

# Modeling U.S. State-Level Policies by Extracting Winners and Losers from Legislative Texts

Maryam Davoodi  
Purdue University

mdavoodi@purdue.edu

Eric Waltenburg  
Purdue University

ewaltenb@purdue.edu

Dan Goldwasser  
Purdue University

dgoldwas@purdue.edu

## Abstract

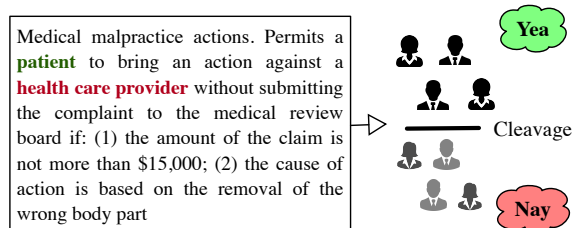
Decisions on state-level policies have a deep effect on many aspects of our everyday life, such as health-care and education access. However, there is little understanding of how these policies and decisions are being formed in the legislative process. We take a data-driven approach by decoding the impact of legislation on relevant stakeholders (e.g., teachers in education bills) to understand legislators' decision-making process and votes. We build a new dataset for multiple US states that interconnects multiple sources of data including bills, stakeholders, legislators, and money donors. Next, we develop a textual graph-based model to embed and analyze state bills. Our model predicts winners/losers of bills and then utilizes them to better determine the legislative body's vote breakdown according to demographic/ideological criteria, e.g., gender.

## 1 Introduction

State-level legislation is the cornerstone of national policies and has long-lasting effects on residents of US states. Thus, decoding the processes that shape state bills is crucial yet involved. State legislatures vote on 23 times more bills than Federal legislatures, exceeding 120K bills per year (King, 2019). In addition, these state bills cover a broader range of local problems as each state possesses lawmaking power effective within its boundaries. E.g., the State of Washington Health Care Committee addresses health service issues including licensing and regulation of health care facilities and providers. Moreover, it regulates pharmacies, pharmaceutical drugs, state public health programs, and private/public insurance markets (House, 2021).

We argue that recent NLP architectures can provide new insights into the state-level legislative efforts. In particular, contextualized graph and text embedding can better represent policies within and across states via a shared political context. However, most of the prior efforts are focused

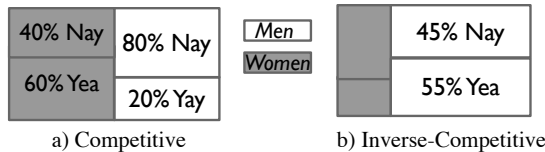
on analyzing congressional bills with traditional techniques, e.g., (Gerrish and Blei, 2011, 2012). A few state-level studies (Eidelman et al., 2018; Davoodi et al., 2020) took great steps in predicting the progression of state bills towards a vote on the floor and the breakdown of votes based on demographic metrics (e.g., gender). But their main downside is they evaluate policies in a limited context and do not capture cross-state patterns.



**Figure 1:** A health bill leading to voting cleavage based on the party metric primarily due to its specific losers (red) and winners (green).

**Winners-Losers analysis.** In this work, we take a new data-driven approach to analyzing state legislation. Our key insight is that each state bill inevitably produces some *winners* and *losers* to provide practical solutions to specific in-state and local problems. Thus, we argue that it is important to examine state bills in the larger context of their impact on different population segments as well as commercial and professional stakeholders. To help clarify this idea, consider the example in Figure 1. This state bill makes it easier for patients (*winners*) to take legal actions against healthcare providers (*losers*). This analysis of *winners and losers* (WLs) can foster transparency in legislative efforts in each state, while interconnecting different states through common stakeholders and revealing cross-state patterns. In addition, the context of WLs can enable a new category of NLP models for predicting the roll-call behavior of legislators.

