The Internet, Policy & Politics Conference 2018 - Oxford Internet Institute

The EPA under the Obama and Trump administrations: Using LDA topic modeling to discover themes, issues and policy agendas on Twitter

Nic DePaula and Teresa Harrison
University at Albany, State University of New York

Abstract

Although studies abound regarding how legislators and other political actors behave on social media, the communicated content of government agencies on these sites is less studied. However, investigating this content may be important to understand how these technologies are being used by executive officials as a matter of communication and information policies. Moreover, text mining techniques such as topic modeling are being increasingly adopted to process text and speech data from several social actors and domains, but less attention is given to government discourse. In this study we thus tested the validity of a LDA topic modeling strategy for producing coherent and meaningful topic models of Twitter posts from the Environmental Protection Agency of the United States over a period of 2.7 years. We examined the themes and issues that could be interpreted from these models and found how specific policy items varied across two distinct political administrations. We contribute to the methodological literature by demonstrating how LDA may be used to produce valid topic models from this source of data. We also contribute to the policy and communication literature by demonstrating how themes and issues broadcast by a government agency on Twitter are not only determined by the nature of the agency and its programs but also by the policy agenda of the administration in power.

KEYWORDS: government communication, social media, Twitter, LDA, topic modeling, text mining, politics, policy agenda setting

1. Introduction

As social media sites have facilitated online public discourse around the world, the institutions of government have also taken advantage of these open platforms of communication on the Internet. The U.S. federal government alone sponsors thousands of accounts distributed across different types of agencies and programs on sites such as Facebook, Twitter and YouTube (Smith 2016). Some of these accounts have thousands of followers; others have millions. Although quantitative and qualitative estimations of this activity by governments have been carried out, aspects of the content communicated by the executive branch on social media is less often studied. Text mining techniques, although often applied to citizen generated text, are not as frequently applied to government discourse—although they seem to be significant actors in these platforms. In this study, we thus sought to examine government communication on Twitter through computational and comparative content analysis. We evaluate the use of LDA for computational topic modeling of the public posts of the U.S. Environmental Protection Agency (EPA) under the Obama and Trump administrations. We test the validity of the topic models produced, explore how themes and specific issues identified vary across the administrations, and point to instances where differences in content can be attributed to the priorities and agendas of the administration in power.

Despite the lack of scholarly attention, there are a number of reasons to be interested in how government agencies and their communicators use social media in general and Twitter in particular. For one thing, Twitter is a major source of news and political commentary; in the U.S. it is reportedly used by an estimated 24% of the adult population (Pew Research Center, 2018). Around the world, Twitter has a total of 330 million active users as of 2017 (Gray, 2017)—although, it's important to mention, almost 50 million or more of these could be bots (Varol et al. 2017). Perhaps more than any other platform in the U.S., Twitter provides an open space for political news and commentary that can also influence the issues that other media channels discuss (Shapiro and Hemphill 2016; Harder, Sevenans, and Van Aelst 2017; Harrison et al. 2017). Moreover, social media, and specially Twitter, are popular online venues for various agencies of the U.S. government (Smith 2016; U.S. Digital Registry 2018). As of July 6th, 2018, the 15 main agency accounts of the federal government had approximately 25 million followers on Twitter combined and together produced over 311,000 posts (compared to 7.5 million followers on Facebook—determined from our own count).

Given the large amount of text data that public social media discourse produces, a number of text mining and machine learning approaches have been developed and applied to answer questions of how political actors—e.g. politicians, legislators, activists—are communicating and exchanging information on social media sites (e.g. Koltsova and Koltcov 2013; Burscher, Vliegenthart, and De Vreese 2015; Burscher, Vliegenthart, and Vreese 2016; Ceron and Negri 2016; Yang et al. 2016; Hagen et al. 2017; Najafabadi and Domanski 2018). LDA, or Latent Dirichlet Allocation, is a commonly used text mining and machine learning algorithm that produces *topic models*, that is, clusters of terms that frequently co-occur in a corpus of text and may yield semantically coherent themes from the text analyzed (Blei, 2012). Topic modeling has been interpreted to produce themes, identify salient issues, and may reveal how the content is being framed (e.g. DiMaggio, Nag, and Blei 2013). Previous studies have used and developed topic modeling techniques for discovering "expressed agendas" from press releases of U.S. legislators (Grimmer, 2010); political opinions on blogs (Koltsova and Koltcov, 2013); topics from e-petition entries (Hagen, 2018); as well as general topics from Twitter content (e.g. Therkelsen and Steinskog 2016; Shi et al. 2018). Although there are limitations to what can be accomplished via these types of analyses, computational and machine learning approaches facilitate the study of large corpora of text, enable systems to learn from the content, can be used to monitor communication in real-time, and may reduce bias in content analysis. Nevertheless, to our knowledge, topic modeling has not been used in the context of government communication on social media. We thus ask:

> *RQ1: Can LDA topic modeling techniques produce valid themes and issues from the Twitter posts of a select U.S. government agency?*

Few studies employ text mining techniques to study government communication on social media; indeed, "government communication" has not been a common term in the social sciences (Canel and Sanders 2012; Hansson 2017). When government communication is studied it is likely more often under more specific domains (e.g. police communication; scientific communication) or as a managerial issue (e.g. Graber 2003; Crump 2011; Meijer and Thaens 2013). Rhetorical, discursive and thematic issues communicated by different types of government agencies in their social media accounts do not seem to be often explored in the scholarly literature. Nevertheless, political communication theory has long recognized that discussion of policy issues in the media can increase the salience of these issues in the minds of the public, in what are referred to as *agenda setting* processes (McCombs and Shaw 1972; Scheufele and Tewksbury 2007). Although we do not explore media effects on public perception, we do explore how and what content is set on the media channels of the government agency at question, what may be conceived as a type of *policy agenda setting*. Themes discussed on social media sites will be related to the domain of each agency (e.g. police agencies will discuss police issues; housing agencies housing issues, etc.). Moreover, laws will probably constrain the type of communication admissible. Nevertheless, we can also expect administrative control and the intention by those in power to communicate the political and policy agenda to influence the themes and issues that are discussed. We thus ask:

> *RQ2: How does the content change across two distinct political administrations? Can changes be attributed to the political administration in power?*

To address these questions, this study proceeds as follows. First, we review studies and theories of government communication and government social media practices to contextualize this phenomenon and provide insights regarding previous studies and expected forms of government discourse on these platforms. We introduce the case of the Environmental Protection Agency (EPA) of the U.S. under the Obama and Trump administrations as a potentially interesting case study in which to examine the interplay between information policies and politics. We then review what topic modeling is, the LDA approach to topic modeling, and how these techniques have been used in related studies. In the methods section, we discuss how the corpus of data has been created, data pre-processing strategies, model building and our human validation task. We evaluate several topic models with different numbers of topics. The best models were selected for further interpretation of themes and issues expressed, and for discussion of how variations of the content can be attributed to changes in the political administration of the agency. Lastly, we address implications of these findings for government policy, politics and future research.

2. Government social media communication

2.1 Forms and functions of communication

Although the subject is not one of the most popular, over the years frameworks have been proposed to interpret and analyze "government communication"—the public communication of executive government actors (Yudof 1979; Hood 1983; Liu and Horsley 2007; Canel and Sanders 2012;

Hansson 2017; DePaula, Dincelli, and Harrison 2018). In the model of "government communication strategies" of Hiebert (1981)—cited as exemplary of government "public relations" and "public information campaigns" by Graber (2003)—four broad strategies are identified: *withholding* information, which may be particularly important, but also not at all publicized*; releasing* information, as the provision of documents and knowledge of governments*; staging* the presentation, which may involve press-conferences and special events; and *persuading* the audience—the latter two of which include meta-textual elements in addition to the explicit semantic content. Similarly, in the well-known model of Hood (1983), the communication activities in the "tool-kit" of governments are used to *advise, inform and persuade*. Hood and Margetts (2007) also note that the communication or release of information by governments may be of three types: *bespoke messages*, which are tailored to specific people; *group-targeted messages*, which are targeted to groups; and *broadcast messages*, which are messages to the world at large, and may best categorize the public social media posts analyzed in this study.

