

Climate Science Communication on Twitter: A Topic Modeling Analysis of U.S. Federal Government Agencies

Nic DePaula

Wayne State University, Detroit MI 48202, USA
ndepaula@wayne.edu

Abstract. Government agencies have a strong role in communicating science and climate change information to the public. However, the characteristics, factors and implications of this behavior have been scarcely examined in the academic literature. In this brief study I address two research questions. First: Can topic modeling analysis via LDA provide valid topics from the Twitter posts of U.S. federal government science agencies? Secondly: How and to what extent is climate change and climate science information being communicated by these agencies under a political administration hostile to climate change science? This study contributes to the literature on the use of topic modeling analysis of social media information in the government context, and to the literature on the politicization of science communication and government communication.

Keywords: Science Communication, Climate Change, Topic Modeling, Social Media, E-government, Government Communication, Politicization of Science.

1 Introduction

Climate change is one of the most important issues of our times, and climate change communication a potentially necessary and useful component for developing sustainable communities around the world [1]. Moreover, social media sites are now large spaces of science communication, and government agencies are increasingly present communicators on these sites, including for environmental and scientific information, and climate change information more specifically [2, 3]. However, there have been few studies examining the extent of climate science communication by government agencies on social media, despite the fact this is a novel, large and potentially impactful phenomenon. At the U.S. federal level alone there are various government agencies communicating about science and the environment, especially on Twitter, including NASA, NOAA and the EPA [4]. Moreover, government agencies that conduct and communicate climate science serve under administrations which may politicize the scientific communication [5]. It was extensively reported how the Trump administration in the U.S. led a number of changes to government websites shortly after it took office in 2017, namely with the removal of references to climate change, global warming and greenhouse gases from sites of the Environmental Protection Agency, the Department of Agriculture and Department of Energy, among others [6, 7].

Given this politically hostile context [8, 9] it may be useful to understand how and to what extent federal government agencies involved in science, and potentially climate

science and climate change, are publishing and broadcasting information on social media sites, such as Twitter. Such an examination can inform theories of bureaucratic autonomy and the politicization of science communication in the context of social media [10]. Moreover, given the large amount of government agencies and their messages broadcast on social media, as well as the availability of computational techniques for text and content analysis, it may also be useful to test how well these methods may be employed in this context for more automated analyses and summaries of large datasets. In this study, I thus raise the following research questions:

RQ1: Can topic modeling via LDA produce valid topic models from the posts of U.S. government science agencies on Twitter?

RQ2: How and to what extent is “climate change” and related information being communicated by the U.S. science agencies under the Trump administration?

In this preliminary paper, I first discuss the adoption of social media in government as communication, engagement with citizens and public relations. I then discuss the potential for bureaucratic autonomy in relation to the politicization of science communication, given the role of political control, values and ideology in this domain. I then describe the method of this study, including details about data collection, the LDA topic modeling technique, and analytical and validation procedures. The analysis of this study is for five Twitter pages, namely two of NASA and two of NOAA and the EPA. The results of this study show the general topics, or topic terms, associated with agency page; the extent to which information related to “climate change” and “global warming” are presented; and assessments of the topic modeling strategy.

2 Government, Social Media and Science Communication

Government agencies have long been associated with public and mass communication [11–13]. In the beginning of the 20th century, as the modern bureaucratic government was taking its form, so were the technologies of global and mass media, including the expansion of the press, and the developments of the telegraph and the radio. Weber recognized that mass and public communication was an essential component of public and private bureaucracies in the turn of the century [12]. Later Lasswell and others would theorize the role of mass media systems specially in relation to wartime communication and propaganda [14]. As the government bureaucracy expanded onto other areas, government agencies became largely involved in environmental protection, studies of the environment as well as science communication. Moreover, with the proliferation of social media sites, and the more general intrusion of networked communication into the lives of most people around the globe—over 60% of the global population uses a mobile phone [15]—some of these sites have become largely populated channels of government information, which has also included science communication [16, 17].

