Targeting and the Timing of Censorship
A case study of Venezuelan Twitter

Ishita Gopal

1 Introduction

This study analyses the content of the tweets posted by the opposition members of the National Assembly (2015-2017) in Venezuela for the time period between 1st to 31st January, 2019. The aim of this analysis is to identify the evolution of topics over time so that the results can be used to support the study of online censorship in the country. The underlying theory used to relate topic evolution with censorship proposes that decisions to block access to online platforms (like, Twitter) in non democratic regimes, depends on a difference between narrow and broad targets. The government is most likely to engage in censorship when the platform is functioning to create widespread negative attention focused on the government itself. Broader negative attention focused on the country as a whole, such as - economic collapse, food shortages, violence - is not likely to result in blocking of the platform. But if these discussions become linked to the government and are negatively discussed, then the possibility of censorship becomes higher. A structural topic model is used to empirically extract the topics and analyze how their prevalence change over time.

2 Media Censorship

The lack of traditional media (Newspapers, Radio stations, Television channels) which allows criticisms of the government (Freedom House, 2014, 2017), combined with the burgeoning use of social media by the population motivates and justifies the use of tweets in this analysis. According to a 2016 report by " Tendencias Digitales" [1], 56% of internet users in Venezuela use Twitter or comparable social media services and according to another statistic, Venezuela ranks amongst the top five countries in terms of Twitter penetration (Schoonderwoerd 2013). It is a key resource for the opposition and has been used to spread information and call upon people to protest. Although, traditional media censorship in Venezuela is not new (Frredom House, 2003), the internet was largely left alone from hostile government actions (OpenNet, 2013). These trends have been shifting over time and Freedom House declared the Internet in Venezuela as 'Not Free' in 2017 (Freedom House, 2017). This was the same year a vaguely worded 'Law Against Hate' was passed by the constituent assembly which provided the government with broad powers to police speech, participation and any activity it deems incites hate in society (Rosati and Zerpa, 2017). Further, NetBlocks has recorded 35 incidents of state backed online censorship incidents since January 2019 till now (NetBlocks: Mapping Net

[1]ee https://www.slideshare.net/wearesocialsg/digital-in-2017-south-america.

Freedom, 2019). These reports conjecture that the blocks are implemented to create barriers for the opposition to reach the population and incite mass action against the government.

3 Research Design

The motivation to censor stems from the possibility of regime destabilizing actions that would force the government to give up its power. In an environment where collective mobilization of the masses has emerged as a rising trend to oust a malign regime, the importance of controlling information is even higher. Traditional literature on what causes people to mobilize have focused on relative deprivation and grievance models. The information which empowers anti-regime actors and forces them to action have been discussed extensively in these studies. The theory of relative deprivation put forward by Gurr (1970) emphasizes the effect of perceived differences between an individual's or group's circumstances to those of more advantaged groups. He foretells that aggregation of these differences, when sharply felt would be a catalyst to violence.

In addition to material mismatch in expectations versus reality, emotional responses to violations of moral code are also powerful sources of protest activation (Ferree et al., 2002; Van Laer, 2017). And with the advent of the internet, coordinating upon these grievances has become easier. It allows citizens to view collective action risk from the conversations and diffusion of messages across social networks and provides the coordinators of movements a medium to broadcast its messages more easily (Lance Bennett et al., 2008; Van Laer, 2017). Following this logic, threats to the government may not only emerge from direct calls to mass action, but also when discussions on topics which heighten these deprivations increase. Anti government messaging is then likely to take forms which incite reflection of these grievances and those which explicitly call upon people to mobilize.

Using this criterion, this study intends to obtain an understanding of the topics discussed by the opposition and hopes to identify the different facets of anti-government messaging. The socio-economic and political crisis of hyperinflation, escalating starvation, absence of medical support, violence and the resulting mortalities are likely sources of pinching frustrations with the government. Broadcasting narratives which emphasize the government's role in failing the people on these issues would generate a kind of implicit target. On the the other hand, calls which ask people to come on the streets to demonstrate, cause disruptions or even incite violence, are more explicit forms of targeting of the government. The increase in either type of discussions would likely be a cause of concern for the government, but more so for the latter category. In all, this study would like to discover anti government messaging and explore how topic prevalence corresponds with the timing of censorship imposed by the state on Twitter.

