Power in Text: Implementing Networks and Institutional Complexity in American Law

Short title: Power in Text

Robert Shaffer
University of Pennsylvania
shafferr@upenn.edu

March 17, 2020

Abstract

How should social scientists measure institutional complexity? Formal (textually-defined) institutional design - and particularly the complexity of formal institutions - is an important object of study across political science, law, and public administration. However, due to measurement constraints, existing work on formal institutional design focuses either on single policy areas or "important" legislation, creating clear selection problems. In this paper, I propose and validate a novel natural language processing approach designed to scalably extract networks of institutional relationships from legal texts. These "implementing networks" offer a straightforward way to represent the institutional content of law, and naturally suggest measures for quantities like institutional complexity. I then apply this method to measure institutional complexity in all American laws enacted from 1993-2014. This approach reveals a surprising disconnect between partisan disagreement and institutional complexity among lower-profile legislation, which would have been difficult to detect without this approach.

Keywords: American politics, natural language processing, machine learning, law, institutional design

This project has benefited immensely from feedback at the 2019 American Political Science Association and the 2018 Society for Political Methodology Annual Meeting. Thanks in particular to Zachary Elkins, Bryan Jones, Stephen Jessee, John Wilkerson, Beth Simmons, Nick Beauchamp, Dan Hopkins, Max Goplerud, Maraam Dwidar, Jonathan Chu, Mike Kenwick, Benjamin Laughlin, and three anonymous reviewers for an array of comments and suggestions. All mistakes, of course, remain my own.

Supplemental material for this project is available in the appendix in the online edition. Replication files are available in the JOP Data Archive on Dataverse (http://thedata.harvard.edu/dvn/dv/jop).

1 Introduction

How do legislators allocate policymaking authority? As any would-be lawyer knows, laws establish relationships between actors, describing who can do what, when, and to whom. These formal institutional design choices represent an important object of study within and beyond political science, and occupy a correspondingly high-profile position in public discourse. For example, consider the Patient Protection and Affordable Care Act (ACA). As a core part of Barack Obama's Presidential agenda, the ACA represented a sweeping overhaul of the American healthcare system, which expanded access to health insurance and allowed regulators to implement new insurance regulations. Supporters argued that these changes would empower experts and protect consumers, while detractors worried that the ACA placed undue authority with a few unelected bureaucrats. The discussion surrounding so-called "death panels" offers a immediate example of this latter set of concerns. Though fantastical, the fear that a single government regulator might be given unchecked authority to make end-of-life decisions for Medicare patients resonated with voters, and represented one of the most durable talking points in the ACA debate (Oberlander 2012).

Because formal institutional design represents such a central object of study, scholars from political science, law, and public administration have developed an array of measurement schemes designed to extract institutional content from text (e.g. Crawford and Ostrom 1995; Epstein and O'Halloran 1999; Huber and Shipan 2002; Franchino 2007; Farhang and Yaver 2016). These schemes have been applied across a variety of national contexts and policy areas, and have produced important insights into the institutional design process. Nearly all, however, also share an important limitation. Because these measurement approaches rely on labor-intensive hand-coding decisions, studies based on the data they produce tend to focus either on single policy areas or "important", high-profile laws.

These measurement-induced scope constraints create clear selection problems, with important theoretical implications. For example, political scientists have argued that formal institutional design - and particularly the complexity of formal institutions - is primarily driven by legislative-executive preference disagreements (see, e.g. Moe 1990b; Farhang and Yaver 2016). Complex institutional structures, these authors argue, slow the implementation of undesirable legislation and expand opportunities for oversight of executive action. However, as I argue in this paper, this pattern is only likely to hold for high-profile legislation. Since designing complex institutions is costly, we should only expect lawmakers to exert the effort required to implement their policy preference through formal institutions on issues which are salient to their constituents. Unfortunately, since measurement constraints have led previous studies to focus on high-profile laws, these studies tend to overstate the relationship between partisanship and institutional design.

To address these measurement shortcomings, I propose a novel natural language processing-based conceptualization and measurement scheme designed to scalably extract institutional information from legal documents. Conceptually, I argue that legal texts are *relational* documents, which describe which actors can do what, when, and to whom. I then implement and validate a scalable neural network-based approach designed to extract these "implementing networks" from legal texts, which combines a customized named entity recognition model with case-specific knowledge regarding Congressional drafting standards. This approach therefore offers a straightforward way to extract institutional information from national-level American legislation, which is easily adaptable to other legal contexts.

As an application of this approach, I generate implementing networks for nearly all enacted national-level American laws passed from 1990-2014. I then use this information to create a network-based measure of the complexity of each law's institutional structures. Like previous scholars, I find that high-importance laws passed under divided government are more complex than those passed under unified government. However, unlike previous work, I find that this relationship is a contingent one. Unlike their high-importance counterparts, more "everyday" laws passed under unified and divided government are similarly complex. These findings both demonstrate the value of the measurement approach I propose, and invite scholars to reconsider the role of partisanship in the institutional design process.

2 Operationalizing Institutional Complexity

2.1 Defining Formal Institutions

When a lawmaker writes a statute, that lawmaker must choose between a range of possible implementing structures. At one extreme, she could create a complicated, interconnected institutional system, which contains many actors and involves many of those actors in each decision. At the other, she could write a simple, vague document, which allows one or a few actors to unilaterally make decisions. The choice between a simple, streamlined implementing structure and a fragmented, diffuse one is consequential. By splitting decision-making authority among many actors, lawmakers reduce implementing efficiency and limit implementer discretion, but offer more opportunities for outside groups to monitor and intervene into the decision-making process. Streamlined implementing structures, by contrast, offer a more efficient implementation process but fewer oversight opportunities.

This basic framework highlights two key characteristics of formal institutional structures: namely, the *number of actors* involved in a formal institutional system, and the *number of collaborative decision points* that system contains. Here, I adopt an inclusive definition of both of these subcomponents; potential "actors", in this usage, include Congressional and judicial actors as well as traditional implementing actors such as administrative agencies and government-sponsored entities, and potential "collaborative" relationships include informal consultation and reporting relationships as well as formalized oversight and approval requirements. Put together, these two components form a concept I term the "complexity" of a law's implementing structure. More complex implementing structures involve more actors and more overlapping decision points, while simpler ones involve fewer actors and a more siloed, independent implementing structure.

This definition - which is similar to Farhang and Yaver (2016)'s definition of administrative "fragmentation" - occupies a middle ground between the highly detailed and highly abstract conceptualization schemes used elsewhere in the literature.[1] On the more detailed side, McCubbins (1985, 724-729), Epstein and O'Halloran (1999, 101), and related authors (see, e.g. Franchino 2004, 2007; Ainsworth and Harward 2009) outline rich multilevel schemes containing as many as 14 types of "structural" powers, relationships, and constraints. By contrast, the definition of "complexity" I employ combines all types of relationships - such as formal joint decision-making relationships, oversight and judicial/administrative review, or reporting/consultation requirements - into a single category which I term a "co-involvement" relationship. On the more abstract side, Crawford and Ostrom (1995) define an abstract "grammar" of instiutional relationships, which they and related authors apply to develop document-specific coding schemes designed to categorize institutional statements made within particular legal documents (see also Basurto et al. 2010; Siddiki et al. 2011, 2019). Huber and Shipan (2002) abstract further still, and lump directive language, institutional constraints, and procedural requirements into a single high-level concept they term the "specificity" of a legal text (see also VanSickle-Ward 2014). By contrast, the definition of "complexity" I adopt does not consider the presence or absence of directive language or procedural requirements, and instead focuses narrowly on the institutional setting and institutional relationships contained within a given legal document.

Despite these limitations, this definition of "complexity" nevertheless captures many of the most important and influential choices lawmakers must make during the institutional design process. By involving more actors in the implementation process and creating a greater number of co-involved decision points, legislators can force relevant actors to cooperate or consult with other (often hostile) actors, which slows implementation and reduces implementer discretion (McCubbins et al. 1987).[2] Slowing implementation of a law can give civil

[1]In particular, Farhang and Yaver (2016) define a law as "fragmented" to the extent that it "[divides] implementation authority over a larger number of distinct actors, over a larger number of different agencies, and [gives] multiple actors the authority to perform the same function with respect to the same statutory provisions." The definition of "complexity" that I adopt is conceptually similar, but includes a broader set of actors - including Congressional committees and some private-sector actors - in addition to the administrative and judicial actors that Farhang and Yaver (2016) focus on.

[2]Ting (2003) makes similar claims with respect to redundant bureaucratic structures, which can help lawmakers to achieve more favorable outcomes when their priorities are not aligned with those of bureaucracy.

society groups more time to mobilize and sound "fire alarms" against undesirable executive activity (McCubbins and Schwartz 1984; Kagan 2009; Farhang 2010). Moreover, at least in the American context, complex decision-making procedures offer interest groups more potential "hooks" on which to hang legal challenges to government action (Kagan 2009; Carpenter et al. 2012). Certainly, these effects are sharper for formalized joint approval structures compared with simple notice or consultation requirements, such as Congressional reporting rules. However, even straightforward, procedural co-involvement relationships can empower outside interest groups and place real constraints on executive authority.

2.2 Measurement and Scalability

Just as scholars have outlined a diverse array of conceptualization schemes to categorize the institutional content of legal texts, they have also proposed a broad set of measurement approaches designed to extract that information from documents of interest. Most, however, can be broadly categorized as hand-annotation schemes.[3] For example, Epstein and O'Halloran (1999) and Franchino (2004) use summaries of legislative texts to identify the presence or absence of the institutional content categories they identify, while Farhang and Yaver (2016) apply a similar procedure to actual legislative texts. VanSickle-Ward (2014) and Basurto et al. (2010) take this procedure further still, and use abstract conceptual categories to define distinct coding schemes for each legal document that they examine.

Though these approaches produce rich information on the content of legal texts, they also share an important limitation: namely, *scalability*. Legal documents are long, complicated, and difficult to interpret, even for experienced experts.[4] As a result, extracting information from them - whether on institutional design or another topic - is highly labor-

[3]Anastasopoulos and Bertelli (2017) offer a promising supervised learning-based alternative designed to address this measurement problem. However, their method relies on the availability of hand-coded training data. Because training data of this kind is labor-intensive to collect, obtaining the necessary training data is likely unrealistic in many research settings.

[4]See Siddiki et al. (2019, 20-21) for a related discussion of the challenges applying Crawford and Ostrom (1995)'s "institutional grammar" tool at scale, and a call for the development of automated or semi-automated coding approaches.

intensive. To underscore the extent of this measurement challenge, consider Farhang and Yaver (2016)'s study of fragmentation patterns in federal US legislation, which is perhaps the closest comparison point for the definition of complexity that I adopt in this paper. As part of their coding efforts, the authors read and annotated some 24,000 pages of legislative text in order to produce data on some 366 "historically significant" laws passed from 1947 to 2008. Though heroic, hand-coding protocols like these are clearly impractical when applied to larger corpora. The few alternative approaches that avoid these scalability problems - such as Huber and Shipan (2002)'s word-count method, which treats statute length as a proxy for the complexity of that statute's legal and institutional structures - are inapplicable for comparisons across policy domains (see also Clinton et al. 2012).[5] As a result, most studies of formal institutional design have been forced to limit the scope of their work, either by focusing on one or a few policy areas (e.g. Huber and Shipan 2002; VanSickle-Ward 2014) or (more commonly) by focusing on high-salience, "important" legislation (e.g. Epstein and O'Halloran 1999; Farhang and Yaver 2016).