2.1 Government and Social Media

In the U.S. federal government, adoption of social media largely began with the Open Government Initiative of Barack Obama, an executive order in his first day in office, January 20, 2009 requesting that executive federal agencies engage in efforts to increase “transparency, collaboration and participation” between agencies and citizens [18]. Although not directly named in the memorandum, social media sites, such as Facebook and Twitter, were noted as new platforms where these principles—assumed beneficial as a matter of public policy—could be implemented, given the potential for “dialogic” and “two-way” communication afforded by these “web 2.0” or “new media”. A large literature was then devoted to examining the extent to which these platforms were being used in this dialogic fashion. Although social media have been institutionalized in various governments [19], unfortunately, for executive government agencies around the world it has been found that the platforms are *not* frequently used for conversations or dialogue [20–22]. However, in many cases the platforms have served distinct purposes, such as “positive self-presentation”, “impression management”, “symbolic acts”, “crisis communication”, and “science communication” [2, 23, 24]. There are now thousands of channels of U.S. federal government agencies, the most popular of which may be NASA on Twitter which has 35 million followers [4].

2.2 Science and Climate Science Communication

Science communication in the mass media has long been a topic of research [25, 26]. More recently, a plethora of research has addressed various dimensions of public communication about climate change and global warming. These have included issues of “balance as bias” in climate change coverage of newspapers [27]; the role of “motivated reasoning” and “identity cues” [28]; how “public health” [29] is discussed in climate change reporting; the role of “ideological cultures” [30] in the news media; and the importance of “framing” more broadly for environmental communication [31]. A number of studies have also examined the “politicization of science communication” and the role of values in the communication of policy-relevant science [10, 32].

Recent works have noted the importance of environmental and science communication by government agencies on social media, principally due to the potential of social media for “dialogic” communication and related benefits such as “positive relationships” and potential increase in “trust in science” [24, 33]. Not surprisingly, the latter work found that there was little dialogic engagement in the NOAA’s official Facebook page examined; however, they also found that “distrust in science” was expressed in approximately 23% of posts associated with “climate change” [33]. Interestingly, employees at NOAA responded to the article noting the difficulties and paradoxes of engagement on social media and the unfair assessment by the authors [34]. Nevertheless, there is a lack of works examining the politicization of government communication on social media, or the use of text mining tools to summarize and analyze government content on these platforms. This brief study works in the direction to fill these gaps.

3 Methods

In this study, I address two empirical research questions: RQ1: Can a topic modeling strategy with LDA produce valid topic terms from U.S. federal government Twitter posts? And RQ2: How and to what extent is “climate change”, “global warming” and related information being communicated during the Trump administration?

3.1 Data

For this study I collected all posts (“tweets”) and shared posts (or “retweets”) from the following five government science pages: the EPA (twitter.com/epa); NOAA (twitter.com/noaa); NOAA Climate (twitter.com/noaaclimate); NOAA NCEI (twitter.com/noaanceclimate); and NASA Climate (twitter.com/nasaclimate). I collected all posts from August 2018 through August 2019. These channels were selected since they seemed to be more relevant for climate change and climate change communication. Since this only yielded 228 posts from the NASA Climate page, I added an extra year of data to this dataset to help the topic modeling technique, which often requires large datasets to produce valid topic models. In these dates, the corresponding agencies are thus all under the Trump administration. The datasets are shown in Table 1.

Table 1. Descriptive statistics of datasets/corpora

Dataset	Followers*	Post Count	Posts + RT	Dates
EPA	604K	545	773	2018-08-01 to 2019-09-01
NASA Climate	329K	292	648	2017-08-01 to 2019-09-01
NOAA	1.05M	397	1233	2018-08-01 to 2019-09-01
NOAA Climate	86.2K	803	1203	2018-08-01 to 2019-09-01
NOAA NCEI	49.2K	2393	2989	2018-08-01 to 2019-09-01
<i>Totals</i>	2.1M	4430	6846	

* Follower count as of Sep 20, 2019

3.2 Topic Modeling via LDA

Latent Dirichlet Allocation (LDA) is a clustering and dimensionality-reduction algorithm which may serve to summarize documents of text [35]. The algorithm cluster terms that frequently co-occur in a corpus of documents, potentially identifying groups of “topic terms” [36]. Topic modeling has been interpreted to produce themes, identify salient issues, and may reveal how the content is being framed [37]. Previous studies have used topic modeling for discovering “expressed agendas” from press releases of U.S. legislators [38]; political opinions on blogs [39]; topics from e-petition entries [40]; as well as general topics from Twitter content [41]. Although issues have been noted with using small, short-text documents for topic modeling, distinct features of the corpus and modeling strategy details may enable the use of LDA in this environment.