**Downstream bill classification tasks.** For instance, the political science community sees tremendous value in predicting voting cleavages,



**Figure 2:** Two types of gender-based voting cleavages.

based on the ideological and demographic identities of legislators (Section 2). Each of such metrics (e.g., party, gender, district, ideology) splits legislators into groups. Measuring lack of consensus within and across these groups, which has political and social benefits, can be done using two classification tasks (Section 5): For a given metric, we say a bill is *competitive* (Figure 2) if the majority vote of legislators from a group (e.g., Democrat, male, urban, liberal) is different from that of the opposite group (e.g., Republican, female, rural, conservative). Similarly, a bill is *inverse-competitive* if there is a tie in votes of members of the same group (e.g., liberals). For instance, the health bill in Figure 1 resulted in a party competitive vote. Another example is a state bill on abortion that “requires... physician performing an abortion to admitting privileges at a hospital in the county” resulted in a gender competitive vote. We show the context of winners/losers of these bills could hint at such cleavages prior to voting (Sections 4, 6).

**Framework overview.** To achieve this goal, we address multiple NLP challenges in our proposed framework: (1) *Data*: The legislative process in US states does not track the stakeholders of bills and the impact of bills on them. Thus, we design a reliable crowd-sourcing pipeline to extract and analyze winners and losers of state bills from their text and form a new annotated dataset. (2) *Modeling*: To automate the WL analysis, next, we provide a nationwide graph abstraction to model the state legislative process, as well as a joint text and graph embedding architecture for predicting winners and losers. Our model captures the interplay of different entities, e.g., bills, stakeholders, legislators, and money donors, while maintaining dependencies between their textual attributes. We leverage RGCN (Schlichtkrull et al., 2018), a relational graph convolutional network, to represent diverse relations. We also adopt the RoBERTa transformer (Liu et al., 2019) after performing domain-adaptive pretraining on political texts using the MLM (Masked Language Model) task. (3) *Application*: Finally, we showcase the ability of our WL analysis and prediction model in

decoding the voting behavior of state legislators. In summary, we make three *technical contributions*:

- We provide the first definition and realization of winners/losers analysis for state bills using the latest NLP advances. (Sections 2, 3, 4).
- We developed a new joint graph and text embedding model that both predicts winners/losers of bills and legislators’ votes. In particular, it incorporates the winners/losers inference into the vote prediction task, to evaluate bills in a broader context (Section 5).
- We operationalized the winners/losers analysis for several legislative topics (e.g., health) and created a new dataset. The extensive evaluation shows our approach delivers a higher F1, than existing models (Sections 3, 6)

## 2 Related Works

Our work is inspired by some promising studies:

**Roll-call classification.** Eidelman et al. 2018 associate the bill’s text with partisan information of its sponsors to predict the likelihood of a member of the U.S. Congress voting in support of a bill. Similarly, Gerrish and Blei 2011 embed the combined topic and text of Congress bills in the ideological space and develop ideal point models for inferring votes. Peng et al., 2016; Kornilova et al., 2018; Kraft et al., 2016; Patil et al., 2019; Karimi et al., 2019; Pujari and Goldwasser, 2021 augment this model using data on social networks, thus generating better embeddings.

**Bill text classification.** Instead of leveraging bill text in models to describe the behavior of each legislator, Yano et al. 2012 include the bill’s text in a model that directly predicts whether a bill comes out from a standing committee. Particularly, they develop features based on the urgency of the problem being solved by the bill and the set of legislators co-sponsoring the bill. Eidelman et al. 2018 conduct a similar study on US states.

**Winners-losers analysis.** Analyzing the impact of bills on its stakeholders is a well-studied topic in the political science literature. Gamm and Kousser, 2010 reveal state legislators are more likely to write bills aimed at a particular local stakeholder when the legislative body is dominated by one party. Similarly, Bagashka and Clark, 2016 show state legislators are motivated to introduce particularistic bills designed to help a specific geographical area within

their district. Pennock, 1979 analyzes legislation based on its generalized and particularized impact on different interest groups. By leveraging recent NLP advances (e.g., contextualized language models, graph embedding, crowdsourcing), our work extends these studies and provides the first automated framework for the stakeholders analysis on state bills.