Scalability limitations like these are problematic in at least two respects. First, from a descriptive standpoint, the labor-intensive nature of textual data collection efforts means that we know very little about the institutional content of most laws, even in an institution as well-studied as the US Congress. Most laws enacted by Congress are not the kinds of transformative legislative actions that receive the bulk of scholarly attention. For example, take Mayhew (1991)'s list of "significant" legislation, an extended version of which forms the sample used by Farhang and Yaver (2016). For the period from 1993-2014 (103rd-113th Congresses), Mayhew's list contains some 128 pieces of legislation, compared with 4,609 total public laws passed during the period.[6] Of course, not all of these enactments were impactful; approximately one-quarter (1,126) were relatively trivial "commemorative" laws, which established monuments or symbols designed to commemorate noteworthy people or

[5]Even within a single policy domain, this method is likely imprecise. As Huber and Shipan acknowledge, statutes can be longer because they describe a single policy program with greater specificity, because they describe a larger number of programs, or because they contain more "filler" language with little legal impact.

[6]See Mayhew's list for details.

places in US history.[7] The remaining three-quarters, however, are highly heterogeneous, ranging from technical, procedural laws[8] to important laws that transformed a particular policy domain. For a concrete example, consider the Enhanced Partnership with Pakistan Act of 2009 (Pub. L. 111-73), which I use as a running example in this paper. Over its five-year lifespan, the Enhanced Partnership with Pakistan Act authorized some $7.5 billion non-military aid to the government of Pakistan, primarily for anti-terrorism and development purposes. Though not significant enough to make a list like Mayhew's, the law nevertheless involved substantial financial outlays and represented an important shift in US-Pakistan relations. Unfortunately, measurement limitations like those I describe above mean that we know little about the content of these influential but not "historically significant" laws.

Second - and perhaps more importantly - focusing on "important" legislation creates a clear selection problem. Scholars have long noted that the lawmaking process for lower-profile legislation follows a different process than the corresponding process for high-profile, historically significant laws. For example, Coleman (1999) and Howell et al. (2000) show that high-importance, "landmark" laws pass at lower rates under divided government, while important but lower-profile laws pass at similar rates in both cases. More recently, Casas et al. (2018) find that majority-party "hitchhiker" laws - or, laws which were proposed as independent bills but enacted as part of unrelated laws - are only 25-50% more likely than their minority-party counterparts to pass, compared with a 200-300% estimated difference for majority-party complete bill proposals (see also Wilkerson et al. 2015). And, VanSickle-Ward (2014) finds that when considering "important" issues, lawmakers are more likely to craft more "specific" laws under divided government than unified government. But, when considering less consequential issues, this relationship vanishes. Results like these suggest that minority-party lawmakers are substantially more successful at shaping the content of legislation for lower-profile lawmaking activities, compared with their higher-profile counter-

[7]As defined by the Congressional Bills Project.

[8]E.g. Pub. L. 108-306, "To provide an additional temporary extension of programs under the Small Business Act and the Small Business Investment Act of 1958 through September 30, 2004, and for other purposes".

parts.

Broadly speaking, we should expect the institutional design process to follow a similar pattern. As Moe (1990a; 1990b) and Moe and Caldwell (1994) argue, complex institutional structures reduce decision-making efficiency, but offer a useful way for lawmakers to check executive-branch activity (McCubbins and Schwartz 1984; Kagan 2009; Farhang 2010). However, compared with other possible approaches to check executive activity, designing complex institutional structures is costly. If a lawmaker wants to create a complicated institutional structure, she will need to carefully research existing legal rules, solicit feedback from interest groups and agency leaders, and design incentives and procedures to encourage bureaucrats to reach favorable decisions. As a result, lawmakers should be *more willing* to exert the energy required to create a complex implementing structure when the policy problem addressed by a law is *more important* to that lawmaker's public and elite constituents. In addition, the effect of partisan disagreement on institutional complexity should be *larger* for laws that address higher-visibility policy problems.

Theoretical expectations like these are straightforward and intuitive, and match related findings reported elsewhere in the literature. Unfortunately, measurement constraints have largely forced scholars to focus on "high-profile" legislation. As a result, existing work likely overstates relationships such as the one between partisanship and institutional complexity, which is likely to be strongest for "important" laws. To avoid these kinds of problems, then, a more scalable and generalizable measurement approach is needed.

3 Extracting Implementing Networks from Legal Texts

To address these measurement challenges, in this section I introduce a novel network- and language-based conceptualization of formal institutional power, which I use to develop a scalable measure of institutional complexity. In social science research, a common way to study the power relationships contained in a document is to examine the *relational* statements that document contains. For example, Franzosi et al. (2012) use a semi-automated method to study portrayals of agency and victimhood in newspaper accounts of lynching episodes in the American South. In international relations, a notable example of this kind of approach is GDELT (Leetaru and Schrodt 2013), which mines news accounts for subject/action/object triples corresponding to international events. Closer to the applications of this paper, Crawford and Ostrom (1995)'s "institutional grammar" approach offers an abstract conceptualization scheme that leverages grammatical information to identify "institutional statements" relating actors and objects of regulation. In their original paper, Crawford and Ostrom use this scheme to develop formal models of institutional interactions, but subsequent work in public policy and public administration has applied their approach to develop coding schemes designed to extract empirical information on the institutional content of legal and administrative documents (see, e.g. Basurto et al. 2010; Siddiki et al. 2011, 2019).

If we view a legal text as a collection of relational statements, then a reasonable way to represent that text is as a *network* of institutional relationships.[9] The "nodes" in this network represent actors involved in the execution of powers outlined in the text, while the "edges" represent the relationships between them. Depending on the scope of the document, these relationships might be directed or undirected, and might include simple connection types or more complex ones. To return to the example from §2.2, the Enhanced Partnership with Pakistan Act provides joint authority to the Departments of State and Defense over the Pakistan Counterinsurgency Capability Fund (an undirected, joint implementation relationship), while empowering the Secretary of State alone to certify to Congress that conditions for the release of aid were being met (a directed oversight relationship). By collecting and combining all such entities and relational ties, we can construct an "implementing network" for laws like the Enhanced Partnership with Pakistan Act, which describes the set of formal

[9]Siddiki et al. (2011, 96) propose a related approach they term a "configuration analysis". Under this approach, researchers organize the institutional statements contained in a given law into a bipartite network composed of actors (or "attributes") and objects of regulation. Because Siddiki et al. (2011) focus on the relationship between actors and regulatory instruments or objectives, their goals are distinct from the ones in this paper. However, their use of network data structures and visualizations as summarizations of the content of law is related to the approach I propose.

institutional relationships envisioned by the document. We can then use this representation to measure downstream quantities of interest such as the overall complexity of the network.

Besides its conceptual advantages, a relational conceptualization of formal power also helps to clarify the measurement tasks involved in extracting quantities such as institutional complexity. At their most fundamental level, networks are constructed from *nodes* (the actors involved in the network) and *edges* (the ties between them). Treating formal legislative texts as implementing networks implies two tasks: in particular, we need to identify the *set of implementing actors* named in a particular legal document, and the *relationships* between those actors. As I show in the following subsections, both of these quantities can be extracted in a straightforward, scalable fashion using natural language processing methods.

3.1 Entity Extraction

To construct an implementing network for a legal text, the first step is to identify the set of government actors named by that text. In political science, perhaps the most common approach to this problem is a dictionary-based system (see, e.g. Leetaru and Schrodt 2013), in which the set of potential government actors is pre-identified using an expert-generated dictionary. Dictionary-based approaches are usually narrowly tailored to particular projects and produce few false positives, but offer a correspondingly high false negative rate, since generating a comprehensive named entity dictionary is impractical for most projects. Alternatively, computer scientists have expended substantial effort to produce general-purpose systems designed to identify so-called "named entities" and organize them into predefined categories, such as people, locations, or geopolitical entities (see, e.g. Manning et al. 2014). Applied to specialized problems - such as recognizing names of government actors - these approaches offer the opposite trade-off to a dictionary based system. Because off-the-shelf named entity recognition models are trained using large-scale, "generic" datasets consisting of internet text, newswire, or other publicly available data, these models will usually capture most entities of interest, and return a low false negative rate. However, off-the-shelf named entity recognition models will also frequently identify irrelevant entries, producing a high false positive rate that offsets these gains.[10]

To address these shortcomings, I therefore opt for a customized machine learning approach. This solution splits the difference between these two alternatives by training a specialized named entity recognition model using expert-generated lists of government entities. To avoid false negatives like those produced by the dictionary method, I use a machine learning approach to identify names of actors. In particular, I use a long short-term memory (LSTM) neural network as my annotation model, which is a machine learning method common in modern natural language processing. Broadly, LSTMs are a type of recurrent neural network, which allow the classification decision for each individual token to be influenced both by characteristics of the token and by characteristics and classification decisions for "adjacent" data points. This approach creates a recursive, context-sensitive prediction structure ideal for classifying sequentially-organized data. Since most linguistic data possess a natural ordering and context, LSTMs are a natural choice in this domain, and have been used for tasks such as language modeling (Sundermeyer et al. 2012), part of speech tagging (Huang et al. 2015; Plank et al. 2016), and classification (Chang and Masterson 2019).

From an implementation standpoint, I rely on the neural network architecture proposed by Lample et al. (2016) and Ma and Hovy (2016).[11] Given a textual excerpt, this implementation predicts token-specific tags based on two sources of information:

1. pre-trained embedding vectors for each word (here, drawn from GloVe, trained on the Google News corpus and described by Pennington et al. (2014)); and
2. embedding vectors and predicted tags for left- and right-adjacent terms.

This approach allows the classifier to infer both simple and complex tagging rules. For example, incorporating predicted tags for adjacent words allows the model to correctly

[10]More specifically, off-the-shelf named entity recognition tools are trained to place recognized named entities into a set of generic categories, such as PERSON, ORG, or LOCATION. As a result, these models usually recognize names of government actors, but tend to place them inside categories containing irrelevant named entities. Without some additional information, separating relevant from irrelevant named entities within a particular category is not possible, which produces a high false positive rate.

[11]As implemented in Tensorflow and Python in the tf_ner library. This implementation is slightly different from the one outlined in the two papers I cite in-text; for details, see the accompanying documentation. See Appendix A.3 for details regarding parameter settings and training.

identify multi-word entities, making recovery of these items straightforward during post-processing. Word embeddings, by contrast, incorporate more subtle information regarding word usage and semantic patterns, which can be used to identify words which are commonly contained in institution names of interest (e.g. "Secretary" or "Agency").

To avoid false positives produced by "generic" entity recognition approaches, I train the LSTM model using a custom-built training set consisting of in-text examples of government actors. I constructed this training set using a three-step process. First, using Wikipedia and US government sources, I built a custom dictionary ($n = 1,346$) of common federal US government entities, which includes executive-branch agencies and departments, Congressional committees, and government-sponsored enterprises.[12] Second, I scraped, cleaned,[13] and split into sentences[14] all non-appropriations bills included in legislation enacted from 1993-2014.[15] Third, for each sentence, I conducted a simple string search for each entity contained in the entity dictionary, and tagged any entities found using a standardized (I)nside/(O)utside/(B)eginning (IOB) tagging scheme (see Table 1 for an example and Ramshaw and Marcus (1999) for discussion).[16] This process left me with some 90,711 sentences containing at least one named entity, which formed the final training set.

How well does this custom-trained LSTM approach perform? Since the purpose of

[12]In particular, I first scraped names from usa.gov, the Federal Register, or one of five Wikipedia sources: specifically, the lists of federal agencies, defunct federal agencies, House committees/subcommittees, Senate committees/subcommittees, and joint committees. I then removed common prefixes and suffixes from these items (e.g. "United States"; "USA"), and stripped names of states and national governments (e.g. "Texas"; "California"; "Federated States of Micronesia") from the list. As an additional quality control measure, an undergraduate research assistant read a random sample of 1,000 sentences, and supplemented this list with a series of additional missing items.

[13]In order to maintain parity between the LSTM training set and the testing and prediction sets, I used identical preprocessing steps for both training and prediction. See §4.1 for a description of the cleaning procedure used for each document.

