

# Online Appendix

Lutscher, Philipp M., Hot Topics: Denial-of-Service Attacks on News Websites in Autocracies. Political Science Research and Methods

October 20, 2021

## Table of Contents

---

A Sample and Measurements	1
B Main Analysis	5
C Categorization of Topics	7
D Sensitivity and Robustness Tests	10
E Text Analysis and Topic Modeling	24

---

## A Sample and Measurements

In this appendix, more details on the used sample of news websites and the measurement of DoS attacks are presented.

**Sample** Table A.1 displays the newspapers included in the analysis and additional information on these outlets, e.g., affiliation, average visitors, and hosting provider. To filter out government or pro-government websites, I followed the following steps:

First, I identified government associated news websites when the website contained a “.gob” address. This led to the exclusion of the websites *vtv.gob.ve* and *rnv.gob.ve*. Second, I relied on secondary sources to identify pro-government websites. In this process, the website *laradiodelsur.com.ve* was added to the pro-government list as evidence suggests that this and the two former websites are distributing pro-government propaganda.<sup>1</sup> Besides, the website *noticias24.com* is perceived as a clear pro-government website.<sup>2</sup> Finally, I classified two websites, *elmundo.com.ve* and *ultimasnoticias.com.ve*, as pro-government websites that belong to the Grupo Últimas Noticias media group. In 2013, pro-government actors bought the group who changed the editorial board and news reporting of the outlets. In this process, more than 120 journalist resigned or were fired.<sup>3</sup> Although there is also evidence that the news outlets *Globovision* and *El Universal* were bought by individuals linked to Maduro regime, Miguel Latouche, a country expert who helped me with the classification, still coded them as private outlets.<sup>4</sup> The same country expert also judged the affiliation of the remaining news outlets and sorted them into private, opposition, or “affiliation unknown” outlets.

**Measurement of DoS attacks** As described in the main text, my measurement approach exploits how web servers communicate. In a nutshell, devices communicate on the Internet in the way that a client A sends a request to a host B, this host B then responds to this request and, if acknowledged, enables client A to communicate with the server. In technical terms, this is the so-called ”Three-Way-Handshake.” In the case of web servers, this communication follows the Hypertext Transfer Protocol (HTTP) where hosts (web servers) answer with a standardized numeric response code to the request made by the client. When the web server returns a “200” status code, the connection can be established. If not, some error likely occurred. Error codes with a “4XX” number

---

<sup>1</sup><https://www.hrw.org/node/256344/printable/print>(accessed 05-12-2021).

<sup>2</sup><https://web.archive.org/web/20140503070256/http://www.diariolasamericas.com/america-latina/portal-chavista-noticias-24-delata-usuarios.html> (accessed 05-12-2021).

<sup>3</sup><https://web.archive.org/web/20150118114140/http://cnpcaracas.org/2014/03/noticias/renuncia-jefa-de-investigacion-de-ultimas-noticias-por-censura/> (accessed 05-12-2021).

<sup>4</sup><https://www.wsj.com/articles/venezuelas-press-crackdown-stokes-growth-of-online-media-1410136494> (accessed 05-12-2021). Since I do not record DoS attacks on these outlets, the inclusion of both outlets to the sample does not alter the results.

ID	Name	Affiliation	Post-processing worked?	Venezuelan readers?	Avg. daily visitors (in million)	Comments	Hosted country	Hoster
32	<a href="http://globotvvision.com/">http://globotvvision.com/</a>	private	yes	<b>82.5%</b>	<b>1.98</b>		US	CLOUDFLARE
33	<a href="https://larradiodelasur.com.ve/">https://larradiodelasur.com.ve/</a>	government	n.a.	n.a.	n.a.		VE	CANTV-NET
34	<a href="http://www.radiofleyalegrianoticias.net/">http://www.radiofleyalegrianoticias.net/</a>	private	yes	<b>78%</b>	<b>0.20</b>		CA	IWEB
35	<a href="http://rmv.gob.ve/">http://rmv.gob.ve/</a>	government	n.a.	n.a.	n.a.		VE	CANTV-NET
36	<a href="http://www.televem.com/">http://www.televem.com/</a>	private	yes	<b>86%</b>	<b>0.13</b>	<b>Due to non-structure might include entertainment/sport news</b>	US	unknown
37	<a href="http://unionradio.net/">http://unionradio.net/</a>	private	yes	68%	0.30		US	MICROSOFT
38	<a href="http://www.noticierovenevision.net/">http://www.noticierovenevision.net/</a>	government	n.a.	<b>80.4%</b>	<b>0.23</b>		US	AMAZON
39	<a href="http://rvv.gob.ve/">http://rvv.gob.ve/</a>	unknown	n.a.	n.a.	n.a.		VE	CANTV-NET
40	<a href="http://www.alrebresha.com/">http://www.alrebresha.com/</a>	private	yes	52%	0.16	Website stopped to work	US	UnitedLayer
41	<a href="http://www.analitica.com/">http://www.analitica.com/</a>	opposition	yes	87%	2.53		US	CLOUDFLARE
42	<a href="https://www.aporreia.org/">https://www.aporreia.org/</a>	unknown	yes	56%	0.09		US	CLOUDFLARE
43	<a href="https://canaldenoticia.com/">https://canaldenoticia.com/</a>	private	yes	37.4%	0.07		US	DREAMHOST
44	<a href="http://confirmado.com.ve/">http://confirmado.com.ve/</a>	private	yes	52%	0.27	<b>Extracting did not work</b>	US	GODADDY
45	<a href="https://dolartoday.com">https://dolartoday.com</a>	opposition	no	n.a.	n.a.		US	CLOUDFLARE
46	<a href="https://elpitazo.com/">https://elpitazo.com/</a>	private	n.a.	n.a.	n.a.		US	unknown
47	<a href="http://www.elmonomilagreto.com/">http://www.elmonomilagreto.com/</a>	private	yes	73.3%	0.08		US	CLOUDFLARE
48	<a href="http://elporodicovenezolano.com/">http://elporodicovenezolano.com/</a>	unknown	n.a.	n.a.	n.a.		CA	OVH
49	<a href="http://entodonoticias.com/">http://entodonoticias.com/</a>	unknown	yes	56%	0.92		VE	Patriclell, C.A.
50	<a href="http://informe21.com/">http://informe21.com/</a>	opposition	yes	75%	14.00		US	CLOUDFLARE
51	<a href="https://www.lapatilla.com/">https://www.lapatilla.com/</a>	unknown	yes	89%	0.39		US	CLOUDFLARE
52	<a href="http://www.notiactual.com">http://www.notiactual.com</a>	unknown	n.a.	n.a.	n.a.		US	Netrouting AS
53	<a href="http://www.noticias24.com/">http://www.noticias24.com/</a>	pro-government	n.a.	n.a.	n.a.		US	MICROSOFT
54	<a href="http://www.noticiasdevenezuela.org/">http://www.noticiasdevenezuela.org/</a>	unknown	n.a.	n.a.	n.a.		US	Backspace
55	<a href="http://noticiasvenezuela.info/">http://noticiasvenezuela.info/</a>	private	yes	83%	2.15	<b>Due to non-structure might include entertainment/sport news</b>	US	CLOUDFLARE
56	<a href="http://www.noticierodigital.com/">http://www.noticierodigital.com/</a>	unknown	n.a.	n.a.	n.a.		US	CLOUDFLARE
57	<a href="http://www.notiven.com/">http://www.notiven.com/</a>	private	yes	n.a.	n.a.		US	CyrusOne
58	<a href="http://todayvenezuela.com/">http://todayvenezuela.com/</a>	private	n.a.	n.a.	n.a.		US	INTERROUTE
59	<a href="https://venezelanalisis.com/">https://venezelanalisis.com/</a>	private	yes	67%	0.04		DE	Dacy Telecom, C.A.
60	<a href="http://www.dinero.com.ve/">http://www.dinero.com.ve/</a>	private	no	n.a.	n.a.		VE	GODADDY
61	<a href="http://enpaiseta.com/">http://enpaiseta.com/</a>	private	yes	92%	0.72		US	MICROSOFT
62	<a href="http://www.2001.com.ve/">http://www.2001.com.ve/</a>	unknown	yes	n.a.	n.a.		CA	OVH
63	<a href="http://www.elmundo.com.ve/">http://www.elmundo.com.ve/</a>	pro-government	n.a.	n.a.	n.a.		CA	AMAZON
64	<a href="http://www.el-nacional.com/">http://www.el-nacional.com/</a>	private	yes	78%	8.57	<b>Due to non-structure might include entertainment/sport news</b>	US	OVH
65	<a href="http://www.eluniversal.com/">http://www.eluniversal.com/</a>	pro-government	yes	80%	2.54		CA	OVH
66	<a href="http://www.ultimashonicias.com.ve/">http://www.ultimashonicias.com.ve/</a>	private	n.a.	n.a.	n.a.		CA	OVH
67	<a href="http://www.av.info.ve/">http://www.av.info.ve/</a>	private	n.a.	n.a.	n.a.		VE	CANTV-NET
68	<a href="http://www.av.info.ve/">http://www.av.info.ve/</a>					Scraping did not work		

Table A.1: Assessment of monitored websites. Note: In the sample included websites are highlighted in bold. In cases when Content Delivery Networks (CDNs) are used the geo-location refers to the delivery network (Cloudflare, Amazon, Microsoft and Rackspace). The estimates for Venezuelan readers are taken from Amazon's Alexa Web Information Service and were only collected for the websites in the sample. For the domain lookup I used OpenIntel (van Rijswijk-Delj et al., 2016).

indicate that the client has problems to reach the host, whereas “5XX” codes mean that the server where the news website is hosted encounters problems.

My university server contacted the Venezuelan news websites every 30 minutes and saved the returned status code using Python’s *urllib2* library with an user agent. I restricted the measurement to every 30 minutes and added an user agent to avoid blocking as well as unnecessary traffic to the servers. Finally, my script returned the code “999” when it could not open the respective website. This may have occurred due to errors on the server-side, my machine, or somewhere in middle. As described in the main text, I primarily relied on a ”503” error code, which indicates a server overload, to infer potential DoS attacks.

Figure A.2 shows the average duration of ”503-codes” for the recorded attack days and different news outlets, highlighting that most attacks had been rather short. This could mean that the attacked website quite quickly defended these attacks or that the perpetrators could or did not want to sustain their attacks. As displayed in Figure A.1, I only record two incidents of potential DoS attacks when focusing on government or pro-government websites.

Finally, while I originally wanted to collect additional information and measure ping times (the time a request takes), this had not been possible as most of the monitored servers were protected and not “pingable.”

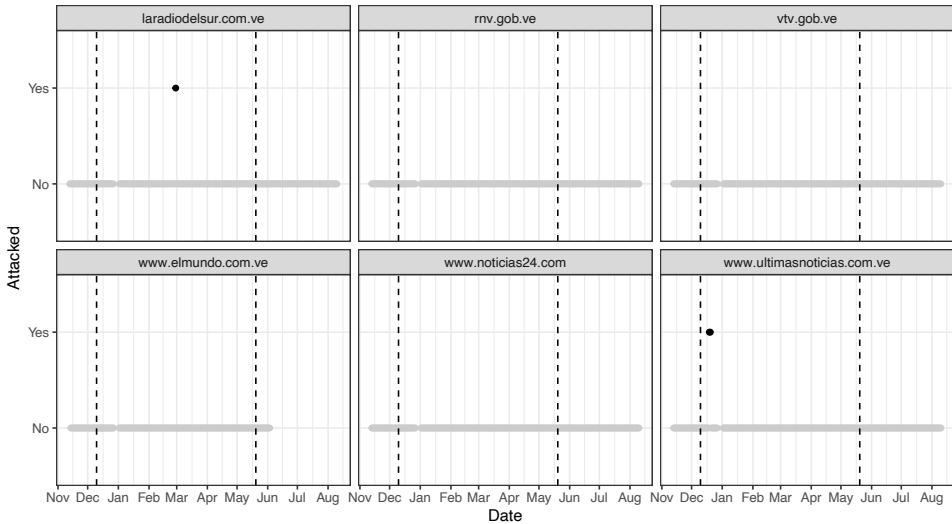


Figure A.1: Incidents of measured DoS attacks against government-affiliated/pro-government websites in Venezuela November 2017 - June 2018. Note: The two vertical lines show the municipal election (December 10, 2017) and presidential election (May 20, 2018). Blank periods indicate time periods where the measurement did not work.

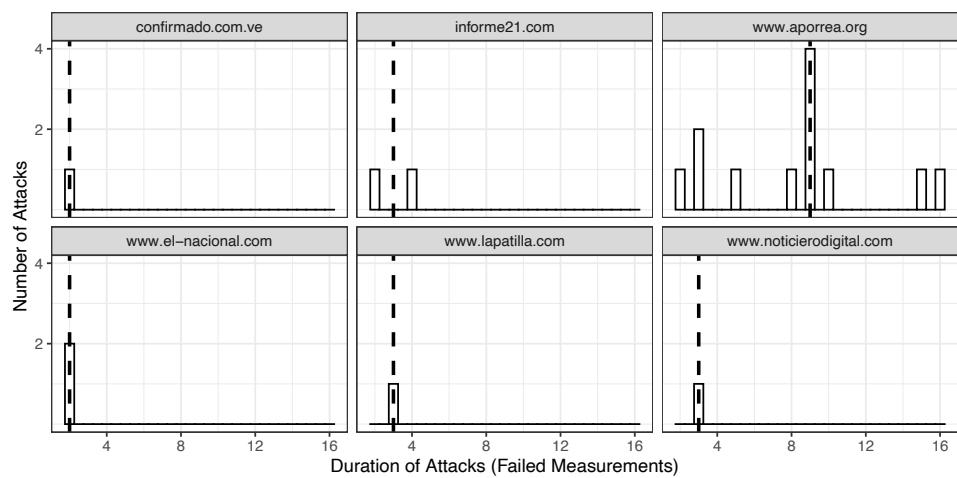


Figure A.2: Number of failed “503” measurements per attack/day. Note: The vertical dashed lines show the respective group means.

## B Main Analysis

	General opinion	Salary/prices	Food policy	Exiled opposition	Court sentences/Lula da Silva	Economy policy	Child mortality	Outages	Election	Government-Opposition dialog
Topic	-1.35** (0.25)	-0.15 (0.05)	n.a. (n.a.)	0.69** (0.04)	-0.47* (0.15)	n.a. (n.a.)	0.45** (0.11)	-0.22 (0.14)	0.52** (0.08)	-0.35* (0.10)
DoS attack (t-1)	2.79** (0.39)	2.35** (0.60)	n.a. (n.a.)	2.65** (0.61)	2.37* (0.71)	n.a. (n.a.)	2.57** (0.60)	2.31** (0.62)	2.38** (0.59)	2.35** (0.58)
AIC	158.50	164.12	n.a.	159.95	163.17	n.a.	162.34	164.11	162.40	164.12
BIC	399.28	405.50	n.a.	408.73	403.95	n.a.	404.12	405.49	403.48	404.80
Log Likelihood	-40.25	-43.36	n.a.	-40.98	-42.59	n.a.	-42.67	-43.35	-42.35	-43.01
Deviance	80.50	86.72	n.a.	81.95	85.17	n.a.	85.34	86.71	84.70	86.02
Num. obs.	3547	3547	n.a.	3547	3547	n.a.	3547	3547	3547	3547
	Music/entertainment	Government ministers	Protest/outages	Colonial border	Oscar Porez/repression	Church/job_other (mixed)	National assembly	Airline (opening/closing)	Political prisoners	Russia
Topic	-0.06 (0.05)	0.07 (0.06)	-0.41* (0.10)	-0.24* (0.11)	0.09 (0.09)	0.10 (0.13)	0.74** (0.15)	0.52** (0.14)	0.52** (0.10)	0.23 (0.15)
DoS attack (t-1)	2.38** (0.61)	2.36** (0.60)	2.46** (0.59)	2.46** (0.66)	2.25** (0.56)	2.38** (0.61)	2.00** (0.49)	2.36** (0.62)	2.36** (0.55)	2.35** (0.58)
AIC	164.67	164.98	164.44	164.37	164.91	164.82	161.50	164.50	163.74	164.84
BIC	405.45	405.76	405.22	405.15	405.69	405.60	402.28	405.28	404.52	405.62
Log Likelihood	-43.32	-43.49	-43.22	-43.19	-43.45	-43.41	-41.75	-43.25	-42.37	-43.42
Deviance	86.67	86.98	86.44	86.27	86.91	86.82	83.50	86.50	85.74	86.84
Num. obs.	3547	3547	3547	3547	3547	3547	3547	3547	3547	3547
	Sanctions	Military	Shortages healthcare	Regulations	USA/Korea	Cuba	International Organizations	Exchange Rate	Maduro	Migration crisis
Topic	0.18 (0.14)	0.12 (0.15)	0.16 (0.10)	-0.48* (0.14)	-0.44 (0.22)	0.56** (0.14)	0.37** (0.13)	0.13 (0.13)	0.65** (0.15)	-0.07* (0.14)
DoS attack (t-1)	2.32** (0.63)	2.33** (0.58)	2.39** (0.64)	2.39** (0.63)	2.31** (0.56)	2.57** (0.58)	2.55** (0.60)	2.37** (0.62)	2.09** (0.62)	2.39** (0.60)
AIC	164.76	164.90	164.84	163.26	163.82	163.25	162.56	165.14	161.22	164.89
BIC	405.54	405.68	405.62	404.04	404.60	404.03	403.34	405.92	402.00	405.67
Log Likelihood	-43.38	-43.45	-43.42	-42.63	-42.91	-42.63	-42.28	-43.57	-41.61	-43.45
Deviance	86.76	86.90	86.84	85.26	85.82	85.25	84.56	87.14	83.22	86.89
Num. obs.	3547	3547	3547	3547	3547	3547	3547	3547	3547	3547
	Petroleum	PDVSA	Resignations	Mining/sport (mixed)	Opposition candidate	Earthquake/accidents	Shortages	Corruption	Education studies	
Topic	0.07 (0.19)	0.46** (0.07)	0.81** (0.06)	0.31* (0.09)	-0.18* (0.06)	-1.11** (0.18)	0.14* (0.06)	0.05 (0.17)	-1.40** (0.21)	
DoS attack (t-1)	2.38** (0.61)	2.27** (0.62)	2.44** (0.67)	2.40** (0.67)	2.37** (0.67)	2.25** (0.67)	2.41** (0.67)	2.35** (0.67)	2.10** (0.67)	
AIC	164.91	160.78	159.55	163.98	164.51	160.01	164.88	164.98	160.32	
BIC	405.69	401.56	400.13	404.76	405.29	400.79	405.66	405.76	401.10	
Log Likelihood	-43.46	-41.39	-40.68	-42.99	-43.25	-41.00	-43.44	-43.49	-41.16	
Deviance	86.91	82.78	81.35	85.98	86.51	82.01	86.88	86.98	82.32	
Num. obs.	3547	3547	3547	3547	3547	3547	3547	3547	3547	

Table B.1: Penalized logistic regression results - main models with newspaper and day fixed effects. Note: Newspaper and time-fixed effects are not displayed. Clustered standard errors in parentheses. Topic variables are scaled. The topics *food policy* and *economy policy* are left out since the models did not converge (reached their iteration limits).

