Understanding the Language of Political Agreement and Disagreement in Legislative Texts

Maryam Davoodi
Purdue University
mdavoodi@purdue.edu

Eric Waltenburg
Purdue University
ewaltenb@purdue.edu

Dan Goldwasser
Purdue University
dgoldwas@purdue.edu

Abstract

While national politics often receive the spotlight, the overwhelming majority of legislation proposed, discussed, and enacted is done at the state level. Despite this fact, there is little awareness of the dynamics that lead to adopting these policies. In this paper, we take the first step towards a better understanding of these processes and the underlying dynamics that shape them, using data-driven methods. We build a new large-scale dataset, from multiple data sources, connecting state bills and legislator information, geographical information about their districts, and donations and donors' information. We suggest a novel task, predicting the legislative body's vote breakdown for a given bill, according to different criteria of interest, such as gender, rural-urban and ideological splits. Finally, we suggest a shared relational embedding model, representing the interactions between the text of the bill and the legislative context in which it is presented. Our experiments show that providing this context helps improve the prediction over strong text-based models.

1 Introduction

Despite the fact that state-level legislation is rarely discussed, it has a dramatic influence on the everyday life of residents of the respective states. The policies enacted at the state-level touch on all aspects, from mundane topics, such as trash removal and state mascots, to highly ideologically-charged topics such as education, religious liberties, and health-care access. Moreover, state-legislatures discuss and vote-on significantly more bills than their Federal counterparts, adding up to over 120,000 bills per year (King, 2019). Also, the lack of general interest, as well as the complexity of the processes that differ across states, often leads to public disengagement from local politics. This results in decisions being made with little understanding of
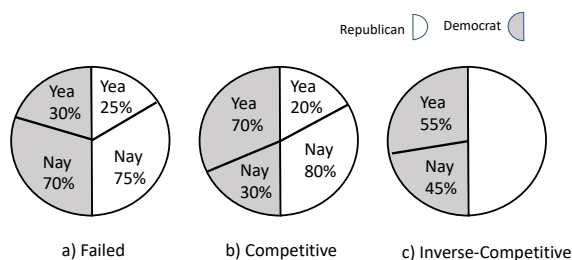

Figure 1: Example of failure and party cleavages.

the processes that shape them and how they are likely to influence different demographics.

Similarly, most effort directed at understanding political processes using data was directed at the Federal level. In the NLP community, several works looked at analyzing political texts (Iyyer et al., 2014) and the resulting behaviors of legislators (Gerrish and Blei, 2011, 2012). The only exception is recent work (Eidelman et al., 2018), predicting whether a bill would pass the preliminary stage, legislative committee, to a full-body vote.

State-level demographic cleavages: Our goal in this paper is to take a first step towards understanding the processes and interests that underlie how decisions are passed using data-driven methods. Our main intuition is that the impact of bills on different demographics will be reflected in the behavior and voting patterns of their representatives. Thus, providing the ability to automatically identify bills, before they are put to a vote, that will have a positive or negative influence on a specific demographic can help inform public responses and increase engagement with local political processes.

To help achieve this goal, we define two novel text classification tasks, characterizing the breakdown of votes, based on different cleavages or demographic indicators such as gender, geography (i.e., rural vs. urban districts), party membership and ideological splits. With respect to each one of these splits, we define two aggregate-level proper-

ties of a vote, *competitive* and *inverse-competitive* cleavages. Both of these measures capture the lack of consensus in the legislature body around a specific bill, but in different ways. We say that a bill is competitive in a vote (Fig. 1b) if the majority of legislators from a logical group (e.g., democrats, women, urban districts, liberals) vote differently from the majority of legislators from the opposite group (e.g., republican, men, rural districts, conservatives). A bill is inverse-competitive (Fig. 1c) if there is a partial or complete tie within the legislators from the same group (e.g., women). To help explain these concepts, consider a bill restricting access to abortion clinics. This bill is likely to results in a *competitive* vote, based on ideology. On the other hand, a bill granting tax breaks for farmers might result in a *inverse-competitive* vote, based on ideology. In that case, a *competitive* vote, based on geography is more likely.

In Table 1, we provide examples of the different splits associated with real bills that were brought to a vote. Unsurprisingly, a "benign" bill, such as #1 is widely accepted and does not result in any contention. A contentious bill, such as #2, touching on the way religion is taught is split ideologically (i.e., the vote is almost unanimous inside each ideological group), but mixed based on economic and gender splits. Bill #4 addressing nepotism issues and regulating public contracts is contentious across all splits. Alerting the public when such bills are brought to a vote can help ensure that legislators take into account the opinions and voiced raised in their constituencies.

Technical Contributions Although a text classification scheme is a reasonable starting point to determine demographic cleavages of bills only based on their content, it is not sufficient. Our key insight in this paper is that the context or relations through which specific information is propagated among different players in the legislative process (e.g., money donors and legislators), can be leveraged to further improve the performance. Thus, we build a shared relational architecture that models the text of a bill and its context into a graph; Our model captures the behavior of individual legislators, language of bills, and influence of contributions on the decision to identify demographic cleavages. While there are different ways to realize our relational model, we chose to build on recent advances in the NLP space, Relational Graph Convolutional Network (RGCN) (Schlichtkrull et al., 2018) and pre-

trained BERT transformers (Devlin et al., 2018). RGCN allows us to define multiple relations between each pair of entities (e.g., a legislator sponsorship and casting a vote on a bill) and BERT enables us to represent the textual information more efficiently. With the help of the attention-based architecture, BERT has been shown to outperform LSTM models. To operationalize our relational settings, we collected information from different sources and introduced a new dataset combining information about legislators, bills, donations, and donors as well as demographic information about the legislators and their districts. In our experiments, we analyze the implication of different relations on the performance and show that our shared architecture outperforms existing text and graph models.