Previous studies of government communication and information practices through social media—here defined as *Internet platforms that facilitate interpersonal, group and mass communication*—have often used the distinction between "one-way" versus "two-way" types of interaction to examine the potential of these tools for interactive relationships between government agencies and the public (Mergel 2013; DePaula and Dincelli 2016; Meijer and Thaens 2013; DePaula and Dincelli 2018). From the terminology of public relations theory (Grunig and Grunig 1992; Waters and Williams 2011), but building on this distinction, the forms of communication, and potentially the semantic content itself, may be defined as follows: *one-way asymmetric*, which refers to tactics of impression management and persuasion; *one-way symmetric*, which refers to the broadcasting of basic and truthful information; *two-way asymmetric*, which is the "pulling" of information without much interaction; and lastly, *two-way symmetric*, which refers to a more dialogic and mutually beneficial exchange of ideas, or what others may refer to as "networking" or "collaboration" (Mergel 2013; Valle-Cruz, Sandoval-Almazan, and Gil-Garcia 2016).

Other more politically and language oriented analyses of how and why governments communicate with the public have been interested in examining the rhetorical and discursive elements of this type of policy behavior (Bates 2013; Hansson 2017; DePaula, Dincelli, and Harrison 2018). For example, Hood (1998) has examined the relationships and implications of rhetoric, culture and world-views in public management, as well as the recurrent phenomenon of "blame avoidance" in modern government contexts (Hood 2011). Hanson has also examined the discursive elements of blame avoidance in government speech and the potential strategies of "calculated overcommunication" by government officials (Hansson 2015a; Hansson 2015b). In a study of local government communication on Facebook, we have explored the use of "symbolic acts" (DePaula, Dincelli, and Harrison 2018) in the form of expressive speech acts, as a principal type of communication on the platform in addition to broadcasting government information.

2.2 Agendas, policies and administrative control

Studies in political communication have long recognized the importance of the media to influence the political and policy issues that the public *thinks about* as well as *how they think about it*, what are often referred to as *first-level* and *second-level agenda setting effects*, or *agenda-setting* and *framing* (McCombs and Shaw 1972; Rogers and Dearing 1988; Weaver 2007; Entman 2007; Ceron, Curini, and Iacus 2016). However, in the context of government communication on social media, the *policy agenda* (i.e. the issues, priorities and concerns) of the government agency at question may be directly communicated to the public without the intervention of traditional media. Although this does not necessarily mean that the policy agenda on social media will "set" or influence what the public or the traditional media agenda talks and thinks about (e.g. Harder, Sevenans, and Van Aelst 2017), government agencies gain direct channels to their followers who may share the content, approve it, or disapprove of it in the form of comments. Moreover, government agencies are able to use these platforms to discuss the issues they care about, frame the communication how they see fit and also have these messages be potentially propagated to other media channels (e.g. Harrison et al. 2017; Conway-Silva et al. 2018)

Government communication is considered a strategic activity, but it is also subject to a number of laws. Bertot, Jaeger, and Hansen (2012) summarized a number of information policies that likely implicate how government agencies adopt social media, and a number of agencies of the U.S. government currently provide specific guidance on how social media is to be used. For example, The U.S. Information Quality Act of 2001 (IQA), which applies to all federal government agencies, requires them to *"issue guidelines ensuring and maximizing the quality, objectivity, utility, and integrity of information (including statistical information) disseminated by the agency"* (Conrad 2002). An EPA social media information policy document found on their website in July

2018 (dated between 2011 and 2014), and which seems to apply some of the IQA principles, suggests the following for social media communicators:

> *"Do not endorse any product, service, company, non-profit organization or any other enterprise. There are some exceptions but … you should be careful about giving an appearance of governmental sanction or endorsement. …*
>
> *Do not engage in any partisan political activity.*
>
> *Do not fundraise for any charitable organization.*
>
> *Do not attempt to directly or indirectly lobby Congress."* (EPA 2014)

Some of these guidelines may indicate that governments will strive to provide accurate, reliable and unbiased communication (Conrad 2002). However, it is unclear how these policies effectively regulate the issues and content that are addressed (or not) in their social media accounts. Since social media communication is controlled by each agency and ultimately by the head of the agency, the administrator in power will likely decide what policy or agenda items to broadcast about, which are likely related to the policy agenda of the political administration that placed the administrator in its position (e.g. in the case of the EPA and other politically appointed administrators). As such, top-level administrators of federal agencies may influence both the content to be communicated as well as how it is expressed (e.g. specific terms to be used or not used; positive, negative or other types of frames employed). Although there is an imperative to be objective and unbiased, existing regulations may not overly restrict them from withholding certain information, or promoting issues of their preference. We thus ask the following research questions:

> *RQ2a. How do themes and issues in the Twitter posts of a select U.S. government agency vary across two distinct political administrations?*
>
> *RQ2b. Can any changes be attributed to what we know about the administration in power?*

3. The EPA under the Obama and Trump administrations

An interesting and potentially revealing government agency in which to examine distinct differences in agenda and policy communication is the U.S. Environmental Protection Agency (EPA) under the Obama and Trump administrations. The EPA is an "independent agency" of the United States Federal government created in 1970 under President Richard Nixon. The agency conducts environmental assessments, research and educational efforts, and is responsible for implementing and enforcing a number of environment related national standards that arise from legislative statutes, along with other related activities (EPA 2013). The organization also creates and implements specific programs to address the environmental issues identified in these laws and has the power to fine and sanction businesses and individuals. Some of its standards and enforcement programs are based on legislation such as the Clean Air Act, Clean Water Act, Endangered Species Act—laws passed prior to the creation of the agency, but which have received several amendments since (Collin 2006). Given these mandates, the agency is responsible for dealing with large and potentially impactful issues such as chemical emissions from vehicles and industries, oil spills and for regulating chemical sources of climate change (C2ES 2018).

Barack Obama was elected president in 2008, and on January 21, 2009, his first day in office, he signed the Open Government Initiative, which directed all federal government agencies to improve their efforts to open government information to the public, foster greater public participation, and encourage collaboration with government (Obama 2009). Although the EPA Twitter account was created in May 2008, other social media accounts were created in part due to this open government initiative (Mergel 2013). In contrast, during the first months of the subsequent administration of Donald Trump, several government agency web sites, including the EPA, removed various pages and documents related to “climate change”, “clean energy” and “greenhouse gases” (EDGI 2018)—in the words of the agency then to “reflect the agency’s new direction under President Donald Trump and Administrator Scott Pruitt”[1]. Although these changes were observed in various web sites (EDGI 2018), no empirical and comparative study of the social media communication across these two administrations seems to have been carried out.

4. Topic modeling in social research

In addition to the political issues regarding changes in communication and information policies across distinct administrations, in this study we are also concerned with the use of text mining and natural language processing techniques to reveal this content as it is broadcast on social media. In particular we are interested in the use of *topic modeling.* Topic modeling refers to dimensionality-reduction and summarization techniques for computational text analysis. These techniques process corpora of documents (e.g. collections of tweets, news articles, book chapters, etc.), and find

[1] "U.S. Environmental Protection Agency, "EPA Kicks Off Website Updates," (April 28, 2017)". EPA. April 28, 2017.

clusters of terms that co-occur, resulting in lists of words that may reveal the thematic structure of the documents analyzed (Blei, 2012). Latent Dirichlet Allocation (LDA) is a popular algorithm used to produce these topic models (Blei, Ng, and Jordan 2003; Hagen 2018) which we adopt and evaluate in this study.

The topic modeling approach to the study of text is considered *corpus-linguistic,* since it is focused on examining large numbers of documents, rather than specific passages of text or speeches (Törnberg and Törnberg 2016). These techniques are also considered generative or unsupervised since the algorithms are not given human labeled data indicating what terms go with what topics; this information is discovered by the algorithm. The textual products of topic models are clusters of words listed in descending order based on some probabilistic or algebraic measure of their association. Although there are difficulties with the use of these tools, they promise to assist in a variety of studies in the social sciences and humanities (Grimmer and Stewart 2013).

An important and valuable area for the study of text mining algorithms is government communication. Studies in digital government, collective action and policy informatics have discussed the use of automated text analysis to help understand how and what citizens are communicating online, such as on social network and e-petition sites (e.g. Ceron and Negri 2016; Margetts et al. 2016; Hagen 2018). However, few studies to our knowledge use these tools for the analysis of government communication. Computational methods may provide more efficient means for studying large datasets or text corpora, and the proliferation of more efficient computers as well as machine learning algorithms have facilitated the training of statistical models for real-time analyses of quantitative and qualitative data, all of which may assist in large-scale monitoring or studying of government discourse and communicative behavior. Moreover, as in other fields, automated text analysis may serve as a potentially more objective alternative to manual content analysis or simply as additional tools with which to conduct social science research (e.g. Hagen 2016; Blei 2012; DiMaggio, Nag, and Blei 2013).