Topic Modeling Strategy. To produce useful topic models, I employ a topic modeling strategy involving: *character manipulation, stop-words removal, tokenization* and *lemmatization* before building the model. All URLs, and @mentions were removed; hashtags were kept to assist identifying similar content. All terms were lower-cased. The basic list of NLTK stop-words was used to remove overly common and non-topic words (e.g. articles, prepositions). Tokenization included the creation of unigrams and bigrams from the corpora. I employed the NLTK Lemmatizer for nouns, verbs and adjectives, which turn some words into one of their root forms depending on their part-of-speech (e.g. “cars” to “car”, “running” to “run”, etc.), helping achieve better models.

Validation Strategy. To produce topic models via LDA it is necessary to determine the number of topics for the algorithm to produce. Since this is context dependent and not known apriori, I examined a few results informally and found models with 15-topics to be best. I then proceeded to validate the quality of the topic models more formally and address RQ1, which involved a basic test of “topic coherence”, a type of face validity check for the topic models [42]. To make the assessment: *the first 5 topic terms identified by the algorithm for each topic cluster* were judged. Similarly to other validations tests [40]: A score of 1 was given if the terms denoted a single, coherent topic; 2 if the terms denoted two topics, or not an entirely clear single topic; and 3 for not referring to a single topic or a clear combination of 2 topics. A mean score was then provided for each topic model.

3.3 Interpretation of Topic Models

To address RQ2, I first summarize the topics (i.e. topic clusters) for each agency and provide a topic name based on the most coherent and revealing combination of the first 2 or 3 topic terms. For example, if a topic cluster contains the following terms: “climate”, “noaa”, “report”, the topic is named as “climate report”. The reason here is because “noaa” is a redundant reference to the agency from which the posts originate. Although this is first an informal and manual process, it may lead to better rules for naming topic clusters in the future. After topic names and topic terms were listed, examined and compared, I also examined the extent to which topics and topics terms referred to “climate change”, “global warming”, or potentially related phenomena, such as wildfires, record high temperatures, hurricanes, sea-level rise, etc. Although this is also a preliminary and to some extent informal analysis, this should help us understand the extent to which these government channels of communication are relatively autonomous for communicating environmental and climate science information or may have been politicized by an administration hostile to climate science.

4 Results

Tables 2, 3 and 4 below show the results of the topic modeling analysis using 15 topics for each dataset. The datasets include both original posts and retweets. The optimum

number of topics will likely not be the same across datasets, especially given their different sizes (see Table 1 above). Nevertheless, in this preliminary analysis this number of topics seemed promising and therefore was used for all 5 agencies of the study. Tables below, however, only show results for 3 of the agencies given space restrictions on this paper. I have also added the 6th and 7th terms identified by the algorithm in the tables below to help the reader understand the overall quality of the topic models.

The lower the score the better the overall topic model. However, it should be noted this is a preliminary examination for further study, which can improve on the topic modeling strategy. In Table 4, in reference to the NOAA Climate page, 9/15 topic clusters were deemed as potentially a single coherent topic or theme (based on the top five topic terms). In Table 3, referring to the EPA page, three topic clusters were too poor to be assessed, and only 4/15 had an assessment as a single coherent topic. This model had the overall worst score. Table 2, in reference to the NASA Climate page, had 7/15 topic clusters which were judged as a single topic, and overall had interpretable topic clusters.

Table 2. Topic models of the NASA Climate Twitter page for 2 years

Topic Name	Top 5 Terms	6th and 7th terms	Score
earth science	earth science observe live eye	love system	2.5
record temperature	temperature warmest record global record keeping	year modern july	1
sea level rise	data sea_level rise view sealevelrise	drought weather	1
global_average carbon dioxide	global_average carbon_dioxide concentration_atmospheric ppm april	last_month per_million	
hurricane storm	ocean hurricane storm track land	map powerful	2
climate change	climate change global model vital_sign	vote understand	1
nasa instrument	nasa time instrument scientist learn	fly earthexpedition	2
water week	world water week warming begin	provide study	2.5
show today	image show today day celebrate	region air	2.5
fire cloud	nasa fire cloud california smoke	impact wildfire	1
carbon measure	carbon measure atmosphere make forest	greenhouse_gas affect	2
sea ice	ice year sea arctic antarctica	decade polar	1
space mission	satellite space mission launch increase	laser icesat	1
glacier melt	study melt glacier greenland warm	ice_sheet antarctic	1
-	planet scientist home happy life	join climatechange	3
		<i>Mean</i>	<i>1.68</i>