4 Data

The study used opposition members of the National Assembly for 2015-17 to identify individuals to track on twitter. This assembly represented the latest legislature of Venezuela until its existence was overridden by the creation of the new National Constituent Assembly in 2017 (Casey, 2017). The former is an important body which has provided evidence of strong opposition against the government. It was the only government sector that was not controlled by Maduro's party, where the opposition held a super-majority with 109 of 164 total seats (National Electoral Council, 2015). The creation of the Constituent Assembly was imposed by decree after months of protests against the regime and many in the international community have condemned it as an illegitimate government (Bronstein and Cobb, 2017; Shifter, 2017). A large portion of the opposition within the country boycotted elections to it for the same reasons. The aim of the new legislature is to draft a new constitution for Venezuela and is unsurprisingly led by Maduro's party. The deputies to the National Assembly continue to identify it and themselves as the legitimate government, providing a direct source of opposition within the country.

To collect the data, a manual search for the Twitter accounts was performed and 112 accounts for each of the opposition members in the legislature were located. 33 of these accounts are authentic - identified by a blue verified badge assigned next to the account names by twitter - and tweets form these accounts are used in the current analysis. [2] Twitter's rest API was used to obtain tweets from each individual's timeline. The restriction on the API allow researchers to obtain at most 3,200 most recent tweets and this count includes retweets. In addition, it is possible to obtain information such as - the number of likes, retweets received and other metadata as well. The study used the tweepy library in the Python programming language to collect this data. The analysis uses 6791 tweets between January 1st and 31st and most of these are in Spanish.

5 Analysis

The structural topic model (STM) (Roberts et al., 2014) is used to model topics. This is an extension to the traditional probabilistic models such as Latent Dirichlet Allocation (LDA) and correlated topic models (CTM). Specifically, the STM is a generative model of word counts where each document in the analysis can be a mixture of multiple topics, each topic is defined as a mixture over words and every word has a probability of belonging to a particular topic. The novelty and usefulness of this model lies in its ability to incorporate document-level metadata in the prior distribution for the discovery of topics and their prevalence. Some examples of metadata which can be used for topic discovery include attributes like publishing date, the number of times a document is shared, the name of the person writing the document. Including the covariate for the date of publishing, for example, will allow the researcher to explore how individual topics relatively become more or less prevalent over time.

For the data pre-processing step, punctuation, spanish stopwords and custom stop words including -`"t.co"`, `"https"`, `"rt"`, `"http"`- were removed. The STM package comes with a built in text processor and uses the snowball stopword list for removing Spanish stopwords. The words in the tweets were not stemmed since it can combine

[2] The study envisions expanding the number of accounts used in the analysis in the future after performing authentication checks.

substantively different forms of words together which is especially problematic for unsupervised techniques (Denny and Spirling, 2018). Further, the path of the links shared were also preserved. If a certain video, newspaper article, social media post was shared by multiple members around the same time, it provide could useful information about the topic within those tweets and help in grouping. The case was also lowered to avoid double counting the same term.

For the analysis, the study decided to use 31 topics (K) after results for K = 5,10,15,25 and 55 were checked. K=55 was obtained when K was not assigned and the model was allowed to choose the appropriate number of topics. The results for K less than 20 seemed to be too broad and included multiple topics which looked like a mixture of two or more seperable groups. For example, there were many instances where the keywords which describe the legitimacy of Guadio's presidency and illegitimacy of Maduro's presidency were grouped together. This would be problematic for the analysis since one of the objectives is to identify how differences in the discussion about each of these themes leads to censorship. Cleaner topics were observed for K greater than 20. For topic = 55, the results seemed to overfit the corpus and provided categories which seemed more granular than necessary. Further, the topic proportions for those with the highest rank did not cross 3%. K=31 was chosen to slightly overestimate the true number of topics which seemed to provide useful results but to do so in a way that overfitting does not increase dramatically. Further, the reasoning for K = 31 is also based on the conjecture that there might be at most one important topic everyday.