Table 1: *Competitive and inverse-competitive bills.*

#	Bill Title	Gen.	Geo.	Ideo.	Party
1	A CONCURRENT RESOLUTION congratulating the Pioneer Junior-Senior High School football team on winning the Indiana High School Athletic Association	None	None	None	None
2	Teaching of the origin of life	Inver.	Inver.	Comp.	Comp.
3	Beer dealer permits	None	None	None	None
4	Officeholder qualifications, nepotism, and public contracts	Both	Inver.	Both	Both

2 Related Work

Bill analysis at the state level has received little attention and our work, while conducting a new in-depth modeling and analysis, is inspired by the following works:

Classification of congress roll calls. (Eidelman et al., 2018) combines the text of the bill with partisan identity of the bill's sponsor(s) in a model predicting the likelihood of a member of the U.S. Congress voting in support of a non-unanimous congress bill or resolution. They find that the models that combine text with sponsorship data significantly outperform several alternative models. Similarly, (Gerrish and Blei, 2011) uses topics associated with congress bills to infer its location in ideological space and then uses ideal point models to predict the likelihood of a U.S. Senator or House member voting in support of a bill. They find that their model increases predictive accuracy by about 4% over a naïve baseline model. (Patil et al., 2019; Kraft et al., 2016; Karimi et al., 2019; Kornilova et al., 2018; Peng et al., 2016) extend this congress model to learn embeddings for legislators and congress bills using other sources of data

(e.g., Twitter, knowledge graphs). More recently, (Budhwar et al., 2018) evaluates different models for predicting roll-call votes based on verbal statements that legislators make during questioning.

Predicting progress of bills Rather than using bill text in models to explain the roll-call behavior of individual legislators, (Yano et al., 2012) include the legislation's text in a model that predicts whether a bill emerges from a standing committee, a point in the legislative process that most bills do not pass. In particular, they use features based on the urgency and importance of the issue being addressed by the bill as well as a set of features extracted from co-sponsors of the bill. Examining the fate of bills between the 103rd and 111th congresses, they find that including features of the bill drawn from the text improves the model's predictive accuracy over their baseline model. (Eidelman et al., 2018) repeat a similar analysis for the states and they show "that combining contextual information about the legislators and the legislatures with bill text consistently provides the best predictions". (Nay, 2017) examines the text of congressional to identify the text structure most associated with a congress bill's enactment and then embeds it using Word2Vec for the classification based on Random Forests; Nay concludes that the full text of a congress bill enables better prediction efficiency.

Demographic bill cleavages. Demographic bill cleavages is a well-studied topic in the political science space. Research has properly differentiated between the multiple ways demographic background of legislators can influence roll-call voting. (Pinney and Serra, 1999) finds that Congressional Black Caucus members vote more consistently with the caucus than they do with fellow partisans or with representatives from their state. (Jenkins, 2012) discusses gender moderates the effect of party and ideology in roll-call voting. Similarly, (Frederick, 2010) discusses gender influences the roll-call vote in the Senate by moderating the effect of partisanship for GOP women. (Broach, 1972) demonstrates that urban-rural cleavages structure vote in less partisan states and on bills that clearly divide urban and rural interests.

NLP applications of GCN. Recently, GCNs have been explored in different NLP tasks. Semantic role labeling (SRL) (Marcheggiani and Titov, 2017), relation classification in clinical narratives (Li et al., 2018), and machine translations (Bastings et al., 2017) are a few instances.

In such tasks, GCN is used to encode syntactic structure of sentences. In a similar context, some works explored the idea of graph neural networks (GNNs) (Peng et al., 2018; Henaff et al., 2015; Defferrard et al., 2016), where each part of a document (e.g., sentences) is collapsed into a graph of words or the citation relations (Kipf and Welling, 2016) creates a network among different documents.

3 Modeling

We model the legislative process as a graph that consists of bills, legislators, and money donors in all states. Building a global graph captures contextual information and relationships that interconnect different states. For instance, money donation by a contributor to two legislators from different states could indicate they have a similar roll call behaviors on abortion bills. Given this intuition, after a brief overview of the legislative process in US states, we describe how we collapse it into a graph structure.
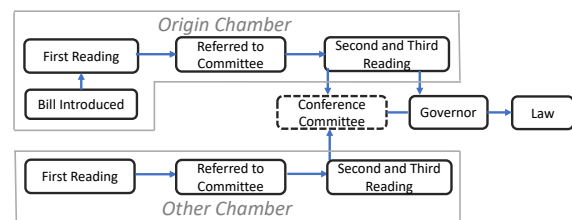

Figure 2: Bill-to-Law stages

3.1 Primer on State-Level legislative Process

Although there are some specific differences across state legislatures, a common process, shown in Figure 2, prevails. This process starts with one or more legislators (Representatives or Senators) who sponsor and file a bill. The idea of a bill could be original or come from a constituent, public official, or an interest group. Each state consists of two "chambers": the House of Representatives ("House") and the Senate. To become law, the bill goes through a reviewing process in the origin chamber, where it can "die" at different stages. If the bill gets a pass vote, it is sent to the other chamber and the same process repeats. Finally, the bill is reviewed by the state Governor for signature. In parallel to these efforts, external contributors, e.g., money donors and lobbyists, play an important yet indirect role in the process. By sourcing information and money into the process, they leave an impact on legislators, which can change the progression of a bill.

