the shape of the probability distribution central to LDA. LDA functions by treating each document as a probability 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 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 29th, 2014. We divided each day's worth of tweets by each coalition into a separate document. 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

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

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, this information is not germane.[8] 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).[9]

To determine the number of topics k, we 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. Although the model fit improves monotonically in the number of topics, the gains from adding more topics 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.[10] For α, we follow Griffiths and Steyvers (2004) and set $\alpha = 50/k = .5$.

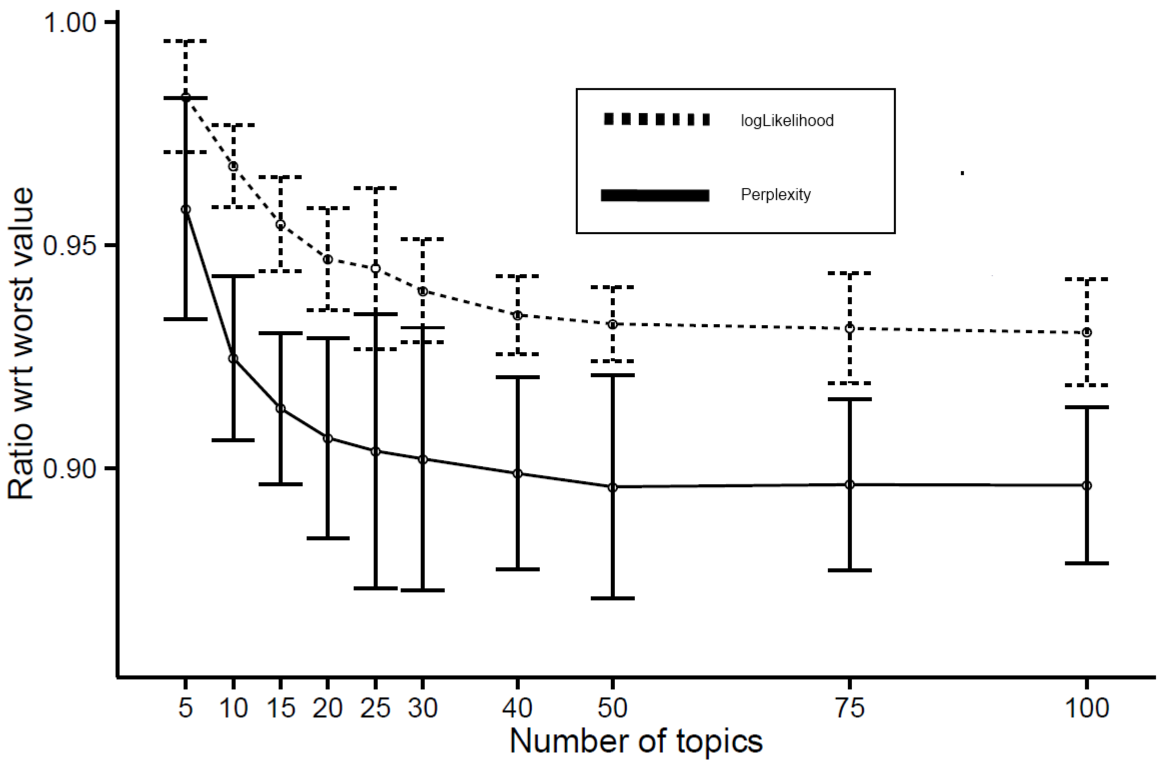
Following the standard in the social science literature, we use the collapsed Gibbs

[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 because the shape of the distribution is similar for the two coalitions, this weakness does not impair our comparative analysis.

[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 using 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 fully validated, they could represent improvements on the naive tweet-pooling technique used in this article.

[10]The main finding of different degrees of *focus* in Figure 2 is statistically significant for k= 30 and k= 50, though the time period for which this is true is narrower than for k =100.

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 we double that recommendation to allow for more variation in *focus*, the variable of interest.

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 consisted of $N = 50,902$ terms.

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

To analyze how *focused* the coalitions are over time, we 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 the imposition of arbitrary thresholds.

The formula for Shannon Entropy is $-\sum_{i=1} p_i \log_2(p_i)$. In this case, because the γ's in each document must sum to 1, $p_i = \gamma_{wk}$ and the Shannon Entropy score for each document is $-\sum_{k=1}^{100} \gamma_{w,k} \log_2(\gamma_{wk})$. The possible Shannon Entropy scores range from 0 (if the γ distribution is unitary) to $\log_2(k = 100)$ (if the γ distribution is uniform). Generally, lower Shannon Entropy scores mean a less uniform distribution, and in the case being analyzed here, a more *focused* message.