\*\* $p < 0.05/281$ , \* $p < 0.05$

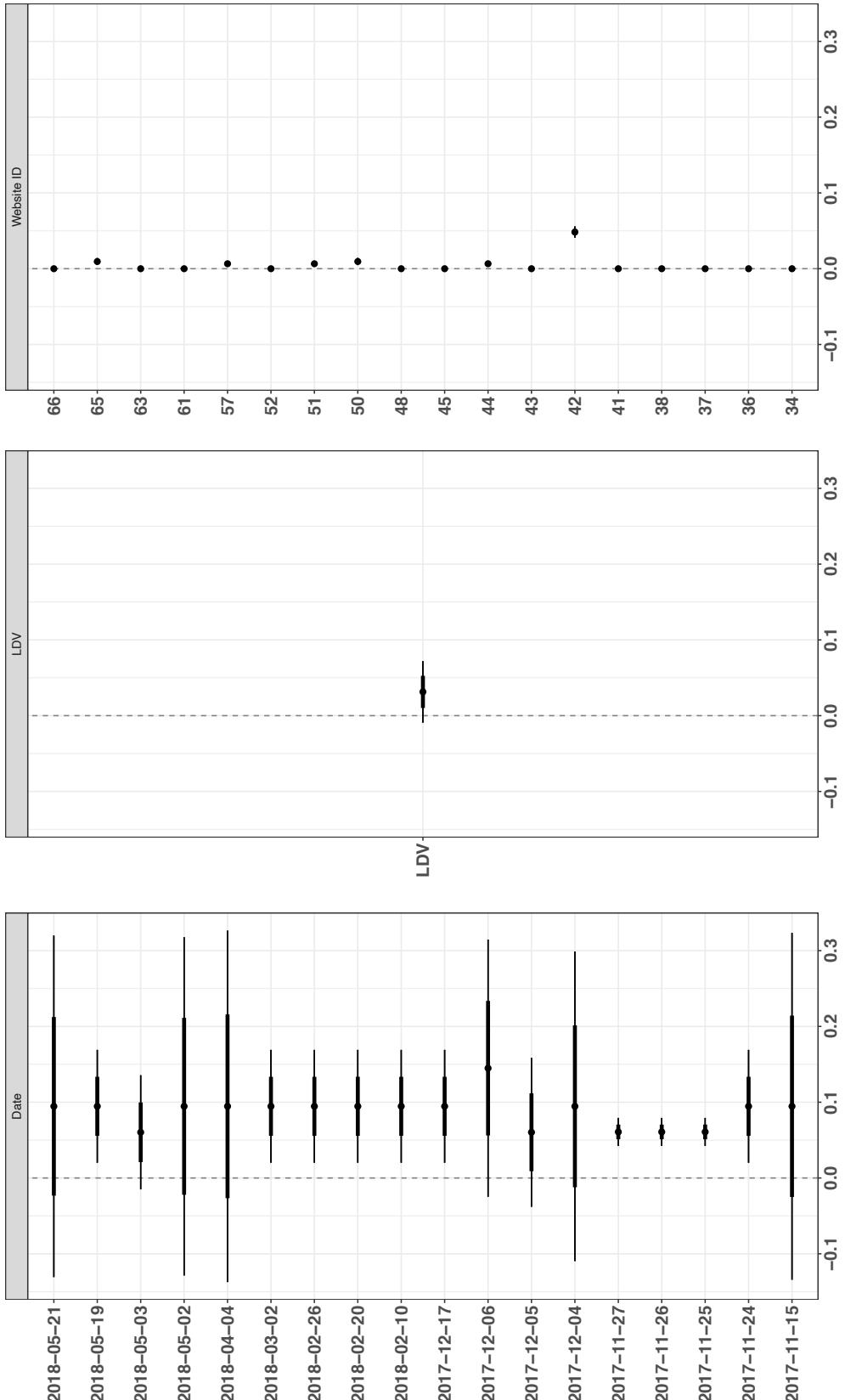


Figure B.1: Average Marginal Effects (AME) of lagged dependent variable, news websites and time effects on a news website's likelihood of receiving DoS attacks Note: Website ID 32 (Globovision) is set as baseline. Simulations based on 1000 draws and a null model only including the newspaper and time dummies as well as the lagged dependent variable. Topics are calculated in individual models and combined in the figure. 95% unadjusted and Bonferroni-corrected confidence intervals displayed.

## C Categorization of Topics

In this appendix, I explain in greater detail how I categorized the topics into broader categories. Table C.1 again displays this categorization.

Table C.1: Categorization of topics.

Category	$P(z)$	Topics	Share of pos. related topics $(\alpha = \frac{0.05}{281})$
Social/economic crisis	0.311	Sanctions, migration crisis, petroleum, salary/prices, shortages healthcare, exchange rate, shortages, outages, <b>child mortality</b> , <b>PDVSA</b> , airline (opening/closing), indigenous groups/diseases, mining/sport (mixed), [recession], [work opinion], [economy opinion], [investment], [financial market], [investment/infrastructure]	16.7%
Domestic politics	0.213	General opinion, <b>Maduro</b> , government-opposition dialog, government ministers, regulations, [economy policy], [food policy], [cryptocurrency], [Spanish development aid]	20%
Political legitimacy	0.212	<b>National assembly</b> , opposition candidate, <b>election</b> , international organizations, <b>resignations</b> , corruption, <b>exiled opposition</b> , [leftist opposition]	57.1%
Protest/repression	0.097	<b>Protest/shortages</b> , <b>Oscar Perez/repression</b> , <b>political prisoners</b> , military	75.0%
International politics	0.088	USA/Korea, Colombia border, <b>Cuba</b> , Russia, court sentences/Lula da Silva	20%
Nonpolitical topics	0.079	Earthquake/accidents, music/entertainment, church/job offer (mixed), education studies, [weather]	0.0%

Note:  $P(z)$  = distribution over whole corpus. Topics within square brackets highly correlate to specific websites ( $r >= 0.6$ ) or did not converge and are excluded from the analysis. Bold topics display significant positive correlations.

The topics were sorted qualitatively orientating myself on the broader topics that are censor-worthy as identified in the theoretical section, the generated topics, between-topic correlations, and newspaper categories in Venezuela.

In a first step, I looked for topics that are related to news on protests and repression. Overall, I identified four topics that fall into this category. The topics *protest/shortages*, *Oscar Perez/repression*, *political prisoners* and *military*—all topics either related to protest or state repression.

In a second step, I looked for topics that cover the social/economic crisis in Venezuela. For this category, I identified the most topics: the topics *sanctions*, *migration crisis*, *petroleum*, *salary/prices*, *shortages healthcare*, *exchange rate*, *shortages*, *outages*, *child*

*mortality, PDVSA, airline (opening/closing), indigenous groups/diseases, mining/sport (mixed), recession, work opinion, economy opinion, investment, financial market, and investment/infrastructure.* While for most topics, their names make it clear why I sorted them into the social/economic crisis category, some topics need more explanation and may be also considered in a different category.

The topic *sanctions* was sorted into this category as it is mainly concerned with economic sanctions that are imposed against the Venezuelan regime. Whereas the imposition of sanctions also questions the political legitimacy of a regime, the economic part appears to be more prominent. The topic *indigenous groups/diseases* mainly reports on healthcare shortages and diseases among indigenous groups within Venezuela—a topic that clearly covers the socioeconomic mismanagement in Venezuela. I put the mixed topic *mining/sport (mixed)* likewise in this category as it is primarily concerned with the contested *Arco Minero del Orinoco* governmental project that faces a lot of resistance and criticism.<sup>5</sup> While the socioeconomic part is present, there had been also considerable street protests against this project. The topic could be therefore also put into the protest and repression category. Finally, the topics *salary/prices, exchange rate, work opinion, economy opinion, investment, financial market, investment/infrastructure* report overall rather negatively on these issues and were thus put into the social/economic crisis category.

In a third step, I searched for topics that question the political legitimacy of the regime by either being related to opposition actors or elections or government criticism more generally. Concerning the former, I included the topic *national assembly*, which is primarily on the in 2017/18 opposition filled national assembly, the topic *opposition candidates*, which reports on candidates that ran against Maduro in the presidential elections, the topic *election*, and the topics *exiled opposition* and *leftist opposition*. With regards to more general topics that may question the political legitimacy of the regime, I identified the topic *international organizations*, which is largely on regional and international organizations criticizing the Venezuelan government, the topic *resignations* that is on resignations of other heads of states and resignation demands of Maduro by domestic and international actors, and the topic *corruption*.

In a last step, the remaining topics were sorted into the categories domestic politics, international politics, or nonpolitical topics. In the domestic politics category, I added mostly topics on the Maduro government for which it is not necessarily clear why they should question the political legitimacy of the regime, i.e., the topics *Maduro, government ministers, government-opposition dialog*, as well as news on food and economic policies, including regulations and cryptomoney policies that aim to solve the Venezuelan socioeconomic crisis. Moreover, I included the topic *Spanish development aid* and the broad

---

<sup>5</sup>The topic is mixed with sports news due to the frequent use of the word “gold” but largely captures the described news.

topic *general opinion* on Venezuela in this category. In the international politics category, I added the topics *USA/Korea*, *Colombia border*, *Cuba*, *Russia*, *Court sentences/Lula da Silva*, for which it is again not necessarily clear why these topics should question the political legitimacy of the Venezuelan regime. Finally, the topics *earthquake/accidents*, *music/entertainment*, *church/job offer (mixed)*, *weather* and *education studies* were sorted into the nonpolitical category.

While compelling, this classification is not necessarily exclusive and some topics may fall into multiple categories. Moreover, some of the topics show considerable correlations between each other—an observation that has to be kept in mind when discussing the results (see E.1 in Appendix E).

## D Sensitivity and Robustness Tests

Here, I outline several additional tests conducted to investigate the robustness and sensitivity of the results. I ran permutation tests, controlled for legitimate traffic, considered single 503-error codes as DoS attacks, operationalized the topics differently, display negative related topics, and ran models considering previous reporting.

**Permutation tests** First, permutation analyses were run to test whether the displayed topics are just a random realization. To achieve this, the occurrence of the measured DoS attacks were randomized across the complete sample. Next, the estimates were retrieved for the main results of presented topics and the exercise was repeated 10,000 times. Figure D.1 shows the density plots of these simulations. The figure highlights that most of the actually observed estimates are above the 95% quantiles in these simulations. This finding thus suggests that the results are not as-if-random but also that up to 5% of the hypothetical DoS attack distributions would have led to similar or even stronger patterns. One explanation for the latter finding may be that previous reporting seems to influence attack patterns; a point I will return to later.

**Controlling for legitimate traffic** Second, to control for the possibility that legitimate interest led to a traffic overload, the models were run again including a variable that should measure daily spikes of legitimate traffic for the respective websites. For this variable, I relied on data provided from [www.alexa.com](http://www.alexa.com) that monitors Internet traffic towards websites worldwide, and retrieved these data for the period of study for all monitored websites at a daily resolution. The marginal effect plots in Figure D.2 show that the results remain consistent.

**Single 503-error to measure DoS attacks** Third, I conducted sensitivity tests to consider one failed 503 measurement as a DoS attack. While this may decrease the likelihood of estimation errors as the number of attack days increases to 42, the chances for measurement errors are greater, which may lead to imprecise estimates and larger confidence intervals. Figure D.3 still shows similar patterns. However, and as expected, the results display smaller effect sizes and higher levels of uncertainty. Topics from the protest/repression category are no longer strongly related to an increased likelihood of DoS attacks, suggesting a higher number of false positives for this operationalization.<sup>6</sup>

**Maximal operationalization of topics** Fourth, I ran the main models again but operationalized the topic variables differently. Instead of focusing on the salience of topics, the maximum distribution of a topic per newspaper/day was considered. Therefore, it

---

<sup>6</sup>Within the social/economic crisis category, in contrast, more topics show positive associations.

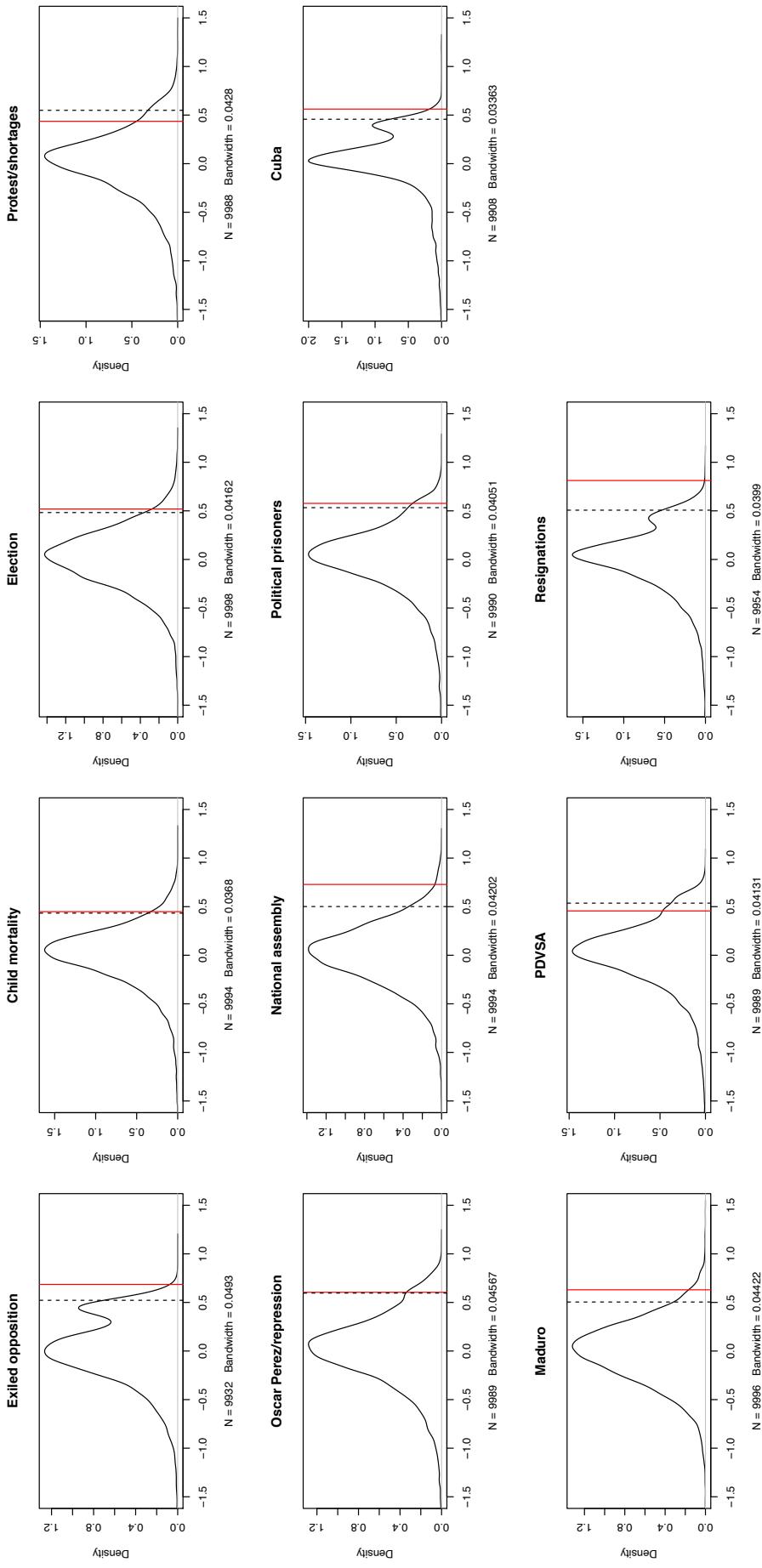


Figure D.1: Permutation tests of significant positively-related topics on a news website's likelihood of receiving DoS attacks. Note: Tests are based on 10 000 runs. The red solid line shows the empirical observed coefficient. The dashed line shows the 95% percentiles of the simulated distributions.