**Voting cleavages.** Research has covered multiple ways that the demographic background of legislators can affect roll-call voting. Frederick 2010 demonstrates gender affects the roll-call vote in the Senate by changing the influence of partisanship for GOP women. Broach 1972 describes that urban-rural voting cleavages happen in less partisan states and on bills that separate urban and rural interests. Similar to us, Davoodi et al. 2020 build a textual graph to predict such cleavages. While our focus is on a different problem, stakeholders analysis, we outperform this prior study by representing bills in a broader context containing their stakeholders.

**Graph embedding in NLP.** Our work uses Graph convolutional networks (GCNs), which have been applied to various NLP tasks, e.g., Semantic role labeling (SRL) (Marcheggiani and Titov, 2017) and relation classification in clinical narratives (Li et al., 2018). In these tasks, GCNs encode the syntactic structure of sentences. Similarly, Defferrard et al., 2016; Peng et al., 2018; Henaff et al., 2015 use graph neural networks (GNNs) to represent a network of documents based on their references. Similar to our work but for a different problem and objective, Sawhney et al., 2020 analyze speech-level stance of members of the parliament, by performing node classification on graph attention networks (GATs), and Pujari and Goldwasser, 2021 analyze social media content generated by politicians using a graph transformer model.

### 3 Modeling

We first provide an overview of key players in the state-level legislative process. Then, we model them using an efficient text-based graph abstraction (Figure 3), which will enable us to embed and evaluate state policies in a broad context and perform the stakeholder and roll-call analysis on them.

#### 3.1 Players in State Legislative Process

Our model, unlike prior works, fully captures the interplay of main players in the lawmaking process:

**1. Legislators.** A state legislature typically con-

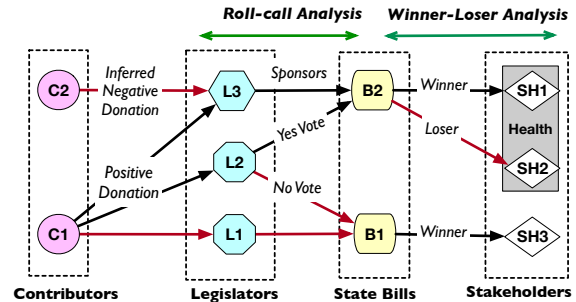


Figure 3: Nationwide, Heterogeneous, and multi-relational graph abstraction for analysis and prediction of winners-losers and roll-calls.

sists of two “chambers”<sup>1</sup>: the House and the Senate. The legislative process starts with legislators sponsoring a bill in a chamber. The idea of a bill can come from different sources. Next, the bill goes through multiple roll-call votes in the origin chamber, where it can fail at any stage. It is first referred to the proper committee by the chamber leader. Committee members, before casting their votes, may set up a public hearing with the sponsors and interested parties. If the bill passes out of the committee, it reaches the second reading, where the full chamber debates, amends, and votes on the bill. If the bill passes by a majority vote, it is scheduled for the third reading and final vote. A bill must go through a similar procedure in the other chamber before it is acted on by the governor.

**2. Contributors.** While legislators navigate through bills, external contributors influence their decisions. Individual and corporate money donors aim at developing changes in the outcome and theme of bills starting from the election times. Lobbyists launch campaigns to persuade legislators towards certain policies. Such efforts inevitably lead to new bills or amendments to existing laws.

**3. Stakeholders.** A state bill cannot benefit everyone and it produces beneficial or detrimental effects on its stakeholders. Identifying winners and losers of a bill from its text is crucial, which can hint at the fate of a bill. Particularly, legislators do not always write bills themselves. Corporations and interest groups (e.g., ALEC) sell fill-in-the-blank bills to legislators. Thus, we can see voting patterns on bills with the same winners and losers.

<sup>1</sup>Nebraska’s legislature is unique in the nation because it has a single-house system.