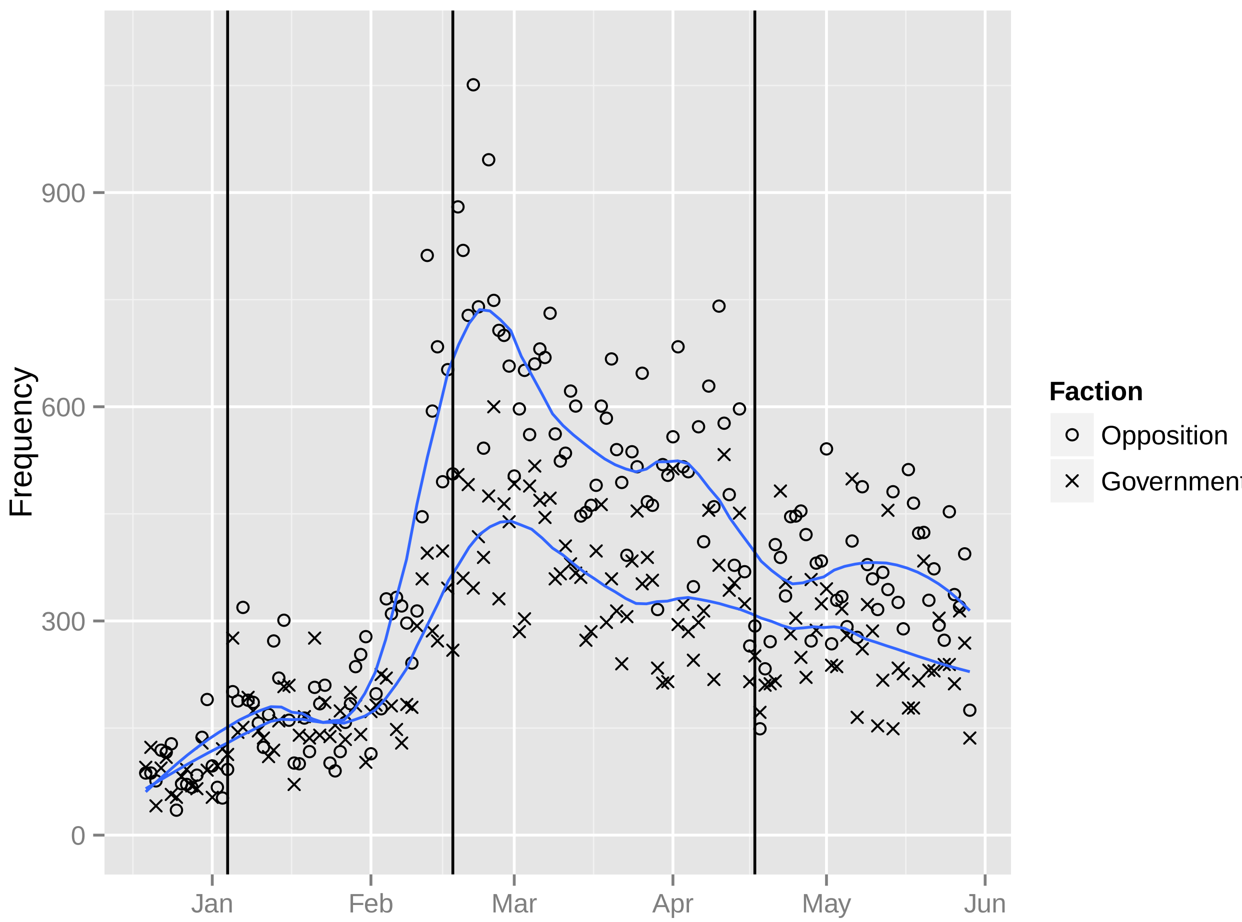
7 Results

In Figures 2, 3, 4 and 5, the dates being analyzed are divided into four periods: (1) December 19-January 6, before the protests; (2) January 7-February 18, protests begin; (3) February 19th-April 19th, ongoing protests; (4) protests subside. For a visual timeline of the events, see Appendix C.

To test H_2, we demonstrate that Twitter activity among *diputados* increased in response to the protests. Figure 2 provides visual evidence that this is indeed the case.

Notice that before the protests, the tweet density was similar and low for both the

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 López), and April 19th (Independence Day).

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, opposition Twitter usage trended downwards.

