Regulation & Governance

Regulation & Governance (2019) doi:10.1111/rego.12282

When does open government shut? Predicting government responses to citizen information requests

Benjamin E. Bagozzi

Department of Political Science & International Relations, University of Delaware, Newark, DE, USA

Daniel Berliner

Department of Government, London School of Economics, London, UK

Zack W. Almquist

Department of Sociology & eScience Institute, University of Washington, Seattle, WA, USA

Abstract

Methods for the analysis of "big data" on citizen-government interactions are necessary for theoretical assessments of bureaucratic responsiveness. Such big data methods also stand to benefit practitioners' abilities to monitor and improve these emerging transparency mechanisms. We consider supervised latent Dirichlet allocation (sLDA) as a potential method for these purposes. To this end, we use sLDA to examine the Mexican government's (non)responsiveness to *all* public information requests filed with the federal Mexican government during the 2003–2015 period, and to identify the request topics most associated with (non)responsiveness. Substantively, our comparisons of the topics that are most highly predictive of responsiveness and nonresponsivess indicate that political sensitivity plays a large and important role in shaping official behavior in this arena. We thus conclude that sLDA provides unique advantages for, and insights into, the analysis of (i) textual records of citizen–government interactions and (ii) bureaucratic (non)responsiveness to these interactions.

Keywords: accountability, information, Mexico, prediction, text analysis, transparency.

1. Introduction

Across the world, citizens and their governments are increasingly interacting through electronic and online mediums. Virtually every national government now offers a range of public services over the web, and over 90 countries currently provide and manage fully integrated online service portals for their citizens (United Nations 2016). Well over 100 countries have now similarly adopted access-to-information (ATI) laws that allow citizens to request information, documents, or data from their governments, and that require officials to respond (Michener 2011; Berliner 2014). An increasing number of these ATI laws also incorporate online systems for filing requests and receiving responses (Fumega 2014). These new forms of citizen–government interaction speak directly to issues of bureaucratic responsiveness, transparency, and accountability, and regularly generate massive amounts of administrative data in the forms of text and outcomes (Connelly *et al.* 2016; Pew 2018). For example, under Mexico's ATI law, citizens now file and receive responses to well over 100,000 requests per year from federal government entities. The US Freedom of Information Act (FOIA) system is larger still, and receives approximately 700,000 annual requests a year, though with a mounting backlog (Noveck 2016).

The availability of citizen request data of this scale opens new and innovative avenues for research, and has the potential to contribute *both* to scholarly understandings of bureaucratic responsiveness *and* to monitoring and evaluation efforts by policymakers and advocacy groups.[1] However, to achieve these aims, researchers require analytic tools that can jointly examine the substance of citizen request texts alongside the responses they generate. Likewise, and given the immense volume and complexity of information handled within modern systems for

Correspondence: Daniel Berliner, Department of Government, London School of Economics, Houghton Street, London, WC2A 2AE, UK. Email: d.berliner@lse.ac.uk

Accepted for publication 29 August 2019.

citizen–government interaction, officials also increasingly need to identify and leverage such tools if they intend to effectively monitor official responsiveness, performance, and compliance. These challenges have not gone unnoticed by past research (Fox *et al.* 2011; Noveck 2016; Lavertu 2017; Pew 2018).

In light of these issues, this article makes three primary contributions. First, we demonstrate the benefits and possibilities of using "big data" on citizen–government interactions to study government responsiveness and accountability in relatively new, technology-enabled settings characterized by high levels of volume, velocity, and variety in available data. Second, we evaluate the utility of an under-utilized method—supervised latent Dirichlet allocation—for exploratory and inductive research in understanding interactions characterized by "open/shut" responses. Third, we illustrate how these results can also shed light on theoretical debates over the motivations for (and against) bureaucratic responsiveness.

To do so, we consider data from one specific ATI regime for which we have access to comprehensive records of every single citizen request for government information: the case of the Mexican federal government. Following Mexico's landmark 2002 ATI law, each and every individual ATI request filed with federal government agencies has been made publicly available—now over 1 million requests in total. Alongside the text of each individual request, related metadata is also stored and made available, including information on the nature of the Mexican government's response to each request. Analyzing the Mexican federal government's responsiveness to these individual ATI request texts accordingly allows us to simultaneously assess (i) existing explanations of bureaucratic responsiveness in this context and (ii) the applicability of several recently developed text-as-data methods for this endeavor.

In undertaking these tasks, we evaluate a supervised machine learning method that we argue is ideal for the interpretation and discovery of *political cues* associated with bureaucratic (non)responsiveness in "big data" settings characterized by large corpora of unstructured texts and associated metadata. Specifically, we propose supervised latent Dirichlet allocation (sLDA) as an optimal method for the tasks described earlier, and evaluate this approach against several more widely used supervised classifiers. While an extensive literature now exists on the development and use of supervised machine learning methods for the classification of political and social text (Hopkins & King 2010; Jurka *et al.* 2013), such methods provide researchers with limited resources for undertaking rigorous evaluations of the text dimensions that are identified as most predictive of a given outcome. At the same time, unsupervised topic modeling methods developed in the social sciences and elsewhere enable users to uncover, and to theoretically evaluate, the common themes underlying large text corpora, but generally do not allow researchers to relate such quantities to an external, document-level *outcome measure* during estimation.

As we demonstrate later, sLDA provides social scientists with a means of jointly leveraging the advantages of both supervised learning and topic modeling approaches. In doing so, it allows one to achieve superior predictive accuracy than more common supervised machine learning approaches, while also offering the benefits of thematic (i.e. topical) interpretability and discovery. Together, we contend that these features make sLDA a promising framework for inductive "needle-in-a-haystack"-type research goals. These tasks commonly arise in instances where researchers or bureaucrats seek to sift through large corpora of texts in search of small but thematically coherent subsets of cases, documents, or themes that uniquely exhibit high predictive leverage vis-à-vis some external outcome. Other possible applications for this type of research include any setting where textual records are subject to positive or negative outcomes, such as complaints, case files, investigations, awards, or even censorship.

We illustrate the methodological and theoretical merits of such a task in the context of Mexican ATI requests through a demonstration of how sLDA allows one to identify a small but potentially politicized subset of ATI request topics that in turn reliably predict the relatively infrequent event of a denied request. Our proposed research framework is particularly appropriate for this setting, given that past research suggests many information requests may be mundane or easily dealt with through ordinary procedures, whereas a select "tip of the iceberg" may be far more politically relevant (Michener & Worthy 2015). Our findings in these regards suggest that instances of highly nonresponsive behavior among government may be shaped by perceptions of political threat or utility, as well as by resource limitations or legal restrictions. Importantly, while many past studies of ATI responsiveness (ben Aaron *et al.* 2017; Lagunes & Pocasangre 2019; Wood & Lewis 2017; Worthy *et al.* 2017; Poole 2018; Špáč *et al.* 2018; Grimmelikhuijsen *et al.* 2019) have used experimental approaches, submitting similar requests across multiple government bodies, this observational approach accordingly allows us to account for the full real-world diversity of citizen uses, many of which are too context-dependent or sensitive to feasibly or ethically deploy in research settings.

This article proceeds as follows. After a brief overview of the ATI system and bureaucratic responsiveness in Mexico, we describe our corpus of Mexican ATI request texts, the steps used to preprocess this corpus for analysis, and the sLDA method. We then use sLDA to predict official (non)responsiveness toward Mexican ATI requests in both an in-sample and out-of-sample context, as measured by the probability of a denied request. Alongside these endeavors, we assess our sLDA-derived topics for political relevance, and find that the topics that are most strongly associated with *nonresponsiveness* do indeed exhibit markedly more political sensitivity than do the topics most associated with high *responsiveness*. One the other hand, we also find that one topic associated with high responsiveness is related to employee absenteeism and potential corruption in the education sector. This association suggests an important distinction between exposing wrongdoing at the national level—where it may be politically damaging—and at the local level—where disclosure may serve the interests of the political center as a form of "fire alarm" monitoring (McCubbins & Schwartz 1984). We also compare the sLDA approach to several alternative classifiers, and find that its performance is superior to these more widely used supervised methods. Our conclusion discusses the implications of our findings, and the broader promise of sLDA methods for big data social science research.

2. Background

Democratic institutions are founded on the notion of responsive government, but responsiveness is usually limited and incomplete. Many scholars have studied why political actors may be more or less responsive in different circumstances—both at a macro-scale in terms of how government policies and spending respond to the preferences of the median voter (Cleary 2007; Golden & Min 2013; Herrera 2017) or to elected legislatures (McCubbins *et al.* 1987; Saltzstein 1992; West & Raso 2012), and at a micro-scale in terms of individual citizen–government interactions (Chaney & Saltzstein 1998; Balla 2000; Yang & Callahan 2007; Butler & Broockman 2011; White *et al.* 2015; McClendon 2016). Often the latter approach focuses on ATI requests (Peisakhin & Pinto 2010; ben Aaron *et al.* 2017; Lagunes & Pocasangre 2019; Wood & Lewis 2017; Worthy *et al.* 2017; Poole 2018; Spáč *et al.* 2018; Grimmelikhuijsen *et al.* 2019), which correspond to individuals' or organizations' requests for public information from their government, and that government's degree of responsiveness to those requests in terms of information provided. In these literatures, explanations for responsiveness usually include capacity, resources, organizational cultures, social barriers or discrimination, as well as—importantly—political incentives.