Since it is possible to incorporate metadata, information on which day each tweet was posted was used as a prevalence co-variate. This decision was motivated by the main theme of this study of trying to measure prevalence of different topics over time. Each date was converted into an integer between 1 and 31. If a tweet was sent out on the 10th of January, the variable was assigned the value of 10, and so on. These integer indicators were entered additively and a spline was used to allow non-linear relationships in the topic estimation stage. Figure 1 summarizes the results of the STM, displaying the prevalence of each of the 31 topics. The study used the FREX statistic to identify key words within each topic. This statistic tries to find words which are both frequent in and exclusive to a topic of interest and ranks them on the basis of these two attributes. The default freweight of 0.5 is used. The first 30 of these key words were translated from Spanish to English using google translate prior to interpretation.

The study was able to identify 4 topics which can be classified as *protest topics* and include explicit targetting of the government. The key words in Topic 21 seem to call upon people to protest and refers to the future dates with words like 'we will','heading','we wait','mobilization','town hall'. Topic 5 contains more inciting language and uses present tense, asking people to protest more aggressively with words like 'let's shout', 'big', 'chavez', 'leave,','streets','went out','rebelled','live'. Topic 29 seems to target the regime and uses language like 'victimizer', 'break', 'defeat', 'tyrant'. And Topic 18 contains many instances of city/individual names possibly giving information about upcoming protests, town halls and other engagements as well as providing live updates as protests develop.