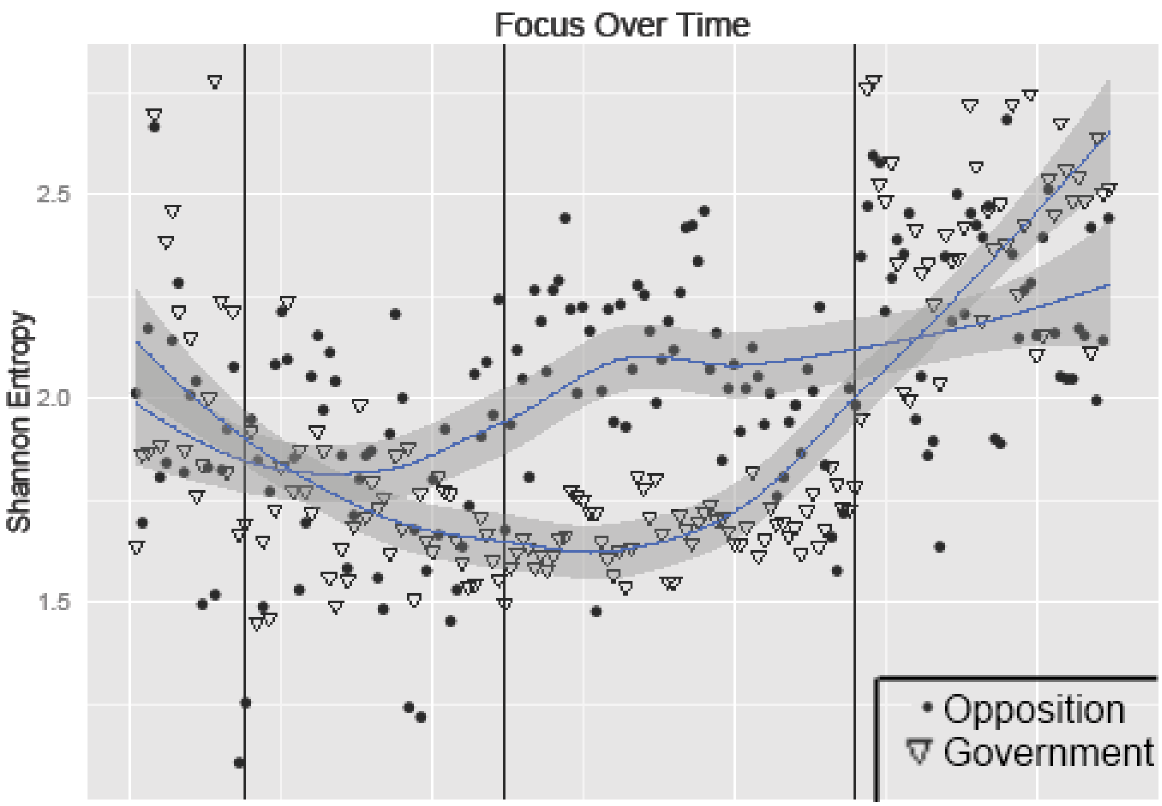
The government also increased its rate of tweeting, but never reached the same level as the opposition. The government was generally less variable in the number of tweets it sent each day. This contrast in variability is consistent with the logic that the opposition was tweeting as much as they could, whereas the government wanted to provide a steady stream of information without giving too much credibility to the protesters.

The trends in Figure 2 contrast with the results from the Shannon Entropy scores, shown in Figure 3. Compare the change in the *focus* of the two coalitions over time. The Shannon Entropy scores for each coalition in the first period are without noticeable trends, but they diverge afterwards. Shannon Entropy scores for the government stay constant in the second period except for the week or so before López's arrest, and then return to roughly their base level in the third period, which is the full-blown protest. This is consistent with H_3: the protests had little effect on the average *focus* in the tweets sent by the government.

The behavior of the opposition also supports H_3: they became more *focused* in the second and especially third periods. The opposition strategy seems to have centered on keeping citizens' attention on the crisis at hand, whatever the particular events of the day. The second and third periods for the opposition are distinct not just because they demonstrate the highest average degree of *focus*. The variance in *focus* is also lowest for these two periods. These are also the only time periods when there is a statistically significant difference in the *focus* of the coalitions. What happens in the fourth period, after Independence Day and Easter, was unexpected: there is a discontinuity in the opposition Shannon Entropy scores. This suggests that the messaging strategy of the opposition may have changed. Because there was no corresponding public event around which they might have coordinated, the discontinuity is consistent with a high degree of coordination among these *diputados*.[11]

[11] 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 3: Focus–as Modeled by Shannon Entropy–Over Time

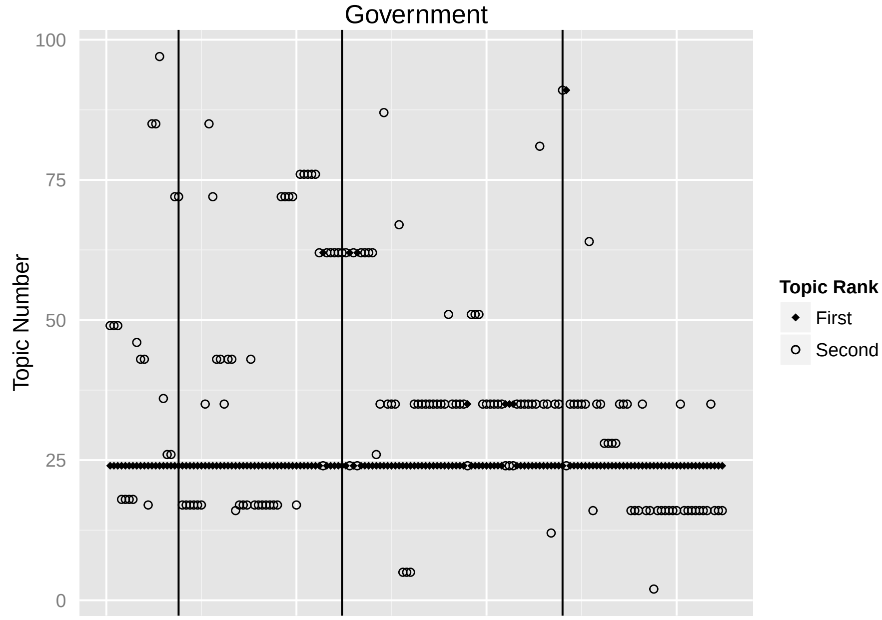


Trend lines and confidence intervals were created with loess. The vertical lines correspond to the murder of Miss Venezuela, the arrest of López, and Independence Day, respectively. The government's *focus* remains constant or even decreases gradually over time, while the opposition's *focus* increases after the onset of the protests and then jumps downward when they subside.