Below we examine this form of responsiveness within one specific ATI system: bureaucratic responses to public information requests in Mexico. To do so, we use a comprehensive data set of over 1 million information requests filed with federal Mexican government agencies. These correspond to queries made by individual citizens, legal representatives, businesses, and NGOs to specific Mexican federal government agencies, and cover, for example, requests for information on government spending, environmental disputes, or police records. Due to the unique online information platform created by Mexico's 2002 *Ley Federal de Transparencia y Acceso a la Información Pública Gubernamental* (LFTAIPG), the text of each of these requests, along with associated metadeta, has been made publicly available going back to mid-2003. One of the most innovative features of the LFTAIPG was its online information platform, which is used to manage requests, responses, and appeals. Citizens file requests and receive responses primarily through this system, which was ultimately named INFOMEX. Where citizens file written requests, these are still managed through INFOMEX. In these cases, agency officials enter the relevant metadata information, and then scan and upload the actual request (text) as an attached image file. Over the 2003–2015 period, nonelectronic requests initially encompassed roughly 7% of all INFOMEX requests, before declining to represent approximately 2% of all INFOMEX requests in more recent years.

While LFTAIPG has improved Mexican citizens' abilities to seek government information, agency officials nevertheless have several reasons to remain strategically nonresponsive to ATI requests. Above and beyond general workload concerns and inapplicable requests, officials charged with responding to requests may often be sensitive to the political potential of particular requests. To the extent that requested information may risk adverse consequences for a given agency or the governing party—through, for example, negative media attention, advocacy group campaigns, or even corruption investigations—agency officials may refuse to provide legitimately requested information.[2] Potential examples include efforts to expose bureaucratic inefficiency, financial sector policies, bias, patronage, or corruption in public procurement or employment practices, or even matters related to the Drug War and cartel activity. Because the officials responding to individual requests are housed within specific federal agencies, they have clear incentives to protect the reputation of their government agency, the federal government on the whole, and/or the governing party.[3] The fact that several official response justifications provide agency officials with an ability to "mask" politically motivated denials (Fox *et al.* 2011) only further incentivizes this potential behavior.

On the other hand, there are also theoretical reasons to expect that, under certain circumstances, officials may actively prefer disclosure. Some scholars have suggested that transparency mechanisms may serve as a form of "fire alarm monitoring" (McCubbins & Schwartz 1984), particularly where it is in the interests of the central government to identify and address wrongdoing at the local level *before* it becomes broadly public (Michener 2015; Distelhorst 2017; Schnell 2017). Busuioc and Lodge (2016) also argue that accountability-enhancing measures can serve bureaucratic organizations' goals of cultivating positive reputations. Empirically, the Mexican case might also be expected to be less prone to politicized responsiveness, given its widely hailed legal and procedural best practices, online request portal, and active and independent information commission (Bogado *et al.* 2007; Bookman & Amparán 2009; Berliner & Erlich 2015). Indeed, Lagunes and Pocasangre (2019) found little evidence of discrimination on the basis of requester identities. These insights suggest that political biases in government nonresponsiveness to ATI requests may not be as straightforward as one might initially expect, making topic modeling an ideal strategy for discovering ATI request themes that are more or less associated with government (non) responsiveness.

3. Measuring (non)responsiveness

We now turn to discuss how our key outcome variable of interest was coded for the Mexico case. As with any study of ATI responsiveness, coding the dependent variable is complicated by the specifics of the ATI legal regime, including the possibilities of both legally valid denials, and noncompliant disclosures (Lagunes & Pocasangre 2019; Wood & Lewis 2017; Worthy *et al.* 2017). We focus on clear denials to ATI requests that could have received a fulfilled response in defining our "denied request" indicator.

This binary indicator accordingly endeavors to encompass only requests that were denied for potentially political or noncompliant reasons. As such, we do not include every potential justification for information not being provided as a "denied request," given that some justifications—such as "this request could not be processed," "this request does not fall under the purview of Mexico's ATI law," or "not the competency of this entity"—predominantly (although not exclusively) arise in cases where the requester failed to properly upload information (e.g. their referenced attachment), had already submitted an identical request that was responded to, requested information that was verifiably not covered under Mexico's ATI laws, or requested information that was held by another government ministry or agency.

Instead, we classify any ATI request refusal that received justifications of "the requested information does not exist," "the requested information is classified or confidential," or "the requested information is partially classified or confidential" as a "denied request." We include the latter ("classified/confidential") categories given the potential for these categories to be overused in cases of agents withholding politically sensitive information. With regards to the former category (i.e. "does not exist"), research suggests that Mexican "agencies have discovered that this is the least risky way to deny requests for information in cases where they did not want it released or when assembling it would be a large burden," given that "in contrast, the burden of proof is on the agency if it claims that information requested is 'confidential' or 'reserved'" (Fox *et al.* 2011, p. 14). Classified or confidential responses may also be for legitimate reasons, such as national security, privacy, or other legal exemptions. However, these too may be misused by officials to avoid disclosure. For example, Almanzar, Aspinwall and Crow (2018, p. 11) find evidence that security exemptions are often inappropriately used, suggesting that some agencies are "not certain (or truthful) about whether an issue is truly a security issue."

After retaining the subset of denied requests that is delineated earlier, we take care to omit any cases still pending as of August 2015, to avoid conflating denied requests with requests still awaiting response. Our final "denied request" indicator is fairly imbalanced with roughly 10% of retained requests receiving a "denied request" by our definition.

4. Information request features

We focus on the ATI request texts themselves as our primary predictors.[4] These texts correspond to each requester's own open-ended description of the specific information that they are requesting. Because public officials are the primary responders to these requests, we believe that the themes found across these requests, and their potential degrees of politicization, will help to predict bureaucratic (non)responsiveness.

We thus downloaded all requests from Mexico's publicly available online information request interface for the period from June 2003 to August 2015. As mentioned earlier, this allows us to recover every federal-level ATI request made within Mexico during this time period, including requests that were originally submitted non-electronically. While most requesters described the nature of their requests within the designated field, a smaller subset (roughly 13%) included a portion or all of their request as an attachment. Because omitting these would yield an unrepresentative sample, we separately downloaded each attachment, digitized the text, and appended it onto the main request text field.[5] Consistent with past research (e.g. Bagozzi *et al.* 2016; Berliner *et al.* 2018), we then truncated all remaining texts from the thousandth string onward. This created our primary corpus, which was further preprocessed using standard steps for the automated analysis of political texts. These steps are described in detail in our Appendix S1, and together produced a corpus of 1,003,756 requests.

We next appended the names of each request's designated federal government agency to our processed texts. Although Mexico's publicly available ATI data omits requesters' individual identities, each information request contains metadata on the federal government agency (hereafter, target agency) that the requester made their request to. As these agencies vary in their levels of politicization and administrative capacity, we anticipate that a request's agency-designation, like a request's textual content, will influence the degree of (non)responsiveness to a given request. Agency information was therefore included as an additional field within our request text input data by appending all agency names as unique features within our request text corpus.[6] Together these agency names encompass roughly 300 distinct Mexican federal ministries and other agencies for our sample. Hereafter, we refer to these combined "request + agency" documents as "request documents" for convenience. In the robustness section, we additionally evaluate the unique contribution of this feature to our predictive tasks. We specifically do so by first estimating comparable models that omit this agency information, and then evaluating changes in predictive accuracy.

5. Supervised latent Dirichlet allocation

Topic models have been shown to be highly valid for the discovery of latent thematic content within information request texts (Berliner *et al.* 2018; Berliner *et al.* 2015). As such, the present article evaluates the utility of sLDA models (Blei & Mcauliffe 2003) for the prediction of bureaucratic (non)responsiveness to these same request texts. As a supervised topic model, sLDA is designed to identify groupings of words (or word-stems) that are most predictive of a document-indexed response variable. These groupings—hereafter referred to as topics—are estimated from a model that treats each document as containing a finite mixture of underlying topics, where the topics themselves are specified as an infinite mixture over a latent set of topic probabilities. One's document-level responses are then regressed on these estimated topic frequencies in a manner that restricts responses to be non-exchangeable with words (Blei & Mcauliffe 2003).

For this sLDA framework, our ATI request texts are treated as mixtures of multiple *latent topics*. Each topic can then be represented by a subset of words contained in within (and across) our ATI request documents. For our Mexican information request corpus, such topics may relate to the thematic area of an information request (e.g. requests pertaining to social/health services) or of the requester's broader agenda (e.g. admonishments about political corruption). We anticipate that these estimated topics will be thematically meaningful, and that *some* will indicate potential politicization in responses rather than denials for straightforward matters of legal compliance. We thus expect that our modeling of all topics across all request documents will aid in the prediction of government (non)responsiveness, as measured by our "denied request" measure. In our sLDA models, we specify the distribution of this response variable to be logistic and employ collapsed Gibbs sampling.