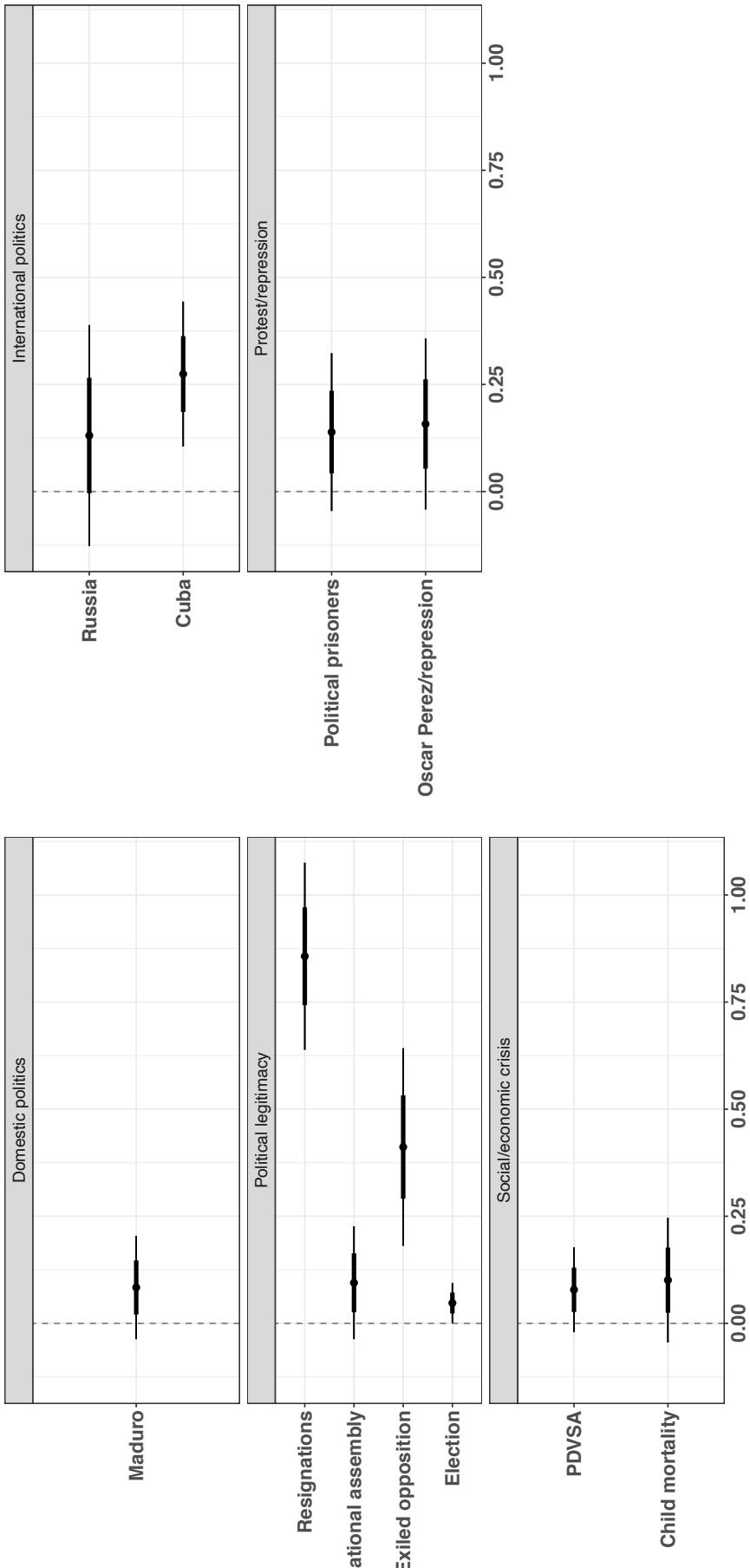


Figure D.2: Average Marginal Effects (AME) of significant positively-related topics on a news website's likelihood of receiving DoS attacks; including web traffic data. Note: Each AME is calculated in separate models. Simulations based on 1000 draws. Topics are calculated in individual models and combined in the figure. Bonferroni-corrected (thin lines) and unadjusted (bold lines) 95% confidence intervals are displayed. Traffic data is retrieved from Amazon's Alexa Web Information Service and measures the number of unique users per day in million. I lagged this variable by one day as DoS attacks may de- or increase the legitimate traffic for a specific website (not lagging this variable leads to almost identical results). The topic *protest/shortages* is left out since the model did not converge (see Table D.1).

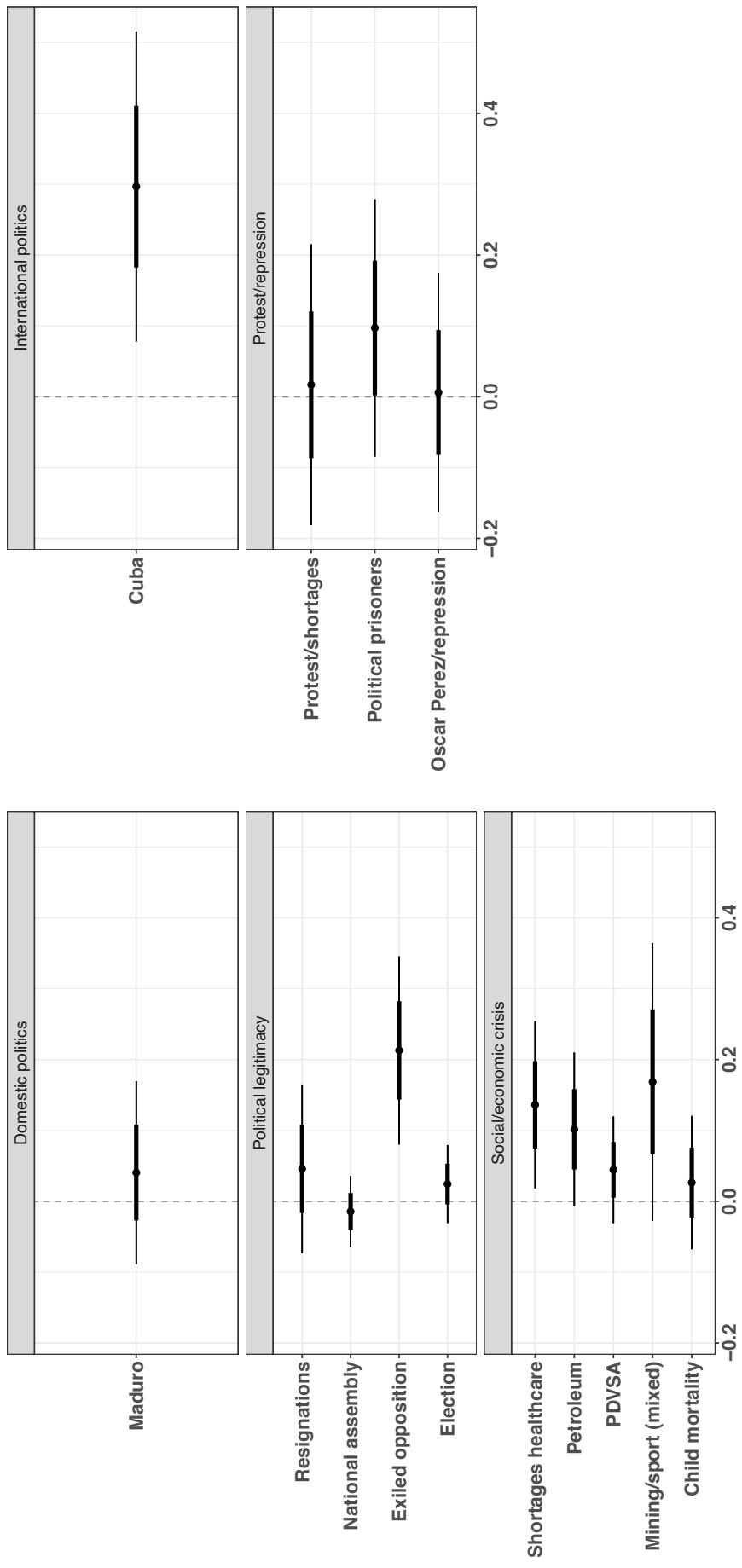


Figure D.3: Average Marginal Effects (AME) of significant positively-related topics topics and topics from the main models on a news website's likelihood of receiving DosS attacks; considering already one 503 measurement as an attack day. Note: Each AME is calculated in separate models. Simulations based on 1000 draws. Topics are calculated in individual models and combined in the figure. Bonferroni-corrected (thin lines) and unadjusted (bold lines) 95% confidence intervals are displayed.

is assumed that it matters when a topic is mentioned at least once. The results show positive AMEs for only a few of the topics that appear in the main results and the effect sizes are much smaller and the levels of uncertainty higher. As expected, it thus seems that the salience of topics is more important in influencing DoS attacks.

**Negatively related topics** Fifth, I simulated the AMEs for significant negatively-related topics as a validity check. For these topics, it is expected that they are either apolitical or not censor-worthy. Figure D.5 shows these negatively-related topics. The AMEs for these topics are much smaller, all close to zero, and the topics seem to be apolitical or not censor-worthy.

**Previous reporting** Sixth, I tested whether previous reporting may influence the likelihood of DoS attacks and ran the main models again considering the average reporting about news topics for the previous seven days.<sup>7</sup> Figure D.6 shows that many of the topics still display the same patterns as in the main analysis. However, some differences remain. Apart from the broad topic *Maduro*, the effect sizes are smaller overall, suggesting that recent content more likely triggers DoS attacks. Besides, the AMEs for most topics from the legitimacy and social/economic crisis categories become weaker and more uncertain, whereas those for topics from the protest/repression category become larger and significant even when considering the Bonferroni-corrected 95% confidence intervals.

**Considering a hundred topics** Finally, it may be that some specific topics escape the detection of my models due to the aggregation to 50 topics. To explore whether this is the case, I ran my main models again and relied on 100 generated topics, as reported in Appendix E. Although it becomes more difficult to compare the results with the main models, Figure D.7 illustrates that the overall patterns remain consistent. The results even highlight stronger associations between an increased likelihood of DoS attacks and more fine-grained news in the political legitimacy, protest and repression, and socioeconomic crisis categories. Due to the creation of new topics, it is, however, worthwhile to discuss some of the findings in detail and put the topics into context.

First, many of the topics that appeared in the main analysis are also showing up in the more fine-grained analysis. This is true for the topics *election(procedure)*, *election(general)*, *national assembly*, *exiled opposition* in the political legitimacy category, the topic *Cuba* in the international politics, and the topics *protest/riots*, *political prisoners* and *Oscar Perez/repression* in the protest/repression category. Moreover, the topic *healthcare* in the social/economic crisis category seems to be closely related to the topic *child mortality* in the main analysis.

---

<sup>7</sup>I decided to use seven days as a threshold as a response should occur relatively close to the publishing date of critical news. Since the downloading of websites sometimes did not work (see Figure E.2), I take the average after removing NAs.

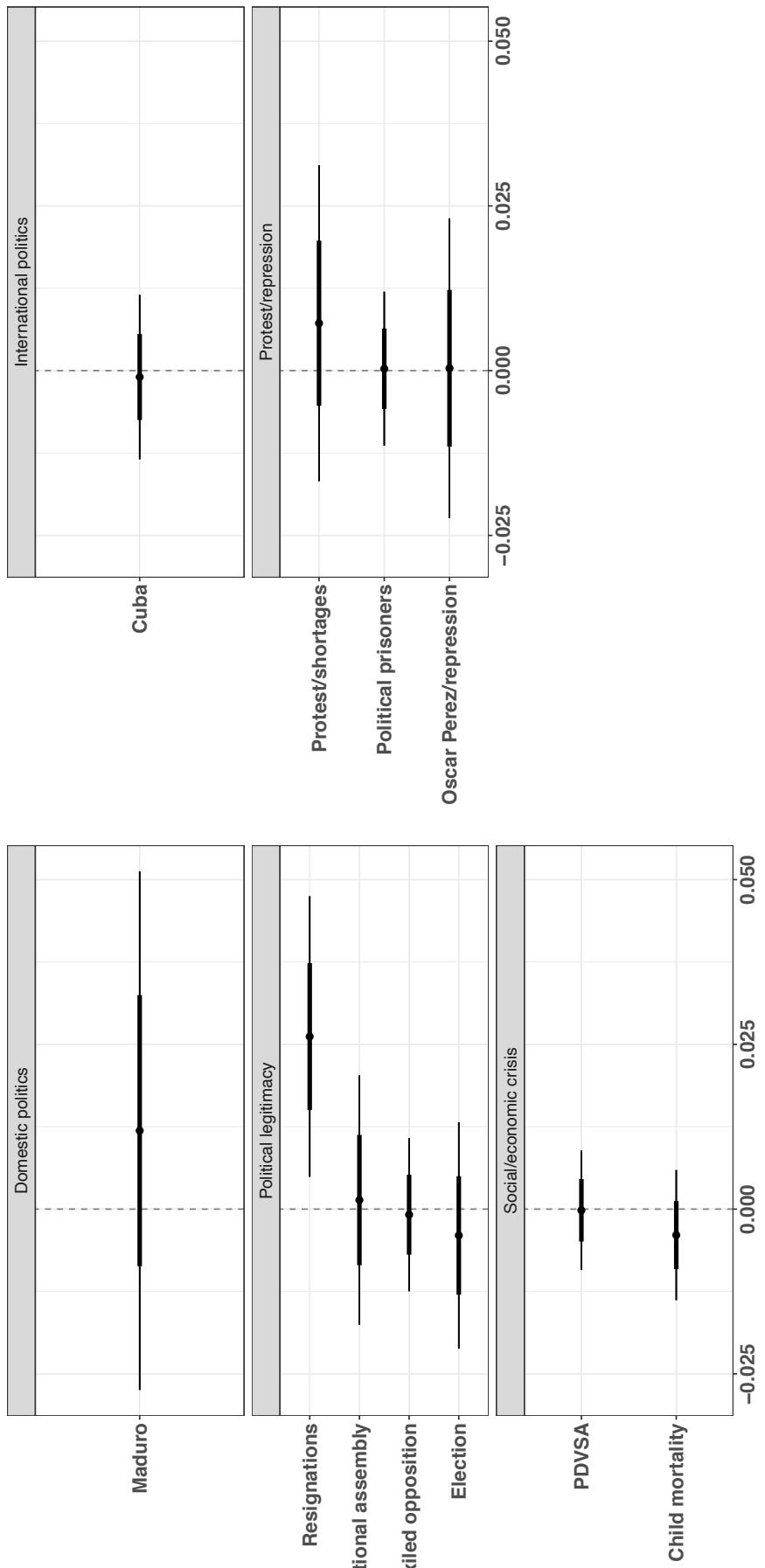


Figure D.4: Average Average Marginal Effects (AME) of significant positively-related topics and topics from the main models on a news website's likelihood of receiving DoS attacks; considering the maximal distribution of topics. Note: Each AME is calculated in separate models. Simulations based on 1000 draws. Topics are calculated in individual models and combined in the figure. Bonferroni-corrected (thin lines) and unadjusted (bold lines) 95% confidence intervals are displayed.