### 3.2 Nationwide, Multi-Relational, and Heterogeneous Legislative Graph

To model these players and their interactions, we design a legislative graph with three important properties (Figure 3). First, since each of the players (e.g., stakeholders, legislators) has different textual attributes, our proposed graph supports heterogeneous textual nodes. Second, we form a nationwide graph to capture cross-state patterns (ablation study in Appendix A.2) by building common entities (e.g., stakeholders in Section 4). Finally, our abstraction supports multiple relations between each pair of entities (e.g., legislators voting and sponsoring a bill). With this overview, we present the nodes and relations that we will realize based on the real data:

**Node types.** The nodes in the legislative graph contain a rich set of textual features: (1) *Bill* nodes embed title, abstract, and body of state bills. (2) *Stakeholder* nodes come with short texts on political interests and constituent entities of stakeholders of policies in bills (will be detailed shortly). (3) *Legislator* nodes contain diverse textual information on legislators, e.g., their biography, political interests, committee assignments, and demographic profile (e.g., party, gender, ideology, and district). (4) *Contributors* nodes have text-based attributes on money donors covering their specific/general business interests, party, and their type (individual or non-individual).

**Relation types.** Based on the legislative process, legislator and bill nodes participate in *Bill Sponsorship*, *'No' Vote*, and *'Yes' Vote* relations in the graph (See Appendix A.4 for handling abstain votes.) A stakeholder node forms *Winner*, *Loser*, or *Neutral* relations with a bill node, which we will extract it based on the bill text. Similarly, we form two types of relations between contributors and legislators: *Positive Donation* realized based on the real donation data, and *Negative Donation*, which we infer when a contributor shows a lack of interest in a demographic of legislators (e.g., never donates to women). We sample and connect such legislators and the contributor via a negative relation.

## 4 Data-Driven Stakeholders Analysis

Next, we describe how we build up the legislative graph, by collecting data on legislators, bills, and contributors. US states do not record the impact of bills on relevant stakeholders. Thus, we explain how to derive stakeholders from bill nodes,

perform winners-losers analysis on them, and interconnect different US states by forming common stakeholder nodes. We highlight how our analysis can be used (1) to inform the public about the dynamic and direction of state policies, and (2) to determine legislators' roll-call behavior with different demographic and ideological profiles.

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 1:** Aggregated statistics of the legislative graph—Cont: Contributor, Leg: Legislator.

Topic	Education	Health	Law	Agriculture
# of bills	957	942	1140	758

**Table 2:** Bills sampled for the stakeholders analysis.

### 4.1 Data Collection & Bootstrapping Graph

**Bills.** From the LegiScan website (LegiScan, 2019), we collected data on bills introduced in Indiana, Oregon, and Wisconsin from 2011 through 2018 (details in Appendix 7). We developed a crawler that uses the LegiScan API to fetch legislative information on every bill, including: (1) bill metadata, e.g., the bill type, title, description, sponsors, and links to its texts; (2) vote metadata, e.g., legislator's roll-call vote; and (3) legislator metadata, e.g., party and district info. Then, our crawler converts bill texts in PDF format to text files. In total, we collected ~35k bills and sampled 58% of them that had both roll-call data and full texts. Our focus is on the 2nd/3rd reading, in which the full chambers vote, so we selected 32% of the bills for building the legislative graph (Table 1). In LegiScan, each bill is associated with a main topic (e.g., health), used for referral to a proper committee. For the four most frequent topics (Table 2), we will define a group of generic stakeholders for the winners-losers analysis.

**Legislators.** Our crawler also used Ballotpedia (Ballotpedia, 2019) to collect text information on each legislator's biography, political interests, and committee assignments. Also, it consumed other publicly available datasets to identify a legislator's demographic profile, e.g., ideology, gender, and district. The ideology scores for legislators (Shor and McCarty, 2011) were grouped into conservatives, moderates, and liberals. The district identifier was combined with GIS census data (Census, 2019) to identify each legislator as representing an urban or