The corresponding topics that are uncovered by this sLDA model have the potential to be qualitatively distinct from those identified by *unsupervised* topic models such as latent Dirichlet allocation (Blei *et al.* 2003) or the structural topic model (STM; Roberts *et al.* 2014). Blei and Mcauliffe (2003) intuitively highlight this potential, in noting that,

> *[W]hen the goal is prediction, fitting unsupervised topics may not be a good choice. Consider predicting a movie rating from the words in its review … good predictive topics will differentiate words like "excellent," "terrible," and "average," without regard to genre. But topics estimated from an unsupervised model may correspond to genres, if that is the dominant structure in the corpus. (121).*

Similarly, for our ATI application, we contend that prediction-oriented supervised topic models, such as sLDA, uniquely offer the potential for the identification distinct topics *within* thematic areas of ATI requests, which may in turn allow one to distinguish between politically sensitive requests and apolitical requests within a given issue area. One example of this potential would be the thematic area of government procurement, where an (unsupervised) LDA model may only group requests based on their (non)correspondence with this theme, whereas an sLDA model trained on denied requests may instead identify separate subgroupings pertaining to (i) anticorruption campaigns conducting oversight of procurement and (ii) government contractors seeking information pertaining to their own procurement contracts. Additionally, our modeling of "denied request" as an outcome variable—via sLDA—rather than as an explanatory variable—as allowed for under an STM—is also the most appropriate temporal approach in our context, given that official responses to request texts arise after (as opposed to prior to) the generation of our request texts themselves.

Researchers must assign the number of topics, k, to be estimated within sLDA as well as a set of associated α and η hyperparameters. We utilize fivefold cross-validation to identify an optimal number of topics for the task of prediction. Herein, we draw a random sample of approximately 250,000 information requests (i.e. roughly 25% of our full request-corpus) and then randomly partition this sample into fivefolds of training and test data. For each set of training data, we hold our α and η hyperparameters fixed at 1.0 and 0.1, respectively, and estimate a series of sLDA models where the number of topics, k, is sequentially set to $k = \{5, 25, 50, 100, 250, 500\}$ and where our outcome variable is assigned as the "denied request" measure described earlier. We then use each resultant model's output to initialize a validation sLDA model using each fold's corresponding test sample. With these results in hand, we calculate the area under each test sample's corresponding receiver operating characteristic (ROC) curve (i.e. the AUC) and precision-recall curve (i.e. AUC-PR) to summarize each model's performance in classifying "denied requests." The AUC aggregates our model's (true positive versus false positive) "denied request" classifications across all possible thresholds into a single measure of classification performance. The AUC-PR provides a similar aggregate measure of classification performance, but in relation to a classification trade-off (precision versus recall) that is more attuned to outcomes with moderate to high class imbalance. Given the relative rarity of denied requests, we favor the latter metric in our assessments.

Figure 1 plots the corresponding AUCs and AUC-PRs for all k's evaluated (dashed lines), along with mean AUCs and AUC-PRs (solid lines). Across both subfigures, we find that an optimal number of topics for the task of predicting "denied requests" falls closest to $k = 250$. This k yields the highest average AUC and AUC-PR for our cross-validation sample while still offering substantial improvement over the next smallest topic number evaluated (i.e. $k = 100$). We hence set $k = 250$ for all sLDA models.

With our topic number identified, we next determine the optimal values for α and η. In this case, we hold our topic number fixed at $k = 250$, and then draw a *second* random sample of approximately 250,000 ATI requests from the remaining request documents in our corpus (i.e. approximately 25% of our full sample; excluding the 250,000 that we drew for the topic number selection routine). As before, we randomly partition this sample into fivefolds of training and test data. For each training set, we estimate sLDA models where α and η are sequentially set to unique pairings within the sets $\alpha = \{0.01, 0.1, 0.5, 1, 5, 10\}$ and $\eta = \{0.01, 0.1, 0.5, 1, 5, 10\}$. We use each resultant model's output, along with its corresponding test sample, to initialize a validation sLDA model, and store that model's accuracy in classifying "denied requests" via AUC and AUC-PR.

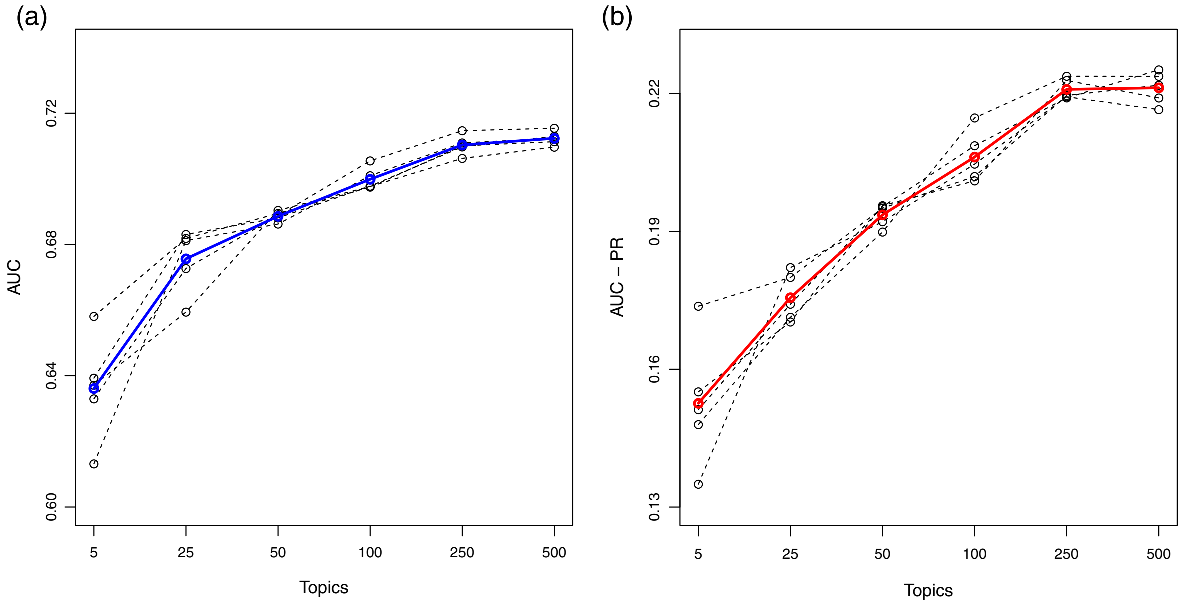

Figure 1 Topic number cross validation results. (a) Using Area under ROC and (b) Using Area under PR-curve.

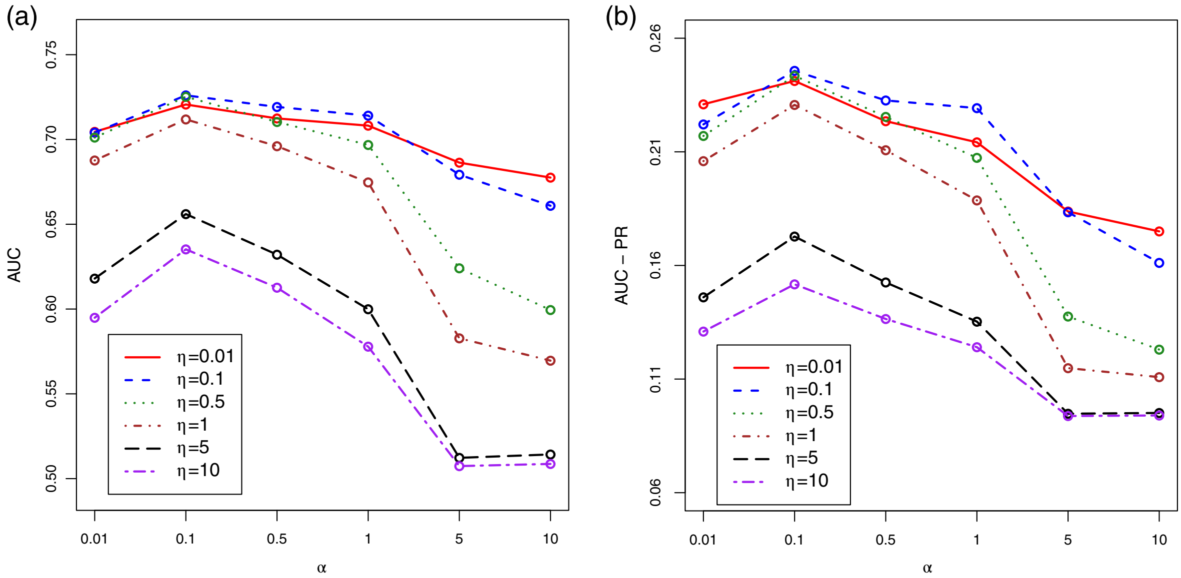

Figure 2 Hyperparameter cross validation Results. (a) Using area under ROC and (b) Using area under PR-curve.