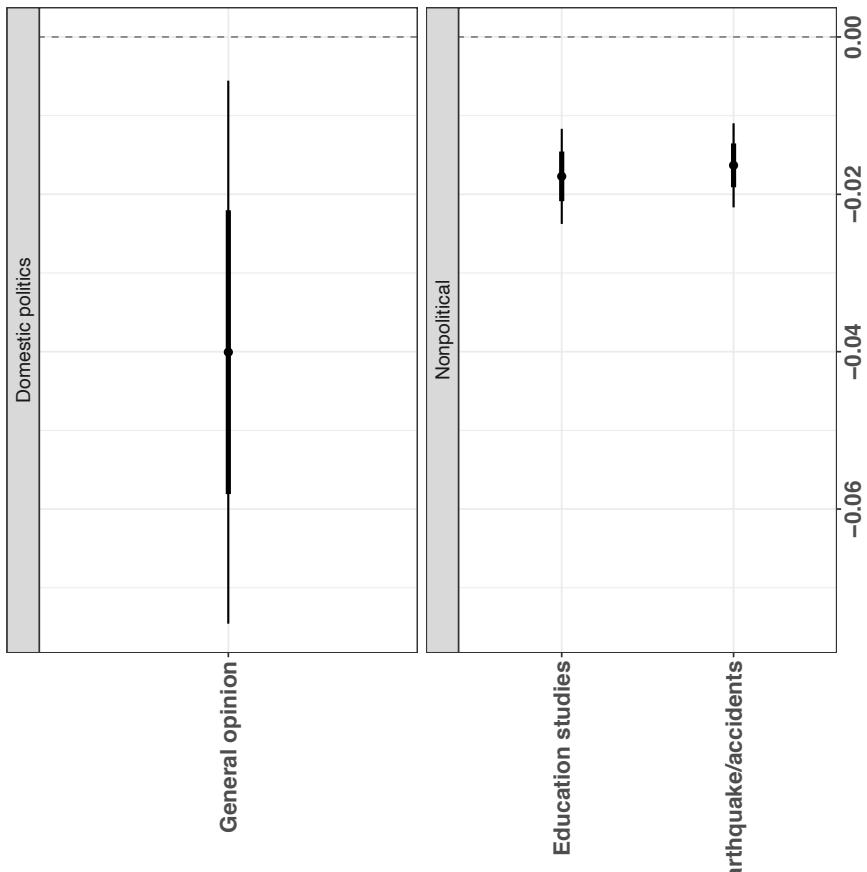


Figure D.5: Average Marginal Effects (AME) of significant positively-related topics on a news website's likelihood of receiving DoS attacks. Note: Each AME is calculated in separate models. Simulations based on 1000 draws. Topics are calculated in individual models and combined in the figure. Bonferroni-corrected (bold lines) and unadjusted (thin lines) 95% confidence intervals are displayed.

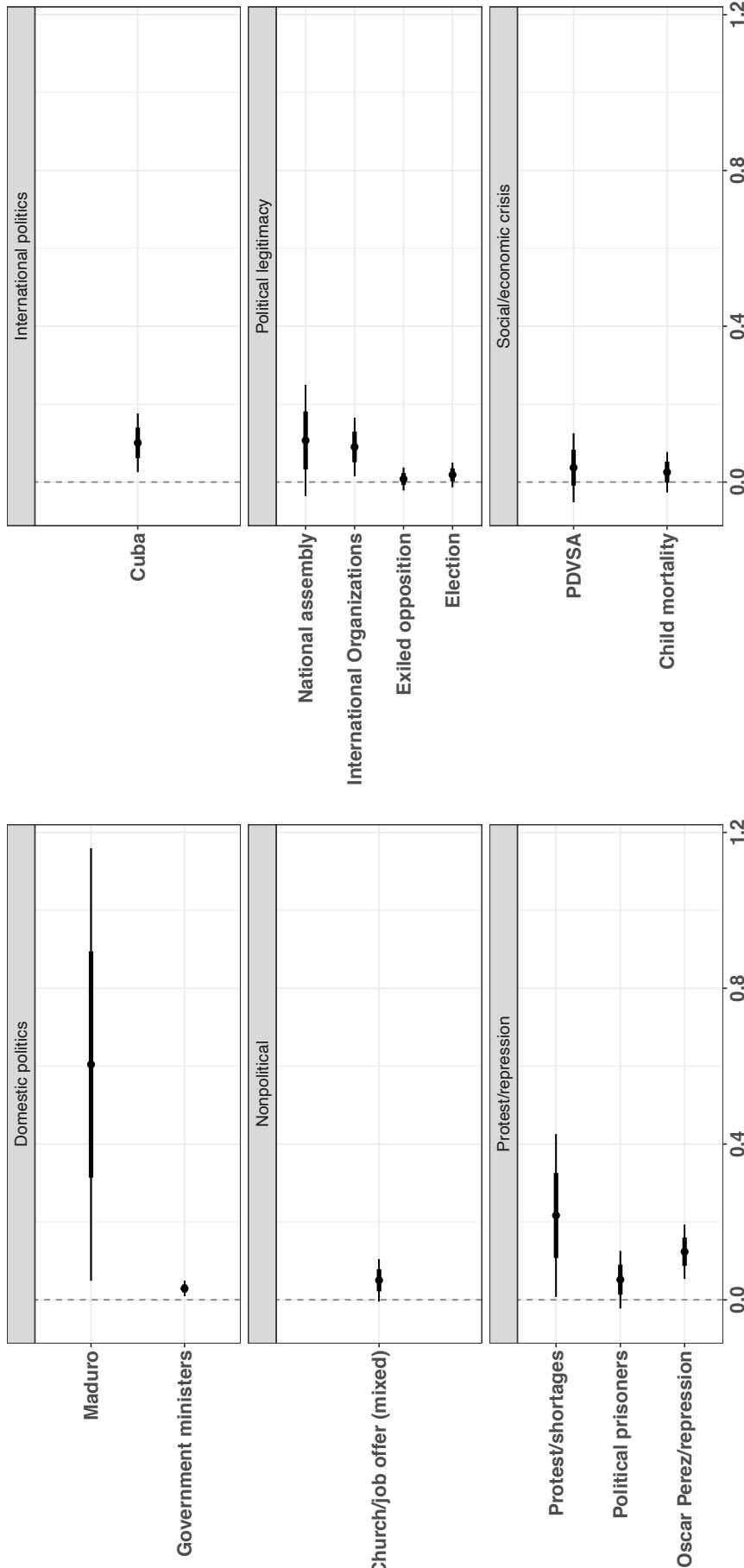


Figure D.6: Average Average Marginal Effects (AME) of significant positively-related topics and topics from the main models on a news website's likelihood of receiving DoS attacks; considering topic reporting up to seven days before. Note: Each AME is calculated in separate models. Simulations based on 1000 draws. Topics are calculated in individual models and combined in the figure. Bonferroni-corrected (thin lines) and unadjusted (bold lines) 95% confidence intervals are displayed. The topic *resignations* is left out as the model did not converge (see Table D.4).

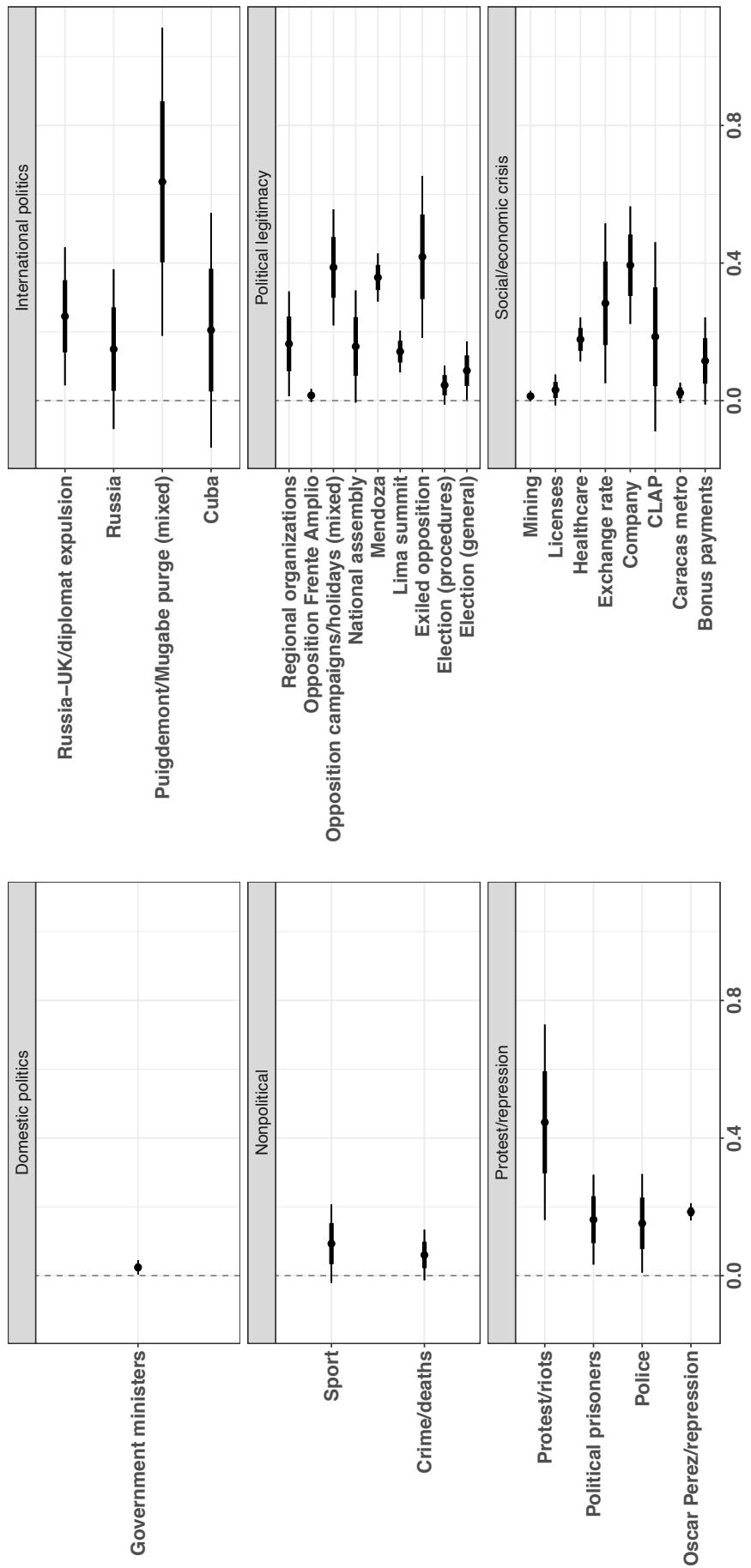


Figure D.7: Average Average Marginal Effects (AMEs) of significant positively-related topics on a news website's likelihood of receiving DoS attacks; considering 100 generated topics. Note: Each AME is calculated in separate models. Simulations based on 1000 draws. Topics are calculated in individual models and combined in the figure. Bonferroni-corrected (thin lines) and unadjusted (bold lines) 95% confidence intervals are displayed.

Second, it appears that the more fine-grained approach divided the topic *resignations*, which is on the resignation of other heads of state and resignation demands for Maduro by regional and international organizations, into several smaller topics. On the one hand, these are the topics *regional organizations* and *Lima summit* on which neighboring countries discussed how to deal with the Venezuelan crisis and the results of the presidential elections. On the other hand, the topic *Pujdemont/Mugabe purge (mixed)* appear to have been captured by the resignation topic in the original analysis.<sup>8</sup> In particular, the Mugabe part may have been perceived as censor-worthy for the Venezuelan regime as it reported on the coup against Mugabe in November 2017—a scenario the Maduro regime wants to avoid. Nevertheless, this finding may be also explained by the fact that this topic correlates to the topic *exiled opposition* ( $r=0.2$ ) and the more general topic *South/Latin America* ( $r=0.2$ ).<sup>9</sup>

Third, supporting the findings from the main analysis, more topics on the opposition and protests are positively and significantly related to an increased likelihood of DoS attacks. These topics are the mixed topic *opposition campaigns/holidays (mixed)*, the topic *Frente Amplio* on the in March 2018 formed opposition group led by Juan Guaidó, the topic *Mendoza* on the Venezuelan billionaire Lorenzo Mendoza who was urged by the opposition to run against Maduro but eventually declined,<sup>10</sup> and the topic *police* in the protest/repression category.

Fourth, some more topics in the social/economic crisis category are now significantly and positively correlated to a higher likelihood of DoS attacks. In particular, topics related to the exchange rate display a robust association; a topic, for which anecdotal evidence indeed highlights an interest of the Venezuelan government to suppress information on (Rueda, 2015). Whereas for the topic *company*, it is not necessarily clear why it should increase the likelihood of DoS attacks, this topic is highly correlated to the topic *petroleum export* ( $r=0.5$ ) and may thus capture news on the PDVSA that showed a positive association in the main analysis.

Finally, the analysis finds a robust relationship for the topic *Russia-UK/diplomat expulsion*. While it is not necessarily clear why this topic should increase the likelihood of DoS attacks, Venezuela expelled several diplomats after the US accused Maduro of electoral fraud.<sup>11</sup>

---

<sup>8</sup>Mixed means that the topic captures two—qualitatively—rather unrelated topics. Since the topic is largely about international news, I sorted it into the category "international politics."

<sup>9</sup>The correlation matrix can be found in the replication material.

<sup>10</sup><https://www.reuters.com/article/us-venezuela-politics-mendoza-idUSKCN1FZ24U> (accessed May 10, 2021)

<sup>11</sup>Moreover, this topic correlates to the *police* topic with  $r=0.2$ . The finding related to the *sport* topic appears to be rather an artifact of the estimations and remains insignificant when considering the Bonferroni correction. Similarly, the topic *crime/deaths* correlates considerably to topics on protest and the security forces ( $r=0.3$ ); see replication files.







	Cloudflare (mean)	Non-Cloudflare (mean)	t-value	p-value
General opinion	0.09	0.06	1.73	0.15
Salary/prices	0.03	0.02	1.35	0.22
Food policy	0.02	0.01	1.75	0.14
Exiled opposition	0.02	0.03	-1.38	0.25
Court sentences/Lula da Silva	0.02	0.01	2.64	0.02
Economy policy	0.01	0.01	0.30	0.77
Child mortality	0.02	0.02	-0.05	0.96
Outages	0.01	0.01	3.16	0.01
Election	0.05	0.06	-0.49	0.66
Government-Opposition dialog	0.03	0.04	-1.30	0.24
Music/entertainment	0.02	0.04	-0.93	0.42
Government ministers	0.04	0.03	1.60	0.16
Protest/shortages	0.03	0.03	-0.34	0.76
Colombia border	0.02	0.01	3.02	0.04
Oscar Perez/repression	0.03	0.03	0.02	0.98
Church/job offer (mixed)	0.01	0.01	0.33	0.76
National assembly	0.03	0.06	-1.32	0.27
Airline (opening/closing)	0.01	0.01	1.76	0.11
Political prisoners	0.02	0.03	-0.67	0.55
Russia	0.03	0.01	3.16	0.03
Sanctions	0.04	0.04	-0.19	0.86
Military	0.01	0.00	2.74	0.01
Shortages healthcare	0.02	0.02	-0.22	0.84
Regulations	0.00	0.00	-0.25	0.81
USA/Korea	0.02	0.01	2.74	0.02
Cuba	0.01	0.01	0.20	0.85
International organizations	0.02	0.04	-0.76	0.50
Exchange rate	0.01	0.01	0.52	0.63
Maduro	0.05	0.07	-1.46	0.24
Migration crisis	0.03	0.03	-0.35	0.74
Petroleum	0.03	0.01	4.71	0.00
PDVSA	0.02	0.04	-0.81	0.48
Resignations	0.02	0.04	-0.70	0.53
Mining/sport (mixed)	0.02	0.02	0.14	0.89
Opposition candidate	0.05	0.04	0.12	0.91
Earthquake/accidents	0.02	0.01	0.86	0.43
Shortages	0.02	0.01	2.56	0.03
Corruption	0.02	0.02	-0.11	0.92
Education studies	0.01	0.01	2.08	0.06

Table D.6: T-tests between topic distributions between Cloudflare protected vs. unprotected servers on attack days. *Note:* Topics that highly correlate to specific websites are excluded (see main analysis and Figure E.1). I thank one of the anonymous reviewer for suggesting this test.

## E Text Analysis and Topic Modeling

In this appendix, I describe in more detail how I retrieved the textual data, pre-processed the obtained headlines, and investigated the validity of the estimated topics.

**News content retrieval and text analysis** To retrieve the newspaper’s content, I again relied on a Python script using the *urllib2* library as specified above and the *BeautifulSoup* library to download the complete HTML code and content of the websites. I run this script every day at 1:00 p.m. Venezuelan time. After extracting the headlines and first paragraphs, if available, I removed punctuation and numbers (dates, etc.), use of lower-casing and Porter stemming, as well as remove Spanish stopwords and other newspaper-related words (“read more” buttons, authors, etc.). Table E.1 reports the summary statistics for the textual data. In a last step, I ran the Bitemr Topic Models as described in the main text. For this task, I relied on an implementation in Python and C++ provided by Yan et al. (2013).<sup>12</sup>

Table E.1: Summary statistics of text corpus.