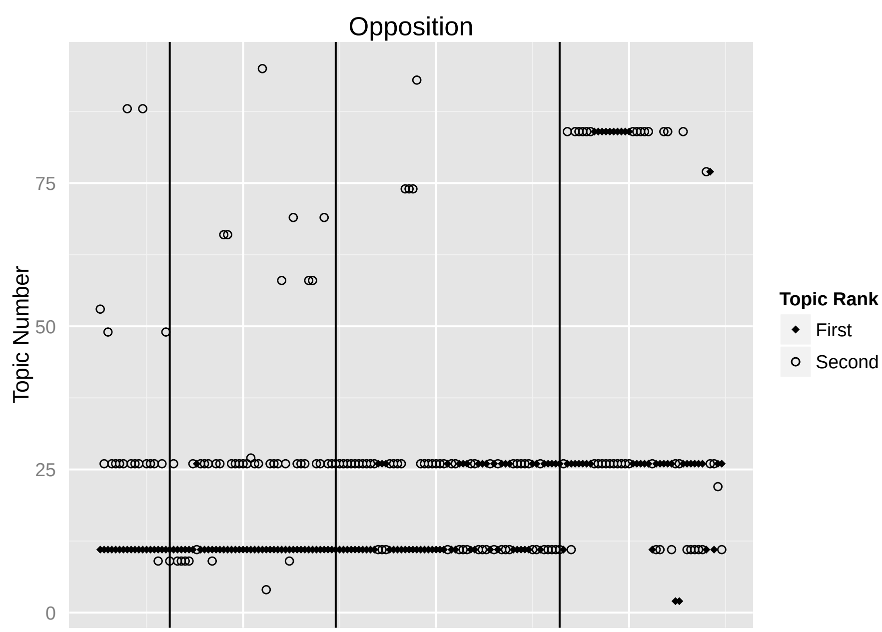
Although we cannot conclusively explain why this happened solely by looking at these data, a reasonable interpretation is that the opposition "gave up" on the aggressive strategy of *La Salida* to oust the Maduro government. Promoted and executed almost entirely by the radical faction of the opposition, *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*, María Machado, was actually removed from her post by the regime. Although the opposition *diputados* were either strategically or ideologically opposed to the street protests and disruptions, they may have 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 10th. 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. The discontinuity in the fourth section of Figure 3 may indicate that the moderate opposition abandoned their strategy of supporting the radical wing of the opposition in favor of a return to their earlier broad-based criticism of the regime.

The above analysis has emphasized the topic *distribution*. The Shannon Entropy approach is agnostic about which topics are the most important 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 Shannon Entropy scores. Figure 4 provides a more detailed look at which topics are most central to each coalition's discussion on each day. In both graphs, it is 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 held more consistently to their number one topic, but it also featured a wider range of second topics (25) in comparison with the opposition (18), which is not much of a difference. H_4 is supported more specifically by the fact that during the height of the protest (period 3) there are only 4 opposition top topics compared to 11 for the government. 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 period 3 even more striking.

Figure 4: Top Topics Over Time



The first and second most popular topics each day for each faction. 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. Although these topics dominated the discussion, the government emphasized a large number of different topics, especially during period 3. The vertical lines correspond to January 6 (the murder of former Miss Venezuela), February 18th (the arrest of López), and April 19th (Independence Day).



The number of the opposition's favorite topics is much smaller, especially in the crucial third period. 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	*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 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.

The critical dates marked by the vertical lines provide less distinction than in the previous figures. 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/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) remained important and became dominant by the end of the 4th period.

Table 2 lists the most common terms used in each of the most important topics.[12] 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

[12] Keep in mind that we over-partition the data by choosing to model 100 topics; these summaries are illustrative, but not central to our analysis.

workers on May 1st, 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 terms in Table 2 demonstrate that the two topics that correspond to specific events–84 for the opposition and 62 for the regime–become prominent precisely when they should (see Figure 4), demonstrating the general validity of our method. There are various other topics that are similarly event-specific[13], 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.

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_2 does not appear to be driven by a change in the content of the messages. This supports the theoretical prediction that the 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 H_4, and it may also provide insight into the specific strategy used by the regime. The distinct shift to Topic 62 in the top panel of Figure 4 in response to an increasingly vocal and aggressive López 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 supporting *La Salida*.

The final piece of evidence of differential Twitter strategies we aim to test is the use of hashtags. Figure 5 presents the trends in the use of hashtags. On Twitter, hashtags are 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.

There are two common types of hashtags: simple topic words that characterize the general subject to which a tweet pertains ("#Venezuela"), and phrases explicitly designed to create a topic ("#MaduroProtectorDeVenezuela"). The latter phrases are often used by social media elites who have enough followers and influence to generate

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

a discussion on Twitter. #MaduroProtectorDeVenezuela was used by the regime to promote the notion that Maduro and his government were Venezuela's only defense against violent protesters.