Topic modeling has been used in a different domains and media contexts, some of which are related to government communication. For example, Grimmer (2010) developed a Bayesian hierarchical approach to topic modeling and used it for the analysis of “expressed agendas” in the press releases of U.S. legislators in a particular year. Quinn et al. (2010) also developed a topic modeling algorithm and used it for the analysis of parliamentary speech data of U.S. legislators. DiMaggio, Nag, and Blei (2013) used LDA to study how U.S. newspapers used different “frames” in their reporting of U.S. government arts funding over a 10-year period, which accompanied actual declines in funding toward the National Endowment for the Arts. More recently, topic modeling has been used to understand themes and issues in communication about Islam (Törnberg and Törnberg 2016), topics in e-petition entries (Hagen, 2016), themes from interview surveys (Baumer et al. 2017) and changes in issues discussed in parliamentary debates (Magnusson et al. 2018). We thus also expect that topic modeling may be used in the context of government social media communication to identify themes and issues from the textual content of the Twitter posts. We thus first address the following question:

RQ1. Can topic modeling with LDA produce valid themes and issues from the corpora of Twitter posts of the selected U.S. government agency?

5. Methods

5.1 Data selection and corpus creation

For this analysis, we retrieved all of the Twitter posts (e.g. tweets or simply posts) of the EPA for the first 16 months of the Administrator Gina McCarthy (under President Obama) and the first 16 months of the Administrator Scott Pruitt (under President Trump). This selection of timeline is justified as follows. At the moment of this writing, Pruitt resigned, or "was told to resign" (Jacobs and Dlouhy 2018), from his position as EPA administrator after a number of ethical probes into his behavior—although none of these seemed to concern his use of the EPA's Twitter account (Davenport, Friedman, and Haberman 2018). His tenure in the position ran from February 17, 2017 to July 6, 2018—16 months and 20 days. Although McCarthy stayed in her position from July 18, 2013 to January 20, 2017—3.5 years—we retrieved the first 16 months of her administration to provide a similar time period to contrast with the subsequent administrator.

Twitter poses challenges for topic modeling given that the document unit—the tweet—has historically been restricted to a maximum of 140 characters, and only recently changed to 280, which is still relatively short in comparison to other documents used for topic models (e.g. news articles, petitions, parliamentary proceedings). Nevertheless, the approach taken here can be considered a type of "author topic model" which aggregates social media posts based on its author or account and has produced better results (Rosen-Zvi et al. 2012; Mehrotra et al. 2013; Therkelsen and Steinskog 2016). In this study we thus create 2 separate corpora or datasets: the McCarthy dataset and the Pruitt dataset, one for each "author". In July of 2018, we retrieved from the official EPA Twitter account: 3,784 posts from July 20, 2013 to December 12, 2014 (501 days), constituting the McCarthy dataset; and 945 posts from February 17, 2017 to July 4, 2018 (501 days), constituting the Pruitt dataset. Although it is unclear if the agency deleted posts during these time periods, given existing regulations there is reason to believe the datasets include the vast majority of posts created by the agency (e.g. McCammon 2018; Stephens 2017; Buckley 2016).

The difference in size of the two datasets is clear and it suggests the account was considerably more active in the McCarthy period selected compared to the Pruitt period selected. We train and evaluate these two distinct datasets separately and thus also gain some information regarding appropriate data sizes for this type of analysis.

5.2 Latent Dirichlet Allocation

Latent Dirichlet Allocation (LDA) refers to a generative probabilistic model of a corpus (Blei et al., 2003). LDA can be seen as an improvement on previous *probabilistic latent semantic indexing* (PLSI) models of natural language processing due to the addition of the *Dirichlet prior* in estimating topic distribution, which improves the fitting of the models to unseen data (Blei et al. 2003; Girolami and Kaban 2003). LDA assumes that the corpus of text analyzed is a *bag-of-words*, that is, that the order of words in a document may be neglected; and that the ordering of the documents in a corpus may also be neglected (Blei et al. 2003). Nevertheless, it does assume the corpus is organized in documents. Each document is viewed as a mixture of latent topics, where each word is assumed to have been drawn from one of these topics. To build the topic models—that is, find clusters of words that may be considered "topics" or "themes"—the LDA algorithm must be trained with the corpus data and provided a number of parameter values, including the number of topics itself, which may change depending on the domain and size of data analyzed (Yau et al. 2014). Moreover, to train and build the models approximation techniques, such as log likelihood maximization and Gibbs sampling are necessary, which vary based on implementation used. Further details about LDA may be found in Blei et al. (2003) and (Blei 2012).

5.3 Data pre-processing strategies

There are a number of data pre-processing strategies that have been used and suggested for topic modeling (Boyd-Graber, Mimno, and Newman 2014; Hagen 2018). Our approach in this study is the following. First, we used 2 separate datasets, in which we applied the same pre-processing (training and validation) steps. We employed *character manipulation, stop-words removal, tokenization* and *lemmatization* before building the model. First we removed URLs, all punctuation, as well as the # and @ symbols—although not the # and @ words (e.g. #Cuba is changed to cuba; @Mario to mario). We lowercased all terms, and kept the symbols % and $ as references to statistics and money. We used the NLTK *stop-words* list to remove terms that are not discriminating in terms of themes and issues (e.g. pronouns, articles, prepositions). The NLTK dictionary is used for being conservative as it only includes 179 stop-words, compared to the 524 from MALLET. However, in pilot analyses we discovered some domain terms emerging as important and frequent but which were not particularly revealing in term of themes, issues nor how the issues were being framed or qualified, such as "work", "year", "today". We thus included these terms in the list of stop words to be removed.

Tokenization refers to the determination of unit terms to be analyzed in the text. A token may be a single word, such as "public", or a sequence of multiple words, such as "public policy" or "private public partnership", which are called *n-grams.* In this study we include both a *unigram* model of the text as well as *bigrams*, in order to capture important two-word terms, such as "climate change" or "job creators" that may frequently appear in the text.

We also employed a basic *lemmatization* strategy. Lemmatization refer to techniques of converting words to their base forms based on syntactic considerations, and is usually discussed in conjunction with or in lieu of *stemming* which refers to the chopping-off of ends of words to achieve some base form of the word (Manning, Raghavan, and Schütze 2008). For example, stemming may cut off the word "ambitiously" to "ambiti", whereas lemmatization may convert the same word to "ambition" or "ambitious", depending on syntax conversion chosen. Some improvements in model evaluation have been observed from using stemming, however, the value of stemming and lemmatization in models evaluated by automated coherence metrics has been put into question (Schofield and Mimno 2016; Hagen 2018). In this study we took a basic lemmatization approach which produces minor modifications to the text. We employed the NLTK Lemmatizer for nouns, verbs and adjectives, but not collapsed onto each other. For example, if used in the sentence as a verb, the terms "winning" and "wins" are converted to "win"; the terms "winner" or "winning" (as nouns) are left as such. The plural nouns (e.g. "cities", "communities") are converted to singular nouns (e.g. to city, community). This is done so that the algorithm can identify that these slightly distinct terms are assumed to be the same or fulfill the same semantic function, and therefore improve training in the smaller size documents that are social media posts.

5.4 Model training and parameters

To train and build the topic models we employed the MALLET LDA implementation, which has been successfully employed in other humanities and social science contexts (McCallum 2002; Baumer et al. 2017; Hagen 2016). LDA requires two special *hyperparameters* to be determined to build the models, and the *K* number of topics to be supplied. For this study, all parameters are left in their default values (see McCallum 2002), *except K*. Identifying the appropriate number of topics for any a particular dataset has been a point of discussion in studies of topic modeling, and the value seems to depend on the domain studied, the medium of communication or genre of discourse, size of the dataset as well the purpose of the study (Boyd-Graber et al. 2014; Grimmer and Stewart 2013; Hagen 2018). Studies which are more computationally focused may attempt to discover hundreds of topics from thousands or millions of documents, which are only evaluated via automated means. Our purpose here is to produce topic models that may be relatively easy to interpret by humans and can serve as valid summarizations of themes and salient issues represented in the corpus. We thus built 5 topic models for each of the 2 datasets with the following values of *K*: {10, 15, 20, 25, 30} and evaluated the validity of all topic clusters produced in each model.

5.5 Model validation and interpretation

Topic model validation is usually carried out with automated *perplexity* and *coherence metrics* (Boyd-Graber et al. 2014). Perplexity is an information theory measure based on the *inverse log likelihood* that a topic matrix can predict a set of unseen collection of documents. Lower perplexity values suggest less uncertainties in evaluating the unseen documents. In a now popular study of Chang et al. (2009) however, they found that perplexity was not correlated and sometimes inversely correlated with human judgement. Therefore, topic coherence metrics have been preferred, a popular one being that of (Mimno et al. 2011), referred to as “u_mass” coherence. Coherence metrics, such as u_mass coherence, are based on calculating *pointwise mutual information*, that is, a measure of how much a term is found with another term (Röder, Both, and Hinneburg 2015). We report perplexity for the 10 different models from the MALLET output and u_mass coherence measures from the Gensim Coherence Model package (Röder, Both, and

Hinneburg 2015). We also validated the coherence of the topics with a human validation task and later explain how we interpret themes and issues from the validated topics.