[14]Using the pretrained Punkt sentence tokenizer, available via NLTK.

[15]As defined using Casas et al. (2018)'s data. This dataset includes both enacted legislation and "hitchhiker" bills (independently-proposed bills included as amendments to other enacted legislation).

[16]Since some named entities are substrings of others - for example, compare "Secretary of Defense" with "Assistant Secretary of Defense" - a naive implementation of this scheme might create incorrect tagging choices. For example, if the string "Assistant Secretary of Defense" was tagged before "Secretary of Defense", both "Assistant" and "Secretary" would receive "B" tags. To avoid this problem, I configured the tagging protocol to only tag tokens that had not already received a named entity tag. I then ordered the entity dictionary from longest to shortest, to ensure that longer named entities would be tagged first.

Table 1: Sample training example

Token	Tag
Funds	O
herein	O
appropriated	O
to	O
the	O
Department	B-MISC
of	I-MISC
Defense	I-MISC
for	O
construction	O
shall	O
be	O
available	O
for	O
hire	O
of	O
passenger	O
motor	O
vehicles	O
.	O

Sample training example, formatted according to a modified version of the CoNLL2003 format. Military Construction Act 1992 §102. For original text see the corresponding congress.gov page.

using a machine learning approach for named entity recognition is to capture named entities not already known to the researcher, the most relevant (and stringent) performance test would be one in which we assess the model's ability to recover *unseen entities* not available during training. In order to assess the model's performance in this scenario, I therefore conducted a five-fold cross-validated predictive accuracy study. First, I randomly split my entity dictionary into five equally-sized sets of entity names. For each set, I then identified all sentences exclusively containing entities from the set in question, and used these sentences to form a held-out test set. I then trained a model using sentences containing entities from the remaining four groups, and calculated precision, recall, and F1 scores for the held-out test set.[17] Finally, I repeated this process for each group, and averaged the performance

[17]Defined as $F_1 = \frac{2PR}{P+R}$, $P = \frac{TP}{TP+FP}$ and $R = \frac{TP}{TP+FN}$, where P and R denote precision and recall, TP and TN denote the count of true positive/negative examples correctly classified and FP and FN denote

statistics to produce the final set of results.

Assessed according to these metrics, the LSTM model I employ achieved a cross-validated F1 score of 0.817, with a cross-validated precision of 0.904 and recall of 0.743. These values are comparable to or higher than those reported elsewhere in the literature; for example, Augenstein et al. (2017) report an overall F1 score of 0.507 for entity recognition on unseen named entities, averaged across three modeling approaches and 19 standard benchmarking datasets. For the standard CoNLL datasets - which contain newswire text that is relatively similar to the language I examine in this paper - these authors report F1 scores of 0.844 and 0.871. Overall, then, the performance results I report match state-of-the-art performance on standard testing datasets, and represent a strong basis from which to work.

3.2 Relation Extraction

The second step to constructing an implementing network is to identify the *relationships* between actors named by a text. As I describe in §2.2, for the purposes of this paper I focus on *co-involvement* relationships, which I define as instances in which two or more actors are jointly involved in the execution of a particular policy decision. This definition is similar to the one used by Farhang and Yaver (2016), and offers both theoretical and practical advantages. Involving more actors in a policy decision creates more opportunities for "fire alarm"-style oversight and more decision points that outside interest groups can use to influence the policymaking process. Details about the nature of these ties - such as the hierarchy of the two entities, or the role assigned to each - are informative, but less significant. So long as we can be reasonably confident that each policy decision under consideration is roughly comparable in its substantive scope, counting the number of actors involved in a law's policy decisions offers useful information about the complexity of that law's institutional relationships, and the ways that those institutional relationships assist or inhibit oversight efforts from civil society groups.

the counts of true positive/negative incorrectly classified.

Fortunately, drafting guidelines for American legislation make these kinds of co-involvement relationships relatively easy to identify. As noted in the drafting guide for the US Consolidated Code, the "basic unit" of federal US law is the *section*.[18] According to the Code, *sections* "shall contain, as nearly as may be, a single proposition of enactment".[19] As a result, if we observe that two actors are co-mentioned in a section of a law, we can reasonably conclude that those two actors are likely involved in the policy decision described in that section.[20] Since sections in modern legislative texts are demarcated using consistent formatting rules, drafting rules like these offer a useful way to identify instances in which decision-making authority is split between one or more actors.

Though useful for the purposes of this article, this relation extraction approach is clearly not applicable for all research questions. Because the co-occurence networks I construct tell us relatively little about the nature of the relationships between actors, this method cannot be used to determine which entities have oversight or veto power over others, or which decisions require formal approval (as opposed to simple advice consultation) from multiple entities (contrast with Siddiki et al. 2011). Moreover, like most machine-assisted measurement schemes, in order to achieve scalability improvements this approach makes simplifying assumptions that may not be accurate in all cases. As a result, the results produced by this method are less useful for studies that require detailed information about the institutional content of a one or a few laws. Rather, this method is most useful for summarizing the broad character of the institutional structures contained in a given law through statistics which describe that structure's complexity or size, or for comparing those statistics across large collections of legislation.

[18] http://uscode.house.gov/detailed_guide.xhtml

[19] 1 U.S.C. §104

[20] As I describe in §5, during preprocessing I remove some sections that obviously violate this assumption, such as short titles, findings of Congress, and "Table of Contents" sections.

3.3 Constructing Implementing Networks

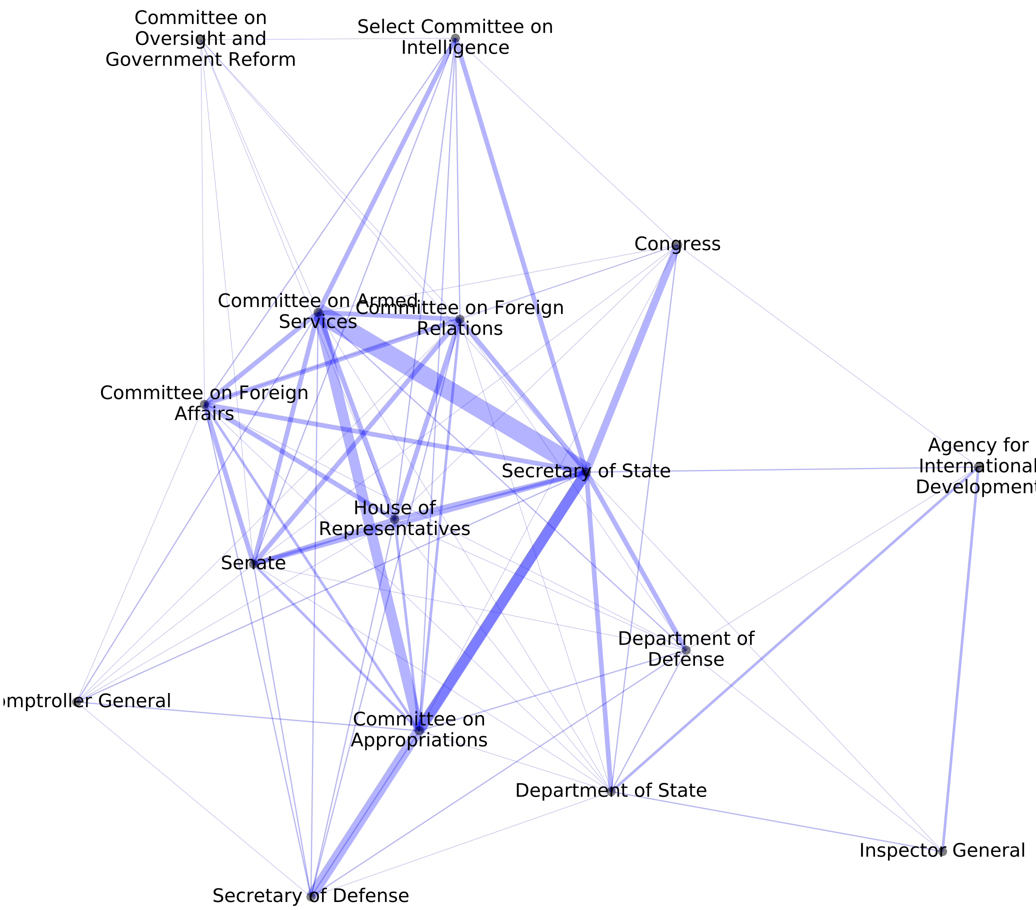
Put together, these named entity and relation extraction approaches offer a simple three-step implementing network construction procedure. First, I split each text into sections.[21] Second, I extract entities from each section using the LSTM model I describe in §3.1. Third, I draw an edge between entities that co-occur in a section. This process yields a weighted implementing network for each law, which can be used both to describe the law's institutional content and to calculate document-level quantities of interest like institutional complexity.

To illustrate this process, I provide a sample output from the Enhanced Partnership with Pakistan Act of 2009 (Pub. L. 111-73). The Enhanced Partnership with Pakistan Act was a relatively straightforward foreign aid law intended to provide military and developmental assistance to the Government of Pakistan. The law authorized the President to provide $1.5 billion in yearly non-military aid from 2010-2014, and provided additional military aid conditional on a certification process implemented by the Secretary of State.[22] Unusually for a defense-oriented law, the Enhanced Partnership with Pakistan Act gave the State Department substantial authority over defense-related aid allocations (Epstein and Kronstadt 2013). As shown in Figure 1, these features are clearly visible in the law's implementing network. The law contains a central cluster consisting of the Secretary of State, the Secretary of Defense, and several Congressional actors. Quantitative assessments of node importance reinforce this visual message; as measured by eigenvector centrality, the Secretary of State is the most central actor in this network (eigenvector centrality of 0.45), followed by the the Committee on Armed Services (0.39), the Committee on Appropriations (0.32), and the Committee on Foreign Relations (0.29). These figures roughly track with qualitative summaries of the law's content, lending this representation substantial face validity.

[21] Using the constitute_tools regular expression parser. This parser separates each document according to a given set of organizational headers (e.g. titles; sections), while maintaining the internal hierarchy of each document. See Appendix A.1 and A.2 for details and sample parsed text.

[22] Enhanced Partnership with Pakistan Act of 2009, §203.

Figure 1: Implementing network, Enhanced Partnership with Pakistan Act of 2009.



Line density is approximately proportional to the number of ties between each node. Node placement is random, but is loosely related to node centrality.

4 Institutional Complexity in American Law

As an application of the procedure I describe in §3, I use the "implementing network" approach to study patterns of institutional complexity for almost all laws passed from 1993-2014. For each law, I first extract an implementing network, and then calculate a complexity value for that network, which I define as the network's *average degree*.[23] I then model this complexity measure using a Bayesian hierarchical regression model. This approach - which I discuss in more detail in §4.4 - allows me to incorporate policy area-specific institutional complexity patterns and separate and model both "standard" legislation and more "technical", procedural laws.

To preview findings, the results from this model suggest that laws proposed by more extremist members, members of the chamber majority, and "high-importance" laws are likely to be more complex. As predicted in §2.2, this approach also uncovers a conditional relationship between issue importance, divided government, and law complexity. Though high-importance laws passed under divided government are more complex than their counterparts passed under unified government, this relationship vanishes for more "everyday" lawmaking activities.

4.1 Developing the Dataset

To investigate institutional complexity in American law, I used Casas et al. (2018)'s data to construct a comprehensive dataset of bill texts (and accompanying metadata) included in laws passed from 1993-2014 ($n = 6,109$).[24] This dataset includes both independently-

[23] As a robustness check, I also repeat all analyses in this section using the *number of actors* as my dependent variable, and draw identical substantive conclusions. See Appendix C.2.5 for details.