No. of website/days	No. of headlines	Avg. no. of headlines per website/day
3,566	123,889	34.74
Avg. term length per headline	Std. dev. term length per headline	Total no. of stemmed terms
13.14	9.41	26,640

**Between topic and website correlations** Figure E.1 shows the pair-wise correlations of all topics and websites for the merged data. The figure points to a large cluster of collinear topics about economic issues. Because these topics are additional highly correlated to one specific website (*dinero.com.ve*), with pair-wise correlations of above 0.6, I do not consider them in the statistical analysis. I exclude these topics because it is very unlikely that these topics are represented well on other websites. Figure E.2 illustrates, for instance, that the average distribution of the topic *recession* is almost zero for every website except for *dinero.com.ve*. Moreover, the topics *leftist opposition*, *weather* and *indigenous groups/diseases* show pair-wise correlations equal or above 0.6 to specific newspapers and are likewise not considered in the statistical models that employ newspaper fixed effects. Although other topics and news websites do not correlate that strongly, some of them show significant correlations; an observation one has to keep in mind when interpreting the empirical results.

**Topic model validity tests** To validate the topics, I first looked at semantic validity, checking whether the generated topics are coherent and overlap with the respective headlines. Second, I also focused on predictive validity, investigating whether the assigned

<sup>12</sup><https://github.com/xiaohuiyan/BTM>

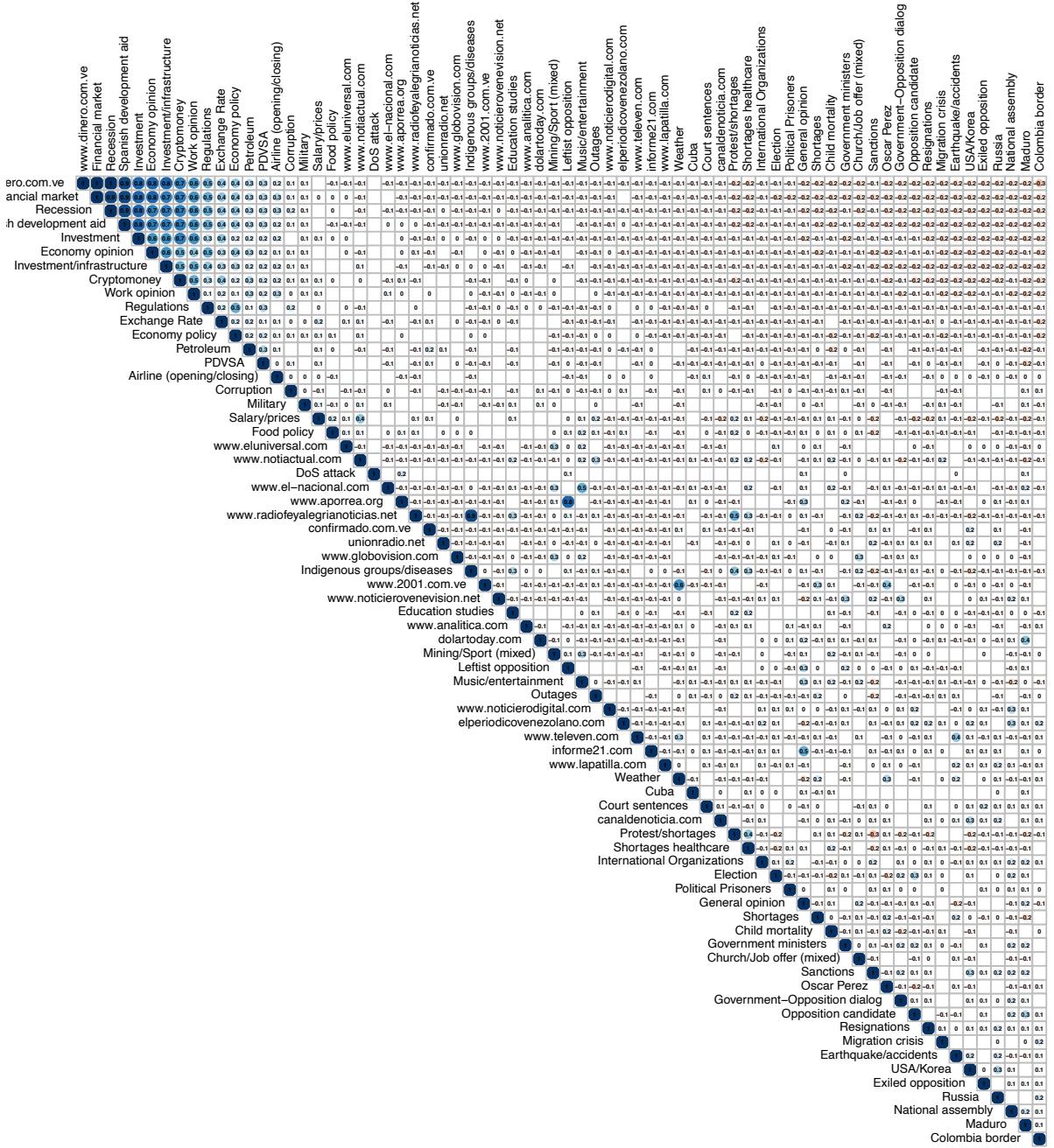


Figure E.1: Newspaper/day correlation between topics and websites. Note: The figure is ordered along a principal component analysis, clustering highly collinear topics and websites. Blue dots display positive correlations, whereas red dots negative correlations. Darker shades indicate a stronger correlation. Blank fields show insignificant correlations.

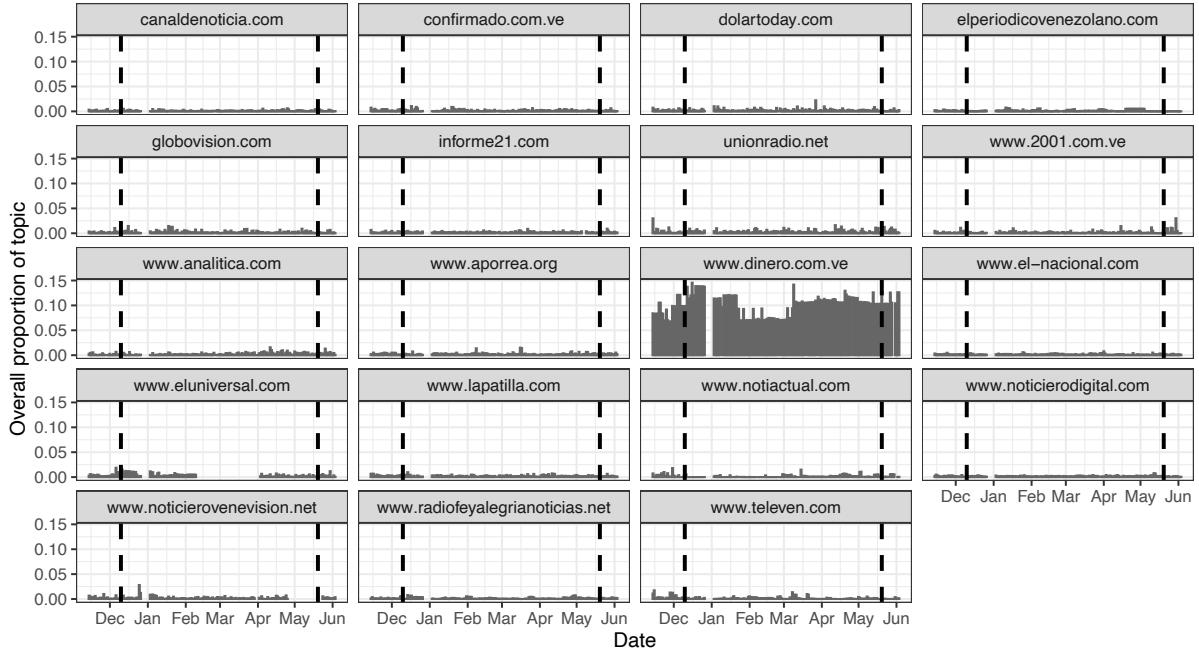


Figure E.2: Development of the topic *recession* in Venezuela November 2017 - June 2018.  
 Note: The two vertical dashed lines show the municipal election (December 10, 2017) and presidential election (May 20, 2018). Blank periods indicate time periods where the measurement did not work.

headlines reflect real-world developments (Grimmer and Stewart, 2013; Quinn et al., 2010).

For the semantic evaluation, I evaluated the top 15 words associated with the topics, as well as the top 25 headlines for each topic. This evaluation shows that most topics appear to be valid and I summarize the results in the last column in Table E.2.

For the predictive validity, I plotted the temporal development of some topics linking them to real-world events. The following figures show the temporal development of the relative proportions of the topics *election* (Figure E.3) and *Óscar Pérez/repression* (Figure E.4) aggregated across all newspapers. For Figure E.3, the relative topic proportion increases sharply close to the election dates on December 10, 2017, and May 20, 2018. In February 2018, the National Electoral Council (CNE) decided about the ban of several candidates, which is reflected by the slight increase during this period. In Figure E.4 one can see that the highest proportion of the topic is in mid-January after a special commando killed Óscar Pérez in the so-called *El Junquito* raid (Casey, 2018). Overall, these tests make me confident that the results of the Biterm Topic Model are valid.

Table E.3 illustrates that many of the terms correspond to mixed topics when I consider K=25 (Table E.3), while for Table E.4 (K=100) it is difficult to distinguish between the topic as some of them are very similar.

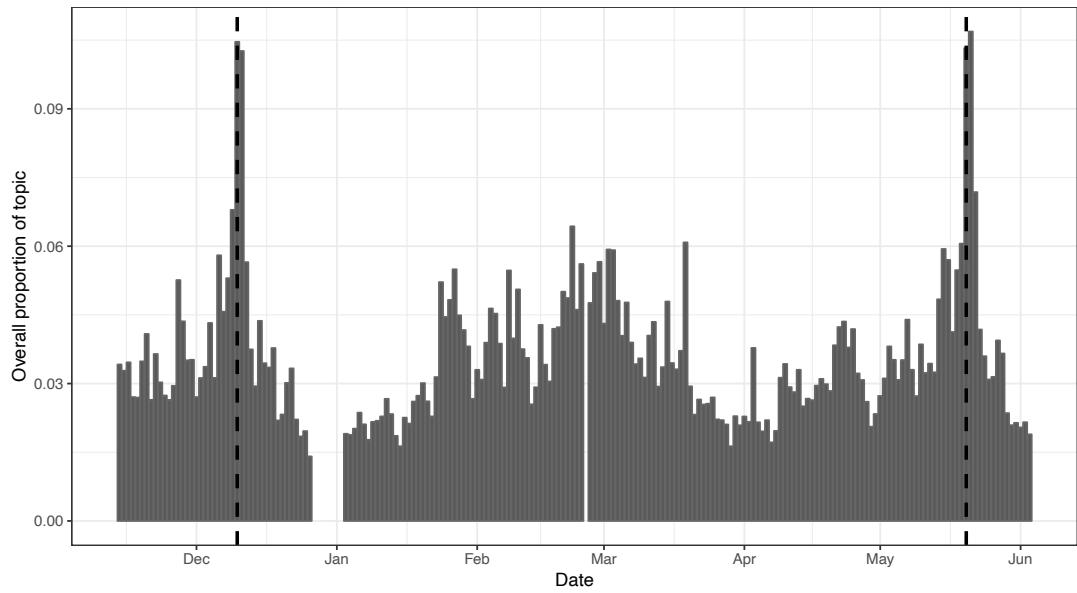


Figure E.3: Temporal development of the aggregated topic *election*. Note: The two vertical lines show the municipal election (December 10, 2017) and presidential election (May 20, 2018).

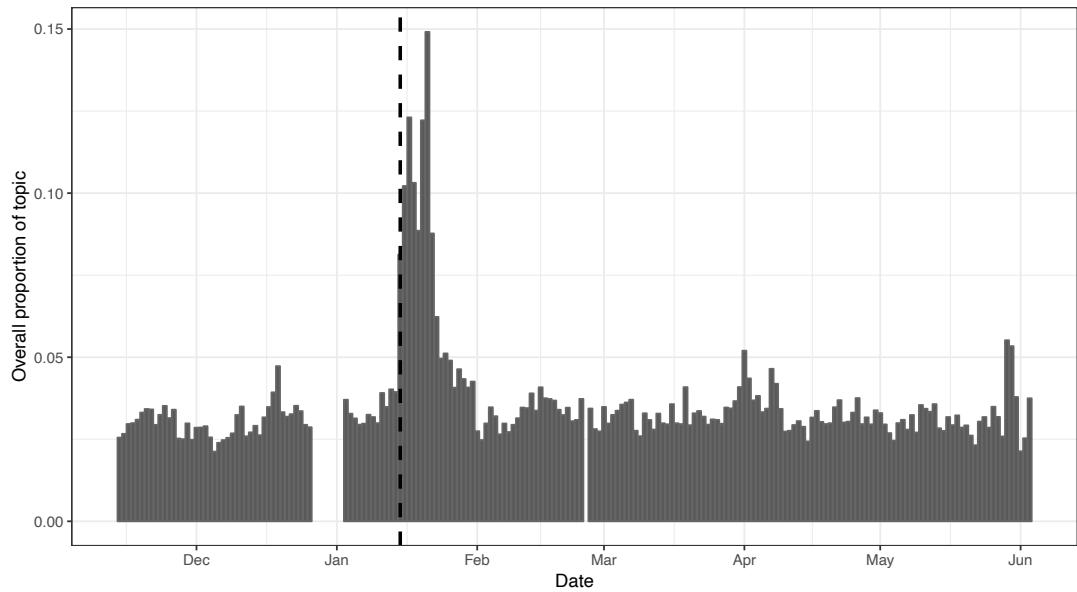


Figure E.4: Temporal development of the aggregated topic *Óscar Pérez/repression*. Note: The vertical line corresponds to the killing of Pérez on January 15, 2018.

p(z)	Label	Top words	Internal valid?
0.066324	General opinion	Venezuela:0.014451 continue:0.007950 Venezuelan:0.006993 will be:0.006266 alone:0.005320 live:0.005056 cial:0.004655 could:0.004332 country:0.010544 can:0.007115 politics:0.006879 should:0.006143 economy:0.005207 so- cial:0.004530 Maduro:0.004530	Yes
0.043895	Sanctions	Venezuela:0.045923 sanction:0.023301 government:0.018773 Venezuelan:0.014024 country:0.012811 Union:0.009741 election:0.008598 national:0.053728 bly:0.035471 dispute:0.024873 president:0.015611 AN:0.012692 politics:0.010413 Venezuela:0.009298 constituent:0.009128 Maduro:0.008510 government:0.007927 commission:0.007764 party:0.007629 new:0.006602 ANC:0.006557 USA:0.023625 United:0.022011 embassy:0.014289 Maduro:0.012839 European:0.012113 EU:0.008805 reject:0.007592 offi- cials:0.006909	Yes
0.041180	National assembly	assem- bly:0.035471 dispute:0.024873 pres- ident:0.015611 AN:0.012692 poli- tics:0.010413 Venezuela:0.009298 con- stituent:0.009128 Maduro:0.008510 gov- ernment:0.007927 commission:0.007764 party:0.007629 ensure:0.007343 new:0.006602 ANC:0.006557 assem- bly:0.053728 dispute:0.024873 pres- ident:0.015611 AN:0.012692 poli- tics:0.010413 Venezuela:0.009298 con- stituent:0.009128 Maduro:0.008510 gov- ernment:0.007927 commission:0.007764 party:0.007629 ensure:0.007343 new:0.006602 ANC:0.006557	Yes
0.040474	Maduro	Maduro 0.089321 Nicolas:0.044773 government:0.014672 Venezuelan:0.014575 ensure:0.009381 announce:0.008243 republic:0.007715 country:0.006637 new:0.005980 presi- dential:0.005932 Santos:0.005134 president:0.048436 Venezuela:0.020501 Venezue- lan:0.014575 ensure:0.009381 an- nounce:0.008243 national:0.008113 election:0.007357 presi- dential:0.005932 Santos:0.005134	Yes

*Continue on the next page*

p(z)	Label	Top words	Internal valid?
0.039453	Opposition candidate	Falcon:0.026931 presidential:0.026221 election:0.017800 opposition:0.010491 candidacy:0.009921 president:0.008389 Venezuelan :0.008193 government:0.007562 candidate:0.026306 Henri:0.019067 Maduro:0.011979 vote:0.010328 party:0.009607 electoral:0.008336 ensure:0.008181	Yes
0.038004	Government-opposition dialog	dialog:0.038177 opposition:0.026216 Venezuelan:0.019267 Venezuela:0.015758 agreement:0.014191 Jorge:0.010237 chancellor:0.008789 Rodriguez:0.008525 election:0.044032 national:0.019208 elections:0.010967 vote:0.010189 process:0.009810 president:0.008820 next:0.007539 municipal:0.007472 Venezuela:0.007011 government:0.036453 Dominican:0.016998 Republic:0.014486 president:0.012837 meeting:0.010039 Maduro:0.008696 national:0.008311	Yes
0.035994	Election	electoral:0.041866 CNE:0.021575 council:0.016683 party:0.010370 realize:0.007101 Venezuela:0.007011	Yes
0.033783	Migration crisis	Venezuelan:0.062987 Venezuela:0.018283 crisis:0.010881 Brazil:0.007488 border:0.007188 help:0.006682 refugees:0.006124 citizens:0.005518 country:0.022501 Colombia:0.015586 international:0.007737 humanitarian:0.007452 migration:0.006905 thousands:0.006681 national:0.005798	Yes