Topic models are *K* clusters of words, where in each cluster words are listed in descending order of importance for the cluster in which they are found, as shown in Figure 1 below. As such, only the first 5, 10 or 20 words are used for humans to code in human validation tasks (e.g. Chang et al. 2009; Hagen 2016). Given the small size of documents for this study we used the first 10 words of each topic. Each cluster should be considered a "topic", but ultimately should be evaluated positively if shown to be "meaningful, interpretable, coherent and useful" (Boyd-Graber et al. 2014, p. 15). There are numerous techniques to help "diagnose" problems with topic models, all of which should consider the specific context and task at hand (Grimmer and Stewart 2013; Boyd-Graber et al. 2014). A common approach is to ask annotators to judge each topic on a 3-point scale, such as whether each topic may be classified as: "good", "intermediate", or "bad" (Mimno et al. 2011); or assigned a 1 for a single topic, a 2 for two topics, and a 3 for three topics or none at all (Hagen 2016). In this study we also carried out this type of *internal topic coherence* evaluation or what may be referred to as *intratopic semantic validity* (Quinn et al., 2010).

We developed a coding scheme for our coherence and thematic evaluation task as further explained below and shown in Figure 1, where each topic cluster may be classified as either having a *single*, *double*, *half* or *none* coherent theme. Each 10-word cluster is presented without its probability values and given two boxes, where to enter the coding values. This is done to facilitate the coding of the first and second components of the cluster, since the top words are more important than the bottom words.

To know if a topic is coherent, annotators are asked to read the first two or three words and think if it suggests a meaningful label (e.g. Hagen 2016). For example, for the first cluster in Figure 1 the first two or three words suggest a theme of "food waste" or "reducing food waste". The following two words seem to strengthen this theme with "recycle" and "reuse". This thus seems like a coherent topic or theme and we thus assign this first component a value of 1. The latter set of words are more ambiguous. However, two strongly associated ideas are present, "landfill" and "greenhouse_gas". The other three words "american", "fight" and "easy" are more ambiguous, but they should not be entirely dismissed since they qualify the theme presented. "Fight" and "easy" may indicate ideas of "fighting food waste" or "ease of recycling". The word "american" is also a qualifier, but may strongly suggest another theme. As part of the annotation rules, if there is only one strong word that seems to put the topic into question, annotators were asked to simply disregard this one word in the evaluation. Given the two strong words in the second component that corroborate the first theme, with other qualifiers not strongly distinct to the concept of “reducing food waste” this is considered a *single* topic cluster.

Figure 1. Examples of individual topic clusters evaluated with our coding scheme

single		*double*		*half*		*none*	
reduce		energy		environment		flood	
food		save		health		home	
waste	1	fuel	1	risk	1	asthma	0
recycle		emission		public		watch	
reuse		money		protect		live	
american		car		prepare		mold	
landfill		standard		lot		plan	
fight	1	economy	2	asthma_trigger	0	radon	0
easy		energystar		wildfire		family	
greenhouse_gas		efficiency		create		summer	

Note: 1/1 refers to a single topic; 1/2 to potentially 2 topics; 1/0 to a cluster where only the first half suggests a coherent topic; and 0/0 to an incoherent cluster or where too many themes and issues may be present.

If a cluster suggests two potential topics or themes, annotators are asked to mark a value of 1 for the first component and a value of 2 for the second component. In the *double* topic example above, the two topics are “fuel/energy” and “money/economy”. The words “car” and “standard” may suggest a third topic, however, they may equally be part of the fuel/energy topic. Indeed, the whole cluster may suggest a single theme of “saving fuel energy for the economy”. Nevertheless, to differentiate between a cluster which is more clearly a single theme from this type of cluster, which suggest two themes or domains (though potentially under an umbrella concept), we classify this and similar clusters as a *double* topic cluster.

The third classification concerns clusters in which only the first half seems to be coherent, but not the latter half of the cluster. In figure one above, this example is given in the “environmental health” or “public environmental health risk” cluster. Although the first component may be considered a single theme as just stated, the last four or five words do not strengthen or more clearly qualify this theme—except perhaps for “wildfire” but which is also a weak term, since there are no specifically wildfire related concepts identified in the first component. Therefore, we assign a value of 0 for this and similar second components.

Lastly, there are clusters which do not suggest any coherent theme or address too many issues and ideas. These clusters may provide words that could be formed into a topic or theme but would require a strong stretch of the imagination to see a single or double set of issues. In this example in Figure 1, “asthma from floods” or “mold and homes” could be one or two related themes. However, words such as “live” and “watch”, which seem to refer to media appearance, in addition to the words “radon” and “summer” led us to classify this cluster as having no consistent topic or theme or too difficult to interpret.

After developing this classification scheme, we went through multiple rounds of coding to develop the coding rules. When we were satisfied with the classification and rules, we classified all of the 200 topic clusters, from the 5 different models of each of the 2 datasets. We then recruited 2 external annotators with basic domain knowledge, but unfamiliar with this specific social media content or topic modeling, to be trained and to independently classify a portion of the topics. We trained the first annotator in half of the topics of the McCarthy dataset, and asked this annotator to independently code the other unseen half. The same procedure was done for the Pruitt dataset. From this training/testing procedure we obtained measures of inter-coder reliability. For each of the annotators/datasets, w report Krippendorff’s alpha and Cohen’s kappa coefficients which are more conservative than simple percent agreement, as they take chance into account.

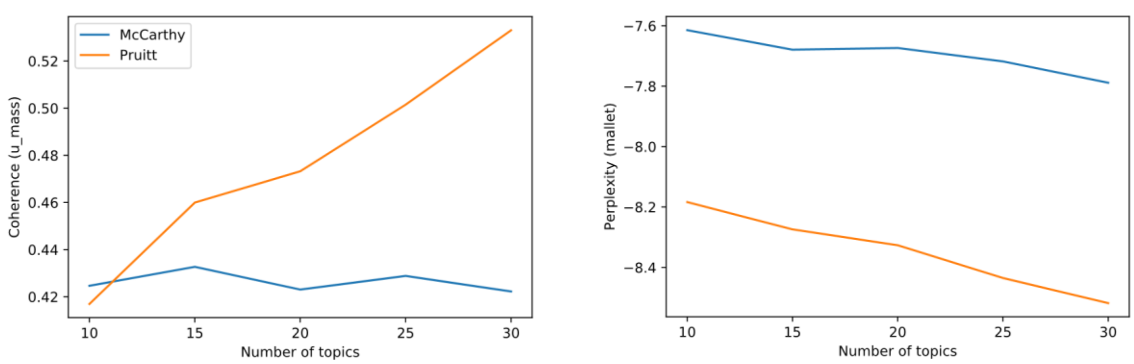
From these results, for all topic clusters in which agreement had not been reached, we discussed the coding with annotators and decided on the best classification, which is what we ultimately report in our evaluation results. From these results we selected the best topic model from each of the datasets from which to first interpret the themes and issues present in the content. We first read through the models with the greatest percentage of *single* and *double* clusters and the least percentage of *halfs* and *none* clusters. We labeled each of the coherent topic clusters in a list, and checked for partially overlapping or a hierarchy of themes (which we indented in the list). We then divided the topic labels as either referring to policies, programs or goals of the agency; or as referring to more general themes and issues—an interpretation that emerged after examining the topics. Lastly, we compared the results from the two distinct administrations, and explored variations in the content in relation to known policy priorities of the distinct administrations.

6. Results and discussion

6.1 Evaluation and reliability results

To address our RQ1, we first report perplexity measures derived from the MALLET application in which the models were trained and built, and u_mass coherence from the Gensim Coherence Model package (since it is not provided by MALLET). As may be observed in Figure 2 below, coherence for the Pruitt dataset increases approximately 25% from the 10 to the 30 topics model. However, the coherence measure remains relatively flat across all models in the McCarthy dataset. For perplexity, where the lower the value the better the model, small improvements are observed for the Pruitt dataset with the increase of the number of topics. Improvements in perplexity values are also observed for the McCarthy dataset but to a lesser extent. Figure 2 also shows that according to the metrics presented, the models of the Pruitt dataset are generally more coherent or valid than the McCarthy dataset as the number of topics increases. Similarly, models from the Pruitt dataset show overall lower perplexity values than the models of the McCarthy dataset, and therefore overall more valid or better models.