Within a chamber the process is as follows: if the leadership in the chamber chooses, the bill gets its

First Reading by title. Then, the chamber president may refer the bill to a committee for review. If the committee casts a vote on the bill, it can be defeated or advance to Second Reading by the full body of legislators. Next, the chamber leadership may decide to approve the bill for Third Reading, where it again comes to a vote by the full body of legislators and a majority vote can advance the bill.
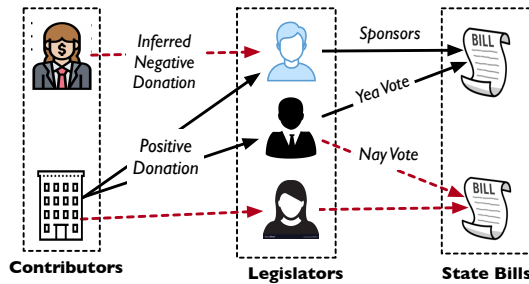

Figure 3: Collapsing the legislative process into a heterogeneous multi-relational legislative graph.

3.2 Legislative Process in a Heterogeneous Multi-Relational Graph

A close look reveals that the legislative process cannot be captured in a simple graph as there can be multiple relations between a pair of nodes (e.g., sponsorship and vote between legislators and bills), and the graph consists of several nodes types with different attributes and labels (e.g., bills with competitive labels). Thus, we model the process using a heterogeneous multi-relational graph, as follows:

Node attributes: The nodes in our proposed legislative graph come with a rich set of features and information: *(1) Bill* nodes contain title, description, and full text of the house and senate state bills. *(2) Legislator* nodes contain diverse textual information abstracting the behavior of a legislator such as his biography, political interests, committee assignments, and demographic attributes (gender, party, and ideology and the district information). *(3) Contributors* nodes come with different information (in the textual format) on money donors such as their specific and general business interests, party, and their type (individual vs non-individual).

Relations: Based on the legislative process, we identify that legislator and bill nodes participate in three main relations: sponsorship *(R1)*, negative ("Nay") vote *(R2)*, and positive ("Yea") vote *(R3)*. Similarly, we establish two types of relations between contributors and legislators: positive donation edges *(R4)*, which are realized based on the real data, and negative or lack of donation edges *(R5)*, inferred when a contributor shows lack of interest in specific legislators (e.g., always donates to Democrats). In this case, we randomly sample such legislators and link them to the contributor. Based on our data analysis, more than 62% of unique contributors always contribute to one party in our dataset. We also conducted an ablation study, not included due to space constraints, and the donor information contributed between 2 to 11 F1 points.

3.3 Bill Inference Problems

For a bill and one of its roll calls in the legislative graph, we seek to predict if (1) it evinces identifiable voting cleavages or (2) it can advance by getting a pass. For voting cleavages, we defined four demographic attributes (gender, party, ideology, and the urban/rural nature of the district) to divide legislators into groups. We assign nine labels to each bill as follows: *(1) Competitive labels*: For an attribute (e.g., party), a voting round of a bill is defined as "competitive" if the majority of legislators from one group (e.g., Democrats) votes differently from the majority of the other group (e.g., Republicans). For example, in Figure 1b, 70% of Democrats vote Yea and 80% Republicans vote Nay on a roll call, then the bill is competitive and the disagreement between the groups is 10% (=80%-70%). *(2) Inverse-competitive labels*: Similarly, for an attribute (e.g., party), we call a voting round as inverse-competitive if there is a partial or full cleavage among the legislators of the same group. For instance, consider a bill with 55% of Democrats voting Yea and 45% of them voting Nay (Figure 1c). In this case, the bill turns out to be inverse-competitive and the disagreement is 45% (the percentage of minority votes). *(3) Survival label*: Depending on the progress, a bill passes a certain voting round if it gets a majority vote (e.g., in 2nd/3rd Reading) or if two-thirds of legislators agree to it (e.g., in amendments).