Figure 2 plots the resultant averaged AUCs and AUC-PRs (across all folds) for each α and η combination evaluated. We find that the optimal α-η combination for the task of predicting "denied requests" corresponds to an α value of 0.1 (which generally yields the highest AUC no matter the α evaluated) and η values of 0.1 or 0.5. For the models presented later, we hence choose an α of 0.1 and an η of 0.1.[7] We ensure that our choices of α, η, and *k* are indeed optimal for our application by repeating these cross-validation routines while optimizing all three simultaneously in our Appendix S1.

6. Application to information requests in Mexico

The previous section identified an optimal set of parameters for our final sLDA model. We now apply this model to our remaining (i.e. unexamined) ATI request texts and "denied request" outcome measure. In these applications, we randomly partition our remaining ATI request text sample into separate training and test data for each evaluation. Subdividing our remaining sample in this incremental manner is critical, in that it serves to guard against overfitting within our models at each stage of our evaluations. It further ensures that the *only* commonality between our training and test data is the data-generating process itself. This helps to guarantee that the findings discussed later are attributable to our underlying "denied request" d.g.p. of interest, as opposed to model choices such as those of α, η, and k.

To this end—and recalling that we use a random sample of 25% of all request texts to identify an optimal number of topics, and a second random sample of 25% of all remaining request texts to identify optimal values for our hyperparameters—we incrementally apply our final sLDA model to random partitions of the remaining 50% of our request text sample. Specifically, we first reestimate a k= 250 sLDA model on a sample of 10% of our total corpus (i.e. 100,000 of our unexamined documents) for our outcome of interest: denied requests. We then generate in-sample and out-of-sample predictions for this outcome variable, where for our out-of-sample predictions we use a random subset of 300,000 of the remaining documents (30% of our total data). Further, we then compare our sLDA model to several alternative classification approaches using the final 10% of our request text corpus.

6.1. In-sample topic results

In order to fully assess our in-sample sLDA results for "denied requests," this subsection first discusses our topic-specific coefficient estimates for our in-sample sLDA model. This is followed by an evaluation of the topics that we find to be most predictive of (non)responsiveness based upon these estimates, and then provide an assessment of in-sample classification.

For our in-sample sLDA model, nearly all 250 topic-specific estimates are statistically significant under traditional thresholds, with the vast majority implying either an increase in responsiveness—or a slight increase nonresponsiveness—when present. However, a small number of topics exhibit uncharacteristically large effects on (non)responsiveness in each model. To view these effects, we recover the logit coefficient estimates from this in-sample model and plot these quantities—along with 95% confidence intervals—in Figure 3.

Figure 3 sorts our sLDA model's logit coefficient estimates by order of magnitude along the y-axis, and presents the magnitudes of these coefficient estimates on the x-axis (in log-odds scale). Here we find that the majority of recovered topics moderately decrease the likelihood of a "denied request." For instance, if we exclude the five largest positive effects in Figure 3 and converting the remaining effects to odds ratios, we find that the average topic-induced estimated effect corresponds to a 58% *decrease* in the odds of a "denied request." However, the five largest positive coefficient estimates in Figure 3 instead imply an average 1,684% *increase* in the odds of a

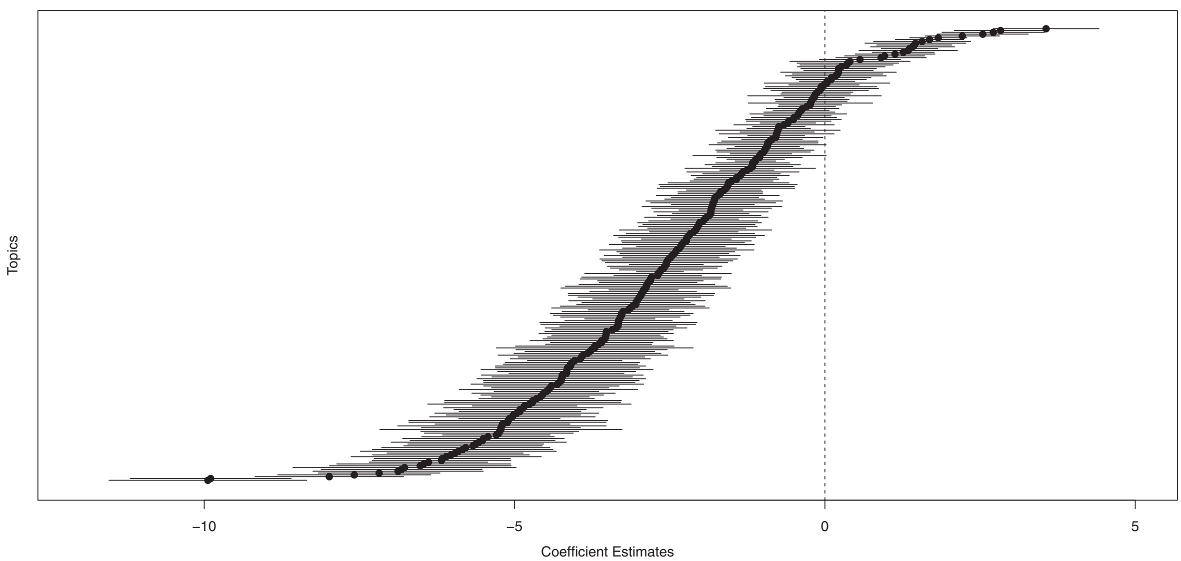

Figure 3 Estimated effect of each topic on the probability of a denied response.

"denied request." The topwords associated with these latter five topics are clearly worthy of further examination. We conduct this assessment further, alongside an evaluation of the topwords associated with the five topics identified as most strongly *decreasing* the odds of a "denied request."

For our substantive assessments of the five topics most positively and negatively associated with a "denied request," we consider (i) the 20 words most highly associated with each topic and (ii) qualitative readings of the 50 documents most highly associated with each topic. With regards to topwords, we extract and report these topwords based upon "word scores". Word scores denote the logged number of times that a given word is assigned to a topic, divided by the logged total number of times that that word is assigned to all other topics. This metric accordingly allows us to interpret topics based upon the words that exhibit both (i) a strong association with a given topic and (ii) a relative uniqueness to that topic.[8] These 20 "word score" topwords are presented for *each* aforementioned high and low leverage topic within Table 1. For these topwords, we have de-stemmed all relevant topwords and have then translated each resultant word to English.[9] With regards to our qualitative readings of the 50 documents most highly associated with each topic of interest, we use our sLDA model's topic-conditioned word assignments to identify and assess each topic's 50 most relevant requests. We also report Spanish and English-translated versions of two highly associated requests for each denied/provided topic in our Appendix S1.

6.2. Denied request topic interpretations

The five topics that are most likely to receive a "denied request" largely appear to be investigative in nature, or to otherwise pose direct threats to a responding agency's resources or reputations. For example, Denied$_{\#1}$—which exhibits topwords such as "police," "federal," "capture," and "security"—contains among its most representative requests several requests seeking copies of government reports related to the search for the escaped cartel leader Joaquín "El Chapo" Guzmán. These requests specifically seek information on Mexican government knowledge and performance during instances where El Chapo was on the verge of capture but escaped, while implying that collusion existed between the government and organized crime in these regards. The remaining representative Denied$_{\#1}$ requests similarly seek information relevant to the police or other security services, including details on firings, demotions, desertions, deaths in the line of duty, and reports on police activity in specific incidents. Together these requests—and our broader readings of the top 50 requests associated with this topic—indicate that this topic encompasses requests that seek information about past police practices or actions, with an eye toward identifying potential abuses of power, corruption, or broader security failures. Such requests are clearly politically sensitive, given media attention and public outrage over government conduct of the drug war. While some of these responses may be legitimate denials in cases of classified information, others are likely to indicate misuse of discretion in order to avoid scrutiny, as has been demonstrated in other cases of security-related requests (Almanzar *et al.* 2018).