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

**Table 3:** Aggregated legislators’ attributes—UR: Urban, RU: Rural, C: Conservative, M: Moderate, L: Liberal.

Topic	Stakeholders	W (%)	L (%)
Education	Edu. companies & service providers	1.4	1
	Educational institutions and schools	23.9	8.7
	State education agencies	6.3	8.6
	Teachers and education workers	13.2	1.3
	Students	34.2	1.6
Agriculture	Agriculture and food-related companies	4.5	4.1
	Agricultural and food producers	24.4	6.9
	End consumers or retail customers	11.6	11.2
	State agriculture and food agencies	14.5	1.4
	Grocery stores or food providers	11.6	9.8
Health	Healthcare facilities	16.7	7.7
	Healthcare providers and professionals	6.8	3.3
	Insurance providers and companies	11.4	10.5
	Patients and insurance owners	16.7	6.3
	Pharma and medical device companies	4.6	0.5
State healthcare agencies	11.7	4	
Law	Law enforcement agencies and officers	15.7	24.7
	Judges	11.5	9.4
	Victims, offenders, suspects	9.9	11.2
	Lawyers	9.8	7.7

**Table 4:** Stakeholders of different bill topics and their frequency distribution as winners (W) and losers (L).

rural district. Table 3 summarizes the distribution of legislators’ demographic profile.

**Contributors:** FollowTheMoney (FollowTheMoney, 2019) records donations to state legislators and candidates. Our crawler collected the information of donors for each legislator in our dataset (See Table 1). This includes multiple textual attributes for each contributor: type (individual or non-individual), general party, and economic and business information. While the contributor data can be utilized in more sophisticated ways, we focused on major contributors by setting a minimum donation threshold and pruning donors who contributed to a single legislator; We set the fraction of negative donations (Section 3) to 30% of the positive ones extracted from the data.

## 4.2 Stakeholders Extraction & Annotation

For each select bill topic, we (authors) randomly sampled 10% of bills and carefully analyzed their texts. We recorded entities discussed in the bill texts as well as the detrimental or beneficial impact of the suggested policies on them (regardless of the legislative outcome, i.e., passed in a vote or not). To interconnect different states and opti-

mize the legislative graph, we deduplicated entities and clustered those whose interests align (e.g., surgeons, doctors, dentists, and etc.) into generic ones (e.g., healthcare providers). Table 4 shows the final list of stakeholders for the select topics. With detailed annotation guidelines, we leveraged Amazon MTurk for labeling ~4k bill texts from these topics (Table 2), where 3-5 workers identified the effect of the suggested policies in each bill on the relevant stakeholders. As will be detailed in Appendix A.1, we ensured the accuracy of the labeled data is 90%+.

## 4.3 Benefits of Winners-Losers Analysis

Based on the outcomes of the previous two steps, we formed a legislative graph for our target states. We briefly provide two results from the winners-losers analysis on the graph to highlight its importance. First, we show the frequency distribution of the stakeholders as a winner vs. a loser for each topic in Table 4, which would inform the public about the dynamics and directions of state-level policies. E.g., under the education topic, students were the largest winners, while educational institutions were the major losers. For law bills, law enforcement agencies were the top losers given the recent nationwide focus on police use of force.

Legislator Profiles		Education Companies	Educational Institutions	State Edu. Agencies	Teachers, Edu. Workers	Students
Party	Democrat	-7.0	-4.0	1.0	6.0	-3.0
	Republican	18	1.0	-14.0	-1.0	24.0
Geography	Rural	-3.3	-4.3	1.0	5.9	-2.8
	Urban	4.7	-2.3	-6.3	2.5	9.9
Gender	Male	1.8	-3.3	-4.2	3.8	5.6
	Female	4.9	-2.2	-7.7	2.2	12.7
Ideology	Liberal	-3.1	-6.3	3.6	4.7	-3.9
	Moderate	14.4	3.2	-5.2	2.0	5.9
	Conservative	-7.4	-5.1	-22.3	5.5	28.1

**Table 5:** Capturing policy preferences of different demographic and ideological groups of legislators on education bills, by measuring the change in the rate (%) of ‘yes’ vote when a stakeholder is a winner and a loser.