Table 3. Topic models of the EPA Twitter page for 1 year

Topic Name	Top 5 Terms	6 th and 7 th terms	Score
watch epa	epa act watch administrator_wheeler event	gt live	2.5
community work	work community epa nation improve	agency federal	2.5
-	emission partner make find check	effort report	3
protect environment	protect environment health child food	family people	2
clean air	clean air energy continue air_quality	provide america	2
epa program	epa program join apply fellowship	application opportunity	1
safe water	water safe drinking area power	level due	1
epa	epa state local plan administrator	address support	2.5
-	today lead day national week	celebrate protection	3
million grant	grant million project issue infrastructure	quality funding	1
hurricane emergency	learn hurricane emergency stay visit	prepare smoke	2.5
environmental award	environmental award congratulation peya student	school president	1
reduce asthma	reduce american asthma food_waste waste	save check	2.5
flood	flood florence home tip affect	update high	2.5
-	read year site today superfund	part superfund_site	3
<i>Mean</i>			2.13

Across the 3 datasets presented a number of valid topics did emerge from the analysis, suggesting that LDA may serve as a valid topic modeling strategy for this type of content, which help address RQ1. Moreover, in addressing RQ2, we observe that the NOAA Climate and NASA Climate pages do seem to explicitly address “climate change” and related phenomena, with information that refer to “record year”, “extreme event”, “record temperature”, “sea_level rise” and “global_average carbon_dioxide”. This type of information makes up a considerable amount of the overall content communicated by those channels. As such, despite the hostile political administration, we observe that some government agencies are communicating about climate change, climate science and related phenomena. The EPA page, however, does not seem to refer to “climate change”, “global warming” or make reference to “extreme events”. These

results thus suggest that while some government agencies were not politicized or influenced by the political administration in restricting climate change information, the same is not true for all government agencies.

Table 4. Topic models of the NOAA Climate Twitter page for 1 year

Topic Name	Top 5 Terms	6th and 7th terms	Score
climate change	climate change check scientist video	effect natural	1
record year	record year high warm sea_ice	arctic heat_wave	2
extreme event	extreme event stateofelimate global july	warmest experience	1
heat	heat noaa people city risk	urban health	2
drought	drought show droughtmonitor water affect	increase country	1
icymi	condition icymi el_niño latest winter	update summer	2
today	today question resource report expert	make assessment	2.5
information webinar	information webinar alaska provide community	hear national	1
weather forecast	weather week forecast scientist answer	noaa recent	1
data	data learn hurricane ocean explore	surface sign	2
fire	smoke impact fire wildfire study	emission understand	1
climate tool	climate tool science time news	latest resilience	1
map	map find rise project sea_level	continue coastal	2
teach climate	climate teach earth energy student	system wind	1
average temperature	temperature average day precipitation contiguous	united_state state	1
<i>Mean</i>			<i>1.43</i>

5 Discussion and Further Studies

The results of this study show that a topic modeling strategy with LDA may be a valid method for summarizing content from Twitter posts of government science agencies. This study also provides a set of topics that show the variety of content being disseminated by government science agencies for a period of one to two years. The topic models point to information related to “climate change” and “global warming” being disseminated, despite the antagonistic administration toward discussions and policies to address man-made climate change. However, there are some shortcomings that could be addressed by a more developed study. First, the topic modeling validation will need further reliability and validity analyses, such as examining the relation between topic

models and actual posts. Secondly, it may be valuable to examine whether or not references to “extreme events” and “record temperature” are discussed as potentially caused by climate change, and the extent to which man-made emissions of carbon dioxide are connected with climate change. This does not emerge so clearly from the topic models themselves and may need to be examined with further detail from the full posts/messages. Future studies may thus show a better picture of how to analyze this content with topic modeling, and the potential consequences of this engagement with science communication on social media.