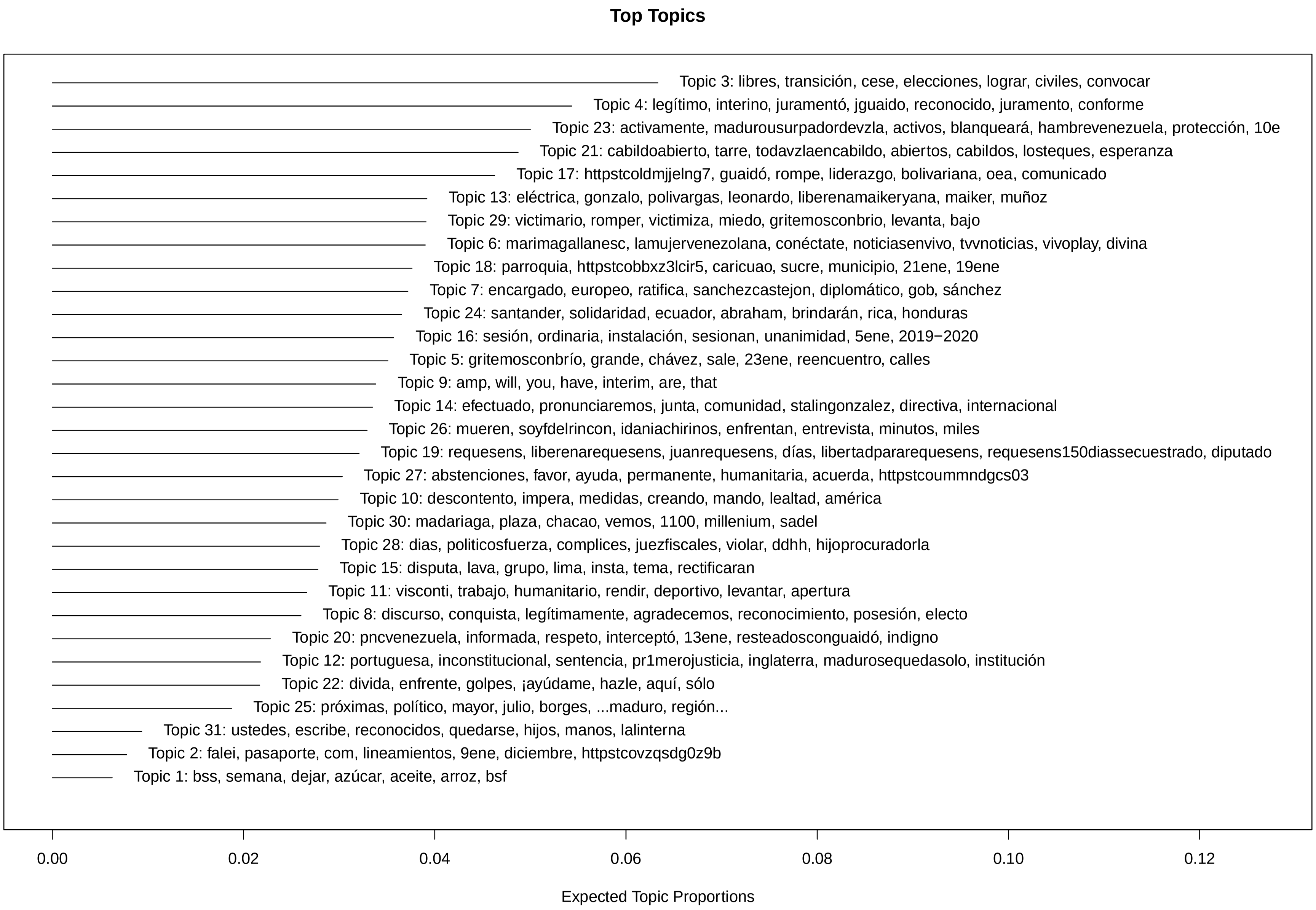

Figure 1: Estimated topic proportions and corresponding words ranked by FREX score.

The translated versions of the actual tweets can be found in Figure 2. These support the findings inferred from the FREX key words. To help better visualize the trend of each of these topics, their the expected topic proportion are plotted with respect to time and can be seen in Figure 3. The instance of explicit targeting can be graphically observed. Topics related to protests and maligning the government with direct references (using language like ’illegitimate dictatorship ’,’Nicolas Maduro’, ’criminal dictatorship’) are not used continuously but see an upsurge during specific time periods.

Topic 21

We continue in #CabildoAbierto this time from Machiques de Perijá where its people are determined to go out to the street # 23ENE. United, pressing and focused we will drive the change that Venezuela so much needs. # NosVemosEl23 https://t.co/ LruiJO42Fm

Today # 17Ene visited my Alma Máter @lacatolica and I met a new generation of Lawyers. It is motivating to see how young people in spite of the crisis are still forming and betting on the country. N nI trust you, trust me and let's take #Venezuela! https://t.co/ e3h8uYZASt

Venezuela was born in a Cabildo and in Open Cabildos, hope is reborn in every corner of our homeland, a town that claims reconciliation and being free! N n # TodaVzlaEnCabildo https:// t.co/8Pb24mAfNR

Topic 5

At the moment #cacerolazo is reported in El Valle, El Cementerio, Andrés Bello, La Candelaria, Catia, Quinta Crespo, Santa Monica, Cotiza, Sabana Grande, Petare, Baralt Avenue, Rocking, El Paraiso, Antimano. The town expresses his rejection of the usurper 10:32 pm

Throwing the tank against a person, which at first sight does not constitute any danger to the official, with premeditation and alevosía, constitutes a flagrant aggression to his physical integrity. The truth always comes out, no matter his desperate attempts to silence it. Https : //t.co/kb6bIaMes6

IMPORTANT. Today the main protagonist is the citizen, we will record in photos and videos the development of the marches and concentrations throughout the country. Today is a day where we all have to make a great effort to communicate the expression of our people # GritemosConBrío

Topic 29

After the illegitimate election of Nicolás Maduro in May 2018, Europe supports the restoration of democracy. I acclaim the courage of hundreds of thousands of Venezuelans who walk for their freedom. Https://t.co/ beTSGVuAyd

21Ene They ask us if this will be worth it, they ask us if we are afraid. N nPlease know that from this Palace of laws, for the future for all Venezuelans, we will be at the forefront of reconciling the people, achieving justice, democracy and freedom. n n # NosVemosEl23 https://t.co/ ywsPj7RRwm

Above all, the 'battle of ideas' is possible in a criminal dictatorship that holds props and has been in power for 60 years, plundering the Cuban people tremendous example Mrs. Polevnsky https://t.co/1S4IHtLc6T

Topic 18

#AVANCE This is how the San José de Cotiza parish is located in Caracas, after the rise of at least 40 troops of the National Guard Command # 21Ene (8:19 am) https://t.co/ BbxZ3lcir5 https: // t .co / dnGbzleW9R

The doctors on call at the Vargas hospital in Caracas, which was taken by the GNB, informed me of the number of wounded who enter by firearm and pellets from the widespread protest throughout #Caracas #Cacerolazo

. @ RedesVP n With the creation of the popular promoter network in the Simon Bolivar parish of the sector Laja of the Caroni municipality. N @ VPBolivar n manages to train and activate more than 100 activists in Popular Networks. N @ RedesVP_Caroni n @ MarianellaPuchi n @ VoluntadPopular n @ leopoldolopez https://t.co/ hjWB0USgcR