Also, our winners-losers analysis captures the policy preferences of different ideological and demographic groups of legislators. For example, Democrats are more likely to support legislation benefiting teachers, compared to Republicans (GOP). This fact is also reflected in our models predicting voting cleavages in Section 6 (e.g., our naive model, WL-Correlation, outperforming other models in its category). Here, to motivate the need for building such models, we are interested in

measuring the rate of 'Yes' votes from each demographic and ideological group of legislators on bills of a given topic, where a stakeholder is a loser and a winner. E.g., on education bills benefiting a stakeholder (e.g., Students) as a winner, we compute,  $A = [\# \text{ of yes votes}] / [\text{total } \# \text{ of votes}]$  in the GOP legislators. Similarly, on education bills, where this stakeholder is a loser, we calculate  $B = [\# \text{ of yes votes}] / [\text{total } \# \text{ of votes}]$  for GOP. We then report the difference,  $A - B$ , in Table 5, where a large positive value indicates the stakeholder is being advantaged by the respective group of legislators. E.g., we see GOP has significantly more Yes votes when students are winners, compared to Yes votes when students are losers. By running queries on the legislative graph containing all players (e.g., donors), we were able to see the voting behavior of GOP could be motivated by major donations to this party from corporations representing students (e.g., School Choice).

## 5 Embedding & Prediction Architecture

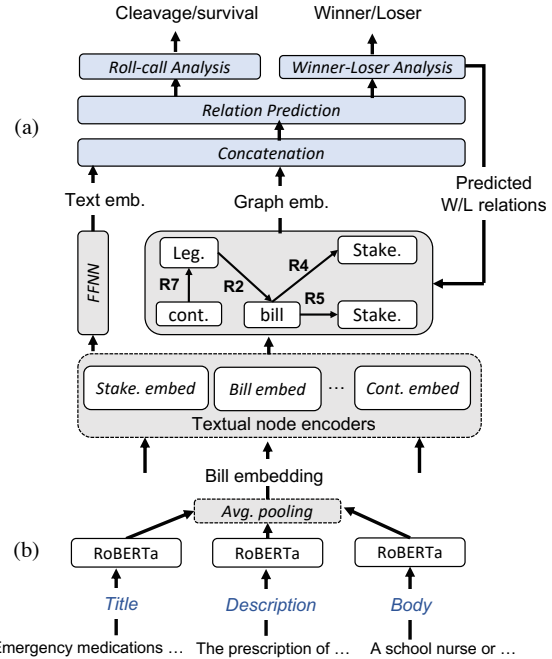
The stakeholder analysis based on human data annotation is expensive and time-consuming. To automate the analysis and better leverage its results in other applications, we build up a contextualized embedding architecture and define two classification tasks on the legislative graph:

### 5.1 Classification Tasks on Legislative Graph

**Task 1: winners-losers prediction.** Our first task is to predict the relation between a bill node and each relevant stakeholder node (based on its topic in Table 4). Such predicted relations will bring valuable insights into the bills, while also clarifying legislators' roll-call behavior (Section 6). Thus, we define the next task to showcase these benefits.

**Task 2: bill cleavage and survival prediction.** For a bill, we predict if (1) it shows identifiable voting cleavages and (2) it can advance by getting a pass. We achieve these by predicting and aggregating roll-call relations (between legislators and bills) in the graph. In particular, we assign 9 labels to each bill: (1) *Competitive labels*: For voting cleavages, we split legislators into groups based on their demographic and ideological profiles (party, gender, ideology, and the urban/rural nature of their district as defined in Section 4). For an attribute (e.g., gender), we say a voting round is "competitive" if the majority of legislators from one group (e.g., Women) and the majority of

the opposite group (e.g., Men) cast different votes (Figure 2a). (2) *Inverse-competitive labels*: Similarly, for an attribute (e.g., gender), a voting round is inverse-competitive if we identify a partial or complete tie (Appendix A.4) in the vote of legislators of the same group (e.g., Men in Figure 2b). (3) *Survival label*: Finally, a bill passes its current voting round by getting a majority vote.