[24] This dataset excludes two types of laws: specifically, "trivial" laws, and appropriations and budget laws. As defined by the Congressional Bills Project, "trivial" laws include commemorative legislation and laws which transfer small quantities of land between government entities. Since these laws are not intended to affect the structure of the administrative state, they are unlikely to include allocation-of-authority language. Appropriations and budget laws are also unlikely to include allocation-of-authority language, but for different reasons. Because these laws are primarily designed to disburse funds to various agencies and governmental entities, most of their language is devoted to detailed descriptions of funding decisions, rather than to the powers and duties of governmental entities. As a result, I retain "hitchhiker" bills that eventually became part of an appropriations or budget law (since these "hitchhikers" are usually unrelated to the budgeting and

enacted bills and separately-proposed "hitchhiker" laws that were eventually included in enacted legislation. As Casas et al. (2018) demonstrate, many bills proposed by members of Congress that fail to pass are later included as "hitchhikers" onto future enacted laws. Because these "hitchhiker" bills are drafted and proposed separately from the laws in which they are eventually included, we should expect their content to be primarily influenced by the characteristics of lawmakers who originally proposed them. In order to enable these kinds of comparisons, I therefore isolated the text of each "hitchhiker" bill from the enacted law in which it was eventually included, and treated these "hitchhikers" as distinct units.[25]

Before proceeding, I conducted some simple preprocessing steps. For each text, I first stripped headers, footers, and editor's notes (e.g. date of passage; legislative history; transcription notes). I then split each law text into sections, and removed sections containing tables of contents, short titles, and related non-legally binding language.[26] Finally, for all remaining sections I removed section titles and law common names from each section's text, to avoid false positives contained in these names.[27]

4.2 The Dependent Variable: Network Complexity

As I describe in §2.1, the formal implementing structures contained in a given text are *complex* to the extent they grant a larger number of actors overlapping responsibility over execution of a particular law. In the network context, the most plausible operationalization of this quantity is the *average degree* of the network. This quantity can be interpreted as the average number of instances in which a given actor is coinvolved with other actors in

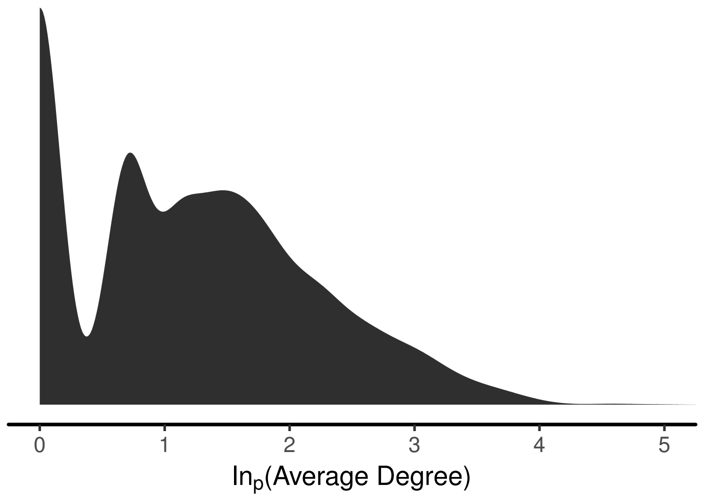
appropriations processes), but remove appropriations and budget laws themselves from the dataset.

[25]In particular, for bills identified by Casas et al. (2018) as "hitchhikers", I compared each section of the hitchhiker to each section of the final "target" law. I then removed all sections from the target law that shared at least 95% of their tokens with a section from the hitchhiker and differed in length by no more than 5%.

[26]In particular, I removed all sections containing any of the terms "Short Title", "Table of Contents", "Finding", "Purpose", "Definition".

[27]For example, the Poison Control Center Enhancement and Awareness Act (Pub. L. 106-174). I used the Cornell Legal Information Institute's list of law common names to identify names to remove.

Figure 2: Density plot, per-law average degree value.



Density plot, log-plus-one transformed average degree value ($n = 4,296$). Laws with undefined average degree values not displayed.

the implementation of the programs contained in a law.[28] To calculate this value, I use the procedure I describe in §3 to extract a distinct implementing network for each law, and calculate an average degree statistic for each law.

Some basic descriptive information for this variable is provided in Figure 2. As this plot suggests, the average degree measure is both widely dispersed and zero-inflated, with a maximum of 639 and some 1,174 observations (approximately 20% of the overall dataset) with an average degree value of zero. Moreover, some 1,813 laws (approximately 30% of the dataset) contain zero named entities, leaving them with an undefined average degree value and reducing the final dataset to $n = 4,296$ observations.

Both of these features are unsurprising from a theoretical standpoint. Not all laws alter the powers or restrictions of government actors. Rather, a non-trivial proportion of the dataset consists of short, technical laws that do not alter the jurisdiction of any actor.[29]

[28]More specifically, the degree of a node in a real-valued network refers to the sum of all edge weights connected to that node. The average degree simply represents the degree value averaged over all nodes in the network.

[29]E.g. Pub. L. 108-306, "To provide an additional temporary extension of programs under the Small Business Act and the Small Business Investment Act of 1958 through September 30, 2004, and for other purposes." This law simply extends authorization for existing provisions of the Small Business Investment

Laws of this latter type - which usually contain zero or a small number of named entities - likely follow a different data-generating process than the one I focus on in this paper, which I address from a technical standpoint in the modeling discussion in §4.4.

4.3 Predictor Variables

As I describe in §2.2, existing work on formal institutional design argue that legislative-executive preference disagreement should be positively associated with institutional complexity. However, I argue that this relationship should be conditioned by the *importance* of a given law. To operationalize preference disagreement, I use a dummy variable indicating whether at least one chamber of Congress was not controlled by the President's party.[30] For law importance, I follow Volden and Wiseman (2014) and use a binary indicator denoting whether the law in question was mentioned in the CQ Almanac's year-end summary of Congressional activity.[31] Since the underlying concept of "importance" is a continuous one, this binary operationalization is necessarily a coarse one. However, it provides a useful, temporally consistent indicator for the public attention paid to a given law, at least among "Congress-watchers" who are heavily invested in legislative activity. As shown in Figure 3, a simple difference-of-means already provides support for the conditional relationship between partisanship and law importance I hypothesize. High-importance bills passed under divided government are approximately 65% more complex than their counterparts passed under unified government ($p < 0.01$), while more "everyday" bills contain indistinguishably complex institutional structures ($p = 0.25$).

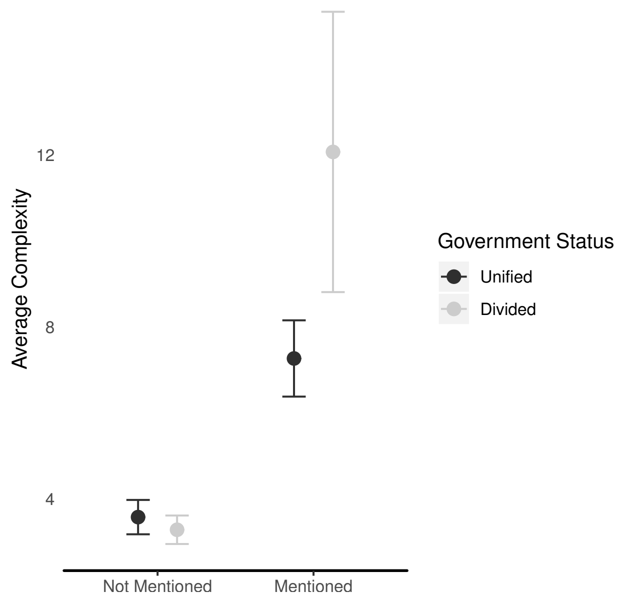
Besides these key theoretically relevant variables, I also collect a series of other law- and individual-level control variables. For the law-level variables, I included a random intercept term corresponding to the Congressional Bills Project's 20 law-level policy codes. These coefficients capture institutional design patterns that are constant within policy areas, and

Act of 1958, and therefore provides no modifications to existing administrative jurisdiction.

[30] See Appendix B for descriptive statistics on all predictor variables.

[31] Using this definition, some 18% of laws fall into the high-importance category.

Figure 3: Average complexity values, by government status and CQ mention.



Difference-of-means plot, showing average complexity values and 95% confidece intervals for the four combinations of the divided government and CQ mention covariates.

should be larger for more complex policy areas (e.g. macroeconomics or defense) and smaller in simpler ones (e.g. public lands). I also included a binary "hitchhiker" indicator variable, and a predictor corresponding to the square root of the number of cosponsors in each law. I expect hitchhikers to contain simpler institutional structures on average compared with independently-enacted legislation, since hitchhikers are by definition subsets of other laws.[32] The relationship between cosponsorship and complexity, by contrast, should be positive, since cosponsorship rates should capture the quality and salience of a given law within Congress. However, since cosponsorship incentives are heterogenous, this expectation is a weak one.[33]

For individual-level predictors, I collected a binary variable indicating whether the proposing member was part of the chamber majority, and continuous variables correspond-

[32]Excluding "hitchhikers" entirely from the model also leaves other model estimates substantively unaffected. See Appendix C.2.4 for details.

[33]For example, Sinclair (2016) notes that time-sensitive laws often bypass the ordinary lawmaking process, which gives lawmakers less time to gather cosponsors. For an instructive example, contrast the Patient Protection and Affordable Care Act (Pub. L. 111-148) with the American Reinvestment and Recovery Act (Pub. L. 111-5). Though both laws received substantial public attention and the latter passed by larger margins than the former, the former law had 40 cosponsors while the latter had only 9. This difference likely reflects the speed with which the bailout law was enacted, compared with the more measured process for the Affordable Care Act.

ing to the proposing member's DW-NOMINATE and squared DW-NOMINATE scores. I expect the relationship between majority status and institutional complexity to be positive, since chamber majority members have a greater ability and incentive to enact ambitious laws that require more complicated institutional structures. For the DW-NOMINATE variables, I expect the relationship between squared DW-NOMINATE - which serves as a measure of extremism[34] - and institutional complexity to be positive, since extremists on both sides are likely to be suspicious of entrenched bureaucratic interests. However, because both liberals and conservatives possess incentives to use institutional complexity as a policymaking tool, I do not have clear expectations for the base-level variable.[35]

4.4 Modeling and Robustness

As I describe in §4.2, the dataset and complexity measure I develop in this paper contain at least three challenging characteristics. First, the operationalization of complexity I use is bounded at zero, with a large number of observations clustered exactly at zero and a long positive tail. Second, the underlying dataset I use consists of two rough "types": a "standard" type, which increases, decreases, or otherwise modifies the named actors' jurisdictions, and a "technical" type, which does not alter the jurisdiction of any actor. "Standard" laws can contain a wide range of actor counts and complexity values while "technical" laws tend to contain very simple structures, which accounts for the cluster of observations with near-zero complexity values. Third, due to historical, path-dependent institutional design patterns, laws in different policy areas potentially contain different levels of institutional complexity. Since policy area is likely correlated with partisanship, law importance, and other factors that are also correlated with complexity, these idiosyncratic policy area-specific institutional design patterns should be modeled alongside more contemporaneous political factors.

To address these modeling constraints, I use a Bayesian gamma regression model with

[34] See, e.g., Theriault and Rohde (2011) for similar usage.

[35] For example, liberals are likely more willing to write laws with more ambitious policy goals and more complex implementing structures, while conservatives tend to be skeptical of the administrative state and might be more willing to use complexity as an oversight mechanism.