References

1. Evans, H.-C., Dyll, L., Teer-Tomaselli, R.: Communicating Climate Change: Theories and Perspectives. In: Leal Filho, W., Manolas, E., Azul, A.M., Azeiteiro, U.M., and McGhie, H. (eds.) *Handbook of Climate Change Communication: Vol. 1: Theory of Climate Change Communication*. pp. 107–122. Springer International Publishing, Cham (2018).
2. DePaula, N., Dincelli, E., Harrison, T.M.: Toward a typology of government social media communication: Democratic goals, symbolic acts and self-presentation. *Government Information Quarterly*. 35, 98–108 (2018). <https://doi.org/10.1016/j.giq.2017.10.003>
3. Welbourne, D.J., Grant, W.J.: Science communication on YouTube: Factors that affect channel and video popularity. *Public Underst Sci*. 25, 706–718 (2016). <https://doi.org/10.1177/0963662515572068>
4. Smith, J.: Which U.S. Government Agency is Most Active on Social Media?, <https://www.spredfast.com/social-marketing-blog/which-us-government-agency-most-active-social-media>
5. Pinto, M.F., Hicks, D.J.: Legitimizing Values in Regulatory Science. *Environmental Health Perspectives*. 127, (2019). <https://doi.org/10.1289/EHP3317>
6. Freedman, A.: The EPA’s climate change website has already changed under Trump, <http://mashable.com/2017/02/03/trump-epa-climate-website-changes/>
7. EDGI: Changing the digital climate: How climate change web content is being censored under the Trump administration, <https://100days.envirodatagov.org/changing-digital-climate/>
8. Jacobson, L.: Yes, Donald Trump did call climate change a Chinese hoax, <https://www.politifact.com/truth-o-meter/statements/2016/jun/03/hillary-clinton/yes-donald-trump-did-call-climate-change-chinese-h/>
9. Lipton, E.: As Trump Dismantles Clean Air Rules, an Industry Lawyer Delivers for Ex-Clients, <https://www.nytimes.com/2018/08/19/us/politics/epa-coal-emissions-standards-william-wehrum.html>, (2018).
10. Suhay, E., Druckman, J.N.: The Politics of Science: Political Values and the Production, Communication, and Reception of Scientific Knowledge. *The ANNALS of the American Academy of Political and Social Science*. 658, 6–15 (2015). <https://doi.org/10.1177/0002716214559004>
11. Hood, C.: *The Tools of Government*. Palgrave Macmillan, London (1983).
12. Weber, M.: *Economy and Society: An Outline of Interpretive Sociology*. Bedminster Press, New York (1968).
13. Graber, D.A.: *The Power Of Communication: Managing Information In Public Organizations*. CQ Press, Washington, D.C (2003).