**Figure 4:** Joint text-graph embedding model for shared prediction of winners/losers and roll-calls in the legislative graph using RGCN & pretrained RoBERTa.

### 5.2 Overview of Embedding & Training

At a high-level, we propose a unified model to jointly embed and classify both roll-call and winner-loser relations in the legislative graph (Figure 4a): (a) We first train our model to predict relations between bill nodes and their stakeholders. One can use the result of this stage for further analysis of state policies (e.g., Section ). (b) Our key insight is that knowing winner-loser relations enhances the embedding of nodes in the legislative graph. Thus, we conduct inference on bills that lack such relations (if any) using the pretrained model from step (a) and add these predicted relations to the graph. (c) Next, continue training on the updated graph to fine-tune the model for the roll-call (vote) prediction task. Finally, we aggregate the predicted votes for the bill cleavages/survival analysis. In all these steps, our model generates and jointly optimizes both text and graph embeddings for each node, and consumes them to classify the two types of relations. Thanks to jointly optimizing

the tasks over the textual and graph information, our architecture outperforms existing models (Section 6). Hereafter, we detail the layers in our model using a bottom-up approach:

### 5.3 Contextualized Text Embedding Layers

The lower half of our model generates a contextual embedding for textual attributes of nodes in the legislative graph. We leverage the RoBERTa architecture (Liu et al., 2019). For improved performance, one of our contributions is that we will pretrain RoBERTa on unlabeled bill texts using the MLM task (Section 6). In more detail, for each *bill* node, we feed three pieces of textual information to RoBERTa: title, abstract, and body. RoBERTa does not support input sequences longer than 512 tokens. Thus, we take the representation of each of these components separately (the embedding of their [CLS] token) and do average pooling to output the final representation of the bill. Similarly, the text embedding of *stakeholder*, *legislator*, and *contributor* nodes are the average of the vectors representing their key textual attributes (Section 4).

### 5.4 Relational Graph Convolutional Layers

On top of the text embedding layer, we place a Relational Graph Convolutional Network (RGCN) to create a graph embedding for each node. The RGCN uses the text embedding of each node to initialize its graph representation. In parallel, we build a feed-forward neural network (FFNN), taking the text embeddings of nodes to a concatenation layer for our joint text-graph optimization. The (non-relational) GCN has multiple layers and each layer performs two operations: *propagation* and *aggregation*. In the propagation, nodes update their neighbors by sharing their features or hidden states. In the aggregation, each node adds up the messages coming to it to update its representation. In GCN, at layer  $l + 1$ , the hidden representation of node  $i$ ,  $h_i^{l+1}$ , with neighbours  $N_i$  is:

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

GCN uses the same weight matrix in each layer,  $W^l$ , and normalization factor,  $c_i = |N_i|$ , for all relation types in a graph. We choose RGCN as it uses unique parameters for each relation type, thus better handling our multi-relational graph. In RGCN, the embedding of node  $i$  in layer  $(l + 1)$  is:

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

A 3-layer RGCN turns out to be sufficient in our case to capture the 3rd order relations between contributors and stakeholder nodes.

### 5.5 Relations Prediction Layers

By combining the outputs of the RGCN and FFNN, we train a relation classification layer by using the DistMult scoring function (Schlichtkrull et al., 2018; Yang et al., 2014). For each relation  $(s, r, d)$  being predicted, this layer computes  $f(s, r, d) = e_s^T W_r e_d$ .  $e_s$  and  $e_r$  are the joint text and graph embeddings of the source and destination nodes and  $w_r$  is a diagonal relational weight matrix. Our loss function is  $L = L_{CLS} + L_{Text} + L_{Graph}$  enabling us to jointly optimize the text and graph embeddings as well as the relation prediction.  $L_{CLS}$  is the cross-entropy loss of the relation classification;  $L_{Graph}$  and  $L_{Text}$  are the L2 regularization of RGCN’s and FFNN’s weights, optimizing the graph and text representations, respectively.