4 Inference on Legislative Graph

We argue for a joint graph and text embedding model to represent the nodes and their textual attributes in the legislative graph, which is used for the roll-call prediction and aggregation. Embedding models that only leverage textual information ignore important relations in the legislative graph. Graph-based models make textual information less distinguishable at the classification stage, where it matters. At a high level, our approach combines the complementary strengths of both approaches.
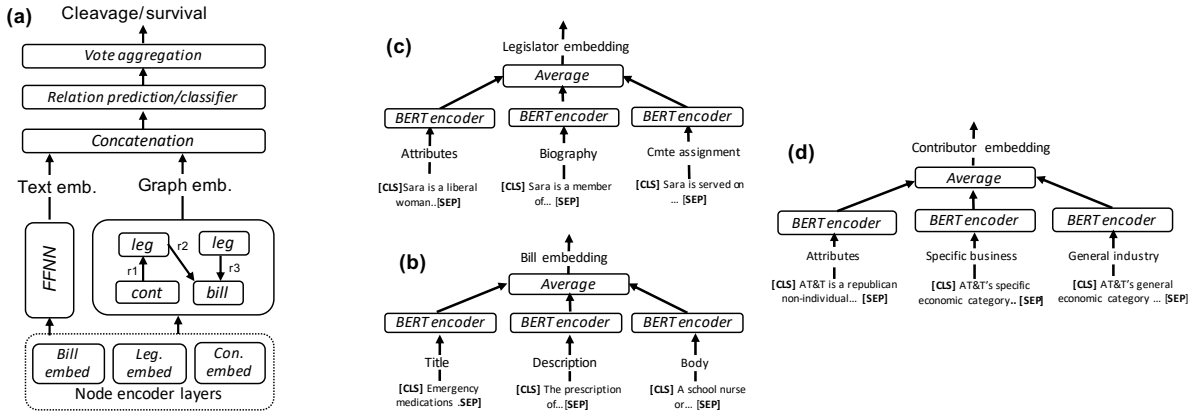
Figure 4: *Joint text-graph architecture for predicting relations in the legislative graph and aggregating vote (roll-call) of individual legislators, by leveraging text-attributed RGCN and BERT's pretrained embeddings.*

Our architecture (Figure 4a) uses BERT's pretrained embedding to represent the textual information of nodes in the graph; and text-attributed RGCN to generate an embedding for them based on their relations. Finally, we combine them to build a representation of edges in the graph for our relation prediction and then aggregate vote relations.

4.1 Text Representation Layer

The lower half of our architecture is based on BERT, which leverages transformers and acts as an efficient replacement for sequential models. In our case, we use the BERT's pretrained embedding to form an initial representation for the textual information of the nodes in the legislative graph.

Bill representation: We represent a bill by averaging three different vectors (Figure 4b) corresponding to: (1) title, (2) description, and (3) body of the bill. For each of these components, we compute the average word vector based on BERT's pretrained word embedding. Thus, the bill representation is $X_{bill} = Avg(e_{title} + e_{description} + e_{body})$.

Legislator representation: To represent a legislator, we compute BERT's pretrained embedding for his textual information: (1) attributes, (2) biography, and (3) committee information. Finally, we take the average of these vectors, $X_{legislator} = Avg(e_{attributes} + e_{biography} + e_{cmte-info})$, as illustrated in Figure 4c.

Contributor representation: Similarly, We transform different pieces of textual information on a contributor, i.e., party- and type-related attributes, business information, and industry data, into separate vectors, $e_{attributes}$, $e_{business}$, $e_{industry}$ and then take their average as the final representation, $X_{contributor}$ (Figure 4d).

4.2 Relational Graph Convolutional Layers

We feed the text representation of the bill, legislator, and contributor nodes, as their initial representation, into Relational Graph Convolutional Network (RGCNs) to better represent them given the legislative graph structure. In parallel, a feed-forward neural network (FFNN) processes these text representations and takes them to a concatenation layer for the joint text-graph optimization. From the message passing perspective, each (non-relational) GCN layer performs two operations: *propagation and aggregation*. In the propagation phase, the neighborhood nodes send their feature/hidden representation to the node that needs to be updated. In the aggregation phase, the node sums up all the messages coming from its neighborhood with its properties. The aggregated message is passed through a non-linear activation function which forms the new representation of the node. If the graph edges are not typed, the hidden representation of each node i, at $(l + 1)$'th layer, is computed by:

$$h_i^{l+1} = \sigma\left(\sum_{j \in N_i} \frac{1}{c_i} W^l h_j^l\right) \quad (1)$$

In which the weight matrix W^l is shared by all edges in layer l. Also, c_i is a normalization factor, which is often set to $c_i = |N_i|$. Relational GCN (RGCN) generalizes GCNs to handle different relations between any pair of nodes, and thus being a better fit for our problem. Unlike GCNs, RGCNs use a different weight matrix and normalization factors (e.g., $c_i^r = |N_i^r|$) for each relation type and thus the hidden representation of nodes in $(l+1)$'th

Table 2: *Statistics of the legislative graphs, aggregated over the 2011-2018 period.*

State	Nodes			Relations	
	# Cont	# Bills	# Leg.	#Cont-Leg	#Leg-Bill
IN	274	4818	226	17729	217026
OR	462	4884	150	29213	102463
WI	175	1320	208	5924	88004
All	911	11022	584	52866	407493

Table 3: *Legislators' attributes across the target states aggregated over the 2011-2018 period—UR: Urban, RU: Rural, C: Conservative, M: Moderate, L: Liberal.*

State	Gender		Party		Geography		Ideology		
	F	M	D	R	UR	RU	C	M	L
IN	50	176	67	159	161	64	125	94	7
OR	47	103	83	67	133	17	28	61	61
WI	51	157	84	124	160	48	78	49	81
All	148	436	234	350	454	129	231	204	149

layer is computed as:

$$h_i^{l+1} = \sigma \left(W_0^l h_i^l + \sum_{r \in R} \sum_{j \in N_i^r} \frac{1}{c_{i,r}} W_r^l h_j^l \right) \quad (2)$$

By having a K-layer RGCN (stacking layers onto each other), we can capture k^{th}-order relations from a node in the graph. However, a 2-layer RGCN turns out to be sufficient in our case as it fully realizes the 2nd order relations between contributors and bills.