Figure 2. Perplexity and coherence measures obtained for the 10 models generated



In Table 1 below, we report evaluation and reliability results from the human annotation task. For the McCarthy dataset the Cohen's kappa and Krippendorff's alpha reliability coefficients were .73, and percent agreement was .90. For the Pruitt dataset, the kappa coefficient was .64 and alpha .63, whereas percent agreement was .83. These coefficients are relatively conservative measures of coding reliability, and, at least for the McCarthy dataset, may be considered acceptable in this type of exploratory study (Lombard et al. 2010).

Table 1. Results from the human coherence annotation task and reliability scores

	McCarthy (3,784 posts)				Pruitt (945 posts)			
	single	*double*	*half*	*none*	*single*	*double*	*half*	*none*
K	%				%			
10	40.0	30.0	10.0	20.0	50.0	20.0	0.0	30.0
15	46.7	6.7	13.3	33.3	46.7	13.3	13.3	26.7
20	65.0	0.0	15.0	20.0	50.0	10.0	10.0	30.0
25	68.0	8.0	8.0	16.0	40.0	8.0	12.0	36.0
30	66.7	6.7	13.3	13.3	43.3	3.3	26.7	26.7
% agreement	*.90*				*.83*			
kappa	*0.7346*				*0.6407*			
alpha	*0.7359*				*0.6362*			

Note: This table presents percentage of topic clusters evaluated as having single, double, half or none coherence for each of the models produced. Horizontally, for each dataset, values should add to 100%. Cohen's kappa and Krippendorff's alpha scores were obtained with Python's nltk.metrics package.

We can also observe in Table 1 that, for the models of the McCarthy dataset, a greater percentage of coherent themes were obtained from higher values of K, with a general decrease of clusters classified as none. In this dataset, although 70% of the topics from the 10-topics model could be considered coherent, as single or double, the 25- and 30-topics model yielded approximately 75% coherent topics, which is, of course, 2 or 3 times more topic clusters than the 10-topics model. This suggests that additional coherent and potentially distinct topics were generated as a result of topic models with a higher number of K for this dataset. This result contradicts the u_mass coherence metrics from Figure 2, but it suggests that our RQ1 can at least partially be answered in the affirmative: LDA topic modeling can produce interpretable clusters of themes and issues from this domain and dataset.

For the Pruitt dataset, we generally observe a smaller percentage of coherent (single) topic clusters and a higher percentage of incoherent (none) topic clusters compared to the McCarthy dataset. We also observe smaller variations across models with different values of *K*. Nevertheless, similarly to the McCarthy dataset, in the Pruitt dataset the smallest model (10-topics) and the largest model (30-topics) have similar proportions of coherent topics (approximately 70% single and double), which also suggests novel coherent topics emerged as a result of having a model with a higher value of K. However, since the Pruitt dataset is considerably smaller than the McCarthy dataset, this difference in size probably explains the greater percentage of incoherent topics being generated. Although this is not a conclusive study, these results suggest that the smaller dataset produces a substantial amount of interpretable topics, but also a substantial amount of none coherent cluster or clusters where only the first few terms form a coherent theme.

6.2 Themes and issues from the administrations

Addressing our RQ2, we report our findings in Table 2 below, which presents the similarities and variations of themes and issues across the two administrations. From the topic models produced, as previously indicated, we labeled each coherent cluster from the models with the greatest number of coherent clusters—which were the 30-topics models in both cases. We used the first two or three words of each cluster and added to the label any other term within the cluster that could be used to specify or clarify what the theme or issue was about. Therefore, except for the "and" conjunctions and slight variations in the verbs (e.g. protect to protecting) all words used to create the labels in the Table 2 below (e.g. "car emission standards", "radon test", "air quality research", etc.) were words present in the same topic cluster—though not always appearing sequentially in the cluster itself. The results of this labeling task produced more generic or thematic labels (e.g. healthy school tips, watch administrator), as well as the names of more specific policies, programs or goals (e.g. the energystar program; the Water Infrastructure Act; the "ditchthemyth" campaign related to publishing information about the Clean Water Act). We thus separated the list of labels into a list of more specific policies, goals and programs, and another of more general themes and issues—not to conclusively define them but to facilitate interpretation.

As can be observed in Table 2 below, in the McCarthy dataset, some of these specific policies and programs are the Clean Water Act and the Clean Power Plan. The word "comments" accompanies this latter issue because this topic cluster also included references to comments and a "comments period". The term ActOnClimate was listed here as a specific program because this term (also a hashtag) seemed to function as a campaign to support different actions on climate change, including climate change discussion. Similarly, SmartWay is a program of the agency which seeks to reduce car emissions by improving the efficiency of freight transportation. As the terms "smartway", "car" and "emissions" appeared in a single topic cluster, we use these defining terms in the label to better describe it. A number of more general themes and issues were also identified in the McCarthy dataset, including references to air quality research, floods and water safety, and protecting children's health, specific policies for which could not be inferred.

Table 2. List of themes, issues and policy agendas identified from the models of both datasets

McCarthy	Pruitt
Policies, goals and programs	***Policies, goals and programs***
ActOnClimate and climate change discussion	*Brownfields grants program*
Clean Power Plan and comments	*Clean Diesel program funding*
Clean Water Act (DitchTheMyth)	*Innovation fellowship*
EnergyStar program and saving energy	*Review WOTUS and Clean Power Plan*
Green small-business grants	*Superfund site cleanup*
SmartWay and car emission standards	*Water Infrastructure Act (EpaBack2Basics)*
General themes and issues	***General themes and issues***
Air quality research	*Asthma prevention*
Celebrating student environmental awards	*Car emission standards*
Chemical labels and safety	*Celebrating student environmental awards*
Climate change and the economy	*Chemical safety*
- *Reduce carbon pollution*	*Community/federal partnership*
Community building	*Deputy Administrator*
Floods and water safety	*Hurricane response (i.e. Harvey, Maria, Irma)*
Meet/Join administrator Twitter/Questions	- *Toxic generator safety*
- *Environmental blog*	*Meet/Watch administrator and discuss agenda*
- *Meet EPAers*	- *American, agenda, forward*
Reducing food waste	- *Agency progress and core mission*
Protecting children's health	*Protecting children's health*
- *Healthy school tips*	*Radon test*
- *Lead exposure prevention*	
Summer fun	
- *Summer health tips*	

In the Pruitt dataset, the more specific policies and programs identified were references to the funding of clean diesel, and thus potentially the Clean Diesel program of the agency; the Water Infrastructure Finance and Innovation Act (WIFIA), which accompanied the hashtag "epaback2basics"; Superfund site clean-up; as well as a "review" of the Waters of the U.S. (WOTUS) and the Clean Power Plan programs. The more general themes and issues included references to air quality, chemical safety, community partnerships, and asthma prevention. As with the McCarthy dataset, the Pruitt dataset also contained a reference to “meeting” or “watching” the administrator online or some other media. However, other similar topic clusters for the Pruitt dataset, also included the terms “agenda”, “forward” and “american” in one case, and references to “agency”, “progress” and “core mission” in another case. As such, we listed them as sub-topics in Table 2.

To address the second part of our RQ2, we reference some of the specific policies and programs identified in Table 2 to known priorities of the administrations documented in other sources. The Clean Power Plan was the hallmark climate change and greenhouse gas policy of the Obama administration (Engel 2015), and it was clearly communicated about on social media during the McCarthy period analyzed. At the same time, our topic models do not show any reference to this policy in the Pruitt dataset, except associated with the term “review”, which probably indicates the explicit intention to replace the Clean Power Plan, as it was eventually announced by the Trump administration (DiChristopher 2018). We also noticed that no mention of “climate change” was found in the Pruitt dataset, either as a specific policy or a general theme, although it was found multiple times in the McCarthy dataset and with explicit calls to “act on climate”. The Water Infrastructure Act (WIFIA), first passed in 2014, the Clean Diesel program, and the agency’s work on superfund sites all existed during the McCarthy administration. Nevertheless, they do not emerge in the topic models of the Obama administrator, but do so in the Pruitt dataset. As it was reported by the agency as well as other new sites, “clean diesel” and “superfund sites” (e.g. Carignan 2018; EPA 2018) were a priority of the Pruitt administration. With these results we can see that our interpretation of the topic models do suggest that changes in policy priorities are reflected in the social media content being published by the agency.

Although stark differences can be observed in the more specific policies, programs and goals expressed, it should be noted that many of the general themes and issues overlapped across the two administrations. In both datasets, general references to water safety, children’s health, community partnerships, and chemical safety are observed. For sure, there are distinct differences in terms used and issues highlighted, such as the notion of “reducing carbon pollution” in the McCarthy dataset, and the “american” theme with an emphasis on “core mission” and “back2basics” in the Pruitt dataset. Nevertheless, references to large environmental issues, as to be expected, were reflected in posts for both administrations.