Research on the use of Twitter in the US indicates that the latter type of hashtag is more common among Republicans than among Democrats (Beauchamp, 2013; Livne et al., 2011). This may be due to the fact that Republicans tend to have tighter networks and so their elites are better able to structure the discourse around topics of their choosing. The situation in Venezuela may be analogous. The median number of characters in the opposition's 100 most-used hashtags was 9. For the regime, it was 15, indicating that most of the hashtags they use are phrases rather than words.

Figure 5 illustrates the evolution of hashtag use during the time period of interest. Before the protests started, the levels and trends of hashtag use were identical, but they diverged in the second and third periods, with the government's rate of hashtag use increasing dramatically. This may provide evidence of the strategic use of Twitter to set the agenda and to promote different topics at different times, and provides support for H_5.

To illustrate, the regime *diputados* used the following hashtags at different times to structure public discourse among their followers: #DesconectateDeLaGuarimba encouraged people to abandon the protest camps in Caracas; #ChavezViveLaPatriaSigue claimed that the spirit of Chavez lived on and that patriots should support Chavistas; #GringosyFascistasRespeten called their opponents US sympathizers and fascists and said that they need to respect Venezuelan sovereignty; #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 a critical frame and to generate discussion of the situation that was unrelated to the demands and complaints of the protesters.

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: divergent strategies by the two coalitions. The regime used longer, phrase-length hashtags to create conversation about a specific subject that would not otherwise have been discussed. This was an attempt to create engagement with their followers, to get their followers to add their own evidence or opinion about "#GringosyFascistasRespeten," for example. If a normal citizen used the hashtags the regime promoted, his followers would see his take on the subject without seeing the regime's direct involvement. Because individual citizens have more diverse follower networks than do political elites, this was a pathway for the regime's messages to spread to people who might not already agree with them while primarily serving to keep loyal followers informed and engaged.

The short hashtags preferred by the opposition indicate the opposite: elite engagement with organically generated, popular subjects. This strategy allowed for greater visibility for the opposition's messages, but did not allow them to structure the discourse. It is also indicative of the opposition engaging directly with tweets by individual citizens. In fact, a large number of the opposition tweets during the protest were actually retweets of popular tweets by non-elites (see Appendix D). The strategy of the opposition may have been to try and broadcast their message and provide evidence that there were many people who already agreed with them[14], while the regime's goal was to provide their loyal followers with new talking points that could be spread to enough people to prevent the opposition from reaching the threshold necessary for a successful revolution.

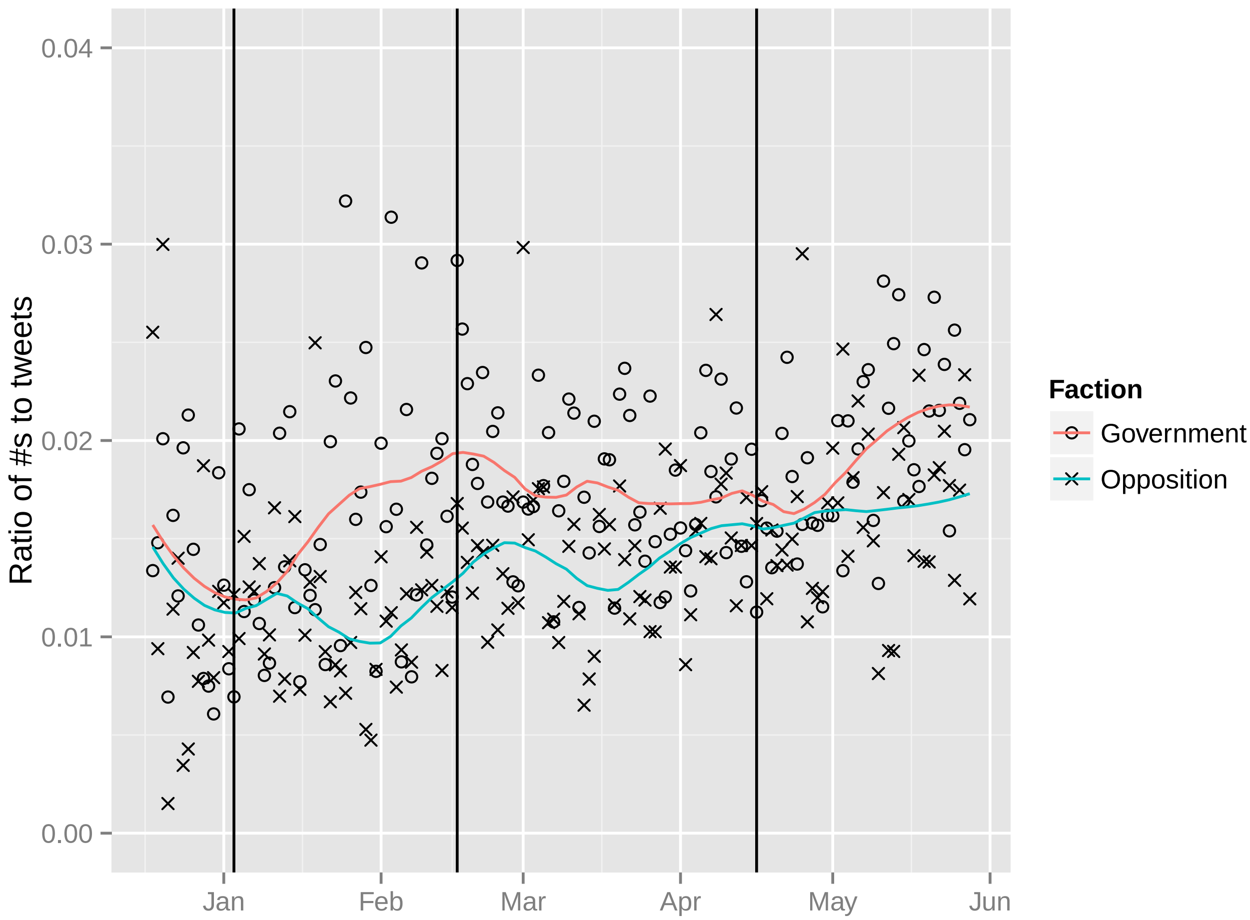
Figures 2, 3, 4, and 5 provide different ways of visualizing the same data, and each provides support for our hypotheses. Figure 2 shows that the tweet frequency for both coalitions was highest as the protests were gaining momentum and higher for the opposition than the regime. Figure 3 shows that this increase coincided with a decrease in the Shannon Entropy, which indicates an increase in *focus* for the opposition while barely changing the *focus* for the government. The lack of 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 the government used lengthy hashtags as a way of generating new and different subjects of discussion as the protests were becoming more serious.