4.3 Roll-Call Classification and Aggregation

By combining the outputs of the RGCN and FFNN, we train a model for predicting relations in the legislative graph through FFNN+softmax. One could leverage DistMult scoring functions (Schlichtkrull et al., 2018; Yang et al., 2014) as well. Next, we post-process the roll-call relations and aggregate them to form the demographic and pass/fail vote breakdowns and determine the final class labels. In more detail, the representation of an edge or relation (s, d) is the dot product of e_s^{joint} and e_d^{joint}, which are the embedding of the corresponding nodes. The representation of a node comes from the concatenation of two components: (1) text embedding (hidden states) coming from the BERT layer after being fine-tuned through FFNN, and (2) the graph embedding (hidden state of the node) from the last RGCN layer.

Loss function: At a high level, our loss function is $L = L_{Cls} + L_{Text} + L_{Graph}$ and jointly optimizes the text and graph embeddings as well as the relation prediction and roll-call aggregation. L_{Cls} is the cross-entropy loss of the relation prediction; L_{Graph} and L_{Text} are the L2 regularizations of RGCN's and FFNN's weights that generate the graph and text representations, respectively.

5 Experiments

In this section, we describe our comprehensive legislative dataset, combining different sources of data (e.g., money donors data, diverse information on

legislators). Table 2 shows the statistics of our dataset after pruning and forming the legislative graph (discussed in Section 3). Next, we focus on our joint embedding model and its great ability in outperforming existing prediction models.

5.1 Data Collection Methodology & Statistics

Bills and legislator data. From the LegiScan website (LegiScan, 2019), we collected data on the text and lower chamber disposition of all bills introduced in Indiana, Oregon, and Wisconsin from the 2011 through 2018 sessions. To do so, we developed a comprehensive crawler in Python that performs multiple operations. First, it uses the LegiScan API to collect legislative information on every bill that covers: (1) bill metadata that includes the bill type, title, description, sponsors, and links to its texts; (2) vote metadata that consists of the individual legislator's vote – "Yea," "Nay," "Absent," or "NV"; and (c) legislator metadata containing party and district information. Then, our crawler accurately converts bill texts that are stored in the PDF format to text files, using open-source libraries. To identify the fine-grained progression of bills in the legislative process, our crawler downloads and processes the "History" section of each bill on the LegiScan Website, which consists of a series of events associated with a bill's history (e.g., committee report, roll-call vote). Such information is not readily available in the LegiScan API. Overall, we collected 34443 bills introduced in the target states from 2011 to 2018. We studied 58% of the bills that had both the votes of individual legislators and full texts, which are necessary for determining vote breakdowns and cleavage labels; However, our focus in this paper is on the 2nd/3rd reading, in which all members of the chambers vote, so we selected 32% of the bills that reached this stage to build the legislative graph (Table 2).

Biography, ideology and geography data. Finally, our crawler uses Ballotpedia (Ballotpedia, 2019) to collect texts on each legislator's biography, political interests, and committee assignments.

Also, it aggregates other publicly available datasets to identify each legislator's attributes such as ideology, gender, and district nature (urban/rural). The ideology scores for legislators were taken (Shor and McCarty, 2011) and they were grouped into conservatives, moderates, and liberals. The district identifier was combined with GIS census data (Census, 2019) to categorize each legislator as representing either an urban or rural district. Table 3 shows the breakdown of legislators' party, gender, ideology, and district information in our target states. For less than 10% of legislators, Ballotpedia profiles were missing. Thus, we used other public textual information about them (e.g., Twitter).

Donors data: FollowTheMoney (FollowThe-Money, 2019) captures and keeps tracks of donations to legislators and candidates in the US states. Our crawler consumes the FollowTheMoney API to collect the information of donors for each legislator and cosponsors of our bills. This includes multiple textual attributes and information for each contributor: type that could be individual or non-individual, general party, and economic and business information. While the contributor data can be used in more sophisticated ways, in this work, we focused on major contributors by setting a donation threshold ($10000) and removing those who contributed to a single legislator; We also separated between ideological contributors and pragmatic ones (donating to both parties) by inferring "negative" (lack of) donation relations (see Section 3); We set the fraction of negative donations to 30% of the positive ones extracted from the real data. Table 2 shows the final per-state statistics of contributors.

5.2 Experimental Setup

We build different graph and textual models on top of PyTorch, DGL (Deep Graph Library), and spaCy. In our joint text-graph model (Figure 4) and other baselines, the initial embedding dimension of both BERT ("bert-large-uncased") and the first-layer RGCN are set to 1024. The FFNN (fully connected layer) and the second-layer RGCN take the initial text and graph embeddings to a 256-dimensional space. We have also experimented with different settings, which while resulting in lower overall performance, retained the same trend when comparing the other models. We used Adam to optimize our model and for each observed relation (Table 2), we sampled a negative example.