Denied$_{\#2}$, on the other hand, primarily encompasses requests for information on financial savings and deposits, in both public savings schemes and in private banks. The topwords associated with Denied$_{\#2}$ include words such as "bank," "value," "deposit," "saving," "account," and "accreditation." The top 50 requests associated

Table 1 English-translated topwords for topics associated with (non)denied request

Denied$_{\#1}$	police, federal, part, fact, authorization, elements, daily, past, security, capture
Denied$_{\#2}$	bank, value, deposit, bank, said, institute, commission, saving, account, accreditation
Denied$_{\#3}$	coordination, administration, general, republic, attorney general's office, work, accurate, federal, position, legislation
Denied$_{\#4}$	request, information, etc., nature, refers, mention, written, documents, documentary, contain
Denied$_{\#5}$	insurance, request, information, delivery, south, confiscated, date, I require, also, specify
Provided$_{\#1}$	education, school, staff, teacher, hours, professors, technology, baccalaureate, DGETI, appointment
Provided$_{\#2}$	education, school, SEP, higher, primary, secondary, level, students, school, teacher
Provided$_{\#3}$	how much, which, history, existence, country, INAH, archeology, each, monuments, they are
Provided$_{\#4}$	budget, assigned, destination, exercise, radio, annual, item, program, expenditures, televisions
Provided$_{\#5}$	wage, salary, position, tabulator, monthly, level, perceptions, salary related, benefits, compensation

with this topic largely focus on bank failures and the oversight role of Mexico's banking regulator in this process, suggesting that this topic exhibits clear political potential.[10] Several of the most representative requests for this topic seek information concerning savings funds set aside during the US-Mexico Bracero Program. A number of additional top 50 requests allude to the 2014 Ficrea credit union fraud scandal in Mexico, or its aftermath. Others seek details on Mexican banks that recently announced bankruptcy, with a potential focus on protecting or recovering savings. These latter requests could legitimately fall under protected exemptions, or they could seek information which no longer exists. On the other hand, they might also threaten to saddle agencies with new financial costs, or direct unwanted scrutiny on the banking regulation system—two clear instances where political biases in government nonresponsiveness may arise.

$\text{Denied}_{\#3}$ relates to requests concerning political appointments, government officials, and oversight thereof. A majority of the topwords associated with this topic are clearly political in nature (e.g. "administration," "federal," "position," and "legislation") and $\text{Denied}_{\#3}$'s most highly associated requests frequently target Mexico's Procuraduría General De La República (i.e. the Attorney General's office). The topic's most associated requests seek very specific—and very extensive—information on the appointments to bureaucratic positions, and their associated salaries and responsibilities; often with reference to specific job codes and titles. These requests are also highly legalistic in language and typically reference very specific laws or statutes as justification for the requested information. Requests of this sort could be perceived by government officials as corresponding to public efforts to investigate patronage or unqualified appointments, and thus as having high political potential. On the other hand, the combined breadth and specificity the majority of top 50 requests for this topic may imply that denials for these requests were just as commonly issued due to the unavailability of (or difficulty in assembling) the actual information requested. As such, we conclude that this topic's political potential is less clear-cut than was the case for $\text{Denied}_{\#1}$and $\text{Denied}_{\#2}$.

$\text{Denied}_{\#4}$ appears to relate to scrutiny of federal procurement, and related compliance issues. $\text{Denied}_{\#4}$'s topwords exhibit a focus on specific documents in these respects, and this topic's most associated requests pertain to (i) an IMSS[11] food-contract corruption scandal, (ii) the relative numbers of direct, public bid, and invited procurement awards for various agencies, or (iii) information on federal agency-imposed fines against various banks and companies, and related information on whether these fines were contested and/or paid. These types of information are highly relevant for investigations of corruption or other irregularities in public procurement, a notoriously corrupt sector in Mexico. Underscoring this point, we can also note that many of the requesters within the top 50 most associated requests for this topic self-identify themselves as a "Contraloría Ciudadana" or "Contraloría Social." These position-titles correspond to social/citizen comptrollers in the context of Mexico, who function as citizen volunteers with the direct responsibility of conducting procurement oversight in the interest of combating corruption. In light of the above, $\text{Denied}_{\#4}$ appears to have significant political potential.

$\text{Denied}_{\#5}$'s topwords include words such as "insurance," "confiscated," "specify," and "date." Its most representative requests encompass detailed queries into the specific properties seized and/or confiscated in various federal drug-crime arrests, including information on the addresses of the properties seized, the type(s) of seizure/foreclosure, the amount of assets seized, and the confiscated items' current status for a specific named individual in each request. For example, the fifth most associated request with this topic seeks this range of information in relation to a Mexican cartel member known as "La Gata," who was captured by Mexican authorities in 2007.[12] Such requests may arise from legal representatives of the arrested individuals seeking to recover confiscated property, or from members of media and related anticorruption campaigns seeking to identify misuse of confiscated resources among government agencies. Interestingly, while the most representative $\text{Denied}_{\#5}$ requests were made using an identical template—with only the affected individual's name changed—there is substantial variation in the justifications for denials among this topic's most highly associated requests, with responses ranging from "reserved/classified" (4), "information does not exist" (16), and "this request does not fall under the purview of Mexico's ATI law" (12).

6.3. Provided request topic interpretations

In contrast to the top five Denied topics, the five Provided topics in Table 1 are much less likely to be interpreted as politically threatening by responding agencies. $\text{Provided}_{\#1}$ and $\text{Provided}_{\#2}$ each encompass sets of requests that seek varied information on public education in Mexico. The topwords across these two topics include words such as "education," "school," "professors," "students," "teacher," and "technology." In the case of Provided$_{\#1}$, the 50 most associated requests are frequently about personnel in educational institutions. This is interesting in that many of these *could* yield information on wrongdoing such as patronage hiring or absentee employees, *but* this is not politically sensitive at a *national* level. The fact that these requests receive routinely *positive* responses suggests that the federal-level education officials in Mexico may have their own interests in rooting out these sorts of bad practices in their subsidiary institutions.[13] On the other hand, Provided$_{\#2}$ encompasses requests for *aggregate statistics* on (i) educational enrollment,[14] (ii) educational spending, or (iii) budgets and spending of Mexican states more generally. Reflecting the apolitical nature of this topic, the requesters within several of these requests were quite clear up-front that they needed the information for their own educational research.

Provided$_{\#3}$ appears to encompass fairly mundane requests concerning land-use or zoning issues, especially in the context of historical preservation. Topwords corresponding to "archeology," "monuments," and "INAH"[15] reinforce this characterization. A closer examination of the top-50 associated requests suggests that the requests underlying this topic are primarily seeking aggregate statistics for archeological sites, urban growth boundaries, or coastal zoning. Provided$_{\#4}$'s top-50 requests largely pertain to simple requests for budgetary-related requests for entire states, state-sponsored scientific research, the national theatre, or the media sector. Topwords such as "radio," "annual," "program," "expenditures," and "televisions" reflect this interpretation. The subset of requests for budgetary information related to the media sector, specifically, could be related efforts to scrutinize the misuse of official advertising budgets, and hence, of having political potential. However, a majority of the most highly associated requests for this topic are very succinct and straightforward, which likely helps to ensure that they are (i) difficult to be denied by officials and (ii) unlikely to be interpreted by officials as being investigative in nature.

Provided$_{\#5}$ contains topwords such as "wage," "benefits," "salary," and "position." These topwords, and many of this topic's top-50 most associated requests, together suggest that the requests associated with this topic are often arising from state employees themselves, who are seeking information about their own salary, benefits, or position—often over a period encompassing their past several years of employment. A number of additional highly associated requests seek more aggregate salary or compensation information, though these again appear to be fairly straightforward and benign. For example, within some of the top requests associated with this topic, a requester provides a Mexican agency with the agency's own publicly available salary or compensation documentation (e.g. spreadsheet), and then asks the agency to corroborate this information or to fill-in any missing information. Hence, this topic appears to reflect fairly apolitical requests for compensation-related information (or assistance in completing an official document), as opposed to efforts by external actors to scrutinize the employment practices of the agency receiving the request.

6.4. In-sample summary statistics

The above evaluations suggest that the five sLDA topics that are most predictive of denied requests tend to each exhibit clear political potential, with corresponding information requests posing the potential to expose policy failures, wrongdoing, or misuse of office in the realms of security, banking, political appointments, or public procurement. By contrast, the five most commonly provided topics appear to instead seek out far more general, and aggregate, information on government employment, budgets, personal job information, or educational statistics. We hence interpret our sLDA model's identified topics as evidence to suggest that variation in government non-responsiveness to ATI requests is at least partly the result of agencies' efforts to protect their own resources, reputation, or personnel.

To validate these findings, we classified all in-sample requests according to their most associated topic, and then created 10 binary indicators for whether (=1) or not (=0) each in-sample request arose from one of our top-5 Denied or top-5 Provided topics. We then separately classified each in-sample request for whether (=1) or not (=0) it was made during a Mexican presidential election cycle, using data on the timing of each request.[16] If our Denied topics are indeed more politicized than our Provided topics, we would expect the former to exhibit a stronger association with presidential election windows—when efforts to uncover policy failures, corruption, or government excesses are potentially more acute.[17] We assess this by examining the association between (i) each of our 10 (Denied or Provided) binary topic indicators and (ii) our binary election period indicator via a series of

χ^2 tests (see Table A.6). We find that four out of five Denied topics exhibit a statistically significant ($p < 0.05$) association with our presidential election indicator; whereas only one of our five Provided topics exhibits such an association. We then summed these Denied and Provided requests to the monthly level,[18] and calculated the share of all in-sample requests arising from (i) the top-5 Denied topics and (ii) the top-5 Provided topics for our monthly time series. Evaluating the difference in means[19] for each proportion across our aforementioned presidential election indicator, we found that our Denied requests encompass a reliably[20] larger proportion of all requests during Presidential election cycles than outside of these cycles, whereas this difference is not statistically significant[21] for our Provided requests. Together, these findings suggest that our Denied topics are more strongly associated with at least one form of temporal variation in political scrutiny than are our Provided topics.