Figure 2: Example tweets corresponding to each topic under the protest category

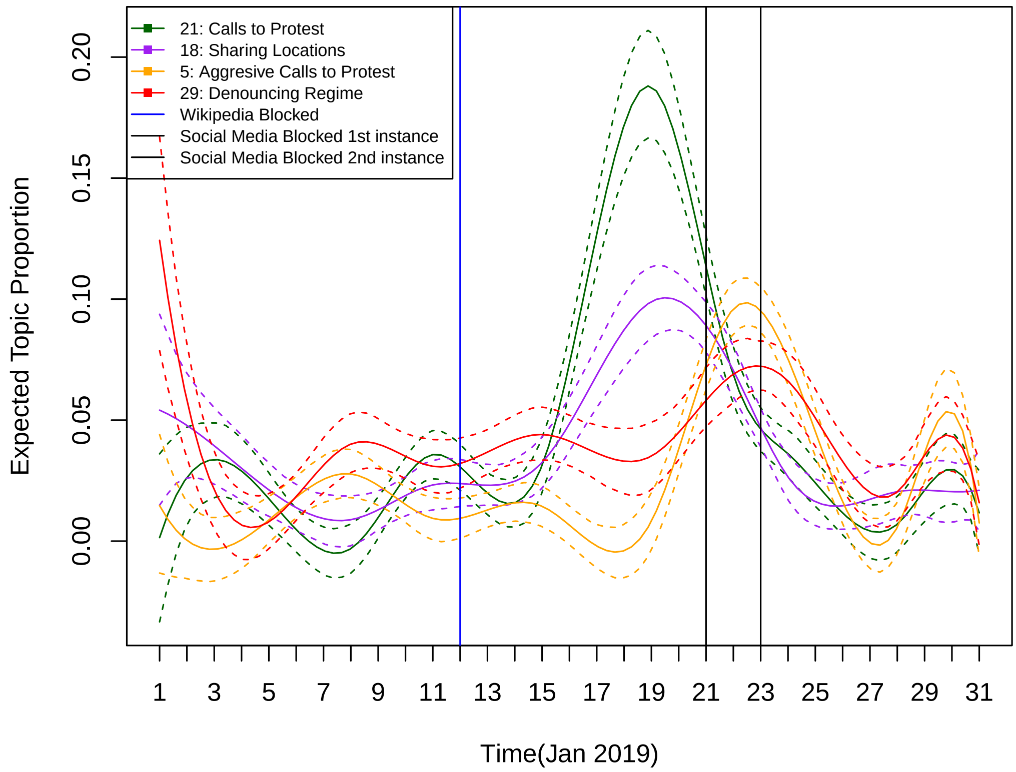

Figure 3: Influence of the date of posting a tweet on the prevalence of protest topics.

Instances of criticisms of the government without active calls to action and discussion of socio-economic and political problems facing the country were found in topics 23 and 13. These topics contain a mix of issues including the collapsing economy, increasing hunger and poverty, declining health services and discussions of journalist repression. The main denominator is criticism of the government as being the cause of these problems. Some of the key words identified using the FREX statistic include 'electric', 'clinical', 'journalist', 'hungervenezuela', 'abuses', 'madurousurp',' failure', 'patients', 'university'. Examples of translated tweets from each topic can be found in Figure 4. In these examples, Topic 13 has instances where regime failure and its impact on the financial assets and the oil industry is discussed and Topic 23 contains discussions on detained or missing journalists and the dire consequences of government led infrastructure failures on individuals in hospitals. The trend in topic proportions with respect to day of posting can be seen in Figure 5.