Data splits. Our focus is on the bill cleavage and survival and thus we split legislative graphs based on bill nodes. To evaluate different scenarios, we have three configurations: *(1) random* where we select 20% of the bills for testing and keep the rest for training and validation. *(2) time-based* where 20% of most recent bills are considered for testing; and *(3) state-based:* where the test bills come from one specific state and train bills from the other states. The test bills and corresponding legislators appear in the test graph, and the difference of the original and test graphs is used for training. Note that vote relations of sponsoring legislators and a bill are known, and appear in training.

Metric. Given the highly skewed data when predicting bill survival and cleavages, we pick Macro F1 as the main metric over accuracy.

5.2.1 Baselines

To demonstrate the benefits of our joint text-graph embedding, we implement a series of text and graph embedding architectures as the baseline.

Category 1: text embedding models: We realize our bill encoder (Figure 4b) using three text embedding models and then train a logistic regression classifier to directly predict if a bill text shows a certain cleavage or passes/fails: *(a) BoW*, where unigram and bigram features (top 10K highest scoring ones using scikit-learn (Pedregosa et al., 2011)) used to represent bill texts. *(b) GloVe* (Pennington et al., 2014) that is a popular word embedding model using the square loss; We used the GloVe-840B-300D pre-trained word vectors in our experiments. (c) *BERT* (Devlin et al., 2018) that is a transformer based architecture capable of capturing contextualized embedding.

Category 2: featureless graph embedding models: We build a edge classifier over edge embeddings generated by models that assume nodes in the legislative graph are homogeneous and featureless, and then aggregate the roll call results: *(a) DeepWalk* (Perozzi et al., 2014) is an embedding model that generates node vectors by running Skip-Gram on random walks formed at different nodes in the graph. *(b) GCN* (Kipf and Welling, 2016) is the basic two-layer GCN model that uses a single weight matrix in each layer and begins with the random node features in the first layer. *(c) RGCN* (Schlichtkrull et al., 2018) is the relational version of the GCN that captures different relations in our legislative graph.

Category 3: text-attributed (TA) graph embedding models: We use the same edge classifier

Table 4: *Macro-F1 in bill survival and cleavage prediction for the random split and known sponsors' relations.*

Embedding		Pass/Fail	Competitive				Inverse-Competitive			
			Party	Gender	Ideology	Geography	Party	Gender	Ideology	Geography
Naive	*Majority*	0.47	0.44	0.46	0.44	0.46	0.48	0.47	0.45	0.47
	Sponsor	0.51	0.43	0.43	0.41	0.43	0.44	0.45	0.41	0.45
Text-based	*BoW*	0.63	0.64	0.64	0.65	0.60	0.58	0.60	0.57	0.62
	GloVe	0.65	0.67	0.66	0.67	0.61	0.57	0.62	0.60	0.63
	BERT	0.68	0.70	0.72	0.69	0.66	0.58	0.64	0.62	0.67
Featureless Graph	*DeepWalk*	0.49	0.52	0.50	0.54	0.56	0.52	0.50	0.52	0.51
	GCN	0.49	0.53	0.51	0.55	0.57	0.52	0.51	0.53	0.52
	RGCN	0.57	0.57	0.53	0.55	0.59	0.54	0.52	0.55	0.56
Text Attributed Graph	*TA-DeepWalk*	0.66	0.67	0.64	0.68	0.60	0.53	0.62	0.55	0.71
	TA-GCN	0.67	0.67	0.65	0.66	0.61	0.52	0.62	0.54	0.72
	TA-RGCN	0.72	0.69	0.65	0.71	0.63	0.56	0.64	0.57	0.72
Joint Graph+Text		**0.82**	**0.83**	**0.79**	**0.82**	**0.73**	**0.64**	**0.78**	**0.65**	**0.78**

but use the graph models that can consume the text-based node features generated by our BERT-based node encoders: *(a) TA-DeepWalk* (Yang et al., 2015) that changes the graph factorization in DeepWalk to support node features. *(b) TA- GCN* (Kipf and Welling, 2016) is the original GCN that takes as input an initial node features. *(c) TA-RGCN* (Schlichtkrull et al., 2018) is a relational GCN that captures node features initialized by our text-based node encoders.

Category 4: naive baselines. We evaluate two other naive classifiers: *(a) Majority*: A baseline predicting the most frequent class in the training data: *(b) Sponsor*: A logistic regression classifier that directly predicts bill survival and cleavages based on the one-hot encoded sponsors' info. encoded.

5.3 Results and Analysis

Performance of different textual and graph models. Table 4 shows macro F1 for different bill cleavages and pass/fail. We first analyze the performance of different models in each category: (1) Among the naive models, the sponsor-based classifier improves the bill survival prediction compared to the majority model but has no positive impact on bill cleavages as expected intuitively. (2) In the textual models, we observe BERT improves the F1 performance by 2%-8% compared to GloVe and BoW. By leveraging a bidirectional operation, BERT more efficiently captures the context of each word in the bill title, summary, and body. (3) In the featureless graph models, RGCN consistently outperforms the standard GCN and DeepWalk models as it treats each of the relations in the legislative graph (e.g., donation and voting) differently and does not mix their weight matrices with each other. This benefit of RGCN is entirely enabled by our new dataset that explicitly tracks

different legislative relations; (4) Unlike the second category, the text attributed graph models capture implicit relations between different nodes in the graph through their text features. By leveraging our node encoders, they begin with better initial representations of the nodes and relations (e.g., particularly votes) and thus provide an improvement by up to 15% in the performance compared to their featureless counterparts. (5) Finally, our proposed model by combining and jointly optimizing the graph and textual representations consistently provides a higher F1 score. Compared to the other models, it improves recall while maintaining high precision, e.g., in the case of the bill survival prediction, the macro precision and recall values for BERT, TA-RGCN, and our model are (0.72, 0.67), (0.92, 0.66), (0.82, 0.84), respectively.