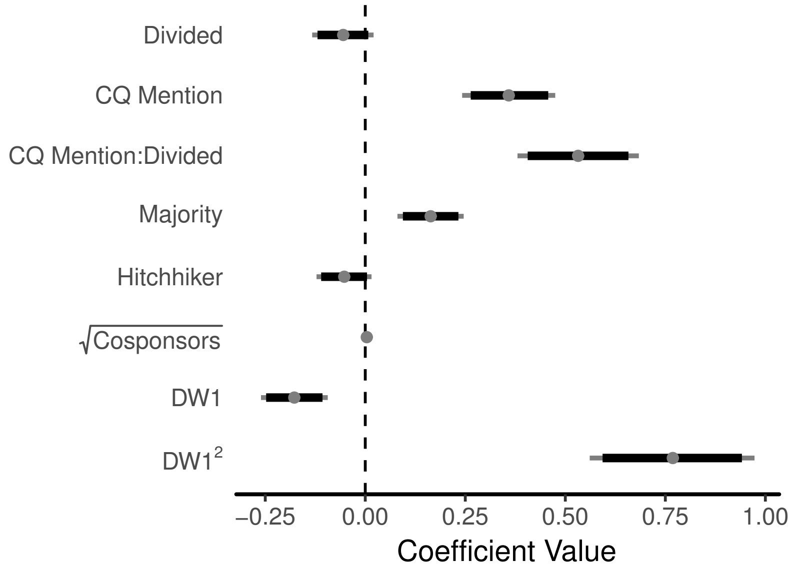
a hurdle component.[36] Modeling the dependent variable as a gamma-distributed random variable incorporates the high dispersion patterns shown in Figure 2. However, because the gamma distribution's support is restricted to the positive real numbers, this distribution cannot incorporate the cluster of observations with complexity values of exactly zero. The hurdle component allows me to model this data feature in a theoretically motivated fashion. Loosely speaking, we can treat hurdle models as a two-step process, in which we first estimate a logistic regression to determine whether a given observation is zero or non-zero and further estimate a gamma regression for the non-zero observations. This structure allows me to estimate separate coefficients for both types of laws, and to handle the distinct data-generating processes that produce each law type. Finally, to handle policy area-specific institutional design patterns, I partially pool the intercept in each model component by policy area, which allows me to capture policy area effects while minimizing loss of explanatory power.

As I note in §4.2, one disadvantage to using the average degree of each law's implementing network as a dependent variable is that it leaves laws with zero named entities with an undefined value on the dependent variable. As a robustness check, I therefore follow Farhang and Yaver (2016) and calculate the *number of unique entities* named by each law as an alternate measure of institutional complexity. This measure is less conceptually precise, but is defined for all laws, which allows me to model all observations. Like Farhang and Yaver (2016) I find that these two measures are highly related (correlation of $r = 0.91$ between log-plus-one transformed versions of each variable), and I draw identical substantive conclusions by fitting an analogous Bayesian negative binomial hurdle model to the gamma regression model I describe above (see Appendix C.2.5 for details).[37]

[36]See Appendix C.1 for details regarding model specification, priors, estimation, convergence diagnostics, and posterior predictive checks.

[37]Since the negative binomial distribution's support includes zero, this operationalization of the dependent variable also makes the hurdle component optional, though still theoreticially desirable. However, as I show in Appendix C.2.6, removing the hurdle component of the model and fitting a simple Bayesian hierarchial negative binomial regression model also returns essentially identical substantive conclusions to the model I describe in-text.

Figure 4: Top-level coefficient estimates, gamma regression model.



Grey dots indicate posterior mean values. Dependent variable is the average degree of each law's implementing network ($n = 4,296$). Thick lines indicate 90% credible intervals, and thin lines indicate 95% credible intervals. Positive estimates indicate that an increase in a given coefficient increases law complexity. Intercept suppressed for readability.

4.5 Results

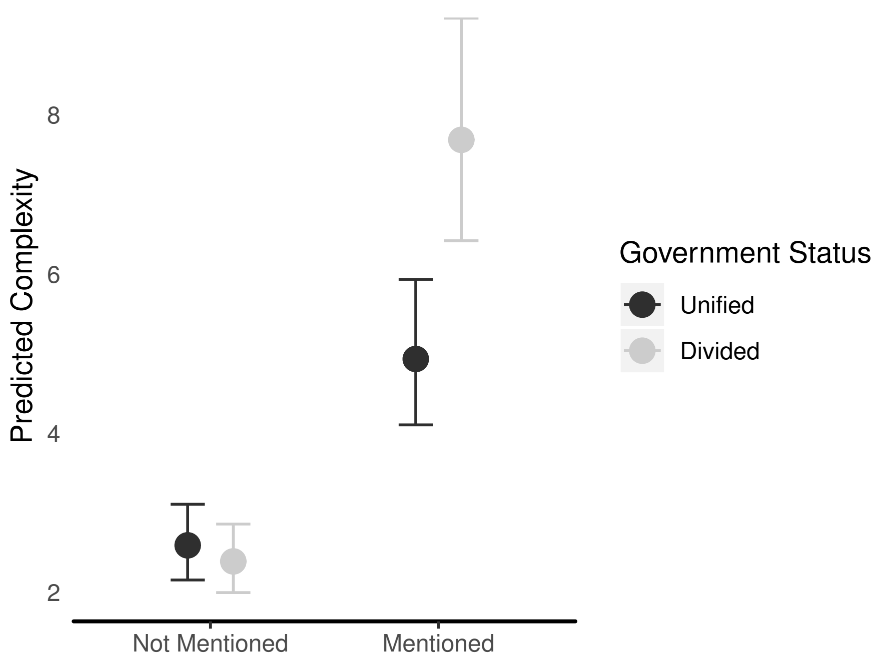
Top-level coefficients for the gamma regression model component - which models the complexity of "standard" laws that actually modify administrative authority - are given in Figure 4.[38] As predicted, the effect of divided government is strongly conditioned by law importance. For more "everyday" laws, the coefficient on the divided government indicator is small and possesses the opposite sign predicted in the literature. By contrast, "high-importance" laws passed under divided government are substantially more complex than their unified-government counterparts. Also as predicted, "high-importance" laws contain consistently more complex implementing structures than their "everyday" counterparts, no matter the background political context.

These findings offer substantial support for the theoretical framework I describe in §2.2, and highlight the selection issues created by focusing on "high-importance" laws. Like

[38]For brevity, I relegate coefficient estimates for the hurdle model - which estimates coefficients for "technical" laws - to Appendix C.2.1.

other authors, I find that executive/legislative preference disagreements are associated with increased institutional complexity. However, this effect is almost entirely constrained to "important" laws that receive attention from legislators' constituents. As shown in Figure 5, for high-importance legislation the effect of executive/legislative preference disagreement is substantial: "important" laws passed under divided government are over 50% more complex than their unified-government counterparts. But, outside of this context, this difference essentially vanishes. Since the measure of "importance" I use in this paper is coarse, probing this finding further with continuous measures of "importance" that differentiate between press coverage, public issue prioritization, and elite, interest group-driven issue prioritization would be a useful direction for future research. However, at the very least, these results complicate findings presented elsewhere in the formal institutional design literature, and suggests that institutional design choices outside of high-importance issues are less influenced by partisan considerations.

Figure 5: Marginal effects plot, divided government/CQ mention interaction.



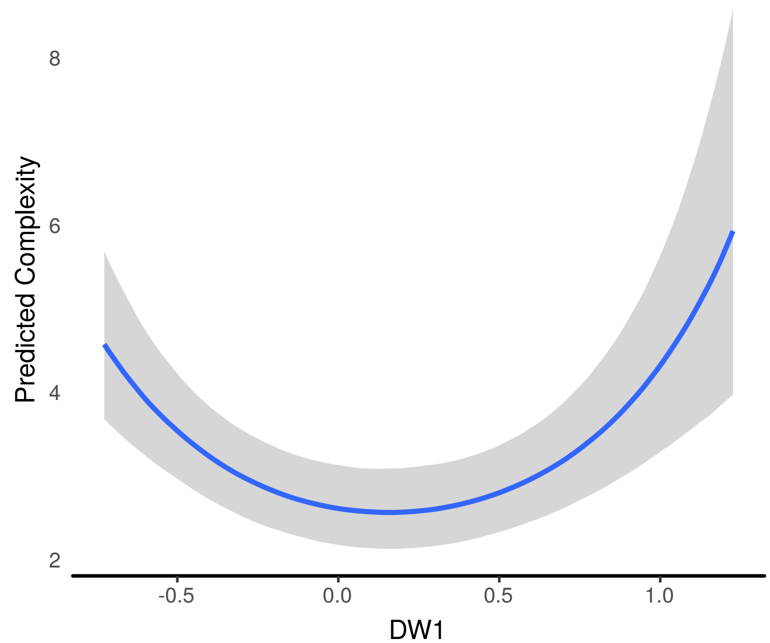
Margins plot, showing predicted values for the four conditions in the divided government/CQ mention interaction. Dependent variable is the average degree of each law ($n = 4,296$). Dots show posterior means and lines show 95% credible intervals. All other predictors in the model are fixed at zero.

Most of the secondary expectations I present in §4.3 are also supported. As predicted, members of the majority tend to pass laws containing more complex implementing structures, which suggests that these members tend to be willing to draft laws addressing more complicated policy problems. Similarly, "hitchhiker" laws tend to be less complex than independently-passed laws, though the coefficient estimate is small and the 95% posterior credible interval crosses zero. Perhaps more surprisingly, the coefficient on the consponsorship variable is essentially zero, which suggests that cosponsorship patterns are essentially unrelated to institutional design choices. However, as I show in the hurdle model results in Appendix C.2.1, laws that receive more cosponsors are more likely to be of the "standard" than the "technical" type. This finding fits with the existing cosponsorship literature. As authors like Wilson and Young (1997) and Box-Steffensmeier et al. (2019) suggest, cosponsorship is generally used by members as a position-taking tool early in the lawmaking cycle, and has a limited relationship with law content or downstream probability of passage. Since members likely derive little utility from taking positions on "technical" laws, we should likely expect these laws to attract few cosponsors.

The relationship between individual-level ideology and institutional complexity also largely matches my expectations. As I describe in §4.3, both liberals and conservatives possess incentives to propose laws containing complex institutional structures, and these incentives are likely to grow for more extreme members in both directions. As a result, it is not clear whether the base-level DW-NOMINATE coefficient should be positive or negative, but the DW-NOMINATE^2 coefficient - which is related to ideological extremism - should be positive. Figure 6 provides support for this set of expectations. For a given DW-NOMINATE magnitude, liberals (members with negative scores) tend to propose slightly more complex laws than conservatives, but the difference between centrists and extremists on either side is substantially larger than the left-right differential.[39] Future work should probe this finding

[39]As I show in Appendix C.2.2, when the squared DW-NOMINATE variable is excluded from the model the corresponding coefficient estimate for the raw DW-NOMINATE variable drops to essentially zero. Substituting the DW-NOMINATE variable for an indicator denoting the party to which the proposing member belongs offers a similar conclusion, suggesting that this finding is robust to measurement and model speci-

Figure 6: Marginal effects plot, DW-NOMINATE score of proposing member.