The above findings notwithstanding, we are also interested in the representativeness of our top five Denied request topics. That is, while our Denied topics are clearly outliers in terms of their respective rates of request-denials, and potentially in their levels of political scrutiny, are these topics also outliers in terms of overall (low) request volume, or in terms of their levels of specificity to only a single target agency (or geographic location)? If the answer to these latter questions is yes, researchers may be concerned that our approach is not identifying cross-cutting thematic request areas, but rather is simply identifying highly specific (and idiosyncratic) problem requesters. To evaluate this potential, we return to our sLDA-estimated word assignments—which we used above to identify the 50-most associated requests with each topic—and use these estimated word assignments to classify each and every in-sample ATI request[22] according to its most associated topic. This allows us to recover the total number of in-sample ATI requests associated with each topic discussed earlier. By comparing these classified in-sample requests to our metadata on each request's target agency, the timing of each request, and each requester's home municipality, we are able to recover and evaluate the relative specificity of each topic over space and time.

In Table 2, we present the counts of *in-sample* requests, target agencies, and requester municipalities identified for each of our Denied and Provided topics via the approach described earlier. For further context, we also report the in-sample percentile ranking of each specific count of requests, target agencies, and sending municipalities in this table. Figure A.3 in Appendix S1 additionally provides plots of the variation in the volume of our topic-indexed in-sample requests over time. Together these summary quantities sharpen our understandings of the topics discussed earlier. First and foremost, the percentiles in Table 2 indicate that our Denied topics exhibit comparable levels of request volume—and comparable levels of diversity in target agency and sending municipality—to our remaining 240 topics. One exception is $\text{Denied}_{\#3}$, which ranks in the third percentile in terms of total requests, target agency diversity, and sending municipalities. Based on these results, we can thus interpret $\text{Denied}_{\#3}$—which largely corresponded to politically tinged requests (to Mexico's Attorney General) for labor and salary information—as uncharacteristically concentrated in request volume. The time series plots in Figure A.3 in Appendix S1 confirm these characterizations, in indicating a concentrated spike in $\text{Denied}_{\#3}$ requests in 2009.[23]

Table 2 In-sample summary statistics for selected topics

	Documents per topic		Agencies per topic		Municipalities per topic	
	# Documents	Percentile	# Documents	Percentile	# Documents	Percentile
$\text{Denied}_{\#1}$	324	51.6	82	44.8	61	48.0
$\text{Denied}_{\#2}$	276	44.4	68	26.0	61	48.0
$\text{Denied}_{\#3}$	99	2.8	27	3.2	18	3.6
$\text{Denied}_{\#4}$	187	20.0	63	21.2	38	14.8
$\text{Denied}_{\#5}$	250	30.0	68	26.0	53	31.6
$\text{Provided}_{\#1}$	1,049	96.4	70	30.4	141	95.2
$\text{Provided}_{\#2}$	1,788	100.0	80	42.8	208	100.0
$\text{Provided}_{\#3}$	724	88.4	152	90.4	120	90.8
$\text{Provided}_{\#4}$	681	84.8	158	95.6	103	83.6
$\text{Provided}_{\#5}$	1,504	99.6	175	97.6	153	97.6

Median documents per topic = 318.5; median agencies per topic = 85, and median municipalities per document = 63.

By comparison, the percentiles reported for the top five Provided topics in Table 2 each appear to be uncharacteristically large in request volume and request scope, as measured by total request volume, diversity in target agency, and diversity in sending municipality. Indeed, in nearly every case, we find that the percentile rankings for our five Provided topics fall in the 80th–100th percentile range. Hence, whereas our top five Denied topics are generally representative of all topics in request volume and scope, the five topics that our sLDA model identifies as most likely to have information provided together represent five of the largest, and most diverse (in terms of sending municipality and target agency) request topics of all 250 topics identified. Figure A.3 in Appendix S1 reaffirms these findings, in demonstrating that each of our top five Provided topics exhibits a high, and sustained, level of request volume across the 2003–2015 period. Taken together, these findings for our five Provided topics are good news for the functioning of the Mexican ATI system, as they imply that the largest and most widespread request topics also happen to be among those that exhibit the highest levels of bureaucratic *responsiveness*.

We further evaluate the uniqueness of our Provided and Denied topics by unpacking the five most middle leverage "denied request" topics from Figure 3 in Appendix S1. The latter topics generally *do not* exhibit strong political potential, and instead largely pertain to straightforward requests in areas such as aggregate government statistics, procurement requests from service providers, tourism research, and higher education. These topics' strengths of association with our previously described Mexican presidential election cycle indicator also fall in between those identified for our Provided and Denied topics. This together suggests that while our sLDA model is identifying theoretically coherent topics across the entire range of coefficient estimates reported in Figure 3, the requests underlying our top-5 Denied topics are atypical in exhibiting high levels of political potential.

6.5. In-sample classification results

We next evaluate the in-sample classification performance of our sLDA model. In doing so, we construct two random "coin-flip" baselines for comparison, hereafter denoted ξ. For our first ξ, we generate random binary "denied request" data with probability $\frac{1}{2}$. For the second ξ, we generate random binary "denied request" data with probability equal to the mean of our true binary response $\overline{y} = 0.1$. As such, $\xi = \overline{y}$ provides us with a random classifier that maximizes overall accuracy, whereas $\xi = \frac{1}{2}$ provides us with a random classifier that instead favors the improved identification of our less common class (i.e. nonresponsiveness).

We compare these two random classifiers against our in-sample "denied request" sLDA results with the aid of AUC-PRs, AUCs, precision, recall, F1-scores, and overall classification accuracy. For our application, precision corresponds to the proportion of our model's "denied request"-predictions that were truly "denied requests" in our request data, whereas recall corresponds to the share of all true "denied request"-cases in our data that were accurately predicted as such by our model. F1-scores are harmonic means of precision and recall, wherein higher values imply superior combined accuracy across both metrics. By contrast, overall classification accuracy simply reports the proportion of all "denied requests" and all "provided requests" that were classified correctly by our model. Given our preference for the accurate prediction of our minority class (i.e. nonresponsiveness), we assign a cutoff of 0.25 for the calculation of precision, recall, F1-scores, and overall classification accuracy.

Table 3's AUC values imply that the sLDA model's in-sample predictions are notably better than chance (AUC=74.09). By comparison, $\xi = \frac{1}{2}$ and $\xi = \overline{y}$ obtain AUCs that are no better than chance (50.36 and 50.00). This superior performance of sLDA is reinforced by our sLDA model's consistently preferable AUC-PR, F1-score and precision values, relative to those obtained under either $\xi = \frac{1}{2}$ or $\xi = \overline{y}$. As expected, $\xi = \overline{y}$ maximizes overall accuracy, with a value (82.21) that is superior to $\xi = \frac{1}{2}$ (49.75). However, the maximized accuracy obtained under $\xi = \overline{y}$ is well below that of sLDA (88.86), and comes at the cost of poorer precision and recall relative to either $\xi = \frac{1}{2}$ or

Table 3 In-sample classification statistics

	AUC-PR	AUC	Precision	Recall	F1-Score	Accuracy
sLDA	26.26	74.09	34.00	15.85	21.62	88.86
$\xi = \frac{1}{2}$	09.62	50.00	09.70	10.05	09.87	49.75
$\xi = \overline{y}$	09.69	50.00	09.70	10.05	09.87	82.21

Table 4 Out-of-sample classification statistics

	AUC-PR	AUC	Precision	Recall	F1-score	Accuracy
sLDA	25.57	73.23	33.34	15.33	21.01	88.83
$\xi = \frac{1}{2}$	09.69	50.03	09.69	50.01	16.24	50.03
$\xi = \bar{y}$	09.75	50.15	09.94	10.25	10.09	83.32

sLDA. Finally, while sLDA underperforms in recall relative to $\xi = \frac{1}{2}$ (49.51 > 15.85), it makes up for this shortcoming with a precision of 34.00 for sLDA versus 9.57 for $\xi = \frac{1}{2}$. Given the latter strength, and those discussed earlier, we can conclude that sLDA outperforms our ξ metrics in the in-sample context.

6.6. Out-of-sample results

We now turn to an evaluation of our sLDA model's out-of-sample classification properties. For this evaluation, we use our primary in-sample sLDA model to generate "denied request" predictions for 30% of our total corpus (i.e. 300,000 of our previously unexamined documents). Using these predictions, we repeat the same steps as above in (re-)generating our same two random classifiers for comparison, $\xi = \frac{1}{2}$ and $\xi = \bar{y}$. We then recalculate the aforementioned set of classification statistics for each approach in Table 4.