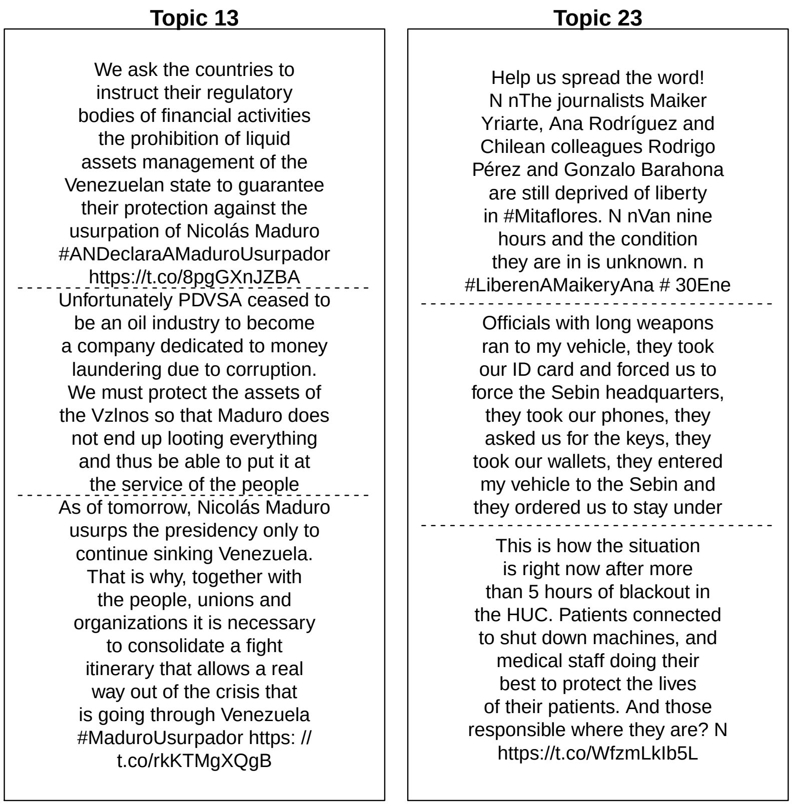

Figure 4: Example tweets corresponding to each topic under explicit targeting

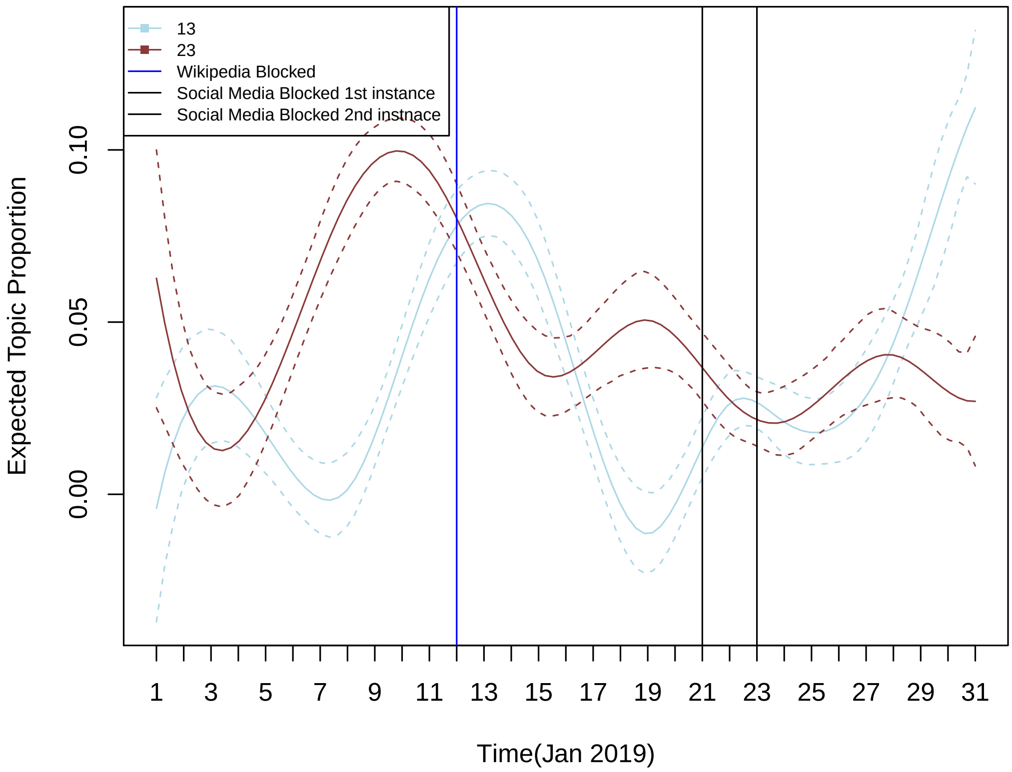

Figure 5

6 Targeting and Blocks

Both Figures 3 and 5 depict a rise in relative proportions of certain topics during specific time periods and this increase seems to occur very close to the time censorship events occurred. Online censorship events took place on the 12th, 21st and 23rd of January. The first incident involved blocking access to Wikipedia. This was subsequently after the re-election of Nicolas Maduro and during the time Jaun Guaido was appointed as the acting president. It prompted a Wikipedia editing war on its Spanish articles where netizens took turns to change the President of Venezuela from Nicolas Maduro to Juan Guaido. On the next two instances of censorship Google search and multiple social media platforms like YouTube, Facebook, Instagram, Twitter were blocked. These blocks coincided with protests happening on the ground, which started on the 21st. Concentrating on Topic 21, the expected proportion of tweets reaches almost 20% on the 19th of January, 2 days before the incident of censorship took place. This is the highest proportion reached amongst all 31 topics on any given day. Other topics under the *protest category* see a less dramatic rise, but there are still perceivable increases in there topic proportions around the time of censorship. Topic 18 and Topic 5, which capture more location related protest tweets and aggressive calls to action, respectively, see their proportions increases from under 3% to 10%. There is a much smaller increase in proportions for Topic 29, found in tweets which explicitly denounce the regime. Topics which capture criticism of the government without calls to mass action (13 and 23) peak nearer to the incident when Wikipedia was blocked. The topic prevalence reaches near 10% for both. However, the rise is not as dramatic as seen for Topic 21.