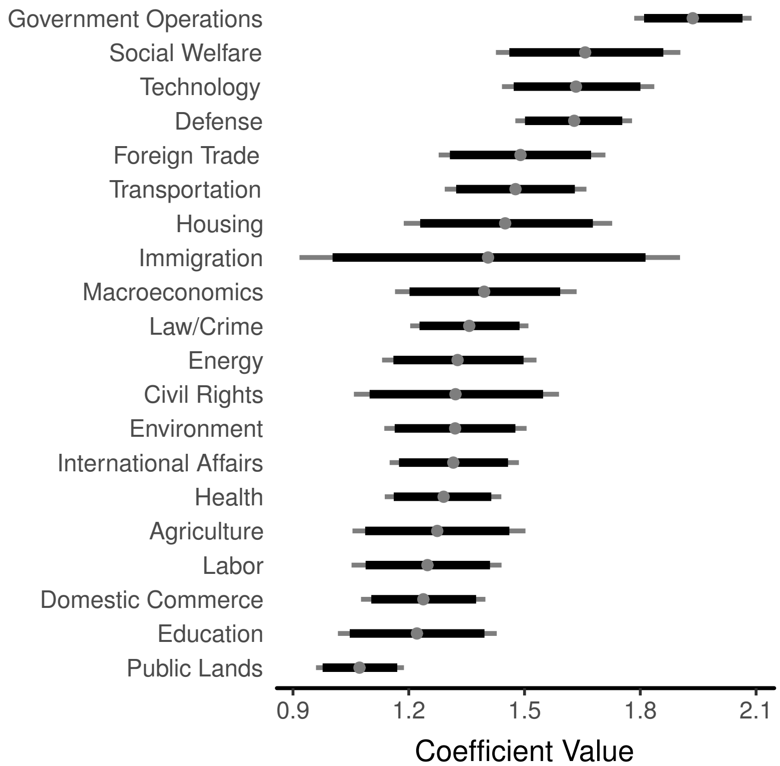
Margins plot, showing predicted values across various levels of the proposing member's DW-NOMINATE score. Dependent variable is the average degree of each law's implementing network ($n = 4,296$). All other predictors in the model were fixed at zero.

further, and examine whether the relationship between ideology and institutional complexity remains constant across policy area and over a longer time horizon.

Finally, the policy area predictions I offer in §4.3 are also at least partially supported. As shown in Figure 7, all else equal public lands laws tend to contain noticeably simpler implementing structures than essentially all other policy areas. This pattern is intuitive; public lands laws often involve small transfers of property between government institutions or small modifications in land regulation or oversight, rather than large-scale modifications of the administrative state. The top end of the scale also fits with existing knowledge, though findings are more mixed. Government operations and defense, for example, are classic high-complexity policy areas, which necessarily involve a large number of government agencies in their implementing process. However, outside of these relatively clear examples, policy area does not appear to have a strong effect on downstream institutional design decisions.

fication choices. However, in all cases, coefficient estimates for other variables in the model are essentially unaffected.

Figure 7: Intercept values by policy area, gamma regression model.



Intercept values, gamma regression model. Coefficient estimates are based on a partially pooled intercept term. Dependent variable is the average degree of each law's implementing network ($n = 4,296$).

5 Conclusion

Overall, this paper offers two primary contributions. Methodologically, I propose, validate, and implement a novel measurement scheme designed to extract "implementing networks" from legal documents. Because extracting institutional design information by hand from legal texts is highly labor-intensive, previous formal institutional design studies have been forced to focus on single policy areas or small subsets of historically significant legislation. These constraints present a clear selection problem, which limits scholars' ability to understand the lawmaking process outside of the high-profile cases studied in existing work. By contrast, the method I present in this paper - which relies on both modern machine natural language processing methods and case-specific knowledge regarding Congressional drafting procedures - offers a more scalable alternative. This approach allows scholars to quickly and easily generate network-based measures of theoretically important quantities of interest such as institutional complexity, which allows them to move beyond these limitations and the selection problems they entail.

As an application of this approach, I extract implementing networks from almost all American laws passed from 1993-2014, which I use to probe the relationship between partisanship, law importance, and institutional complexity. Existing political science scholarship argues that institutional complexity is primarily an administrative oversight tool, which legislators are primarily inclined to use when executive and legislative preferences differ. By contrast, I demonstrate that this relationship only holds for high-visibility, "important" legislation. For more "everyday" laws, institutional complexity and partisanship are essentially unrelated. These findings dovetail with a growing literature on "submerged" bipartisan collaboration over the *content* of legislation in American politics (e.g. Wilkerson et al. 2015; Casas et al. 2018), and demonstrate the utility of the measurement approach I propose.

These results offer a number of directions for future research. Methodologically, augmenting the network extraction procedure I propose with a relation extraction procedure designed to identify types of relationships between actors would enrich the "implementing networks" produced by this method, and allow researchers to move beyond the simple co-involvement relationships I focus on in this paper. Substantively, comparing original and amended versions of bills might help researchers to identify the point in the drafting process in which institutional complexity is introduced. Studying patterns of complexity in historical legislation might help researchers to identify whether the differences between laws enacted under divided and unified government remain constant as the ideological gap between the two parties grows and shrinks. The framework and tools I present could also be applied to legal documents from state governments or governments from outside the United States. Finally, the measurement framework I present also offers opportunities to study other quantities of interest, such as the centrality of particular actors or the frequency with which pairs of actors collaborate on policy implementation.

References

Scott H Ainsworth and Brian M Harward. Delegation and discretion in anticipation of coalitional drift. *American Politics Research*, 37(6):983–1002, 2009.

Jason Anastasopoulos and Anthony M Bertelli. Understanding delegation in the european union through machine learning. 2017.

Isabelle Augenstein, Leon Derczynski, and Kalina Bontcheva. Generalisation in named entity recognition: A quantitative analysis. *Computer Speech & Language*, 44:61–83, 2017.

Xavier Basurto, Gordon Kingsley, Kelly McQueen, Mshadoni Smith, and Christopher M Weible. A systematic approach to institutional analysis: applying crawford and ostrom's grammar. *Political Research Quarterly*, 63(3):523–537, 2010.

Janet M Box-Steffensmeier, Dino P Christenson, and Alison W Craig. Cue-taking in congress: Interest group signals from dear colleague letters. *American Journal of Political Science*, 63(1):163–180, 2019.

Daniel Carpenter, Jacqueline Chattopadhyay, Susan Moffitt, and Clayton Nall. The complications of controlling agency time discretion: Fda review deadlines and postmarket drug safety. *American Journal of Political Science*, 56(1):98–114, 2012.

Andreu Casas, Matthew Denny, and John Wilkerson. More effective than we thought: Accounting for legislative hitchhikers reveals a more inclusive and productive lawmaking process. 2018. Forthcoming: *American Journal of Political Science*.

Charles Chang and Michael Masterson. Using word order in political text classification with long short-term memory models. *Political Analysis*, pages 1–17, 2019.

Joshua D Clinton, Anthony Bertelli, Christian R Grose, David E Lewis, and David C Nixon. Separated powers in the united states: The ideology of agencies, presidents, and congress. *American Journal of Political Science*, 56(2):341–354, 2012.

John J Coleman. Unified government, divided government, and party responsiveness. *American Political Science Review*, 93(4):821–835, 1999.

Sue ES Crawford and Elinor Ostrom. A grammar of institutions. *American Political Science Review*, 89(3):582–600, 1995.

David Epstein and Sharyn O'Halloran. *Delegating powers: a transaction cost politics approach to policy making under separate powers*. Cambridge University Press, New York, 1999.

Susan B Epstein and K Alan Kronstadt. Pakistan: Us foreign assistance. *Current Politics and Economics of the Middle East*, 4(3):575, 2013.

Sean Farhang. *The litigation state: Public regulation and private lawsuits in the United States*. Princeton University Press, 2010.

Sean Farhang and Miranda Yaver. Divided government and the fragmentation of american law. *American Journal of Political Science*, 60(2):401–417, 2016.

Fabio Franchino. Delegating powers in the european community. *British Journal of Political Science*, 34(02):269–293, 2004.

Fabio Franchino. *The powers of the union: delegation in the EU*. Cambridge University Press, New York, 2007.

Roberto Franzosi, Gianluca De Fazio, and Stefania Vicari. Ways of measuring agency: an application of quantitative narrative analysis to lynchings in georgia (1875–1930). *Sociological Methodology*, 42(1):1–42, 2012.

William Howell, Scott Adler, Charles Cameron, and Charles Riemann. Divided government and the legislative productivity of congress, 1945-94. *Legislative Studies Quarterly*, pages 285–312, 2000.

Zhiheng Huang, Wei Xu, and Kai Yu. Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*, 2015.

John D Huber and Charles R Shipan. *Deliberate discretion? the institutional foundations of bureaucratic autonomy.* Cambridge University Press, New York, 2002.

Robert A Kagan. *Adversarial legalism: The American way of law.* Harvard University Press, 2009.

Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. Neural architectures for named entity recognition. *arXiv preprint arXiv:1603.01360*, 2016.

Kalev Leetaru and Philip A Schrodt. Gdelt: Global data on events, location, and tone, 1979–2012. In *ISA annual convention*, volume 2, pages 1–49. Citeseer, 2013.

Xuezhe Ma and Eduard Hovy. End-to-end sequence labeling via bi-directional lstm-cnns-crf. *arXiv preprint arXiv:1603.01354*, 2016.

Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60, 2014.

David R Mayhew. *Divided we govern.* Yale University, 1991.

Mathew D McCubbins. The legislative design of regulatory structure. *American Journal of Political Science*, pages 721–748, 1985.

Mathew D McCubbins and Thomas Schwartz. Congressional oversight overlooked: police patrols versus fire alarms. *American Journal of Political Science*, pages 165–179, 1984.

Mathew D McCubbins, Roger G Noll, and Barry R Weingast. Administrative procedures as instruments of political control. *Journal of Law, Economics, & Organization*, 3(2): 243–277, 1987.

Terry M Moe. Political institutions: the neglected side of the story. *Journal of Law, Economics, & Organization*, 6:213–253, 1990a.

Terry M Moe. The politics of structural choice: toward a theory of public bureaucracy. In O.E. Williamson, editor, *Organization theory: from Chester Barnard to the present and beyond*, pages 116–153. Oxford University Press, New York, 1990b.

Terry M Moe and Michael Caldwell. The institutional foundations of democratic government: a comparison of presidential and parliamentary systems. *Journal of Institutional and Theoretical Economics*, 150(1):171–195, 1994.

Jonathan Oberlander. The future of obamacare. *New England Journal of Medicine*, 367(23): 2165–2167, 2012.

Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.

Barbara Plank, Anders Søgaard, and Yoav Goldberg. Multilingual part-of-speech tagging with bidirectional long short-term memory models and auxiliary loss. *arXiv preprint arXiv:1604.05529*, 2016.

Lance A Ramshaw and Mitchell P Marcus. Text chunking using transformation-based learning. In *Natural language processing using very large corpora*, pages 157–176. Springer, 1999.

Saba Siddiki, Christopher M Weible, Xavier Basurto, and John Calanni. Dissecting policy designs: an application of the institutional grammar tool. *Policy Studies Journal*, 39(1): 79–103, 2011.

Saba Siddiki, Tanya Heikkila, Christopher M Weible, Raul Pacheco-Vega, David Carter, Cali Curley, Aaron Deslatte, and Abby Bennett. Institutional analysis with the institutional grammar. *Policy Studies Journal*, 2019.

Barbara Sinclair. *Unorthodox lawmaking: New legislative processes in the US Congress*. CQ Press, 2016.

Martin Sundermeyer, Ralf Schlüter, and Hermann Ney. Lstm neural networks for language modeling. In *Thirteenth Annual Conference of the International Speech Communication Association*, 2012. `https://www.isca-speech.org/archive/interspeech_2012/i12_0194.html`.

Sean M Theriault and David W Rohde. The gingrich senators and party polarization in the us senate. *The Journal of Politics*, 73(4):1011–1024, 2011.

Michael M Ting. A strategic theory of bureaucratic redundancy. *American Journal of Political Science*, 47(2):274–292, 2003.

Rachel VanSickle-Ward. *The Devil Is in the Details: Understanding the Causes of Policy Specificity and Ambiguity*. SUNY Press, 2014.

Craig Volden and Alan E Wiseman. *Legislative Effectiveness in the United States Congress: The Lawmakers*. Cambridge University Press, 2014.

John Wilkerson, David Smith, and Nicholas Stramp. Tracing the flow of policy ideas in legislatures: A text reuse approach. *American Journal of Political Science*, 59(4):943–956, 2015.

Rick K Wilson and Cheryl D Young. Cosponsorship in the us congress. *Legislative Studies Quarterly*, pages 25–43, 1997.

Biographical statement: Robert Shaffer is a postdoctoral fellow at the University of Pennsylvania's Perry World House, Philadelphia, PA, 19104.