Our out-of-sample classification findings are highly consistent with our in-sample findings. As earlier, the sLDA model outperforms both random classifiers in AUC-PR, AUC, precision, F1-score, and overall accuracy, and performs second best (to $\xi = \bar{y}$) in recall. The results reported in Table 4—across all classifiers—suggest that our out-of-sample sLDA predictions perform comparably to, albeit slightly worse than, our in-sample sLDA results. For example, our sLDA model accurately classifies 88.83% of all out-of-sample cases, whereas in the in-sample context our sLDA model's overall accuracy was 88.86%. Differences between these two sets of sLDA predictions are slightly larger when one examines AUCs (74.09 vs. 73.23), or AUC-PRs (26.26 vs. 25.57) though these differences are again fairly negligible, especially relative to the effect of one's choice of k on the in-sample AUCs obtained in Figures 1 and 2.

7. Robustness tests

We further assess the robustness of our sLDA application in Appendix S1, and summarize these assessments here. As a baseline check, we first compare our sLDA approach to three widely used alternatives: logistic regression with a LASSO penalty, logistic regression with a ridge penalty, and standard logistic regression. In all cases, we use a document term matrix[24] that includes all unique (processed) wordstems alongside each request's intended federal agency name as features. We leverage the remaining 10% of our full 2003–2015 Mexican request text corpus for these model-based comparisons, which is equivalent to roughly 100,000 total requests. Herein, we randomly subdivide this sample into new sets of training requests (n = 25,000) and test requests (n = 75,000). We then use the 25,000 training documents to reestimate a new sLDA model alongside our logit, LASSO, and ridge estimators, so as to ensure that the out-of-sample predictions generated by (i) our sLDA model and (ii) our comparison models are comparable in terms of the size of the training sample used. Details on tuning- and hyper-parameter selection for these models, along with a table of out-of-sample classification statistics, appear in Appendix S1. Stated briefly, we find in these comparisons that sLDA outperforms LASSO, ridge, and logistic regression across our most relevant metrics, including AUC-PR, recall and AUC.

Next, we compared our primary sLDA model's in-sample and out-of-sample predictive results to a "requests only" sLDA model that omits our *target agency* names as features in Appendix S1. For these comparisons, we repeat the analyses performed further above when excluding the *target agency* names that are included as additional strings within our main sLDA classification routines. We find in Appendix S1 that applying sLDA models to our "requests only" text yields in-sample and out-of-sample predictions of "denied requests" that perform slightly worse in classifying our true "denied request" outcomes, relative to the primary sLDA model earlier that uses both the text of our requests and these requests' *target agency* names as features. For example, the "requests only" model's out-of-sample AUC is 71.47; noticeably worse than that of our primary sLDA model. Our other comparison metrics yielded similar conclusions: The addition of *target agency* names as features leads to small but consistent improvements in classification accuracy.

Our Appendix S1 also compares our primary sLDA model's in-sample Denied and Provided topics to the topics obtained from a 250-topic STM. As noted earlier, we believe that sLDA's treatment of "denied request" as an outcome variable, rather than as an explanatory variable—as is the case in an STM—is most appropriate in our context. This is because official responses to ATI requests arise *after* the generation of these requests by citizens. Even so, estimating an STM that treats "denied request" as an independent variable has the potential to (i) provide further insights into the stability of our sLDA model's Denied and Provided topics, (ii) sharpen our overall understandings of high leverage topics in the Mexico ATI request context, and (iii) offer evidence of the STM's effectiveness for "needle-in-a-haystack"-type research tasks more generally. We identify a notable degree of topical overlap among our top five (sLDA) Denied and Supplied topics and the STM's most high leverage (Denied and Provided) topics, while also identifying a number of additional relevant Denied topics from our STM. These findings help to (re)affirm the usefulness of both the sLDA and the STM for the discovery of topics associated with "denied requests." However, we continue to favor the sLDA model given that its assumed d.g.p. more closely matches the actual temporal sequencing of requests and responses in our application.

8. Conclusion

Many forms of citizen–government interaction now take place electronically and generate large amounts of data. Online ATI systems are a prime example, and have proliferated across the world in recent years. To study these new and complex forms of "big data," researchers need tools that can jointly account for (i) the nature, size, and variety of request texts and (ii) the linkages between these texts and key outcomes of interest. Governments likewise require similar "big data" tools to handle the immense volume and complexity of these systems, especially in their efforts to monitor ATI responsiveness, performance, and compliance.

We analyze over 1 million publicly available ATI request records from Mexico (2003–2015) in order to demonstrate *both* the utility of these types of data sources for understanding how bureaucratic responsiveness operates *and* the merits of supervised topic models for analyzing these citizen–government interactions. We find that the topical content of Mexico's ATI requests can predict responsiveness to these requests at levels noticeably better than a variety of alternatives. This lends support to the use of supervised topic modeling methods like sLDA for the study and monitoring of ATI systems. It also underscores the importance of thematic word clusters (i.e. relative to simpler term-based predictive models) in understanding organizational behavior across large numbers of interactions.

The results from the sLDA analysis of our Mexico sample suggest, moreover, that politicization plays a role in the Mexican ATI system, and in Mexican government agencies' responsiveness to ATI requests more specifically. To this end, it appears that requests investigating specific forms of corruption and inefficiency receive lower responsiveness. This finding complements recent research into analytical frameworks for, and empirical assessments of, government corruption in public procurement elsewhere across the globe (Fazekas & King 2018). However, we also find some evidence of unusually high responsiveness related to local corruption in the education sector, suggesting a role of ATI systems for fire-alarm monitoring of local problems by the political center.

For those interested in applying text analysis methods to citizen–government interactions, our application provides four additional insights. First, our findings demonstrate that integrations of citizen request data with supervised topic models can allow one to reliably identify "needle-in-a-haystack"-topics that are particularly distinctive in their association with rare bureaucratic outcomes. Second, we find that researchers can use the topwords and top-associated-requests from these distinctive topics to glean substantive insights into the nature of a political or administrative process of interest—specifically with regards to the *extremes* of a process rather than its usual operation. Third, we show how one can then associate one's identified topics with metadata—such as information on a topic's relative concentration over space and time—to better understand when those extremes arise. Fourth, we demonstrate that this framework also allows one to generate accurate *predictions* of a rare outcome of interest, which could be useful to future citizen users of ATI systems, to advocacy groups that monitor the performance of such systems, or to officials themselves.

Acknowledgements

The authors declare no conflicts of interest. Bagozzi's research was supported in part by the University of Delaware's General University Research Fund. Almquist's research was supported in part by an ARO YIP Award #W911NF-14-1-0577.

Endnotes

1 Such insights also have the potential to contribute to understandings of ICT-enabled public participation more generally (e.g. Lodge & Wegrich 2015; O'Brien *et al.*, 2017).

2 This may also involve the use of different procedures applying greater political scrutiny, such as the "amber-lighting" documented by Roberts (2006) in Canada.

3 Despite merit-based policies for some appointments in Mexico, "personal loyalties and even party affiliations continue to play a significant role in hiring and promotion" (Dussauge Laguna 2011, p. 62).

4 We exclude requests for personal data as those are made under a different legal regime and handled through different procedures that are also part of the INFOMEX system.

5 Fewer than 1% of attachment files were corrupted, meaning that we were unable to extract their corresponding text for inclusion here.

6 When doing so, we transform agency names into unigrams such that, for example, Instituto Nacional de Desarrollo Social appears as Instituto-Nacional-de-Desarrollo-Social.

7 Given that $\eta = 0.1$ generally performs better than $\eta = 0.5$ across the range of α; especially for AUC-PR.

8 We also report topwords based on posterior probability of word-to-topic assignment in Appendix S1.

9 The original Spanish versions (both stemmed and de-stemmed) appear in Appendix S1.

10 Though some top 50 requests may have received denials not due to concerns over political sensitivity, but rather because the requested information was genuinely limited to the public (e.g. individual bank accounts).

11 Mexico's Social Security Institute.

12 See: https://www.proceso.com.mx/207490/cae-el-narco-oscar-alonso-candelaria-escajeda. Accessed 19 May 2018.

13 Though other top requests associated with this topic appear to be from educational personnel themselves, and seek either information on those individuals' own employment or a list of available job opportunities.

14 For example, total student enrollments or enrollments separated across various demographics or grade-levels.

15 Mexico's National Institute for Anthropology and History.

16 Defined as equal to one for January–July during either 2006 or 2012, given that presidential elections occurred in July during these two sample years. Importantly, the timing of each request was not included as a feature within our estimated models.

17 Importantly, the binary topic measures described here do not incorporate any information on the government's actual response to each request.

18 Creating a data set of 147 total observations.

19 Via a two-sided *t*-test.

20 $t = 1.95$, $df = 16.16$, $pP - value = 0.068$.

21 $t = 1.13$, $df = 16.29$, $pP - value = 0.275$.

22 That is, the requests contained within the sample that was used to train our in-sample sLDA model, which corresponded to 10% of our total corpus, or, approximately 100,000 total documents.

23 Interestingly, Figure A.3 also reveals a substantial spike in requests associated with $\text{Denied}_{\#5}$ in 2013, potentially indicating that this topic was similarly narrow in temporal scope.

24 That is, document-level counts.

References

Almanzar T, Aspinwall M, Crow D (2018) Freedom of Information in Times of Crisis: The Case of Mexico's War on Drugs. *Governance* 31(2), 321–339.