Amongst other topics, Topic 3 discusses need for free elections Topic 4 discusses the legitimacy of Juan Guadio and Topic 17 seems to contain instances of international reactions to Venezuela's crisis and Topic 7 contains examples of endorsement of Guaido by diplomats such as Carlos Scull. Thus, topics 4, 17 and 7 can probably be grouped under an *endorsement category*. Topic 9 contains English words and a deeper analysis did not reveal substantial influence of time over topic prevalence, possibly indicating that the opposition tries to appeal for international support on a regular basis. To further understand how these topics cluster together and if these groupings are meaningful, the study plotted topic correlations. Positive correlations between topics indicate that both topics are likely to be discussed within a tweet and can be seen in Figure 6. Ties are formed between two topics if correlation crosses the threshold of 0.01. Most of the topics are not tied and this is expected since tweets are short and contain at most 280 characters. It would be difficult to comment on multiple topics within such limited space. However, instance of some ties can still be observed between Topics 5 and 29; 21 and 18; 4,17 and 7. These are the same topics the study earlier identified and clubbed under categories of protest and endorsement.

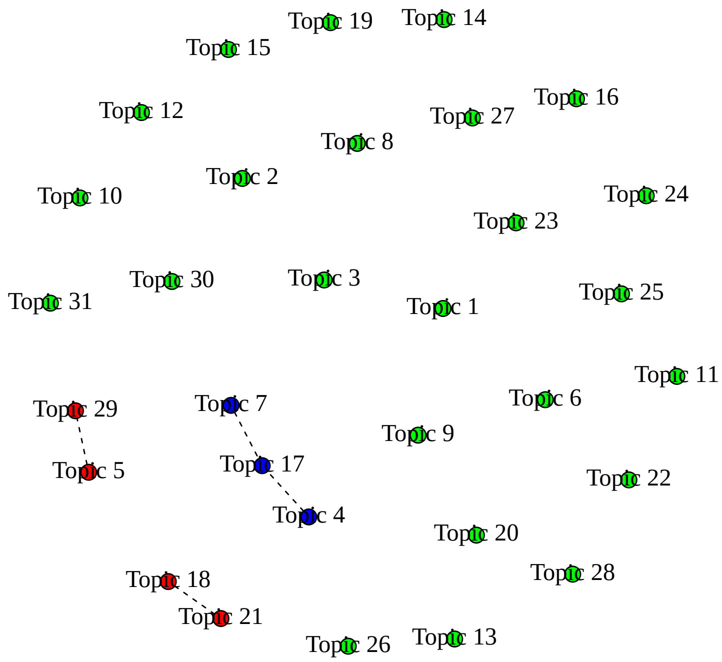

Figure 6: Topic Correlations. Nodes were manually colored for easier interpretation. Red represents the protest category and blue, criticism without protest. A threshold of 0.01 was used to create ties between topics.