*Continue on the next page*

p(z)	Label	Top words	Internal valid?	
0.031420	Protest / Short-ages	protest:0.027537 ing:0.010183 city:0.008252 bor:0.007388 new:0.007128 nounce:0.006803 transport:0.006715 habitat:0.006276	San:0.015367 municipal:0.010159 sector:0.007497 neighbor:0.007379 bolivar:0.006942 de- street:0.006725 regional:0.006701	miss- Yes
0.031162	Oscar Perez / repression	Perez:0.022184 sassinate:0.016213 official:0.009250 police:0.008558 two:0.008316 CPC:0.007948 dead:0.007656	Oscar:0.020064 body:0.011489 steal:0.008872 men:0.008435 kill:0.007951 CI- national:0.007672 year:0.006789 Venezuelan:0.006496	as- (incl. other killings)
0.028106	Earthquake / accidents	dead:0.031197 son:0.023265 leave:0.017647 magnitude:0.009730 result:0.008966	hurt:0.025463 at least:0.020505 two:0.013686 after:0.012195 international:0.011957 died:0.009862 accident:0.009463 three:0.008569 registered:0.008514	per- Yes
0.027333	International Organizations	Venezuela:0.033977 international:0.018806 denounce:0.012866 UN:0.011290 organisation:0.010889 protest:0.009684 magro:0.008946 Luis:0.008695	rights:0.019264 human:0.018235 OAS:0.011484 general:0.011158 government:0.010716 inform:0.009191 Al- Nicaragua:0.008819	Yes (fo- cus on Nicaragua)

*Continue on the next page*

p(z)	Label	Top words		Internal valid?
0.025594	Government ministers	minister:0.022420 national:0.015720 vicepresident:0.012889 Venezuela:0.010987 party:0.009915 Diosada:0.009055 PSUV:0.007680 Jorge:0.007581	Rodriguez:0.020474 Cabello:0.013887 social:0.012862 plan:0.010562 president:0.009407 ensure:0.008984 Lopez:0.007622	Mixed (not only minis- ters)
0.025438	Political Prisoners	prisoners:0.028654 Penal:0.013729 national:0.012490 free:0.011634 Foro:0.011215 Venezue- lan:0.009659 denounce:0.009321 di- rector:0.007523 SEBIN:0.007302 Romero:0.006385	political:0.026987 liberty:0.013115 arrest:0.012011 Helicoid:0.007405 inform:0.006795	Yes
0.024928	USA / Korea	Trump:0.051455 dent:0.028576 USA:0.019123 United:0.018715 American:0.013238 Jong:0.009752 China:0.008804 summit:0.008452	presi- Donald:0.025267 Korea:0.018888 North:0.015156 US- Kim:0.012052 Jerusalem:0.009233 international:0.008607 announce:0.006899	Yes
0.024925	Petroleum	Venezuel:0.025539 production:0.016822 Venezue- lan:0.014950 prize:0.014382 country:0.013712 economy:0.013455 million:0.013265 dollar:0.012994 year:0.012678 barrel:0.011584 infla- cion:0.010082 OPEC:0.009600 accord- ing:0.009005 oil:0.008977	petroleum:0.021110 Venezue- la:0.014382 coun- try:0.013455 dollar:0.012994 infla- tion:0.010082 accord- ing:0.009005 oil:0.008977	Yes

*Continue on the next page*

p(z)	Label	Top words		Internal valid?
0.022507	Resignations	president:0.042073 Peru:0.022901 Kuczynski:0.013540 Lima:0.011251 Chile:0.009862 Pedro:0.009171 Pablo:0.009053 after:0.007998	America:0.023391 summit:0.019598 Venezuela:0.012032 resignation:0.010724 Madur:0.009205 country:0.009132 congress:0.008785	Yes
0.021977	Corruption	prosecution:0.024146 general:0.018316 Tarek:0.014059 Ramirez:0.013662 PDVSA:0.011895 republic:0.011018 accuse:0.008794	Saab:0.019186 Rafael:0.014747 corruption:0.013842 William:0.013165 arrest:0.011211 house:0.010497 investigation:0.009558 president:0.009064	Yes
0.020878	Salary / prices	Bolivar:0.025164 payment:0.021588 transport:0.018087 bs:0.015475 minimum:0.010897 work:0.008890 new:0.008118	increase:0.023093 million:0.018980 salary:0.017085 thousand:0.012006 passanger:0.011831 public:0.009525 venezolan:0.008314	Yes
0.019830	Colombia border	Colombia:0.044504 Venezuelan:0.014170 Venezuela:0.012413 Santos:0.010908 napping:0.009307 Juan:0.007949 peace:0.007581 Manuel:0.007463	Colombian:0.019540 border:0.013419 two:0.011176 ELN:0.009826 FARC:0.008024 international:0.007826 national:0.007511	Yes

*Continue on the next page*

p(z)	Label	Top words		Internal valid?
0.019772	Shortages health-care	patient:0.026464 health:0.020674 doctors:0.015288 supplies:0.013279 medicine:0.009601 work:0.009059 denounce:0.008976 crisis:0.007010	hospital:0.023871 missing:0.018602 medicines:0.013541 protest:0.010530 shortage:0.009472 Venezuela:0.009036 country:0.007416	Yes
0.019346	Exchange rate	new:0.024480 etary:0.019644 Venezuela:0.015966 exchange:0.015453 bolivar:0.013158 money:0.012494 central:0.011134	bank:0.019940 Dicom:0.018367 change:0.015510 auction:0.013899 BCV:0.013088 reconversion:0.011268 dollar:0.010789	Yes
0.019185	Shortages	Caracas:0.040046 water:0.018338 national:0.012043 municipal:0.008800 informe:0.007954 work:0.007480 transport:0.007008 santo:0.006578	metro:0.018553 services:0.015670 Vargas:0.008916 station:0.008300 Miranda:0.007838 hydrocapital:0.007179 failure:0.006609	Mostly (also includes advertising)
0.018979	Russia	Syria:0.027777 tack:0.017285 Russian:0.012889 United:0.008763 Frace:0.008277 president:0.007088 USA:0.006439 queen:0.005758	Russia:0.017495 international:0.015124 Argentina:0.010564 submarine:0.008446 Putin:0.008244 after:0.006721 chemical:0.006237	Mostly (includes miss world election)

*Continue on the next page*

p(z)	Label	Top words	Internal valid?
0.018727	Outages	electricity:0.044275 failure:0.019987 out:0.011093 Corpoelec:0.009591 affected:0.009134 Tachira:0.008877 Dominguez:0.008650 minister:0.008070	Zulia:0.021333 service:0.016423 Maracaibo:0.010084 inform:0.009560 sector:0.008918 Motta:0.008758 energy:0.008320
0.018245	Exiled opposition	Ledezma:0.040742 Venezuela:0.021408 government:0.014410 president:0.014013 metropolitan:0.013841 mayor:0.013214 Caracas:0.012902 Venezolan:0.011921 international:0.010339	Antonio:0.027935 Spain:0.018790 Spanish:0.014403 Puigdemont:0.012514 Carl:0.010896 Maria:0.009350
0.018224	Music / entertainment	new:0.014320 prize:0.005712 woman:0.005373 world:0.005027 know:0.004871 first:0.004397 music:0.004352 lic:0.004266 first:0.004119	Venezuelan:0.009153 present:0.005545 international:0.005192 here:0.004939 ano:0.004569 pub- lic:0.004115 be:0.004115
0.017594	Child mortality	years:0.039488 Venezuelan:0.021119 mother:0.009790 year:0.009221 son:0.008137 Jose:0.007089 tional:0.006506 then:0.006092	children:0.028425 died:0.010578 death:0.009289 two:0.008149 deaths:0.008097 after:0.007072 na- tional:0.006401 baby:0.006401

*Continue on the next page*

p(z)	Label	Top words		Internal valid?
0.017069	Court sentences / Lula da Silva	Lula:0.025963 court:0.018376 ex-president:0.016171 da:0.015788 Brazil:0.015525 justice:0.014857 Silva:0.014425 years:0.013726 sentence:0.013651 supreme:0.012469 prisoner:0.011982 Brazilian:0.011207 TSJ:0.009975 prison:0.009128 president:0.009089		Yes
0.015727	PDVSA	PDVSA:0.042174 Venezuela:0.031467 debt:0.023113 payment:0.020954 oil:0.017441 bond:0.016686 petroleum:0.011817 million:0.010063 state:0.008553 firm:0.008492 Venezuelan:0.007550 government:0.007102 dollar:0.006775 international:0.006668 Quevedo:0.006163		Yes
0.014823	Church / job offer (mixed)	Pope:0.038525 Francis:0.027935 holy:0.014790 search:0.014390 journalist:0.014237 journalist:0.014237 community:0.012937 social:0.012187 digital:0.011932 Venezuela:0.010226 churchi:0.008999 celebrate:0.008560 Venezuela:0.008235 international:0.007716 Vatican:0.006471 like:0.006349		Mixed (only church)
0.014736	Food policy	food:0.017743 prize:0.013776 CLAP:0.012731 product:0.010879 production:0.010024 million:0.009108 thousand:0.009091 meat:0.008608 food 0.008527 buy:0.008381 deliver:0.008323 national:0.007934 kilo:0.006853 ton:0.006778 sell:0.006158		Mixed (includes recipes)

*Continue on the next page*

p(z)	Label	Top words		Internal valid?
0.013925	Airline (opening/closing)	Venezuela:0.039317 Venezuelan:0.018155 flight:0.017685 aounce:0.014610 airline:0.012854 raza:0.011048 Aruba:0.010010 end:0.009159	airlin:0.024934 Panama:0.018075 open:0.017185 Copa:0.014332 air:0.012139 country:0.010516 suspend:0.009868	Yes
0.012826	Indigenous groups / diseases	Delta:0.023141 community:0.018882 indigenous:0.015425 almost:0.013647 Venezuela:0.010598 malnutrition:0.009221 health:0.008951 affect:0.007555	Amacur:0.019440 boy:0.016134 Wara:0.014064 denounce:0.012809 measle:0.009564 missing:0.009012 malaria:0.008236	Yes
0.012719	Education Studies	study:0.025545 national:0.012353 profesor:0.010790 University:0.010201 institute:0.009981 educated:0.009735 crisis:0.006421 month:0.006062	universal:0.023648 education:0.012114 public:0.010599 school:0.010072 schools:0.009835 class:0.006813 year:0.006140	Yes
0.011546	Mining / sport (mixed)	mining:0.019540 Arco:0.012923 Venezuelan:0.011401 world:0.007034 two:0.006945 nal:0.006496 Orinicoc:0.006385 year:0.005729	Venezuela:0.013718 gold:0.012574 win:0.007119 football:0.006970 play:0.006856 great:0.006403 years:0.005932	Yes (mixed topic)

*Continue on the next page*

p(z)	Label	Top words	Internal valid?
0.009951	Economy policy	Sundde:0.033807 prize:0.027698 se- cure:0.023546 personal:0.020700 debt:0.015170 national:0.013595 credit:0.013442 opinion:0.013152 su- perintendent:0.012685 finance:0.012598 trade:0.012375 will be:0.011353 fi- nancing:0.010902 import:0.010844 represent:0.010840	Yes
0.009760	Recession	economy:0.069400 income:0.037787 Yes economy:0.032495 petroleum:0.028953 Venezuela:0.022918 problem:0.020949 product:0.020823 recession:0.020721 money:0.020089 less:0.019971 flow:0.019570 abroad:0.019566 con- tinue:0.019026 import:0.018562 im- pact:0.017875	
0.008951	Cryptomoney	Petro:0.056755 cryptomoney:0.039489 Yes market:0.033936 would change:0.024953 Venezuelan:0.024112 economic:0.019942 service:0.016793 permit:0.015340 dol:0.015023 exchange:0.014862 well:0.013810 exterior:0.013092 libera- tion:0.012853 it:0.012633 flow:0.011332	
0.008886	Economy opinion	politics:0.044283 economy:0.041736 Yes Venezuelan:0.029005 measure- ments:0.028426 crisis:0.028222 take:0.027625 economy:0.025953 neces- sary:0.023656 opportunity:0.022387 president:0.021933 governmen- tal:0.021697 new:0.020312 deterio- ration:0.019154 ideal:0.019131 man- age:0.018914	

*Continue on the next page*

p(z)	Label	Top words	Internal valid?
0.008552	Military	forces:0.024590 armed:0.024374 na- tional:0.017667 account:0.016883 (in- military:0.016291 Banesco:0.015305 cludes Venezuela:0.014304 crisis:0.012803 in- some tervencion:0.012259 official:0.012019 in other stitution:0.011701 government:0.009945 head- also:0.009526 FANB:0.009493 bar- lines) racks:0.009180	Mostly
0.007179	Work opinion	work:0.041266 company:0.034561 Yes benefits:0.031346 can:0.030131 (about Venezuelan:0.024452 open:0.022773 distant cost:0.020917 infrastructure:0.018911 work) result:0.016888 reduce:0.015983 better:0.015616 labor:0.015438 point:0.015352 well:0.014166 man- age:0.014092	
0.007131	Cuba	Diaz:0.043007 Cuba:0.036575Luis:0.035531 Cas- tro:0.028181 Cuban:0.016626 Or- tega:0.016372 president:0.016216 Raul:0.016193 channel:0.013845 Vicent:0.012640 children:0.011630 Miguel:0.011417 way:0.011174 Fi- dell:0.011105 Marquez:0.009449	Yes
0.005473	Weather	Inameh:0.062016 country:0.057035 Yes national:0.046779 rain:0.039776 precipitate:0.037470 part:0.036974 great:0.035524 cloud:0.033585 weather:0.025155 forecast:0.022413 Venezuela:0.021691 dispersion:0.021248 after:0.020752 predict:0.018987 re- gion:0.018153	

*Continue on the next page*

p(z)	Label	Top words	Internal valid?
0.005210	Leftist opposition	program:0.031889      oligarchy:0.024329 horror:0.024171      alone:0.018900      gov- ernment:0.018821      pact:0.016619 fight:0.014528      saman:0.014252      elec- toral:0.014181      occupy:0.013920      to- wards:0.013589      direct:0.012989      con- trol:0.012295      travel:0.011222      sen- tence:0.010827	Yes
0.004710	Regulations	law:0.117708      sport:0.083348      Yes vem:0.046064      fond:0.045505 right:0.035748      embargo:0.028618 project:0.027029      treaty:0.026523 can:0.026305      organize:0.026287 company:0.025161      tax:0.024812      re- spect:0.024795      impose:0.024612      regula- tion:0.022936	
0.004517	Investment / infrastructure	society:0.042481      Venezuela:0.034329      Yes will      be:0.033000      corporate:0.029470 every:0.029416      stay:0.029361      lim- ited:0.022601      time:0.022328      Cara- cas:0.019580      pass:0.018725      of- fice:0.018343      opinion:0.018224 square:0.017406      comunnity:0.017378 build:0.017333	
0.004514	Investment	group:0.068389      intention:0.054658      Yes person:0.043887      service:0.042439      ac- quire:0.039626      condition:0.036157 achieve:0.035501      debt:0.035091 basic:0.034336      product:0.033880 a      few:0.032050      certain:0.029710 meeting:0.028417      time:0.028390      fi- nance:0.026633	

*Continue on the next page*

p(z)	Label	Top words	Internal valid?
0.004283	Spanish development aid	Spanish:0.081387 achieve:0.044364 technical:0.041265 Miraflores:0.041063 organization:0.041006 last:0.040065 cystic:0.039154 similar:0.038712 could:0.038319 circle:0.038204 word:0.037091 opinion:0.036573 try:0.035057 result:0.032908	Yes
0.004234	Financial market	circulation:0.050091 ticket:0.046529 was:0.041919 Venezuelan:0.040599 total:0.040463 recent:0.039386 traditional:0.039299 oscillation:0.038318 liquidity:0.038027 quantity:0.037590 measurement:0.036901 continue:0.036765 descendent:0.036688 opinion:0.036290 yet:0.036018	Yes

Table E.2: Generated Topics (K=50). Note: Words are translated and unstemmed.  $P(z)$  shows the distribution of the topics over the whole corpus of 123,889 (123 574) headlines. Probabilities after words display the likelihood of a specific word appearing in the respective topic. Top 15 words for each topic are displayed. To further check the semantic validity, I investigated whether the top 30 headlines per category correspond to the identified topic (see last column and replication files).

p(z)	Label	Top words
0.079420	Maduro	madur:0.015353 venezuel:0.011715 gobiern:0.007832 venezolan:0.007441 si:0.007168 polit:0.006651 pais:0.006362 president:0.006061 sol:0.006018 nicol:0.005448