Supplemental Information

Robert Shaffer

University of Pennsylvania

shafferr@upenn.edu

March 17, 2020

Contents

A Measurement

A.1 Header Regular Expressions

Table 1 gives the set of regular expressions used as inputs to the constitute_tools parser, which I used to parse the American legislative text database I use in this paper. Note that not all of these levels are present in all documents.

Table 1: Regular expressions used to parse American legislative texts

<table>
<tr><th>Regular Expression</th><th>Sample Plain-Text Match</th></tr>
<tr><td><code>(SECTION|SEC\.) [0-9]+[-A-Z]*\.</code></td><td>SECTION 101; SEC. 446a</td></tr>
<tr><td><code>\(([ivx])?([a-hj-uwyz])?\)\s*</code>
<code>(?=(?(1)[\s\S]*?\n\s*\([jwy]\)))</code></td><td>(a)</td></tr>
<tr><td><code>\([0-9]+\)\s*</code></td><td>(32)</td></tr>
<tr><td><code>\(([IVX])?([A-HJ-UWYZ])?\)\s*</code>
<code>(?=(?(1)[\s\S]*?\n\s*\([JWY]\)))</code></td><td>(B)</td></tr>
<tr><td><code>\([ivx]+\)\s*</code></td><td>(ii)</td></tr>
<tr><td><code>\([IVX]+\)\s*</code></td><td>(IV)</td></tr>
<tr><td><code>‘‘\([A-Z0-9a-z]+\)</code></td><td>“(A)</td></tr>
</table>

A.2 Sample Parsed Text

Table 2: Sample parsed document

Title	Text
SEC 416	FOREIGN STUDENT MONITORING PROGRAM.
(a)	Full «NOTE: 8 USC 1372 note.» Implementation and Expansion of Foreign Student Visa Monitoring Program Required.–The Attorney General, in consultation with the Secretary of State, shall fully implement and expand the program established by section 641(a) of the Illegal Immigration Reform and Immigrant Responsibility Act of 1996 (8 U.S.C. 1372(a)).
(b)	Integration «NOTE: 8 USC 1372 note.» With Port of Entry Information.–For each alien with respect to whom information is collected under section 641 of the Illegal Immigration Reform and Immigrant Responsibility Act of 1996 (8 U.S.C. 1372), the Attorney General, in consultation with the Secretary of State, shall include information on the date of entry and port of entry.

USA PATRIOT Act §416(a-b). For original text see the corresponding congress.gov page.

A.3 LSTM Parameter Specification

As described in §3, I used an LSTM to extract named entities from legislative texts. Where not otherwise specified, I left parameter settings at their defaults in the tf_ner library.

For model construction, I used a 300-node layer for word embeddings and a 100-node layer for the LSTM encoder. For the word embedding layer I relied on pre-trained embeddings drawn from Pennington *et al.* (2014)'s GloVe dataset. Like virtually all neural network applications, I trained this model using stochastic gradient descent.[1] I trained the model for up to 25 epochs, with the model set to halt training if no improvement was observed in a held-out development set every 500 batches (with a minimum of 8,000 batches run). I used sentences containing 90% of pre-identified named entities for training and 10% as a held-out development set. To avoid overfitting, I used a dropout rate of 0.5 and a batch size of 20 for the gradient descent algorithm used to fit the model.

[1]Specifically, an ADAM optimizer. See Kingma and Ba (2014) for details.

B Descriptive Statistics

Figure 1: Predictor and dependent variable descriptive statistics, for bills with at least one named entity.

Variable	Mean	SD	Source
Average Degree	4.76	14.69	Author
Divided	0.61	-	Casas et al. (2018)
CQ Mention	0.21	-	CQ Almanac
Majority	0.78	-	Casas et al. (2018)
Hitchhiker	0.51	-	Casas et al. (2018)
$\sqrt{\text{Cosponsors}}$	2.81	2.74	Casas et al. (2018)
DW-NOMINATE	0.10	0.44	Casas et al. (2018)
Republican	0.56	-	Casas et al. (2018)

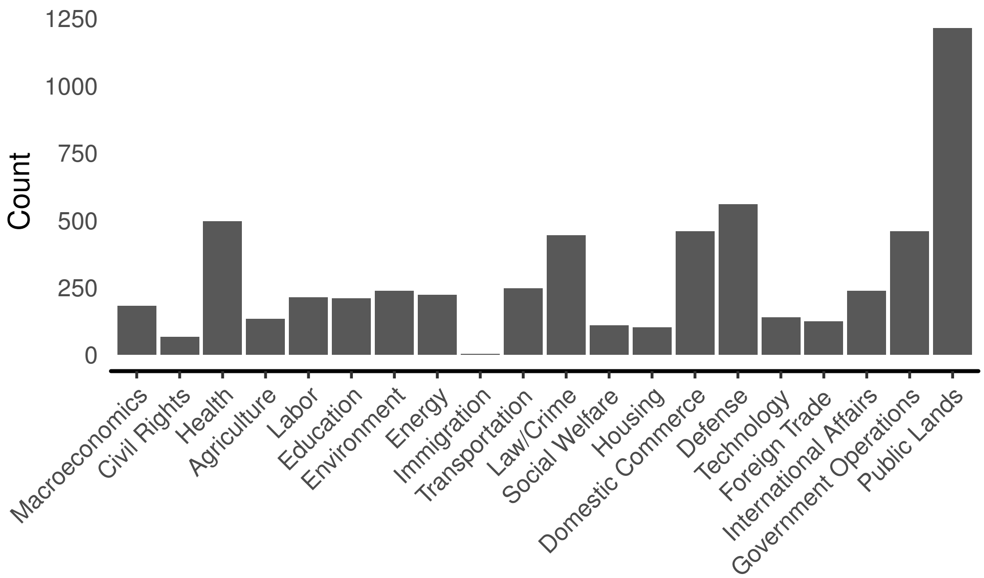
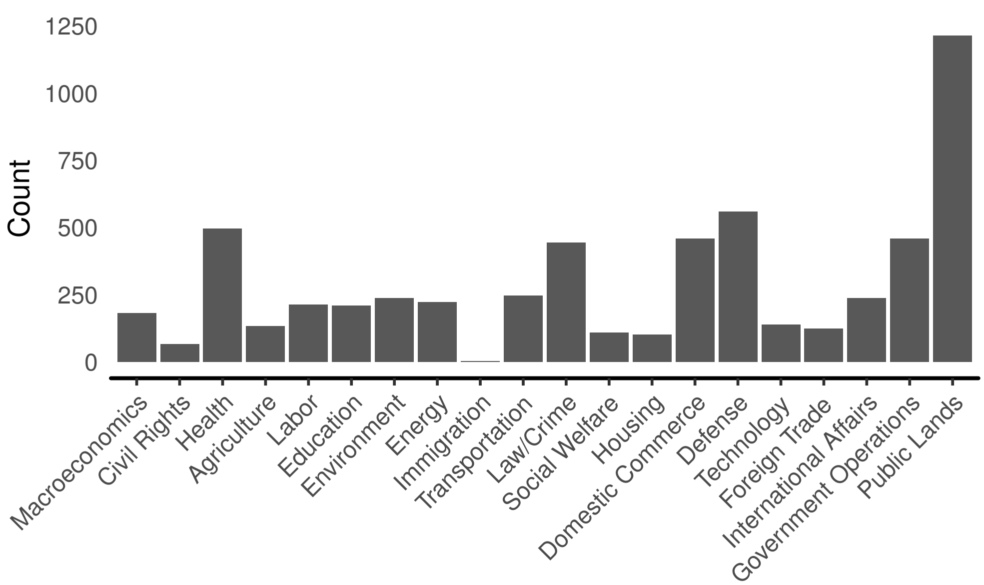


Figure 2: Predictor and dependent variable descriptive statistics, for all bills.

Variable	Mean	SD	Source
Total nodes	5.68	14.83	Author
Divided	0.61	-	Casas et al. (2018)
CQ Mention	0.18	-	CQ Almanac
Majority	0.77	-	Casas et al. (2018)
Hitchhiker	0.52	-	Casas et al. (2018)
$\sqrt{\text{Cosponsors}}$	2.75	2.67	Casas et al. (2018)
DW-NOMINATE	0.11	0.45	Casas et al. (2018)
Republican	0.57	-	Casas et al. (2018)



C Bayesian Model Details

C.1 Specification and Model Fit

To fit all regression models I present in this paper, I use the `brms` library's interface to the Stan programming language (Carpenter *et al.* 2016). For the gamma regression models, I use the likelihood and implementation specified in the `brms` library's `hurdle_gamma()` function. Similarly, for the negative binomial hurdle model I present as part of my robustness checks in Appendix C.2.5, I use the `brms` library's `hurdle_negbinomial()` function.

For priors, I selected weakly informative prior distributions for all variables. These prior values are intended to be uninformative in all cases where a non-trivial quantity of data is present, but restrict parameters from attaining implausible values when data is sparser. Following Ghosh *et al.* (2018), for all regression coefficients I assign a $t(3, 0, 10)$ prior for all intercept coefficients, a $t(3, 0, 2.5)$ prior for all other regression coefficients, and a $Cauchy(0, 5)$ prior on all standard deviation parameters. I additionally placed a $gamma(0.01, 0.01)$ prior on the gamma and negative binomial distribution distribution dispersion parameters.

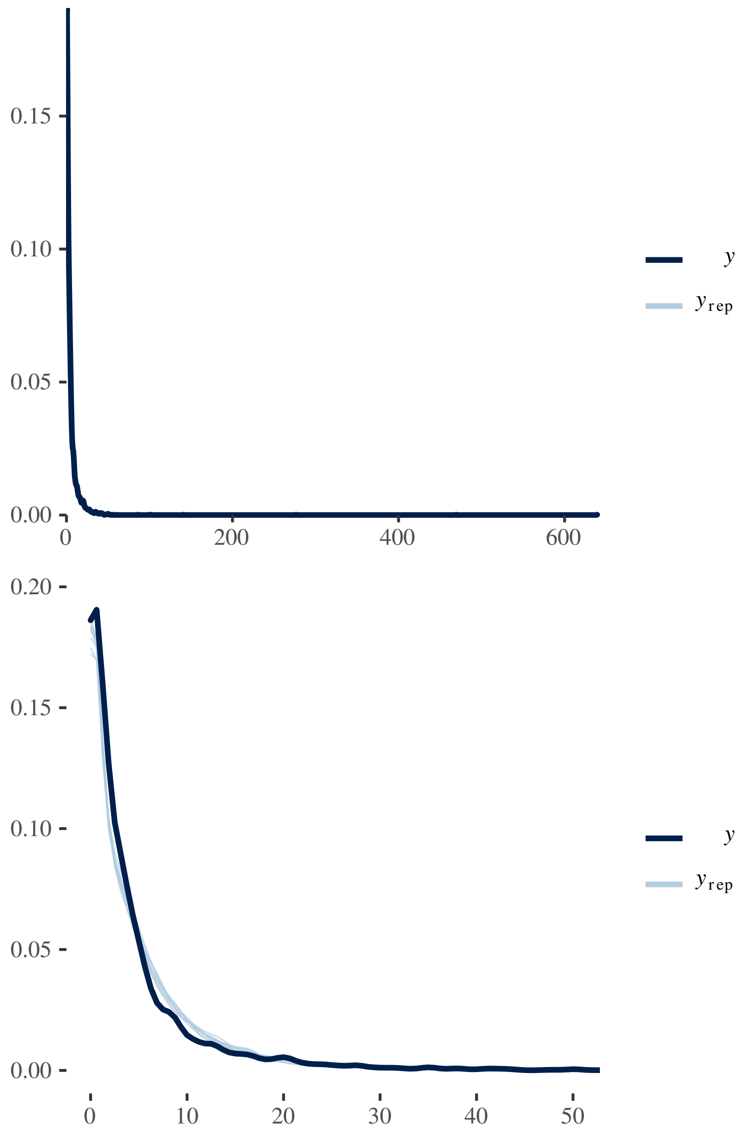
For all models, I ran four chains with 1,000 warmup iterations, 1,500 post-warmup iterations in each chain, and random initializations for all parameter values. For the negative binomial hurdle model, I additionally set the *adapt_delta* value in the sampler to 0.85 to avoid divergent transitions. Visual plots suggested good mixing across chains in all models, with $\hat{R} \leq 1.01$ for all parameters and $n_{eff} \geq 1000$ for all parameters.[2]

Following Gelman *et al.* (2014), in Figure 3 I visually assessed model fit for the "main" model I present in-text using posterior predictive checks. In each plot, I provide the observed density of the node count dependent variable, overlaid on density plots for 10 simulated dependent variable datasets based on randomly-selected post-warmup posterior parameter draws. As shown in the top panel, across the whole dataset the model fit is excellent. Zooming in on smaller values (where most posterior density is located) reveals that the

[2] With $\hat{R}$ a diagnostic quantifying the consistency of an ensemble of Markov chains, and n_{eff} a rough effective sample size calculation (Gelman *et al.* 2014).

model slightly overpredicts at moderate values of the dependent variable ($5 \leq y \leq 15$). Even in this range, however, model fit remains strong.