Bagozzi BE, Berliner D, Zack Almquist W. (2016) Predicting Government (Non)Responsiveness to Freedom of Information Requests with Supervised Latent Dirichlet Allocation. Proceedings of the International Conference on the Advances in Computational Analysis of Political Text-PolText, 15 Jul 2016, Dubrovnik, Croatia.

Balla SJ (2000) Political and Organizational Determinants of Bureaucratic Responsiveness. *American Politics Quarterly* 28(2), 163–193.

ben Aaron J, Denny M, Desmarais B, Wallach H (2017) Transparency by Conformity: A Field Experiment Evaluating Openness in Local Governments. *Public Administration Review* 77(1), 68–77.

Berliner D (2014) The Political Origins of Transparency. *Journal of Politics* 76(2), 479–491.

Berliner D, Bagozzi BE, Palmer-Rubin B (2018) What Information Do Citizens Want?: Evidence from One Million Information Requests in Mexico. *World Development* 109, 222–235.

Berliner D, Bagozzi BE, Palmer-Rubin B, Erlich A (2020) The Political Logic of Government Disclosure: Evidence from Information Requests in Mexico. *Journal of Politics*.

Berliner D, Erlich A (2015) Competing for Transparency: Political Competition and Institutional Reform in Mexican States. *American Political Science Review* 109(1), 110–128.

Blei DM, Mcauliffe JD (2003) Supervised Topic Models. *Advances in Neural Information Processing Systems* 20, 121–128.

Blei DM, Ng A, Jordan M (2003) Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3, 993–1022.

Bogado BF, Martinez-Morales E, Noll BD, Bell K. (2007) The Federal Institute for Access to Information and a Culture of Transparency: Follow Up Report. Technical report Annenberg School of Communications, University of Pennsylvania.

Bookman Z, Amparán J-PG (2009) Two Steps Forward, One Step Back: Assessing the Implementation of Mexico's Freedom of Information Act. *Mexican Law Review* 1(2), 25–49.

Busuioc EM, Lodge M (2016) The Reputational Basis of Public Accountability. *Governance* 29(2), 247–263.

Butler DM, Broockman DE (2011) Do Politicians Racially Discriminate against Constituents? A Field Experiment on State Legislators. *American Journal of Political Science* 55(3), 463–477.

Chaney CK, Saltzstein GH (1998) Democratic Control and Bureaucratic Responsiveness: The Police and Domestic Violence. *American Journal of Political Science.* 42, 745–768.

Cleary MR (2007) Electoral Competition, Participation, and Government Responsiveness in Mexico. *American Journal of Political Science* 51(2), 283–299.

Connelly R, Playford CJ, Gayle V, Dibben C (2016) The Role of Administrative Data in the Big Data Revolution in Social Science Research. *Social Science Research* 59, 1–12.

Distelhorst G (2017) The Power of Empty Promises: Quasi-Democratic Institutions and Activism in China. *Comparative Political Studies* 50(4), 464–498.

Dussauge Laguna MI (2011) The Challenges of Implementing Merit-Based Personnel Policies in Latin America: Mexico's Civil Service Reform Experience. *Journal of Comparative Policy Analysis* 13(1), 51–73.

Fazekas M, King LP (2018) Perils of Development Funding? The Tale of EU Funds and Grand Corruption in Central and Eastern Europe. *Regulation & Governance* 13(3), 405–430.

Fox J, Haight L, Palmer-Rubin B (2011) Proporcionar transparencia ¿Hasta qué punto responde el gobierno mexicano a las solicitudes de información pública? *Gestión y Política Pública* 20(1), 3–61.

Fumega S. (2014) Information and Communication Technologies and Access to Public Information Laws. Technical report Red de Transparencia y Acceso a la Información and World Bank Institute, Santiago, Chile.

Golden M, Min B (2013) Distributive Politics around the World. *Annual Review of Political Science* 16(1), 73–99.

Grimmelikhuijsen S, John P, Meijer A, Worthy B (2019) Do Freedom of Information Laws Increase Transparency in Government? A Replication of a Field Experiment. *Journal of Behavioral Public Administration* 2(1), 12–34.

Herrera V (2017) *Water and Politics: Clientelism and Reform in Urban Mexico*. University of Michigan Press, Ann Arbor.

Hopkins DJ, King G (2010) A Method of Automated Nonparametric Content Analysis for Social Science. *American Journal of Political Science* 54(1), 229–247.

Jurka TP, Collingwood L, Boydstun AE, Grossman E, van Atteveldt W (2013) RTextTools: A Supervised Learning Package for Text Classification. *The R Journal* 5(1), 6–12.

Lagunes P, Pocasangre O (2019). Dynamic transparency: An audit of Mexico's Freedom of Information Act. *Public Administration* 97(1), 162–176.

Lavertu S (2017) We all Need Help: 'Big Data' and the Mismeasure of Public Administration. *Public Administration Review* 76 (6), 865–872.

Lodge M, Wegrich K (2015) Crowdsourcing and Regulatory Reviews: A New Way of Challenging Red Tape in British Government? *Regulation & Governance* 9(1), 30–46.

McClendon G (2016) Race and Responsiveness: A Field Experiment with South African Politicians. *Journal of Experimental Political Science* 3(2), 60–74.

McCubbins MD, Schwartz T (1984) Congressional Oversight Overlooked: Police Patrols Versus Fire Alarms. *American Journal of Political Science.* 28, 165–179.

McCubbins MD, Noll RG, Weingast BR (1987) Administrative Procedures as Instruments of Political Control. *Journal of Law, Economics, & Organization* 3(2), 243–277.

Michener G (2011) FOI Laws around the World. *Journal of Democracy* 22(2), 145–159.

Michener G (2015) How Cabinet Size and Legislative Control Shape the Strength of Transparency Laws. *Governance* 28(1), 77–94.

Michener G, Worthy B (2015) The Information-Gathering Matrix A Framework for Conceptualizing the Use of Freedom of Information Laws. *Administration & Society* 50(4), 476–500.
Noveck BS (2016) Is Open Data the Death of FOIA? *Yale Law Journal: Forum* Available from URL: https://www.yalelawjournal.org/forum/is-open-data-the-death-of-foia.
O'Brien DT, Offenhuber D, Baldwin-Philippi J, Sands M, Gordon E (2017) Uncharted Territoriality in Coproduction: The Motivations for 311 Reporting. *Journal of Public Administration Research and Theory* 27(2), 320–335.
Peisakhin L, Pinto P (2010) Is Transparency an Effective Anti-Corruption Strategy? Evidence from a Field Experiment in India. *Regulation & Governance* 4(3), 261–280.
Pew. (2018). How States Use Data to Inform Decisions: A National Review of the Use of Administrative Data to Improve State Decision-making. *The Pew Charitable Trusts*.
Poole EG (2018) How Institutional Culture Trumps Tier Effects: Evidence from Government Responsiveness to FOI Requests. *Journal of Public Administration Research and Theory* 29(2), 210–226.
Roberts A (2006) *Blacked out: Government Secrecy in the Information Age*. Cambridge University Press, New York, NY.
Roberts ME, Steward BM, Tingley D *et al.* (2014) Structural Topic Models for Open Ended Responses. *American Journal of Political Science* 58(4), 1064–1082.
Saltzstein GH (1992) Bureaucratic Responsiveness: Conceptual Issues and Current Research. *Journal of Public Administration Research and Theory* 2(1), 63–88.
Schnell S (2017) Cheap Talk or Incredible Commitment? (Mis)Calculating Transparency and Anti-Corruption. *Governance* 31 (3), 415–430.
Spáč P, Voda P, Zagrapan J (2018) Does the Freedom of Information Law Increase Transparency at the Local Level? Evidence from a Field Experiment. *Government Information Quarterly* 35(3), 408–417.
United Nations (2016) *United Nations E-Government Survey 2016: E-Government in Support of Sustainable Development*. United Nations, New York, NY.
West WF, Raso C (2012) Who Shapes the Rulemaking Agenda? Implications for Bureaucratic Responsiveness and Bureaucratic Control. *Journal of Public Administration Research and Theory* 23(3), 495–519.
White AR, Nathan NL, Faller JK (2015) What Do I Need to Vote? Bureaucratic Discretion and Discrimination by Local Election Officials. *American Political Science Review* 109(1), 129–142.
Wood AK, Lewis DE (2017) Agency Performance Challenges and Agency Politicization. *Journal of Public Administration, Research, and Theory* 27, 581–595.
Worthy B, John P, Vannoni M (2017) Transparency at the Parish Pump: A Field Experiment to Measure the Effectiveness of Freedom of Information Requests in England. *Journal of Public Administration, Research, and Theory* 27, 485–500.
Yang K, Callahan K (2007) Citizen Involvement Efforts and Bureaucratic Responsiveness: Participatory Values, Stakeholder Pressures, and Administrative Practicality. *Public Administration Review* 67(2), 249–264.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Appendix S1. Supplementary Material