*Continue on the next page*

p(z)	Label	Top words	
0.072853	National assembly	nacional:0.034377 venezuel:0.020858 madur:0.012585 gobiern:0.010190 derech:0.008179	asamble:0.021469 diput:0.014546 president:0.012233 venezolan:0.008637 luis:0.008086
0.072600	Election	eleccion:0.031827 electoral:0.024757 falcon:0.013889 part:0.011961 henri:0.009997	presidencial:0.026818 candidat:0.015666 nacional:0.013662 cne:0.011839 vot:0.010245
0.056434	Venezuela crisis	venezuel:0.022298 venezolan:0.017166 social:0.008796 busc:0.007480	pais:0.018288 crisis:0.016369 econom:0.008700 period:0.007145 na- cional:0.006426 comun:0.006372
0.055839	Protest / shortages	protest:0.011668 san:0.009214 carac:0.006122 falt:0.005786 tachir:0.005255	nacional:0.010548 municipi:0.007373 com:0.006047 denunci:0.005287 clap:0.005213
0.054978	Political prisoners / corruption (mixed)	/ deten:0.011021 fiscal:0.010656 nunci:0.009013	nacional:0.012884 pres:0.011297 general:0.010812 polit:0.010234 saab:0.008222 pe- nal:0.008128 inform:0.007873
0.053011	Government- opposition dialog	gobiern:0.031470 venezuel:0.024858 venezolan:0.014016 rodriguez:0.012235 republ:0.011602	dialog:0.028227 oposicion:0.018906 dominican:0.012328 acuerd:0.012049 president:0.011427
0.050881	Lima summit	madur:0.048480 nicol:0.024117 amer:0.013052 peru:0.010864 nuev:0.007887	president:0.048141 venezuel:0.018958 cumbr:0.011171 pais:0.008163 cub:0.007640

*Continue on the next page*

p(z)	Label	Top words	
0.046849	International	trump:0.030378 eeuu:0.019987 ald:0.014716 core:0.010330 dounidens:0.008925	unid:0.020193 president:0.017360 internacional:0.010707 siri:0.009236 rusi:0.008639
0.046246	Electricity / transportation (mixed)	electr:0.019691 vici:0.015214 trabaj:0.013992 cional:0.010044 zuli:0.008331	carac:0.017878 transport:0.014657 fall:0.011721 agu:0.009924 inform:0.007629
0.045831	Cryptomoney	venezuel:0.017274 banc:0.013845 petr:0.012055 ser:0.009946 dolar:0.009414	millon:0.015026 pag:0.012227 bolivar:0.010272 nuev:0.009860 criptomoned:0.009399
0.042986	Oscar Perez / crime (mixed)	anos:0.018961 perez:0.014209 venezolan:0.011470 muert:0.008305 nin:0.007411	asesin:0.014375 oscar:0.012655 dos:0.010226 cuerp:0.007612
0.042433	Border colombia / airline (mixed)	venezolan:0.041813 colombi:0.025387 fronter:0.008809 internacional:0.008599 oper:0.007977	venezuel:0.029001 pais:0.012944 colombian:0.008618 airlin:0.008195 sant:0.007390
0.040694	Entertainment / sport	venezolan:0.012949 venezuel:0.007257 ano:0.006221 dia:0.005572 premi:0.004595	nuev:0.007546 anos:0.006350 mund:0.005819 celebr:0.004839 internacional:0.004187
0.039588	Accident / crime	muert:0.027149 her:0.019255 internacional:0.012456 tras:0.011412 mur:0.008952	person:0.020144 men:0.018134 dej:0.014590 dos:0.011956 ataqu:0.009300

*Continue on the next page*

p(z)	Label	Top words	
0.034902	Sanctions / Exiled opposition	venezuel:0.035857 sancion:0.024881 ledez:0.017487 madur:0.014967 unid:0.014545	venezolan:0.025098 polit:0.020451 gobiern:0.016127 eeuu:0.014627 med:0.013407
0.031581	Petroleum / PDVSA	economi:0.024733 petrole:0.021327 petroler:0.018159 duccion:0.016063 millon:0.010307	pdvsa:0.023938 venezuel:0.020809 econom:0.017750 venezolan:0.012232 dolar:0.010142
0.030265	Healthcare / indigenous groups (mixed)	pacient:0.017276 falt:0.014856 medic:0.010417 nin:0.009707 lul:0.016735	hospital:0.016628 salud:0.014053 denunci:0.010216 medicament:0.008791 expresident:0.013682
0.027576	Court sentences	president:0.013223 justici:0.011018 da:0.010680 silv:0.009785	tribunal:0.011833 brasil:0.010788 internacional:0.010241 puigdemont:0.009403
0.020850	Prices	economi:0.031018 product:0.020256 fluj:0.014106 merc:0.013454 sundd:0.013363	preci:0.026914 moned:0.015374 aument:0.013981 salari:0.013424 ingres:0.013197
0.014110	Church / Sport (mixed)	ley:0.035633 port:0.028841 vem:0.015481 fond:0.014608 trat:0.010121	pap:0.035589 francisc:0.025013 derech:0.014921 venezuel:0.011986 embarg:0.009946
0.013142	Financial opinion	deb:0.026109 grup:0.024117 intencion:0.019462 person:0.018366 gur:0.017395	pued:0.024622 product:0.020311 servici:0.019393 finanz:0.017752 cumpl:0.017285

Continue on the next page

$p(z)$	Label	Top words	
0.012554	Weather / Earthquake (mixed)	pais:0.029430 sism:0.024840 lluvi:0.020678 part:0.017504 gran:0.016648	inameh:0.027056 nacional:0.023426 magnitud:0.019721 venezuel:0.017271 precipit:0.016415
0.009417	Inflation	billet:0.036833 venezolan:0.025601 cion:0.023447 total:0.020851 trecient:0.020536 monetari:0.020108 aun:0.019430 espanol:0.075121 organizacion:0.039371 ultim:0.037258 pod:0.035750 parec:0.034822	circulacion:0.028586 hiperinflacion:0.021934 habi:0.020536 mone:0.018731 polit:0.046258 logr:0.038758 tecnic:0.037250 miraflor:0.035750 palabr:0.034615
0.004961	Financial market		

Table E.3: Generated Topics ( $K=25$ ). Note: Words are not translated and only stems are displayed.  $P(z)$  shows the distribution of the topics over the whole corpus of 123,889 headlines. Probabilities after words display the likelihood of a specific word appearing in the respective topic. Top 10 words for each topic are displayed.

p(z)	Label	Top words
0.039521	Venezuela	venezuel:0.011246 pais:0.009467 polit:0.008965 madur:0.008425 ser:0.006103 venezolan:0.005884 si:0.005500 pas:0.005333 pod:0.005256 anos:0.005061
0.039510	Crisis economy	venezuel:0.024993 venezolan:0.017998 madur:0.016847 asegur:0.015867 pais:0.015593 gobiern:0.014904 deb:0.010713 polit:0.010574 presi- dent:0.010235 nacional:0.008999
0.027087	Election (general)	eleccion:0.063479 presidencial:0.042912 venezuel:0.025052 electoral:0.019293 vot:0.015919 madur:0.014580 presi- dent:0.013670 comici:0.012022 vene- zolan:0.011346 proxim:0.010210
0.025517	Maduro	madur:0.112147 president:0.063265 nicol:0.058887 gobiern:0.015828 anunci:0.013820 venezuel:0.013817 republ:0.013465 venezolan:0.013005 nacional:0.011495 nuev:0.010287
0.024949	National assembly	nacional:0.070420 asamble:0.050943 diput:0.034085 president:0.019754 an:0.018184 constituyent:0.012861 comision:0.011987 polit:0.009855 anc:0.009600 nuev:0.009036
0.023068	Sanctions	sancion:0.041392 venezuel:0.034480 gobiern:0.030520 eeuu:0.026573 unid:0.024340 venezolan:0.019695 europe:0.018256 union:0.014794 madur:0.013463 ue:0.012964
0.022643	Migration crisis	venezolan:0.082273 colombi:0.019235 pais:0.019217 brasil:0.010298 mil:0.009712 venezuel:0.009115 mi- gracion:0.008053 fronter:0.007928 refugi:0.007731 migrant:0.007457

Continue on the next page

p(z)	Label		Top words
0.019803	Opposition dates	candi-	falcon:0.050927 henri:0.036621 pres- idencial:0.033537 candidat:0.032173 candidatur:0.017465 madur:0.017228 eleccion:0.013922 electoral:0.012042 part:0.011275 oposicion:0.009925
0.019589	Government- opposition dialog		dialog:0.058814 gobiern:0.048190 oposi- cion:0.040616 dominican:0.030076 republ:0.021649 acuerd:0.016642 vene- zolan:0.016009 mes:0.014300 negocia- cion:0.011706 reunion:0.011555
0.018125	Election (procedure)		electoral:0.056043 cne:0.037798 na- cional:0.030492 consej:0.027514 elec- cion:0.025455 part:0.015853 pres- idencial:0.012211 proces:0.012029 rector:0.010048 municipal:0.009989
0.017358	Accidents / deaths		muert:0.040736 her:0.032104 per- son:0.031024 men:0.025398 dej:0.024484 accident:0.015471 tras:0.015079 dos:0.014161 internacional:0.012508 result:0.011303
0.017170	Ministers meetings		venezuel:0.048255 ministr:0.020004 cancill:0.018796 pais:0.018654 ar- reaz:0.017493 relacion:0.016820 ex- terior:0.013371 jorg:0.013170 presi- dent:0.012338 reunion:0.011156
0.017159	Security Forces		funcionari:0.020229 deten:0.017995 nacional:0.017241 dos:0.016821 polici:0.015998 presunt:0.011103 bolivarian:0.011048 rob:0.010004 tres:0.009003 gnb:0.008322
0.016973	Lima summit		amer:0.041863 venezuel:0.038382 cumbr:0.034537 madur:0.025896 president:0.023184 lim:0.019703 peru:0.018553 pais:0.013844 eeuu:0.010189 grup:0.009684

*Continue on the next page*

p(z)	Label	Top words
0.016819	Inflation 1	venezuel:0.028376 econom:0.026383 economi:0.020831 ano:0.019394 inflacion:0.017886 pais:0.012255 segun:0.011650 crisis:0.010703 in- dic:0.010603 preci:0.008803
0.016546	Political prisoners	polit:0.036993 pres:0.036077 pe- nal:0.019577 for:0.016874 liber:0.015498 libert:0.013171 denunci:0.012552 na- cional:0.012519 deten:0.011925 vene- zolan:0.010120
0.016196	Protest / shortages	protest:0.040749 agu:0.020930 san:0.020351 falt:0.014297 vecin:0.013789 habit:0.012142 sec- tor:0.010995 municipi:0.010962 ser- vici:0.009948 exig:0.009635
0.015207	Shortages	electr:0.055677 fall:0.024223 ser- vici:0.024055 agu:0.017029 zuli:0.016079 carac:0.013916 apagon:0.013534 sis- tem:0.011804 inform:0.011782 corpo- elec:0.011344
0.015075	Crisis	venezuel:0.037642 crisis:0.034431 pais:0.028876 venezolan:0.022676 hu- manitari:0.018408 situacion:0.016924 econom:0.012995 salud:0.011427 emer- gent:0.010455 internacional:0.007950
0.014888	Crime / deaths	asesin:0.028809 anos:0.024711 vene- zolan:0.018506 hombr:0.013348 mat:0.012613 jov:0.010511 muj:0.010044 hall:0.009928 dos:0.009350 rob:0.008266
0.014071	Shortages healthcare	pacient:0.035797 hospital:0.030630 falt:0.023351 medic:0.020176 in- sum:0.017836 salud:0.017207 medica- ment:0.015568 protest:0.013936 medicin:0.011534 denunci:0.011010

Continue on the next page

p(z)	Label	Top words	
0.013895	USA / Korea	trump:0.065755 core:0.032521 nort:0.025294 jong:0.017466 ledez:0.061761 venezuel:0.026591 metropolitan:0.019856 zolan:0.018610 madur:0.014400 colombi:0.065391 sant:0.022480 fronter:0.016881 president:0.014302 paz:0.011723 ciud:0.016612 liv:0.013453 report:0.011350 puert:0.008981 ciut:0.035583 subast:0.024052 cambi:0.019799 derech:0.043942 human:0.032367 subast:0.024052 cambi:0.019799 derech:0.043942 human:0.032367 organizacion:0.011296 nacion:0.010732 luis:0.036148 diaz:0.025146 oea:0.022763 madur:0.020780 almagr:0.020769 nacion:0.017310	donald:0.033339 president:0.032067 kim:0.021528 unid:0.016693 estadounidens:0.015055 eeuu:0.013846 antonio:0.043044 carac:0.021391 vene:alcald:0.017413 espan:0.013081 go-biern:0.011335 colombian:0.029758 venezuel:0.019069 eln:0.015251 farc:0.012983 manuel:0.011583 centr:0.016073 saque:0.013243 protest:0.009080 tras:0.008683 person:0.007750 regional:0.007611 criptomoned:0.026084 divis:0.025836 dolar:0.020102 bolivar:0.013659 tas:0.013112 venezuel:0.037971 comision:0.016747 cidh:0.011129 nacion:0.010683 orteg:0.036145 general:0.024115 venezuel:0.022301 almagr:0.020769 nicaragu:0.016350
0.013514	Exiled opposition		
0.013264	Colombian politics		
0.012132	Protest / riots		
0.011973	Cryptomoney / exchange rate		
0.011825	International organizations		
0.011661	Regional organizations		

*Continue on the next page*

p(z)	Label	Top words	
0.011196	Corruption	fiscal:0.043043	saab:0.040293
		general:0.030793	tarek:0.029433
		william:0.027404	republ:0.018959
		deten:0.013103	nacional:0.011525
		inform:0.011158	corrupcion:0.010959
0.010905	Transportation	transport:0.071232	public:0.024469
		pasaj:0.023680	aument:0.012070
		carac:0.011836	falt:0.009655
		unidad:0.008953	servici:0.008383
		par:0.008073	maracaib:0.008058
0.010796	Airline (opening / closing)	venezuel:0.043501	airlin:0.031911
		vuel:0.021430	panam:0.020527
		venezolan:0.020508	oper:0.019966
		cop:0.018174	anunci:0.016135
		aero-line:0.016051	aere:0.014953
0.010796	Petroleum production	petrole:0.046322	produccion:0.029486
		barril:0.026257	dolar:0.025834
		venezuel:0.025437	opep:0.022189
		millon:0.019928	venezolan:0.018274
		crud:0.018266	petroler:0.015563
0.010693	Protest / education	trabaj:0.048218	nacional:0.014950
		colegi:0.011082	pag:0.010928
		ucacion:0.010727	protest:0.010516
		anunci:0.009773	educ:0.009673
		institu:0.009500	in-
			stitu:0.008980
0.010355	Oscar Perez / repression	perez:0.071974	oscar:0.068265
		cuerp:0.020557	junquit:0.016622
		cicpc:0.011352	familiar:0.010982
		oper:0.010167	investig:0.010072
		nacional:0.009861	na-
			cional:0.009313
0.009982	Russia	siri:0.028676	president:0.024078
		eeuu:0.017511	unid:0.017230
		rusi:0.015878	iran:0.013688
		internacional:0.013396	inter-
			nacional:0.013322
		acuerd:0.012922	putin:0.012860
			franci:0.012860

*Continue on the next page*

p(z)	Label	Top words
0.009941	Child mortality	nin:0.061709      anos:0.033434      desnutricion:0.013310      edad:0.011592 beb:0.011430      muri:0.010569 muert:0.010482      falleci:0.010449      hospital:0.010412      madr:0.010105
0.009915	Peru / resignations	president:0.031834      peru:0.029464 kuczynski:0.027455      tribunal:0.024258 tsj:0.022474      pedr:0.018991 suprem:0.018248      pabl:0.018223      justici:0.018202      renunci:0.015583
0.009596	Music / Royal wedding	venezolan:0.016000      premi:0.013093 nuev:0.010528      internacional:0.009465 princip:0.008495      anos:0.007895 megh:0.006385      music:0.006313 present:0.005849      asi:0.005584
0.009569	Municipal candidates	candidat:0.026760      alcaldi:0.026441 municipi:0.026007      alcald:0.018214 social:0.013096      carac:0.011956 barut:0.010559      libert:0.010408 gomez:0.010352      gobern:0.010038
0.009553	Salary	bolivar:0.045225      millon:0.040482 aument:0.038436      salari:0.034011 bs:0.029510      minim:0.024452 mil:0.019084      preci:0.018395 canast:0.013424      madur:0.011089
0.009549	Lula da Silva	lul:0.047289      brasil:0.029670      da:0.028545 silv:0.026321      expresident:0.024144 brasilen:0.019498      prision:0.016279 inaci:0.015693      luiz:0.015343      condens:0.014180
0.009545	Opposition Frente Amplio	frent:0.025002      gobiern:0.022010 popul:0.021489      ampli:0.020661 oposicion:0.020273      luch:0.019604 eleccion:0.018393      venezuel:0.014875 movimient:0.012650      particip:0.012568