14. Lasswell, H.D., Casey, R.D., Smith, B.L. eds: Propaganda and Promotional Activities: An Annotated Bibliography. University of Chicago Press (1969).
15. Silver, L.: Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally, <https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/>, (2019).
16. DePaula, N.: #Supporting the cause: An analysis of how government agencies use Twitter hashtags. *Proceedings of the Association for Information Science and Technology*. 55, 788–789 (2018). <https://doi.org/10.1002/pr2.2018.14505501117>
17. Hagen, L., Keller, T., Neely, S., DePaula, N., Robert-Cooperman, C.: Crisis Communications in the Age of Social Media: A Network Analysis of Zika-Related Tweets. *Social Science Computer Review*. 0894439317721985 (2017). <https://doi.org/10.1177/0894439317721985>
18. Obama, B.: Transparency and Open Government: Memorandum for the heads of executive departments and agencies, <https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government>
19. Mergel, I.: Social media institutionalization in the U.S. federal government. *Government Information Quarterly*. 33, 142–148 (2016). <https://doi.org/10.1016/j.giq.2015.09.002>
20. Bonsón, E., Royo, S., Ratkai, M.: Citizens' engagement on local governments' Facebook sites. An empirical analysis: The impact of different media and content types in Western Europe. *Government Information Quarterly*. 32, 52–62 (2015). <https://doi.org/10.1016/j.giq.2014.11.001>
21. Hofmann, S., Beverungen, D., Räckers, M., Becker, J.: What makes local governments' online communications successful? Insights from a multi-method analysis of Facebook. *Government Information Quarterly*. 30, 387–396 (2013). <https://doi.org/10.1016/j.giq.2013.05.013>
22. DePaula, N., Dincelli, E.: Information strategies and affective reactions: How citizens interact with government social media content. *First Monday*. 23, (2018). <https://doi.org/10.5210/fm.v23i4.8414>
23. Graham, M.W., Avery, E.J., Park, S.: The role of social media in local government crisis communications. *Public Relations Review*. 41, 386–394 (2015). <https://doi.org/10.1016/j.pubrev.2015.02.001>
24. Lee, N.M., VanDyke, M.S.: Set It and Forget It: The One-Way Use of Social Media by Government Agencies Communicating Science. *Science Communication*. 37, 533–541 (2015). <https://doi.org/10.1177/1075547015588600>
25. Griffin, R.J., Dunwoody, S., Gehrman, C.: The Effects of Community Pluralism on Press Coverage of Health Risks from Local Environmental Contamination. *Risk Analysis*. 15, 449–458 (1995). <https://doi.org/10.1111/j.1539-6924.1995.tb00337.x>
26. Miller, D.: Risk, science and policy: definitional struggles, information management, the media and BSE. *Social Science*. 17 (1999).
27. Boykoff, M.T., Boykoff, J.M.: Balance as bias: global warming and the US prestige press. *Global Environmental Change*. 14, 125–136 (2004). <https://doi.org/10.1016/j.gloenvcha.2003.10.001>
28. Hart, P.S., Nisbet, E.C.: Boomerang Effects in Science Communication: How Motivated Reasoning and Identity Cues Amplify Opinion Polarization About Climate Mitigation Policies. *Communication Research*. 39, 701–723 (2012). <https://doi.org/10.1177/0093650211416646>
29. Weathers, M.R., Kendall, B.E.: Developments in the Framing of Climate Change as a Public Health Issue in US Newspapers. *Environmental Communication*. 10, 593–611 (2016). <https://doi.org/10.1080/17524032.2015.1050436>

30. Carvalho, A.: Ideological cultures and media discourses on scientific knowledge: re-reading news on climate change. *Public Underst Sci.* 16, 223–243 (2007). <https://doi.org/10.1177/0963662506066775>
31. Lakoff, G.: Why it Matters How We Frame the Environment. *Environmental Communication.* 4, 70–81 (2010). <https://doi.org/10.1080/17524030903529749>
32. Druckman, J.N.: Communicating Policy-Relevant Science. *APSC.* 48, 58–69 (2015). <https://doi.org/10.1017/S1049096515000438>
33. Lee, N.M., VanDyke, M.S., Cummins, R.G.: A Missed Opportunity?: NOAA’s Use of Social Media to Communicate Climate Science. *Environmental Communication.* 12, 274–283 (2018). <https://doi.org/10.1080/17524032.2016.1269825>
34. Lindsey, R., Herring, D., Greenhalgh, E.: Response by Lindsey, Herring, and Greenhalgh to “A Missed Opportunity?: NOAA’s Use of Social Media to Communicate Climate Science.” *Environmental Communication.* 12, 284–286 (2018). <https://doi.org/10.1080/17524032.2017.1394659>
35. Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent Dirichlet Allocation. *Journal of Machine Learning Research.* 3, 993–1022 (2003).
36. Blei, D.M.: Topic Modeling and Digital Humanities. *Journal of Digital Humanities.* 2, (2012).
37. DiMaggio, P., Nag, M., Blei, D.: Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics.* 41, 570–606 (2013). <https://doi.org/10.1016/j.poetic.2013.08.004>
38. Grimmer, J.: A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases. *Political Analysis.* 18, 1–35 (2010). <https://doi.org/10.1093/pan/mpp034>
39. Koltsova, O., Koltcov, S.: Mapping the public agenda with topic modeling: The case of the Russian livejournal. *Policy & Internet.* 5, 207–227 (2013). <https://doi.org/10.1002/1944-2866.POI331>
40. Hagen, L.: Content analysis of e-petitions with topic modeling: How to train and evaluate LDA models? *Information Processing & Management.* (2018). <https://doi.org/10.1016/j.ipm.2018.05.006>
41. Therkelsen, J.F., Steinskog, A.O.: Characterizing Twitter Data using Sentiment Analysis and Topic Modeling, <https://brage.bibsys.no/xmlui/handle/11250/2415333>, (2016).
42. Grimmer, J., Stewart, B.M.: Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Political Analysis.* 21, 267–297 (2013).