8 Conclusion

Ultimately, the Venezuelan protests of 2014 failed. The regime maintained its grip on power. Although underlying macroeconomic and political concerns persist, the regime effectively demonstrated their repressive capacity, at least for the time being. Part of their strategy seems to have been to target specific aspects of the opposition with

[14] Interestingly, the involvement of popular social media users to promote messages or causes that others are trying to spread has recently become known in the US as "signal boosting." Though this term and Edmond's "signal-jamming" have very different origins, their symmetry nicely summarizes the differences in the strategies of the two coalitions.

Figure 5: Hashtag Use



Although they begain at similar levels, the hashtag use of government *diputados* eclipsed that of opposition *diputados* in the second and third periods. The y axis is the total number of hashtags tweeted divided by the total number of tweets per day for each coalition. This division is necessary to account for the divergence in the rates of tweeting.

messages on Twitter while de-emphasizing the protests themselves. The opposition sought to center attention on the protests, but their rapid reversion to more specific economic criticisms after Easter weekend may have represented a tipping point for their cause; the precise timing and degree of coordination could not have been observed without analyzing Twitter usage. This change in strategy by the opposition might have been one of necessity rather than preference. The *diputados* represented the moderate faction of the opposition, and they may have sought to balance their desire to provide a unified front with skepticism of López and Machado's more radical "*La Salida*" to oust Maduro.

Analysis of the kind presented in this article could not have been conducted without social media data. New information technology always increases the amount of information in society (at least in "hybrid" democratic-authoritarian regimes like Venezuela in 2014) and, relative to traditional media, typically makes it more difficult for the regime to control the information that is being disseminated. We find support for the theory that different groups of elites use social media strategically according to their respective goals, and that the strategy used by the regime concords fairly well with predictions from the game-theoretic literature on global games and regime change.

The potential of social media as a catalyst for revolution has been recognized by scholars of protest, but the literature has thus far been too optimistic about the degree of change social media makes possible. To study the social media use of just the protesters or opposition is to miss half the story, especially as social media becomes more widespread instead of being concentrated among the young and educated.

The quantitative textual analysis developed in this paper is a novel implementation of well-established machine learning techniques. Although looking at the Shannon Entropy scores and distribution of top topics is less qualitatively interesting 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 make inferences about their incentives and that of the regime more generally. That institution is peculiar to

Top Terms for Relevant Topics

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

Vietnam, but social media is not, and it offers a similar avenue for analysis.