*Continue on the next page*

p(z)	Label	Top words	
0.009374	Puigdemot / Mugabe (mixed)	president:0.037797 mont:0.028187 gobiern:0.019201 mugab:0.017669 carl:0.016128 catal:0.014513	puigde- espanol:0.020189 espan:0.018025 zimbabu:0.016598 independent:0.014732
0.009211	Military 1	lopez:0.027120 padrin:0.018957 nac- ional:0.018104 militar:0.016755 armad:0.014869	ministr:0.024113 milit:0.018479 defens:0.017389 fuerz:0.015664 reverol:0.011340
0.009098	Church	pap:0.067079 sant:0.013270 icani- an:0.011194 chil:0.009987 venezolan:	francisc:0.047852 iglesi:0.013175 vat- obisp:0.010240 internacional:0.009566 pid:0.009286
0.008654	Syria	ataqu:0.039876 muert:0.021759 men:0.017670 bombarde:0.013600 grup:0.012140	siri:0.030528 atent:0.020309 internacional:0.014579 mur:0.013447 terror:0.011693
0.008591	Licenses	banc:0.030396 carnet:0.014269 nuev:0.011448 bon:0.010159	patri:0.017746 pag:0.013584 venezuel:0.010825 sistem:0.009757 document:0.009647
0.008527	Corruption 2	rafael:0.039066 rupcion:0.023210 odebrecht:0.014182 venezuel:0.012864 investig:0.008864	ramirez:0.038115 pdvs:0.017849 cas:0.013598 andorr:0.011827 ex:0.008859
0.008488	Sport	venezolan:0.016303 gan:0.009594 venezuel:0.009109 mundial:0.008571 madr:0.007674	final:0.010718 jueg:0.009235 futbol:0.008643 segund:0.008202 lig:0.006069

*Continue on the next page*

p(z)	Label	Top words
0.008128	Women / birth	si:0.014047 muj:0.012290 pued:0.010539 hac:0.010296 person:0.009218 deb:0.008454 sol:0.006145 eme:0.006019 sangr:0.005538 mejor:0.005361
0.008102	Exchange rate 1	monetari:0.047876 nuev:0.045784 re-conversion:0.029382 billet:0.026447 moned:0.022928 con:0.019398 boliv:0.019073 anuncio:0.013213 bcv:0.012970 president:0.012761
0.008045	South / Latin America	argentin:0.047878 president:0.027194 chil:0.024571 submarin:0.021717 juan:0.017236 piner:0.016679 hon-dur:0.016536 macri:0.013400 interna-cional:0.011283 san:0.011094
0.007800	Tourism / security	varg:0.017407 segur:0.015207 na-cional:0.012095 activ:0.011314 pais:0.011188 sant:0.011061 tur-ist:0.010813 carnaval:0.010449 mar:0.009880 civil:0.008571
0.007539	Pension	pension:0.034534 pag:0.024673 col:0.019794 efect:0.017824 cobr:0.017611 gasolin:0.016563 tachir:0.015761 banc:0.012978 ban-cari:0.011695 sudeb:0.011166
0.007501	Government nouncement	an-rodriguez:0.088261 jorg:0.033296 ministr:0.023303 torr:0.018378 delcy:0.018219 nacional:0.015675 venezuel:0.014265 miguel:0.014090 president:0.013048 gobiern:0.011128
0.007388	Petroleum	pdvsa:0.073377 petroler:0.027401 empres:0.022523 petrole:0.022406 queved:0.014672 gas:0.013096 mil-lon:0.012517 industri:0.012334 cono-cophillips:0.011170 venezuel:0.010490

*Continue on the next page*

p(z)	Label	Top words
0.007360	Internet	red:0.024941 social:0.018765 em- pres:0.010565 public:0.010375 usuari:0.010134 cuent:0.009665 crip- tomoned:0.009206 trav:0.009162 per- son:0.007764 telecomun:0.007306
0.007289	Prices 1	preci:0.022147 compr:0.021464 com:0.014042 product:0.013010 carn:0.011452 venezolan:0.010250 vid:0.008980 efect:0.008472 caf:0.008381 hiperinflacion:0.007682
0.007228	Government ministers	cabell:0.063287 vicepresident:0.048255 diosd:0.042296 part:0.024304 psuv:0.023667 unid:0.021361 prim:0.019233 social:0.017525 ais- sami:0.015721 venezuel:0.013985
0.007180	Prices 2	sundd:0.048453 preci:0.046854 com- erci:0.018975 nacional:0.018758 super- merc:0.014467 superintendent:0.014272 product:0.013653 orden:0.012438 derech:0.012393 defens:0.012238
0.007120	US school shooting/Trump (mixed)	escuel:0.027226 flor:0.025805 tirote:0.023707 eeuu:0.017750 nuev:0.013631 sexual:0.011897 mi- ami:0.011331 york:0.010268 vene- zolan:0.009280 estudi:0.009112
0.007020	Israel	jerusal:0.039353 embaj:0.026655 israel:0.026251 palestin:0.022208 eeuu:0.021927 trump:0.019682 unid:0.019085 capital:0.018417 re- conoc:0.016325 israeli:0.013923
0.006963	CLAP	clap:0.028395 produccion:0.026031 aliment:0.024448 entreg:0.014971 tonel:0.014162 nacional:0.012992 millon:0.011875 bernal:0.011149 freddy:0.011072 gobiern:0.010960

Continue on the next page

p(z)	Label	Top words
0.006609	Police	carabob:0.029581 motin:0.018074 policarabob:0.017370 carcel:0.015254 pres:0.014732 muert:0.014532 pen- itenciari:0.012510 comand:0.011993 lacav:0.011383 polici:0.010661
0.006526	Christmas	navid:0.026825 celebr:0.014856 tambi:0.012341 nin:0.011579 dia:0.011446 madr:0.011377 tradic- cion:0.010747 naviden:0.010431 vene- zolan:0.009436 regal:0.009360
0.006276	Recession	economi:0.059044 ingres:0.033189 product:0.032415 petroler:0.032022 problem:0.031767 recesion:0.031701 moned:0.030633 menor:0.030626 ex- tranjer:0.030357 venezuel:0.030272
0.006250	Government	juan:0.043546 president:0.023613 pabl:0.022869 conindustri:0.017775 capril:0.015722 carl:0.014478 go- biern:0.012458 guanip:0.011602 olalquiag:0.010707 gremi:0.009904
0.006156	Education system	univers:0.046346 estudi:0.045464 uni- versitari:0.018794 profesor:0.014424 na- cional:0.013449 public:0.012360 cien- tif:0.012193 fundacion:0.011785 interna- cional:0.011317 relev:0.009995
0.006124	Indigenous groups	delt:0.050959 amacur:0.042873 wara:0.029610 comun:0.027964 de- nunci:0.022276 indigen:0.017186 falt:0.016185 tucupit:0.014788 mu- nicipi:0.012753 sarampion:0.010618
0.005961	Bonus payments	deud:0.052211 pag:0.039248 bon:0.038986 venezuel:0.038365 pdvsa:0.026830 default:0.014812 in- cumpl:0.012646 amp:0.012391 mil- lon:0.011639 calificacion:0.011369

*Continue on the next page*

p(z)	Label	Top words	
0.005951	Joshua Holt	venezuel:0.037401 holt:0.027587 joshu:0.021057 marc:0.020684 senador:0.017249	eeuu:0.031658 estadounidens:0.024789 unid:0.020967 rubi:0.019841 trump:0.014485
0.005878	Drug trade	anos:0.030671 trafic:0.021716 conden:0.015874 venezuel:0.014902 brin:0.014230	cocain:0.023570 dos:0.017672 flor:0.015224 drog:0.014384 unid:0.013740
0.005389	Cuba	cub:0.050715 diaz:0.024438 anos:0.020059 president:0.016763 cuban:0.014993	castr:0.041002 raul:0.021890 canel:0.017961 fidel:0.016152 miguel:0.012414
0.005132	Opposition campaigns / holidays (mixed)	pastor:0.020349 sant:0.013083 campan:0.012113 bertucci:0.011753 dia:0.010896	divin:0.013692 sam:0.012490 javi:0.011857 carac:0.011601 realiz:0.009085
0.004934	Mining	miner:0.055085 orinoc:0.021611 oro:0.019453 iern:0.014305 curs:0.012113	arco:0.035499 venezuel:0.020936 sur:0.014479 reserv:0.012605 conoc:0.011297
0.004918	Diseases	cas:0.031026 salud:0.020528 digen:0.017001 venezuel:0.014151 malari:0.012738	palud:0.020770 sierr:0.017945 perij:0.016257 comun:0.013958 brot:0.012646
0.004916	Earthquake	sism:0.073383 registr:0.032870 sacudi:0.017618 inform:0.013638 funvisis:0.013295	magnitud:0.056492 sacud:0.020093 terremot:0.016615 internacional:0.013580 escal:0.013111

*Continue on the next page*

p(z)	Label	Top words
0.004762	Weather	inameh:0.071270 pais:0.062034 nacional:0.046377 lluvi:0.043891 precipit:0.042268 part:0.040645 gran:0.038306 nub:0.037978 tiemp:0.028069 prev:0.024988
0.004752	Job offer	period:0.078783 busc:0.049933 so- cial:0.045374 comun:0.038462 digi- tal:0.037173 si:0.021921 gust:0.020061 trabaj:0.018072 venezuel:0.017734 globovision:0.017535
0.004659	Interview	chavez:0.023916 entrev:0.021040 pin:0.016559 hug:0.016276 viv:0.015862 president:0.015103 anos:0.014891 martinez:0.014794 eulogi:0.013824 nacional:0.012659
0.004597	Financial products	grup:0.058372 intencion:0.051693 cumpl:0.043073 person:0.034569 deb:0.034399 servici:0.034310 condic- cion:0.033514 vez:0.032307 ba- sic:0.032253 merc:0.031753
0.004509	Caracas metro	carac:0.084729 metr:0.074311 es- tan:0.023807 oficin:0.023379 mu- nicipi:0.022850 chaca:0.022705 esta- cion:0.022048 opinion:0.019250 con- stru:0.018548 merced:0.018548
0.004422	Opposition	jos:0.044893 luis:0.034613 vi- cent:0.032633 ram:0.026062 leon:0.023125 allup:0.022112 presi- dent:0.018636 general:0.014416 demo- crat:0.013226 velasquez:0.012920
0.004383	Politics / economy opinion	polit:0.069490 tom:0.045392 econom:0.042936 oportun:0.042795 med:0.042167 gubernamental:0.039335 ideal:0.038294 desperdici:0.037535 economi:0.037010 venezolan:0.036897

*Continue on the next page*

p(z)	Label	Top words		
0.004357	Russia-UK / diplomat expulsion	rus:0.054262 mat:0.022561 tan:0.016921 skripal:0.015544 mar:0.013969	rusi:0.036445 exespi:0.017355 envenen:0.016761 putin:0.013997 unid:0.013629	diplo- bri- venen skripal putin
0.004285	Mendoza	banesc:0.043987 banc:0.023984 lorenz:0.020263 intervencion:0.018402 carl:0.013119	venezuel:0.031905 president:0.023821 mendoz:0.019975 escotet:0.013349 direct:0.011450	
0.004006	Business opening	pued:0.049812 trabaj:0.039567 empres:0.029393 venezuel:0.028768	benefici:0.045208 cost:0.032131 infraestructur:0.031855 reduc:0.029014 result:0.027824	in- fraestruc oper result
0.003760	Spanish development aid	espanol:0.089523 logr:0.049072 enquist:0.043926 mari:0.047294 mach:0.031258 ordin:0.023932 lic:0.018056 necesit:0.009896	polit:0.049619 tecnic:0.046406 organizacion:0.044658 parec:0.043762 corrill:0.043511 ultim:0.043227 venezuel:0.034389 corin:0.028820 vent:0.019363 public:0.011543 internacional:0.009479 nacional:0.009479	mi- raflor enquist parec corrill ultim venezuel corin vent public internaci nacional
0.003740	Unclear			
0.003608	Regulations	ley:0.135799 vem:0.059998 derech:0.040200 trat:0.031957 part:0.029703	deport:0.105959 fond:0.051266 embarg:0.033563 respect:0.030079 organ:0.029396	
0.003532	Venezuela / Ecuador	ecuador:0.029007 pecial:0.017574 hac:0.015783 peligr:0.013910 part:0.013457	asi:0.017597 capac:0.017423 corre:0.014783 venezolan:0.013875 lent:0.013457	es- pecial hac peligr part

*Continue on the next page*

p(z)	Label	Top words
0.003250	Inflation 2	billet:0.054713      circulacion:0.051555 total:0.050961      habi:0.050898      tradicional:0.049850      oscil:0.049471      liquidez:0.049041      recient:0.048536      cantidad:0.047854      medicion:0.047841
0.003158	Exchange rate 2	economi:0.087529      merc:0.065191      cambiari:0.061031      venezolan:0.035898      intercambi:0.032570      bien:0.031738      permit:0.031348      fluj:0.030867      dol:0.030256      liberacion:0.030073
0.003136	Company	venezuel:0.067945      sociedad:0.060823 ser:0.050088                corpor:0.042770 pas:0.032755                comandit:0.032519 vez:0.031223                cad:0.029835      co- mun:0.027531 firm:0.025869
0.002753	Oligarchy opinion	program:0.050508      horror:0.045887 oligarqui:0.045663      pact:0.031635 ocup:0.031024            sol:0.029354      cuartel:0.024017      haci:0.023257      elec- toral:0.023212 direct:0.021989
0.002672	Diplomat expulsion	embaj:0.062819      venezuel:0.037707 negoci:0.028891            encarg:0.028046 nuev:0.027001            grat:0.024713      miniaturiz:0.023223      inalambr:0.022639 marcapas:0.021933 robinson:0.021395
0.002633	Financial products	segur:0.089919            ser:0.052883 deb:0.045700            personal:0.036741      im- port:0.035541            proteccion:0.034559 utiliz:0.032752            inteligent:0.032191 financ:0.031817 finanz:0.031521
0.002541	Petroleum export	ingres:0.067402      economi:0.043460 sensatez:0.043331      divis:0.037438      ex- port:0.036873            represent:0.035646 petroler:0.034597 hidrocarbur:0.033693 asoci:0.033047 approxim:0.032208

*Continue on the next page*

p(z)	Label	Top words	
0.002511	Military 2	armad:0.039584	fuerz:0.039323
		cuent:0.033278	crisis:0.029423 insti-
		tucion:0.026596	nacional:0.021712
		oficial:0.021074	cuartel:0.020062
		fanb:0.019882	solicitud:0.019653
0.002420	Healthcare	llev:0.036435	dia:0.033757 cad:0.030114
		cab:0.029283	estan:0.026131
		riesg:0.020386	centr:0.020370
		men:0.020031	especi:0.019844
		sol:0.018065	

Table E.4: Generated Topics (K=100). Note: Words are not translated and only stems are displayed.  $P(z)$  shows the distribution of the topics over the whole corpus of 123,889 headlines. Probabilities after words display the likelihood of a specific word appearing in the respective topic. Top 10 words for each topic are displayed.

---

## References

- Casey, Nicholas. 2018. “Venezuela’s Most-wanted Rebel Shared His Story, just before Being Gunned Down.” *The Independent*, February 02. [https://www.independent.co.uk/news/long\\_reads/oscar-perez-death-venezuela-helicopter-attack-caracas-muerte-suspendida-a8177051.html](https://www.independent.co.uk/news/long_reads/oscar-perez-death-venezuela-helicopter-attack-caracas-muerte-suspendida-a8177051.html) (accessed 07-29-2019).
- Grimmer, Justin and Brandon M Stewart. 2013. “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts.” *Political Analysis* 21(3):267–297.
- Quinn, Kevin M, Burt L Monroe, Michael Colaresi, Michael H Crespin and Dragomir R Radev. 2010. “How to Analyze Political Attention with Minimal Assumptions and Costs.” *American Journal of Political Science* 54(1):209–228.
- Rueda, Manuel. 2015. “Coordinated DDoS Attack During Russian Duma Elections.” *Splinter*, June 23, note = <https://splinternews.com/meet-the-venezuelan-rebel-whose-crime-is-publishing-exc-1793848615> (accessed 03-24-2020).
- van Rijswijk-Deij, Roland, Mattijs Jonker, Anna Sperotto and Aiko Pras. 2016. “A High-Performance, Scalable Infrastructure for Large-Scale Active DNS Measurements.” *IEEE Journal on Selected Areas in Communications* 34(6):1877–1888.
- Yan, Xiaohui, Jiafeng Guo, Yanyan Lan and Xueqi Cheng. 2013. A Biterm Topic Model for Short Texts. In *Proceedings of the 22Nd International Conference on World Wide Web*. WWW ’13 New York, NY, USA: ACM pp. 1445–1456.