Figure 3: Posterior predictive plots for the node count dependent variable



C.2 Numerical Coefficient Estimates

For all models in this section, estimates prefixed with (hu) indicate hurdle coefficients. Numerical intercept estimates for all models besides the model presented in-text are suppressed for brevity, but are available in replication materials. For all model variants, key findings regarding the `CQ Mention` and `Divided` coefficients are replicated, at roughly the same scales. Findings regarding all other variables are approximately replicated, though effect sizes are more variable, particularly for the negative binomial and no-hitchhiker model variants.

C.2.1 Gamma Hurdle Model (in-text)

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	1.41	0.08	1.25	1.57
$\sqrt{\text{Cosponsors}}$	0.00	0.01	-0.01	0.02
DW-NOMINATE	-0.18	0.04	-0.26	-0.09
DW-NOMINATE^2	0.77	0.11	0.56	0.97
Majority	0.16	0.04	0.08	0.25
CQ Mention	0.36	0.06	0.24	0.47
Divided	-0.06	0.04	-0.13	0.02
Hitchhiker	-0.05	0.03	-0.12	0.02
CQ Mention:Divided	0.53	0.08	0.38	0.68
(hu) Intercept	-0.44	0.13	-0.70	-0.18
(hu) $\sqrt{\text{Cosponsors}}$	-0.05	0.02	-0.08	-0.02
(hu) DW-NOMINATE	0.15	0.10	-0.04	0.34
(hu) DW-NOMINATE^2	0.08	0.24	-0.40	0.55
(hu) Majority	-0.26	0.08	-0.42	-0.09
(hu) CQ Mention	-1.16	0.17	-1.50	-0.83
(hu) Divided	0.06	0.08	-0.09	0.22
(hu) Hitchhiker	-0.16	0.07	-0.30	-0.03
(hu) CQ Mention:Divided	0.14	0.22	-0.28	0.57

	Estimate	Est.Error	Q2.5	Q97.5
Macroeconomics	1.40	0.12	1.16	1.63
Civil Rights	1.32	0.14	1.06	1.59
Health	1.29	0.08	1.14	1.44
Agriculture	1.27	0.11	1.05	1.50
Labor	1.25	0.10	1.05	1.44
Education	1.22	0.11	1.02	1.43
Environment	1.32	0.09	1.14	1.51
Energy	1.33	0.10	1.13	1.53
Immigration	1.41	0.25	0.92	1.90
Transportation	1.48	0.09	1.29	1.66
Law/Crime	1.36	0.08	1.20	1.51
Social Welfare	1.66	0.12	1.43	1.90
Housing	1.45	0.14	1.19	1.73
Domestic Commerce	1.24	0.08	1.08	1.40
Defense	1.63	0.08	1.48	1.78
Technology	1.63	0.10	1.44	1.84
Foreign Trade	1.49	0.11	1.28	1.71
International Affairs	1.32	0.09	1.15	1.49
Government Operations	1.94	0.08	1.78	2.09
Public Lands	1.07	0.06	0.96	1.19
(hu) Macroeconomics	-0.34	0.21	-0.75	0.07
(hu) Civil Rights	-0.40	0.22	-0.83	0.05
(hu) Health	-0.64	0.16	-0.97	-0.32
(hu) Agriculture	-0.55	0.21	-0.96	-0.16
(hu) Labor	-0.46	0.19	-0.83	-0.09
(hu) Education	-0.29	0.19	-0.66	0.09
(hu) Environment	-0.37	0.17	-0.71	-0.03
(hu) Energy	-0.07	0.19	-0.46	0.31
(hu) Immigration	-0.33	0.27	-0.85	0.23
(hu) Transportation	-0.48	0.18	-0.83	-0.13
(hu) Law/Crime	-0.61	0.16	-0.93	-0.29
(hu) Social Welfare	-0.36	0.22	-0.80	0.09
(hu) Housing	-0.30	0.22	-0.73	0.15
(hu) Domestic Commerce	-0.29	0.17	-0.62	0.03
(hu) Defense	-0.32	0.15	-0.62	-0.02
(hu) Technology	-0.66	0.21	-1.09	-0.27
(hu) Foreign Trade	-0.47	0.20	-0.88	-0.06
(hu) International Affairs	-0.66	0.19	-1.03	-0.31
(hu) Government Operations	-0.59	0.16	-0.89	-0.29
(hu) Public Lands	-0.53	0.12	-0.77	-0.30

C.2.2 Gamma hurdle model (base DW-NOMINATE coefficient only)

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	1.56	0.08	1.41	1.72
$\sqrt{\text{Cosponsors}}$	0.00	0.01	-0.01	0.02
DW-NOMINATE	-0.02	0.04	-0.10	0.05
Majority	0.17	0.04	0.09	0.25
CQ Mention	0.36	0.06	0.25	0.48
Divided	-0.06	0.04	-0.14	0.02
Hitchhiker	-0.07	0.03	-0.14	0.00
CQ Mention:Divided	0.52	0.08	0.37	0.67
(hu) Intercept	-0.42	0.12	-0.65	-0.18
(hu) $\sqrt{\text{Cosponsors}}$	-0.05	0.01	-0.08	-0.02
(hu) DW-NOMINATE	0.17	0.08	0.00	0.33
(hu) Majority	-0.26	0.08	-0.42	-0.09
(hu) CQ Mention	-1.15	0.17	-1.48	-0.83
(hu) Divided	0.06	0.08	-0.09	0.22
hu_outcomeHitchhiker	-0.17	0.07	-0.31	-0.02
(hu) CQ Mention:Divided	0.13	0.22	-0.30	0.56

C.2.3 Gamma hurdle model (party specification)

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	1.59	0.08	1.44	1.75
$\sqrt{\text{Cosponsors}}$	0.00	0.01	-0.01	0.02
Republican	-0.08	0.03	-0.15	-0.02
Majority	0.18	0.04	0.10	0.26
CQ Mention	0.35	0.06	0.24	0.47
Divided	-0.05	0.04	-0.13	0.02
Hitchhiker	-0.07	0.03	-0.14	-0.01
CQ Mention:Divided	0.53	0.08	0.38	0.68
(hu) Intercept	-0.47	0.13	-0.72	-0.22
(hu) $\sqrt{\text{Cosponsors}}$	-0.05	0.02	-0.08	-0.02
(hu) Republican	0.10	0.08	-0.04	0.25
(hu) Majority	-0.25	0.09	-0.42	-0.08
(hu) CQ Mention	-1.16	0.17	-1.50	-0.83
(hu) Divided	0.07	0.08	-0.09	0.22
(hu) Hitchhiker	-0.17	0.07	-0.31	-0.03
(hu) CQ Mention:Divided	0.14	0.22	-0.29	0.57

C.2.4 Gamma hurdle model (no hitchhikers)

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	1.37	0.11	1.15	1.59
$\sqrt{\text{Cosponsors}}$	0.00	0.01	-0.02	0.01
DW-NOMINATE	-0.19	0.07	-0.32	-0.05
DW-NOMINATE^2	1.15	0.15	0.86	1.46
Majority	0.10	0.07	-0.04	0.24
CQ Mention	0.35	0.08	0.20	0.50
Divided	-0.08	0.06	-0.20	0.04
CQ Mention:Divided	0.63	0.10	0.43	0.84
(hu) Intercept	-0.31	0.19	-0.69	0.07
(hu) $\sqrt{\text{Cosponsors}}$	-0.10	0.02	-0.15	-0.05
(hu) DW-NOMINATE	0.11	0.14	-0.17	0.40
(hu) DW-NOMINATE^2	0.16	0.34	-0.52	0.82
(hu) Majority	-0.24	0.14	-0.51	0.04
(hu) CQ Mention	-1.69	0.25	-2.19	-1.21
(hu) Divided	0.10	0.11	-0.13	0.32
(hu) CQ Mention:Divided	0.74	0.30	0.15	1.36

C.2.5 Negative binomial hurdle model (total node dependent variable)

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0.58	0.12	0.35	0.81
$\sqrt{\text{Cosponsors}}$	0.03	0.01	0.01	0.05
DW-NOMINATE	-0.20	0.06	-0.33	-0.08
DW-NOMINATE^2	0.35	0.16	0.05	0.66
Majority	0.37	0.06	0.25	0.49
CQ Mention	1.18	0.09	0.99	1.36
Divided	-0.05	0.06	-0.16	0.06
Hitchhiker	0.04	0.05	-0.06	0.14
CQ Mention:Divided	0.40	0.12	0.16	0.65
(hu) Intercept	-0.61	0.16	-0.91	-0.30
(hu) $\sqrt{\text{Cosponsors}}$	-0.03	0.01	-0.06	-0.01
(hu) DW-NOMINATE	0.15	0.08	-0.01	0.32
(hu) DW-NOMINATE^2	0.53	0.20	0.14	0.92
(hu) Majority	-0.23	0.08	-0.37	-0.08
(hu) CQ Mention	-1.58	0.18	-1.95	-1.24
(hu) Divided	0.05	0.07	-0.08	0.19
(hu) Hitchhiker	-0.08	0.06	-0.20	0.05
(hu) CQ Mention:Divided	0.65	0.22	0.23	1.09

C.2.6 Negative binomial model (hurdle component excluded)

	Estimate	Est.Error	l-95% CI	u-95% CI
Intercept	0.92	0.09	0.75	1.09
$\sqrt{\text{Cosponsors}}$	0.02	0.01	0.01	0.04
DW-NOMINATE	-0.20	0.05	-0.29	-0.10
DW-NOMINATE^2	0.16	0.12	-0.07	0.40
Majority	0.32	0.05	0.23	0.41
CQ Mention	1.24	0.07	1.09	1.38
Divided	-0.05	0.04	-0.14	0.03
Hitchhiker	0.03	0.04	-0.05	0.10
CQ Mention:Divided	0.25	0.09	0.07	0.44

References

Bob Carpenter, Andrew Gelman, Matt Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Michael A Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *J Stat Softw*, 2016.

Andrew Gelman, John B Carlin, Hal S Stern, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis.* Taylor & Francis, Boca Raton, FL, 3 edition, 2014.

Joyee Ghosh, Yingbo Li, Robin Mitra, et al. On the use of cauchy prior distributions for bayesian logistic regression. *Bayesian Analysis*, 13(2):359–383, 2018.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Jeffrey Pennington, Richard Socher, and Christopher Manning. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*, pages 1532–1543, 2014.