References to both general themes and specific policies broadcast about on social media cannot always indicate that the issues or actions to which they refer are being de facto addressed by the administration, or addressed in the way they are communicated about. That is to say, whether or not any particular administration acts or implements policies according to the way they communicate about it, of course, may need to be evaluated in a more detailed and potentially in a case by case basis. Nevertheless, at least for some instances above, we were able to link references to specific issues and policies to actions and *de facto* priorities of the administrations as documented elsewhere. These findings thus do suggest that the administrations are using the social media not to hide their agendas and goals but to explicitly express them—and control what is not to be expressed.

6.3 Limitations and future work

This study is exploratory rather than conclusive, and there are limitations that may be addressed in future research. One of the issues concerns the use of computational coherence and perplexity measures. Although the coherence metric is the preferred method for automated topic modeling validation (Boyd-Graber et al. 2014; Fang et al. 2016), there is no guarantee that the metrics obtained from these methods and software packages will produce valid estimations of the quality of topics (Grimmer and Stewart, 2013). Perplexity has been observed as problematic in validating topic models (e.g. Chang et al. 2009) and computational coherence may depend on the corpus used, variations in parameters and the specific metric use (Röder et al., 2015). In our results, whereas the automated metrics indicated that the Pruitt models were better compared to the McCarthy models, our human validation task suggested otherwise. Nevertheless, further tests are needed to better conclude if computational coherence measures can reliably produce valid results in this domain and from this type of data.

In our human validation task, the models with higher levels of *K*, in both of the datasets, increased the overall number of coherent topics produced. However, in the best of instances, these models also generated a substantial number of topic clusters (e.g. ~25 %) that are only half coherent or not coherent at all according to our classification scheme and coding rules. This suggests that there is room for improvement with the topic modeling strategy. Variations in data pre-processing strategies (e.g. stemming, stop-words removal) can generate substantial differences in the results produced (e.g Hagen 2016). Moreover, other topic modeling approaches beyond

LDA could also be more appropriate to analyze short text documents (e.g. Shi et al. 2018). Our reliability scores can also be improved, which may be related to the number of coherent topics itself that are produced. If topic models present better models with more coherent topics, this should in turn also help improve intercoder validation agreement.

We should also note that we were relatively liberal in the interpretation of “topics” from the clusters produced, allowing themes to emerge and be qualified by the words present in the clusters. For example, we did not discard the first *single* topic listed in Figure 1 as incoherent for having the words “american”, “fight” and “easy” associated with the “reduce food waste” topic, especially given their location within the cluster. We assumed that the presence of these words in the cluster were at least partially due to their co-occurrence with the topic of food waste in the corpus. However, if applications to topic modeling in this context desire topic clusters to be composed of only a number of topic defining nouns (e.g. waste, trash, landfill) further considerations need to be made regarding the vocabulary of the corpus and the nature of the social media post as document. Ultimately, further validation could more strongly suggest if this liberal approach in interpreting the topic models is reasonable. There are a number of evaluation and validation techniques for topic models that were not discussed here (e.g. see Boyd-Graber et al. 2014; Hagen, 2016), and future work may explore other validity checks to show if these topic models can serve as representative summaries of the documents of the corpus.

7. Conclusions

Despite some limitations and the exploratory nature of this study there are some clear and interesting patterns observed from our analyses, many of which follow expectations discussed in this paper. For one, the topic models produced a substantial number of coherent and semi-coherent topic clusters (i.e. clusters with a single coherent theme, two themes, or with only the first 5 or 6 words forming a coherent theme) which could be interpreted and labeled. Moreover, our kappa and alpha reliability scores in the human validation task are potentially too conservative, since they do not recognize that some values are more likely to be generated than others (i.e. most of the models were more likely to have a coherent component rather than not, e.g. see Hagen 2016, who suggests that Gwet’s coefficient may compensate for this tendency in topic model data). Therefore, the validation task may be better than those reliability scores suggest. Also, although computational validation measures employed here did not conform with results from the human validation task, these findings are not particularly surprising and reinforce the principle that topic modeling should or may always need to be validated by humans (Grimmer and Stewart 2013). Lastly, in terms of methods, we could see clear improvements in the validation results with an increase in the number of documents in the corpus, suggesting that less than one thousand Twitter posts may not be sufficient for producing the best topic models.

Following a simple and manual approach to labeling these clusters, distinct themes and issues were interpreted and labeled as referring to more specific policies, programs as well as agency goals, in addition to generic themes and issues. The labeled content reflected categories of topics expected from a federal environmental agency, as well as more specific policies and priorities of the administrations in power. Interpretation of this content required some domain knowledge, nevertheless differences in themes and issues addressed across the two political administrations were relatively clear to observe. We supported some of this interpretation with references to knowledge about the EPA in general as well as knowledge about the agenda of these administrations as they were communicated in other contexts or known to have been adopted. This study thus showed that topic modeling can be a successful technique to summarize government social media content and retrieve valid themes and issues from a relatively large corpora of Twitter posts. Moreover, we showed that Twitter has been used, at least to some extent, communicate the policy agenda of the administration in power, in a type of policy agenda setting behavior. In this case the agenda message also involved the elimination of content explicitly referring to climate change—a potentially problematic situation since this may be one of the most important environmental issues in the world (UNDP 2007).

REFERENCES

Bates, Jo. 2013. “The Domestication of Open Government Data Advocacy in the United Kingdom: A Neo-Gramscian Analysis.” *Policy & Internet* 5 (1): 118–137. doi:10.1002/poi3.25.

Baumer, Eric P. S., David Mimno, Shion Guha, Emily Quan, and Geri K. Gay. 2017. “Comparing Grounded Theory and Topic Modeling: Extreme Divergence or Unlikely Convergence?” *Journal of the Association for Information Science and Technology* 68 (6): 1397–1410. doi:10.1002/asi.23786.

Bertot, John Carlo, Paul T. Jaeger, and Derek Hansen. 2012. “The Impact of Polices on Government Social Media Usage: Issues, Challenges, and Recommendations.” *Government Information Quarterly* 29 (1): 30–40. doi:10.1016/j.giq.2011.04.004.

Blei, David M. 2012. “Topic Modeling and Digital Humanities.” *Journal of Digital Humanities* 2 (1). http://journalofdigitalhumanities.org/2-1/topic-modeling-and-digital-humanities-by-david-m-blei/.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan. 2003. “Latent Dirichlet Allocation.” *Journal of Machine Learning Research* 3 (Jan): 993–1022.

Boyd-Graber, Jordan, David Mimno, and David Newman. 2014. “Care and Feeding of Topic Models: Problems, Diagnostics, and Improvements.” In *Handbook of Mixed Membership Models and Their Applications*, edited by Edoardo M. Airoldi, 40.

Buckley, Madeline. 2016. “Can Police Delete Facebook Posts? Experts Say No.” *Indianapolis Star*, July 6. https://www.indystar.com/story/news/2016/07/06/can-police-delete-facebook-posts-experts-say-no/86766648/.

Burscher, Bjorn, Rens Vliegenthart, and Claes H. De Vreese. 2015. “Using Supervised Machine Learning to Code Policy Issues: Can Classifiers Generalize across Contexts?” Edited by Dhavan V. Shah, Joseph N. Cappella, and W. Russell Neuman. *The ANNALS of the American Academy of Political and Social Science* 659 (1): 122–131. doi:10.1177/0002716215569441.

Burscher, Bjorn, Rens Vliegenthart, and Claes H. de Vreese. 2016. “Frames Beyond Words: Applying Cluster and Sentiment Analysis to News Coverage of the Nuclear Power Issue.” *Social Science Computer Review* 34 (5): 530–545. doi:10.1177/0894439315596385.

C2ES. 2018. “Regulating Power Sector Carbon Emissions.” *Center for Climate and Energy Solutions*. https://www.c2es.org/content/regulating-power-sector-carbon-emissions/.

Canel, Maria Jose, and Karen Sanders. 2012. “Government Communication: An Emerging Field in Political Communication Research.” *The Sage Handbook of Political Communication*, 85–96.

Carignan, Sylvia. 2018. “Superfund Program Could Lose Prominence Without Pruitt.” *Bloomberg BNA*, July 5. https://www.bna.com/superfund-program-lose-n73014477103/.