Language and implications of different cleavages. We can make a few observations: it is slightly more challenging to identify inverse-competitive bills compared to competitive ones. This happens across different graph and text models, and thus indicating the language of these bills and the dynamics of relations behind them is rather complex. To help provide an intuition, we summarized in Table 6 the top bigrams and unigrams used in competitive and inverse-competitive bills across the different cleavages. Interestingly, the top n-grams of competitive bills align better with the cleavages (e.g., *"abortion"* is competitive both based on ideology and gender) compared to the top inverse-competitive n-grams, which often focus on mundane issues such as taxes and services, suggesting that when non-polarizing legislation is discussed, group agreement takes a secondary role.

From another angle, Figure 5 further illustrates the differences between these two categories of cleavages. Overall, there are 10%-20% more com-

Table 5: *Macro F1 for bill survival and party cleavages for the best model in each category based on the state- and time-based data splits.*

Embedding	State-based (Test: IN)			State-based (Test: OR)			Time-based (Test: 20%)		
	Pass/fail	Comp.	Inverse. Comp	Pass/fail	Comp	Inverse Comp	Pass/fail	Comp.	Inverse. Comp.
Naive (Majority)	0.47	0.44	0.45	0.46	0.45	0.44	0.48	0.45	0.46
Text-based (BERT)	0.63	0.64	0.53	0.61	0.64	0.54	0.67	0.67	0.57
Featureless Graph(RGCN)	0.52	0.52	0.50	0.51	0.50	0.51	0.54	0.54	0.52
Text-Attributed Graph (TA-RGCN)	0.60	0.62	0.53	0.62	0.61	0.52	0.67	0.68	0.55
Joint Graph+Text	**0.70**	**0.72**	**0.58**	**0.70**	**0.70**	**0.58**	**0.73**	**0.76**	**0.61**

Table 6: *Most frequent unigrams and bigrams of competitive and inverse-competitive bills.*

|	Type	Unigram/Bigram
Comp.	**Party**	law, fund, abortion, political subdivision, providing penalty, badger-care plus, parental choice
	Gender	abortion, child, medical, school, providing penalty, motor vehicle, minimum wage, parental choice
	Ideology	income, abortion, insurance, drugs, local government, retirement system, natural resources, political subdivision
	Geography	county, service, commission, district, transportation, housing, residential, state financial, criminal history, restroom facility, greenhouse gas
Inv-comp.	**Party**	state, program, motor vehicle, real estate, study committee, education matters,
	Gender	financial, emergency, permits, legislative council, economic development, criminal penalty
	Ideology	tax, services, county, criminal, alcoholic beverages, board education, commission declaring
	Geography	law, school corporation, property tax, unemployment insurance
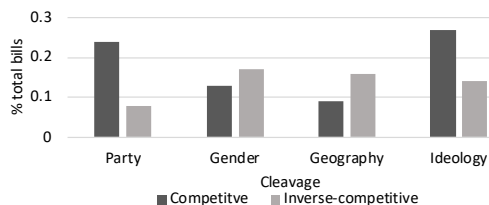
Figure 5: *Distribution of competitive and inverse-competitive bills before split over 2011-2018.*

petitive bills compared to inverse-competitive ones under the party and ideology attributes, indicating cross-group disagreements (e.g., conservatives VS. moderates VS. liberals) are more likely than intra-group disagreement. This pattern is reversed for the gender and geography attributes.

Implication of state- and time-based data splits. For the pass/fail and party cleavages with the best model in each category, Table 5 shows a sharp drop in the F1 score for the state-based and time-based data split, particularly for graph-based models (RGCN and TA-RGCN). By training the model with the two states and testing it with an-

other one, the graph-based embedding models are challenged with representing many unseen legislators. While GCN-based solutions are capable of creating such representations in the test time (using the same weight matrix), they are sub-optimal particularly in featureless GCN settings. One interesting observation is that when the model is tested with the OR data, the drop is even sharper as OR tends to be a democratic state; While WI and IN are often republican states. For the time-based data split, we observe a similar but slightly better performance as the number of unseen nodes are fewer. In all these different configurations, our joint model still improves the F1 score but it is limited on how the underlying graph model behaves.

6 Summary

In this paper, we take the first step towards understanding the dynamics of state-level legislative processes in the US through a data-driven approach. We proposed to collapse the legislative process into a heterogeneous multi-relational graph and suggest several tasks for capturing disagreement over several ideological and demographic cleavages, as well as predicting the outcome of the legislative process. We approach these problems by formulating them as aggregate roll-call prediction.

To fully realize the potential of graph-based modeling, we created a new dataset, used to characterize the real-world context in which the legislative process takes place, consisting of bills, donors, and legislators and their behavior. We model the rich relationship between these entities and the content of the bills using a joint text and graph prediction model on top of BERT and RGCN, outperforming each one of the models in isolation.