Appendix

A Original Terms for Table 2

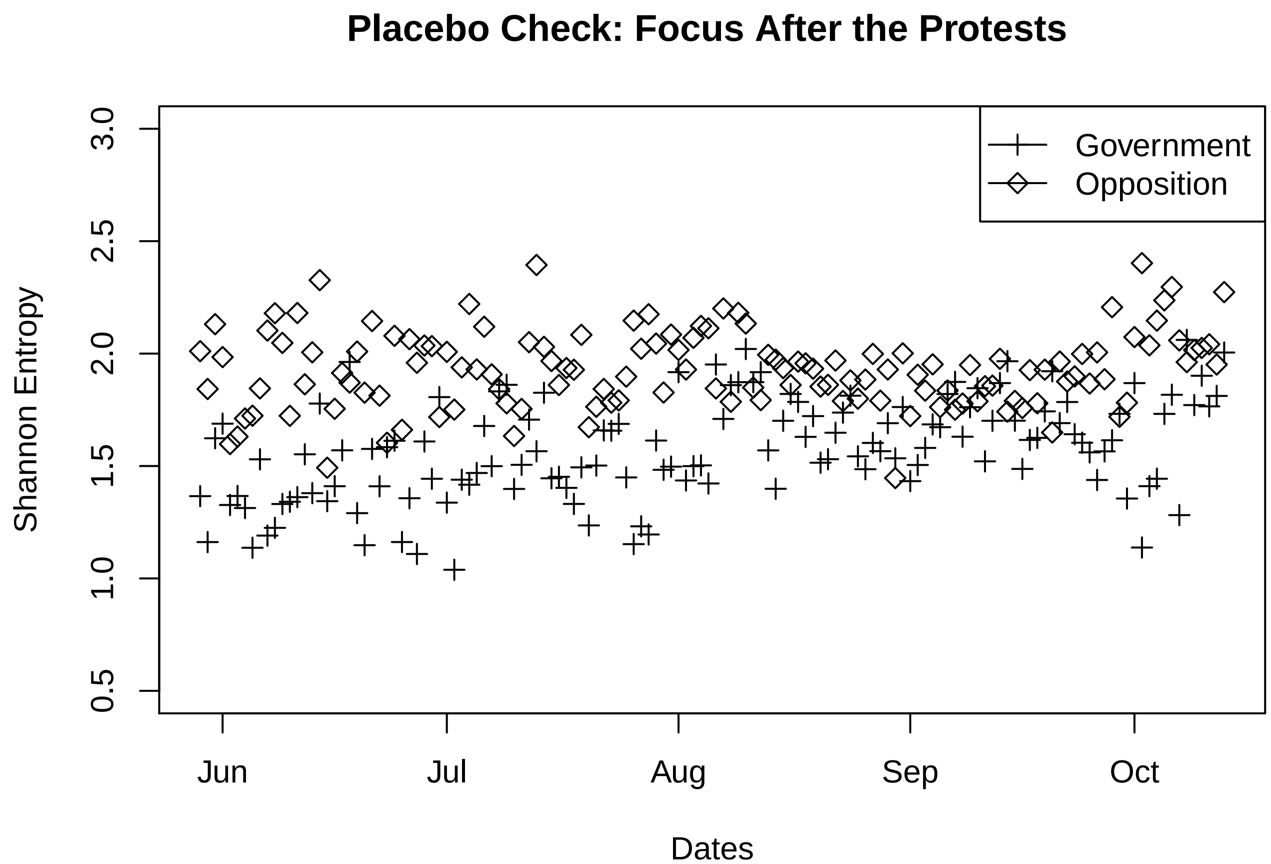
These are the untranslated terms in Table 2.

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 Shannon Entropy to track *focus*. The results are depicted in Figure A 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 A, the opposition tends to have slightly higher Shannon Entropy scores and thus be less *focused*, further marking their



high *focus* during the protests as significant.