Ceron, Andrea, Luigi Curini, and Stefano M. Iacus. 2016. “First- and Second-Level Agenda Setting in the Twittersphere: An Application to the Italian Political Debate.” *Journal of Information Technology & Politics* 13 (2): 159–174. doi:10.1080/19331681.2016.1160266.

Ceron, Andrea, and Fedra Negri. 2016. “The ‘Social Side’ of Public Policy: Monitoring Online Public Opinion and Its Mobilization During the Policy Cycle.” *Policy & Internet* 8 (2): 131–147. doi:10.1002/poi3.117.

Chang, Jonathan, Sean Gerrish, Chong Wang, Jordan L. Boyd-graber, and David M. Blei. 2009. “Reading Tea Leaves: How Humans Interpret Topic Models.” In *Advances in Neural Information Processing Systems 22*, edited by Y. Bengio et al., 288–296. Curran Associates, Inc. http://papers.nips.cc/paper/3700-reading-tea-leaves-how-humans-interpret-topic-models.pdf.

Collin, Robert W. 2006. *The Environmental Protection Agency : Cleaning up America’s Act*. Understanding Our Government. Westport, Conn. : Greenwood Press.

Conrad, James W. Jr. 2002. “The Information Quality Act - Antiregulatory Costs of Mythic Proportions Information Disclosures by Government: Data Quality and Security Concerns Symposium.” *Kansas Journal of Law & Public Policy* 12: 521–558.

Conway-Silva, Bethany A., Christine R. Filer, Kate Kenski, and Eric Tsetsi. 2018. “Reassessing Twitter’s Agenda-Building Power: An Analysis of Intermedia Agenda-Setting Effects During the 2016 Presidential Primary Season.” *Social Science Computer Review* 36 (4): 469–483. doi:10.1177/0894439317715430.

Crump, Jeremy. 2011. “What Are the Police Doing on Twitter? Social Media, the Police and the Public.” *Policy & Internet* 3 (4): 1–27. doi:10.2202/1944-2866.1130.

Davenport, Coral, Lisa Friedman, and Maggie Haberman. 2018. “E.P.A. Chief Scott Pruitt Resigns Under a Cloud of Ethics Scandals.” *The New York Times*, July 7, sec. Climate. https://www.nytimes.com/2018/07/05/climate/scott-pruitt-epa-trump.html.

DePaula, Nic, and Ersin Dincelli. 2016. “An Empirical Analysis of Local Government Social Media Communication: Models of e-Government Interactivity and Public Relations.” In *Proceedings of the 17th International Conference on Digital Government Research*, 348–356. New York, NY, USA: ACM.

DePaula, Nic, and Ersin Dincelli. 2018. “Information Strategies and Affective Reactions: How Citizens Interact with Government Social Media Content.” *First Monday* 23 (4). doi:10.5210/fm.v23i4.8414.

DePaula, Nic, Ersin Dincelli, and Teresa M. Harrison. 2018. “Toward a Typology of Government Social Media Communication: Democratic Goals, Symbolic Acts and Self-Presentation.” *Government Information Quarterly* 35 (1): 98–108. doi:10.1016/j.giq.2017.10.003.

DiChristopher, Tom. 2018. “Trump Administration to Replace Obama’s Clean Power Plan.” *CNBC*. https://www.cnbc.com/2018/08/21/trump-administration-to-replace-obamas-clean-power-plan.html.

DiMaggio, Paul, Manish Nag, and David Blei. 2013. “Exploiting Affinities between Topic Modeling and the Sociological Perspective on Culture: Application to Newspaper Coverage of U.S. Government Arts Funding.” *Poetics*, Topic Models and the Cultural Sciences, 41 (6): 570–606. de

EDGI. 2018. “Changing the Digital Climate: How Climate Change Web Content Is Being Censored under the Trump Administration.” https://100days.envirodatagov.org/changing-digital-climate/.

Engel, Kirsten H. 2015. “EPA’s Clean Power Plan: An Emerging New Cooperative Federalism?” *Publius: The Journal of Federalism* 45 (3): 452–474. doi:10.1093/publius/pjv025.

Entman, Robert M. 2007. “Framing Bias: Media in the Distribution of Power.” *Journal of Communication* 57 (1): 163–173. doi:10.1111/j.1460-2466.2006.00336.x.

EPA. 2013. “Our Mission and What We Do.” Overviews and Factsheets. *U.S. Environmental Protection Agency*. https://www.epa.gov/aboutepa/our-mission-and-what-we-do.

EPA. 2014. “Using Social Media to Communicate with the Public.” EPA Information Procedures. *U.S. Environmental Protection Agency*. https://www.epa.gov/sites/production/files/2013-11/documents/comm_public.pdf.

EPA. 2018. “EPA Provides $1.3 Million to States for Diesel Reduction Efforts.” Speeches, Testimony and Transcripts. *U.S. Environmental Protection Agency*. https://www.epa.gov/newsreleases/epa-provides-13-million-states-diesel-reduction-efforts.

Fang, Anjie, Craig Macdonald, Iadh Ounis, and Philip Habel. 2016. “Topics in Tweets: A User Study of Topic Coherence Metrics for Twitter Data.” In *Advances in Information Retrieval*, 492–504. Lecture Notes in Computer Science. Springer, Cham. doi:10.1007/978-3-319-30671-1_36.

Girolami, Mark, and Ata Kaban. 2003. “On an Equivalence between PLSI and LDA.” In *Proceedings of the ACM Conference on Research and Development in Information Retrieval*, 433–434. Toronto, CA. http://www.cs.bham.ac.uk/~axk/sigir2003_mgak.pdf.

Graber, Doris A. 2003. *The Power Of Communication: Managing Information In Public Organizations*. 1st edition. Washington, D.C: CQ Press.

Gray, Naomi. 2017. “Why Twitter Adjusted Its Monthly Active User Numbers.” *Market Realist*, November 24. https://marketrealist.com/2017/11/why-twitter-adjusted-its-monthly-active-user-numbers.

Grimmer, Justin. 2010. “A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in Senate Press Releases.” *Political Analysis* 18 (1): 1–35. doi:10.1093/pan/mpp034.

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 (03): 267–297.

Grunig, James E., and L. Grunig, eds. 1992. “Models of Public Relations and Communication.” In *Excellence in Public Relations and Communication Management*, 285–326. Hillsdale, N.J: Routledge.

Hagen, Loni. 2016. “Topic Modeling for E-Petition Analysis : Interpreting Petitioners’ Policy Priorities.” Thesis (Ph. D), Department of Informatics, University at Albany, State University of New York.

Hagen, Loni. 2018. “Content Analysis of E-Petitions with Topic Modeling: How to Train and Evaluate LDA Models?” *Information Processing & Management*, May. doi:10.1016/j.ipm.2018.05.006.

Hagen, Loni, Thomas Keller, Stephen Neely, Nic DePaula, and Claudia Robert-Cooperman. 2017. “Crisis Communications in the Age of Social Media: A Network Analysis of Zika-Related Tweets.” *Social Science Computer Review*, August, 0894439317721985. doi:10.1177/0894439317721985.

Hansson, Sten. 2015a. “Discursive Strategies of Blame Avoidance in Government: A Framework for Analysis.” *Discourse & Society* 26 (3): 297–322. doi:10.1177/0957926514564736.

Hansson, Sten. 2015b. “Calculated Overcommunication: Strategic Uses of Prolixity, Irrelevance, and Repetition in Administrative Language.” *Journal of Pragmatics* 84 (Supplement C): 172–188. doi:10.1016/j.pragma.2015.05.014.

Hansson, Sten. 2017. “Government Communication.” In *The Routledge Handbook of Language and Politics*, edited by Ruth Wodak and Bernhard Forchtner, 1 edition, 326–341. Milton Park, Abingdon, Oxon ; New York, NY: Routledge.

Harder, Raymond A., Julie Sevenans, and Peter Van Aelst. 2017. “Intermedia Agenda Setting in the Social Media Age: How Traditional Players Dominate the News Agenda in Election Times.” *The International Journal of Press/Politics* 22 (3): 275–293. doi:10.1177/1940161217704969.

Harrison, Teresa M., Catherine Dumas, Nic DePaula, Tim Fake, Will May, Akanksha Atrey, Jooyeon Lee, Lokesh Rishi, and S. S. Ravi. 2017. “E-Petitioning and Online Media: The Case of #BringBackOurGirls.” In *Proceedings of the 18th Annual International Conference on Digital Government Research*, 11–20. Dg.o ’17. New York, NY, USA: ACM.

Hood, Christopher. 1983. *The Tools of Government*. London: Palgrave Macmillan.

Hood, Christopher. 2011. *The Blame Game: Spin, Bureaucracy, and Self-Preservation in Government*. Princeton University Press.