References

Ballotpedia. 2019. State-level political encyclopedia data. https://ballotpedia.org/.

Joost Bastings, Ivan Titov, Wilker Aziz, Diego Marcheggiani, and Khalil Sima'an. 2017. Graph convolutional encoders for syntax-aware neural machine translation. *arXiv preprint arXiv:1704.04675*.

Glen T Broach. 1972. A comparative dimensional analysis of partisan and urban-rural voting in state legislatures. *The Journal of Politics*, 34(3):905–921.

Aditya Budhwar, Toshihiro Kuboi, Alex Dekhtyar, and Foaad Khosmood. 2018. Predicting the vote using legislative speech. In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*.

GIS Census. 2019. Gis census data. https://www.nhgis.org/.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. 2016. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in neural information processing systems*, pages 3844–3852.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Vlad Eidelman, Anastassia Kornilova, and Daniel Argyle. 2018. How predictable is your state? leveraging lexical and contextual information for predicting legislative floor action at the state level. *arXiv preprint arXiv:1806.05284*.

FollowTheMoney. 2019. State-level contributor data. https://www.followthemoney.org/.

Brian Frederick. 2010. Gender and patterns of roll call voting in the us senate. In *Congress & the Presidency*, volume 37, pages 103–124. Taylor & Francis.

Sean Gerrish and David M Blei. 2011. Predicting legislative roll calls from text. In *Proceedings of the 28th international conference on machine learning (icml-11)*, pages 489–496.

Sean Gerrish and David M Blei. 2012. How they vote: Issue-adjusted models of legislative behavior. In *Advances in Neural Information Processing Systems*, pages 2753–2761.

Mikael Henaff, Joan Bruna, and Yann LeCun. 2015. Deep convolutional networks on graph-structured data. *arXiv preprint arXiv:1506.05163*.

Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. 2014. Political ideology detection using recursive neural networks. In *Association for Computational Linguistics*.

Shannon Jenkins. 2012. How gender influences roll call voting. *Social Science Quarterly*, 93(2):415–433.

Hamid Karimi, Tyler Derr, Aaron Brookhouse, and Jiliang Tang. 2019. Multi-factor congressional vote prediction. *Advances in Social Networks Analysis and Mining (ASONAM)*.

Kevin King. 2019. State Legislatures Vs. Congress: Which Is More Productive? http://bit.ly/30YsKwT. [Online; accessed 19-July-2019].

Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Anastassia Kornilova, Daniel Argyle, and Vladimir Eidelman. 2018. Party matters: Enhancing legislative embeddings with author attributes for vote prediction. In *Proceedings of ACL*.

Peter Kraft, Hirsh Jain, and Alexander M Rush. 2016. An embedding model for predicting roll-call votes. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*.

LegiScan. 2019. State-level legislative data. https://legiscan.com/.

Yifu Li, Ran Jin, and Yuan Luo. 2018. Classifying relations in clinical narratives using segment graph convolutional and recurrent neural networks (Seg-GCRNs). *Journal of the American Medical Informatics Association*, 26(3):262–268.

Diego Marcheggiani and Ivan Titov. 2017. Encoding sentences with graph convolutional networks for semantic role labeling. *arXiv preprint arXiv:1703.04826*.

John J Nay. 2017. Predicting and understanding lawmaking with word vectors and an ensemble model. *PLoS ONE*, 12(5):e0176999.

Pallavi Patil, Kriti Myer, Ronak Zala, Arpit Singh, Sheshera Mysore, Andrew McCallum, Adrian Benton, and Amanda Stent. 2019. Roll call vote prediction with knowledge augmented models. In *Proceedings of the 23rd Conference on Computational Natural Language Learning (CoNLL)*, pages 574–581.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. 2011. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830.

Hao Peng, Jianxin Li, Yu He, Yaopeng Liu, Mengjiao Bao, Lihong Wang, Yangqiu Song, and Qiang Yang. 2018. Large-scale hierarchical text classification with recursively regularized deep graph-cnn. In *Proceedings of the 2018 World Wide Web Conference*, pages 1063–1072. International World Wide Web Conferences Steering Committee.

Tai-Quan Peng, Mengchen Liu, Yingcai Wu, and Shixia Liu. 2016. Follower-followee network, communication networks, and vote agreement of the us members of congress. *Communication Research*, 43(7):996–1024.

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

Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*.

Neil Pinney and George Serra. 1999. The congressional black caucus and vote cohesion: Placing the caucus within house voting patterns. *Political Research Quarterly*, 52(3):583–607.

Michael Schlichtkrull, Thomas N Kipf, Peter Bloem, Rianne Van Den Berg, Ivan Titov, and Max Welling. 2018. Modeling relational data with graph convolutional networks. In *European Semantic Web Conference*, pages 593–607. Springer.

Boris Shor and Nolan McCarty. 2011. The ideological mapping of american legislatures. *American Political Science Review*.

Bishan Yang, Wen-tau Yih, Xiaodong He, Jianfeng Gao, and Li Deng. 2014. Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575*.

Cheng Yang, Zhiyuan Liu, Deli Zhao, Maosong Sun, and Edward Chang. 2015. Network representation learning with rich text information. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.

Tae Yano, Noah A Smith, and John D Wilkerson. 2012. Textual predictors of bill survival in congressional committees. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 793–802. Association for Computational Linguistics.