C Timeline of Events

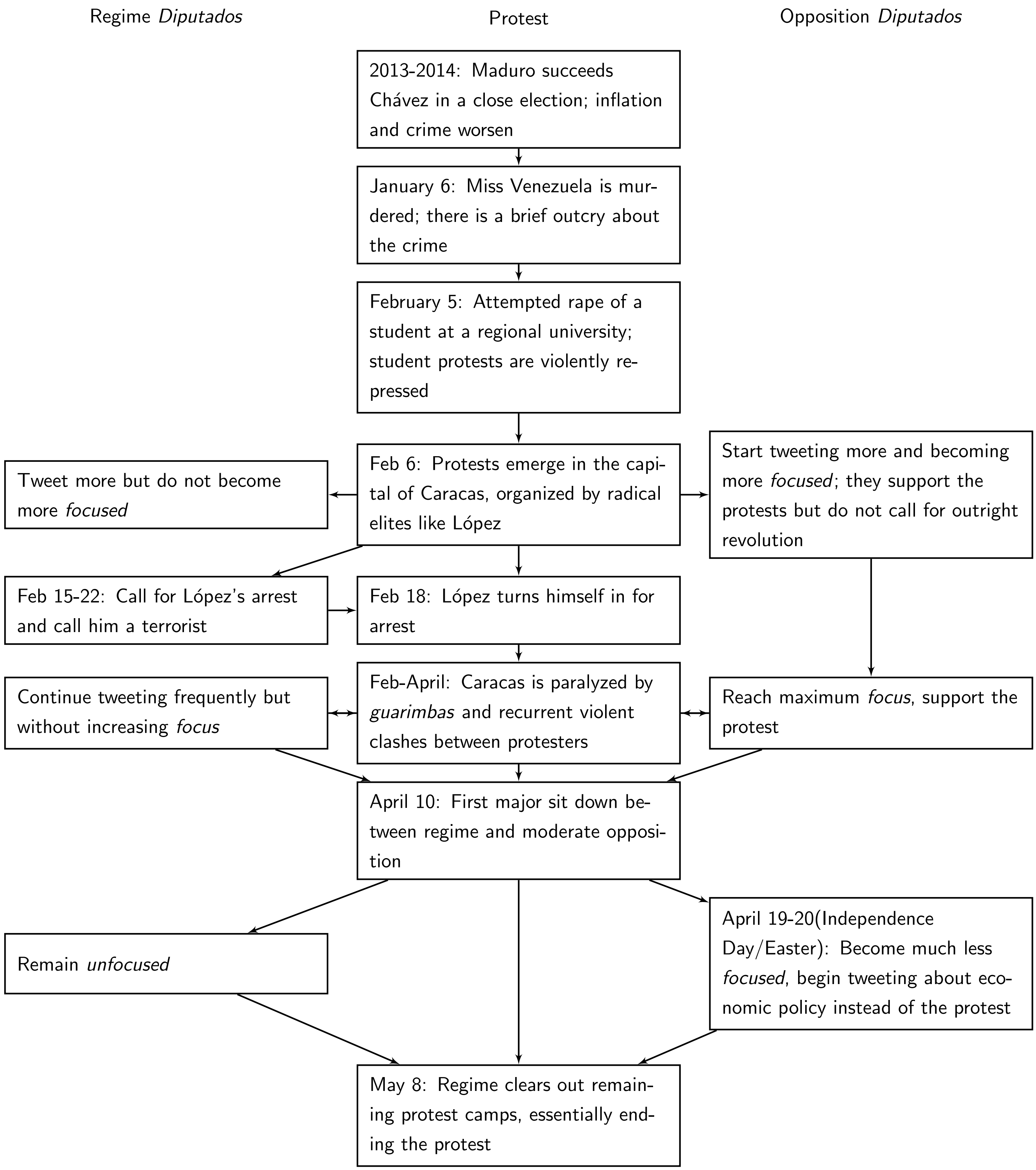
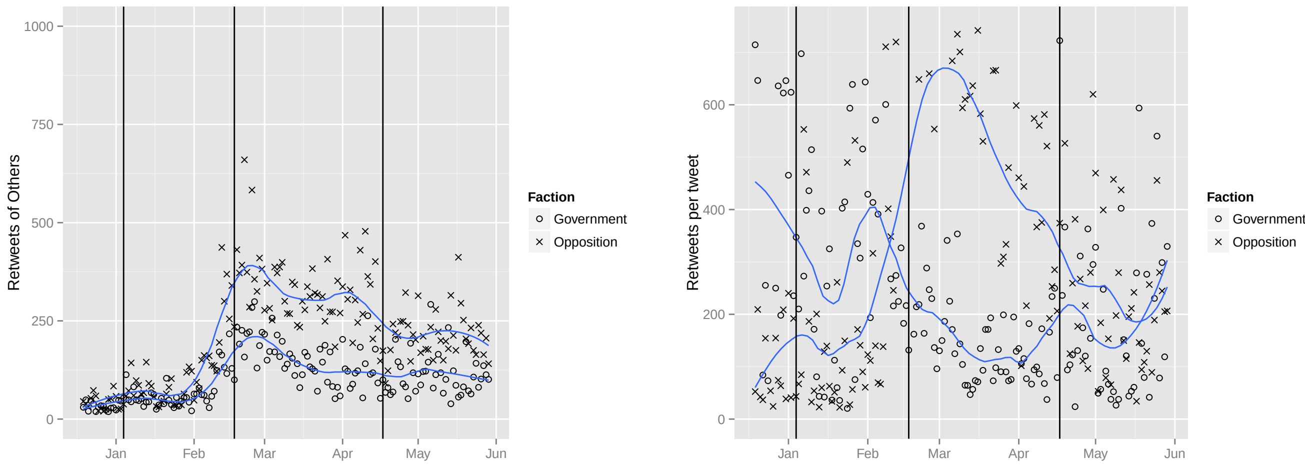


Figure 6: Retweet Frequency and Popularity



D Retweet Patterns

Although most tweets in the dataset (and in general) are original to the user who tweets them, another type is the "retweet," in which another user's tweet is shared with that user's followers. Differential rates of use of original tweets and retweets by the two coalitions allow us to make inferences about their respective strategies, as discussed in the section about hashtag use. More retweets of non-elite users demonstrates an interest in spreading a message as far as possible and using the words of someone else to do it to create more engagement and show other citizens how much support the coalition has. Although each user can be selective about what she retweets, leaving the content creation up to others does entail a decreased degree of control.

The left panel of Figure 6 shows the frequency of retweets by each coalition. A substantial portion of the gap in tweet frequency shown in Figure 2 comes from differential rates of retweeting, with the opposition engaging in substantially more of it. The right panel shows the number of retweets that each tweet had when the elite retweeted it. Essentially, the opposition was retweeting more popular tweets after the protest started, except for a brief period in early February when the regime was focused on demonizing López.

References

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

Beauchamp, Nick. 2013. Predicting and interpolating state-level polling using twitter textual data. In *Meeting on automated text analysis, London School of Economics, London*.

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.

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* .

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.

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* .

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.

Livne, Avishay, Matthew P Simmons, Eytan Adar, and Lada A Adamic. 2011. "The Party Is Over Here: Structure and Content in the 2010 Election.".

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.

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

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.

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

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.