Hood, Christopher, and Helen Margetts. 2007. *The Tools of Government in the Digital Age*. 2nd edition. Basingstoke: Palgrave.

Jacobs, Jennifer, and Jennifer A. Dlouhy. 2018. “Pruitt Was Told to Resign in Call From White House, Sources Say.” *Bloomberg.Com*, July 6. https://www.bloomberg.com/news/articles/2018-07-06/pruitt-said-to-have-been-told-to-resign-in-call-from-white-house.

Koltsova, Olessia, and Sergei Koltcov. 2013. “Mapping the Public Agenda with Topic Modeling: The Case of the Russian Livejournal.” *Policy & Internet* 5 (2): 207–227. doi:10.1002/1944-2866.POI331.

Liu, Brooke Fisher, and J. Suzanne Horsley. 2007. “The Government Communication Decision Wheel: Toward a Public Relations Model for the Public Sector.” *Journal of Public Relations Research* 19 (4): 377–393. doi:10.1080/10627260701402473.

Lombard, Matthew, Jennifer Snyder-Duch, and Cheryl Campanella Bracken. 2010. “Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research Projects.” *Matthew Lombard: Intercoder Reliability*. http://matthewlombard.com/reliability/.

Magnusson, Måns, Richard Öhrvall, Katarina Barrling, and David Mimno. 2018. “Voices from the Far Right: A Text Analysis of Swedish Parliamentary Debates.” *SocArXiv*, April. doi:10.31235/osf.io/jdsqc.

Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze. 2008. *Introduction to Information Retrieval*. New York: Cambridge University Press.

Margetts, Helen, Peter John, Scott Hale, and Taha Yasseri. 2016. *Political Turbulence: How Social Media Shape Collective Action*. Princeton, New Jersey: Princeton University Press.

McCallum, Andrew K. 2002. “MALLET: A Machine Learning for Language Toolkit.” http://mallet.cs.umass.edu.

McCammon, Muira. 2018. “Federal Agencies Have Far Too Much Leeway to Delete Tweets.” *Slate Magazine*, April 17. https://slate.com/technology/2018/04/can-federal-agencies-really-just-delete-tweets.html.

McCombs, Maxwell E., and Donald L. Shaw. 1972. “The Agenda-Setting Function of Mass Media.” *Public Opinion Quarterly* 36 (2): 176–187. doi:10.1086/267990.

Mehrotra, Rishabh, Scott Sanner, Wray Buntine, and Lexing Xie. 2013. “Improving LDA Topic Models for Microblogs via Tweet Pooling and Automatic Labeling.” In , 889. ACM Press. doi:10.1145/2484028.2484166.

Meijer, Albert, and Marcel Thaens. 2013. “Social Media Strategies: Understanding the Differences between North American Police Departments.” *Government Information Quarterly* 30 (4): 343–350.

Mergel, Ines. 2013. “Social Media Adoption and Resulting Tactics in the U.S. Federal Government.” *Government Information Quarterly* 30 (2): 123–130. doi:10.1016/j.giq.2012.12.004.

Mimno, David, Hanna M. Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum. 2011. “Optimizing Semantic Coherence in Topic Models.” In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 262–272. EMNLP ’11. Stroudsburg, PA, USA: Association for Computational Linguistics. http://dl.acm.org/citation.cfm?id=2145432.2145462.

Najafabadi, Mahdi M., and Robert J. Domanski. 2018. “Hacktivism and Distributed Hashtag Spoiling on Twitter: Tales of the #IranTalks.” *First Monday* 23 (4). doi:10.5210/fm.v23i4.8378.

Obama, Barack. 2009. “Memorandum for the Heads of Executive Departments and Agencies: Transparency and Open Government.” *The White House*. https://obamawhitehouse.archives.gov/the-press-office/transparency-and-open-government.

Pew Research Center. 2018. “Social Media Use in 2018.” *Pew Research Center: Internet, Science & Technology*. http://www.pewinternet.org/2018/03/01/social-media-use-in-2018/.

Quinn, Kevin M., Burt L. Monroe, Michael P. 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.

Röder, Michael, Andreas Both, and Alexander Hinneburg. 2015. “Exploring the Space of Topic Coherence Measures.” In , 399–408. ACM Press. doi:10.1145/2684822.2685324.

Rogers, Everett, and James Dearing. 1988. “Agenda-Setting Research: Where Has It Been, Where Is It Going?” *Annals of the International Communication Association* 11 (1): 555–594.

Rosen-Zvi, Michal, Thomas Griffiths, Mark Steyvers, and Padhraic Smyth. 2012. “The Author-Topic Model for Authors and Documents.” *ArXiv:1207.4169 [Cs.IR]*, July. https://arxiv.org/abs/1207.4169.

Scheufele, Dietram A., and David Tewksbury. 2007. “Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models.” *Journal of Communication* 57 (1): 9–20. doi:10.1111/j.0021-9916.2007.00326.x.

Schofield, Alexandra, and David Mimno. 2016. “Comparing Apples to Apple: The Effects of Stemmers on Topic Models.” *Transactions of the Association for Computational Linguistics* 4: 14.

Shapiro, Matthew, and Libby Hemphill. 2016. “Politicians and the Policy Agenda: Does Use of Twitter by the U.S. Congress Direct New York Times Content?” *Policy & Internet* 9 (1): 109–132. doi:10.1002/poi3.120.

Shi, Tian, Kyeongpil Kang, Jaegul Choo, and Chandan K. Reddy. 2018. “Short-Text Topic Modeling via Non-Negative Matrix Factorization Enriched with Local Word-Context Correlations.” In , 1105–1114.

Smith, Jason. 2016. “Which U.S. Government Agency Is Most Active on Social Media?” *Spredfast*. https://www.spredfast.com/social-marketing-blog/which-us-government-agency-most-active-social-media.

Stephens, Chelsea. 2017. “Don’t Delete That Post! Why Government Agencies Must Archive Social Media.” *Digital Services Georgia*. https://digitalservices.georgia.gov/blog/2017-09-20/don%E2%80%99t-delete-post-why-government-agencies-must-archive-social-media.

Therkelsen, Jonas Foyn, and Asbjørn Ottesen Steinskog. 2016. “Characterizing Twitter Data Using Sentiment Analysis and Topic Modeling.” Thesis, Master of Science in Computer Science, Norwegian University of Science and Technology.

Törnberg, Anton, and Petter Törnberg. 2016. “Muslims in Social Media Discourse: Combining Topic Modeling and Critical Discourse Analysis.” *Discourse, Context & Media* 13 (Part B): 132–142.

UNDP. 2007. “Fighting Climate Change: Human Solidarity in a Divided World.” In *Human Development Report 2007/2008*, 1–18. United Nations Development Programme. London: Palgrave Macmillan UK.

U.S. Digital Registry. 2018. “U.S. Digital Registry.” *DigitalGov*. https://usdigitalregistry.digitalgov.gov/.

Valle-Cruz, David, Rodrigo Sandoval-Almazan, and J. Ramon Gil-Garcia. 2016. "Citizens' Perceptions of the Impact of Information Technology Use on Transparency, Efficiency and Corruption in Local Governments." *Information Polity* 21 (3): 321–334. doi:10.3233/IP-160393.

Varol, Onur, Emilio Ferrara, Clayton A. Davis, Filippo Menczer, and Alessandro Flammini. 2017. "Online Human-Bot Interactions: Detection, Estimation, and Characterization." *ArXiv:1703.03107 [Cs]*, March. http://arxiv.org/abs/1703.03107.

Waters, Richard D., and Jensen M. Williams. 2011. "Squawking, Tweeting, Cooing, and Hooting: Analyzing the Communication Patterns of Government Agencies on Twitter." *Journal of Public Affairs* 11 (4): 353–363. doi:10.1002/pa.385.

Weaver, David H. 2007. "Thoughts on Agenda Setting, Framing, and Priming." *Journal of Communication* 57 (1): 142–147. doi:10.1111/j.1460-2466.2006.00333.x.

Yang, Xinxin, Bo-Chiuan Chen, Mrinmoy Maity, and Emilio Ferrara. 2016. "Social Politics: Agenda Setting and Political Communication on Social Media." *ArXiv:1607.06819 [Physics]* 10046: 330–344. doi:10.1007/978-3-319-47880-7_20.

Yau, Chyi-Kwei, Alan Porter, Nils Newman, and Arho Suominen. 2014. "Clustering Scientific Documents with Topic Modeling." *Scientometrics* 100 (3): 767–786. doi:10.1007/s11192-014-1321-8.

Yudof, Mark G. 1979. "When Governments Speak: Toward a Theory of Government Expression and the First Amendment." *Texas Law Review* 57: 57.