7 Conclusion & Future Directions

This study is an exploratory analysis of the tweets sent by the opposition during January 2019, a politically charged time in Venezuela. The study finds that the structural topic model is useful in identifying topics of interest, specifically instances of anti-regime messaging and these categories seem to observe an increase in prevalence very close to the time the platform was censored. This provides an indication about the applicability of the proposed theory of targeting. The use of an unsupervised technique (STM) is justified since it allows all the information in the corpus to be incorporated and removes some of the bias which may creep in if the researcher decides to observe only specific topics before the study begins. It also allows the discovery of new topics. Since labelling the posts would require subjective decisions, it further removes the issue of inter-coder reliability that may emerge in case supervised learning approaches are used.

Although the results seem promising, further analysis and validation of topics is required. For K=31, there were instances - Topic, 11, 12, 15 - which seemed to contain a mixture of more than one topics. The study checked their topic proportions over time but did not find dramatic surges. Nevertheless, these groups contained some instances of relevant information on the humanitarian crisis and discussion of illegitimacy. Future steps envision validating the number of topics after checking more values of K and using topics with clearer distinctions. Aggregation techniques proposed specifically for dealing with coherence problems in short texts will also be explored (Blair et al., 2019). Further, instead of tracking the prevalence of a subset of topics, the targeting variable could also track how the proportion of each topic changes everyday and how many of these topics fall under the anti-regime messaging category.A similar strategy is used in Munger et al. (2019), where increase or decrease in daily topic diversity near incidents of protests is used to explain strategic communication decisions made by the elites and the opposition. According to the idea of targeting presented in this paper, it is expected that nearer to the incident of censorship of Twitter, the proportions of topics having anti-regime messaging will be higher than father to the censorship event. Information on the reach of the post - embedded in the number of retweets and likes the posts received - would also be used to account for multiplicative effects.

References

Blair, S. J., Y. Bi, and M. D. Mulvenna (2019). Aggregated topic models for increasing social media topic coherence. *Applied Intelligence*.

Bronstein, H. and J. S. Cobb (2017). Venezuela faces outrage after new assembly takes legislative power.

Casey, N. (2017). Venezuela's New, Powerful Assembly Takes Over Legislature's Duties.

Denny, M. J. and A. Spirling (2018, 4). Text Preprocessing for Unsupervised Learning: Why It Matters, When It Misleads, and What to Do about It. *Political Analysis 26*(2), 168–189.

Ferree, M. M., J. Goodwin, J. M. Jasper, and F. Polletta (2002). Passionate Politics: Emotions and Social Movements. *Contemporary Sociology*.

Freedom House (2014). Freedom of the Press 2014. Technical report.

Freedom House (2017). Freedom on the Net 2017: Manipulating Social Media to Undermine Democracy.

Frredom House (2003). Freedom of the Press 2003: A Global Survey of Media Independence. In *Freedom House*, pp. 135.

Gurr, T. (1970). Relative Deprivation and the Impetus to Violence. In *Why Men Rebel*.

Lance Bennett, W., C. Breunig, and T. Givens (2008). Communication and political mobilization: Digital media and the organization of anti-Iraq war demonstrations in the U.S. *Political Communication*.

Munger, K., R. Bonneau, J. Nagler, and J. A. Tucker (2019). Elites tweet to get feet off the streets: Measuring regime social media strategies during protest. *Political Science Research and Methods*.

National Electoral Council (2015). Elections to the National Assembly 2015. Technical report.

NetBlocks: Mapping Net Freedom (2019). Internet services restricted in Venezuela amid uprising.

OpenNet (2013). OpenNet Initiative.

Roberts, M. E., B. M. Stewart, and T. Dustin (2014). Journal of Statistical Software stm: R Package for Structural Topic Models. *Journal of Statistical Software 10.2*, 1–40.

Rosati, A. and F. Zerpa (2017). Venezuela Passes 'Anti-Hate Law' to Clamp Down on the Media.

Shifter, Michael; Raderstorf, B. (2017). Venezuela After the Constituent Assembly. *Foreign Affairs*.

Van Laer, J. (2017). The mobilization dropout race: Interpersonal networks and motivations predicting differential recruitment in a national climate change demonstration. *Mobilization*.