## 6 Experiments

We first evaluate the efficiency of our legislative graph abstraction and text+graph embedding model in the winners-losers prediction. Then, we show the benefits of our combined inference of stakeholders and roll-calls in decoding state bills.

### 6.1 Experimental Setup

**Data Split and metric.** We split the legislative graph (Formed in Section 4) based on bill nodes. We randomly select 20% of the bills for testing and keep the rest for training and validation. We study three settings in terms of the winners-losers (stakeholders) information in the graph: (a) *Unknown* winners-losers relations. (b) *Known* relations based on our human-labeled annotation. (c) *Predicted*: 30% of bills in the train graph come with such relations and we predict them for the rest of bills. In Appendix A.3, we will report the results of state- and time-based splits. Finally, given our data is highly skewed, we choose Macro F1 as the main metric over accuracy.

**Settings/parameters.** We build our joint model (Figure 4) on top of PyTorch, DGL, and spaCy. We set the initial embedding dimension in RoBERTa and RGCN to 1024. The FFNN and RGCN take

the embeddings to a 256-dimensional space. We also used Adam optimizer, and for each observed relation (Table 1), we sampled a negative example.

## 6.2 Baseline Models

We devise robust baselines for both of our tasks:

**1. Text-based models.** We build a logistic regression (LR) classifier that takes the text embedding of a bill and predicts if it shows a certain cleavage or passes/fails. A similar classifier takes the text embeddings of a bill and a stakeholder to classify their relation. We evaluate three embedding architectures: *(a) BoW*, where unigram and bigram features (top 5K highest-scoring) are used to represent textual information. *(b) RoBERTa* (Liu et al., 2019). *(c) Pretrained RoBERTa* that we adapted its domain by applying MLM on 10K unlabeled state bills (39K sentences) (Gururangan et al., 2020). We study two additional variations of these models (only for winners-losers prediction due to limited space): *Sponsors*, where the bill sponsors are represented using a one-hot vector and concatenated to the bill text representation. *Roll-Call*, where we concatenate a vector containing cleavage/survival info. of each bill to its text embedding.

**2. Graph-based models:** We build a relation classifier over edge embeddings, generated by three widely-used graph models, to predict roll-call and winner-loser relations (for the bill cleavages/survival task, we aggregate votes): *(a) DeepWalk* (Perozzi et al., 2014) that generates embeddings for nodes and edges by running Skip-Gram on random walks on the graph. *(b) GCN* (Kipf and Welling, 2016) is a basic 3-layer GCN model with random node features in its first layer. *(c) RGCN* (Schlichtkrull et al., 2018) is the relational GCN handling different relation types in the legislative graph.

**3. Naive models.** We evaluate three naive classifiers: *(a) Majority*: A baseline predicting the most frequent class in the train data. *(b) Sponsor*: An LR classifier that uses the one-hot embedding of bill sponsors to determine bill survival/cleavages (similarly winner/loser relations). *(c) WL-Correlation* (solely for the survival/cleavage task) predicts a legislator’s vote on a test bill with known winners/losers based on his historical votes on train bills with the same winners/losers.

## 6.3 Exp 1: Winners-Losers Prediction

We compare these models in predicting relations between bills and their relevant stakeholders (Table 6).

Model	Variation	Embedding	F1
Naive	-	Majority	40.3
		Sponsor	55.1
Text-based	Vanilla	BoW	66.2
		RoBERTa	69.1
		Pretrained RoBERTa	69.8
	Sponsors	BoW	67.7
		RoBERTa	70.2
		Pretrained RoBERTa	70.8
	Roll-call	BoW	66.5
		RoBERTa	68.3
		Pretrained RoBERTa	69.1
Graph-based	-	DeepWalk	57.4
		GCN	57.7
		RGCN	63.9
Our Joint Text + Graph	-	RoBERTa + RGCN	73.2
		Pretrained RoBERTa + RGCN	74.1

Table 6: Performance of different models in predicting winner-loser relations between bills and stakeholders.

*(1) In the vanilla text-based category*, RoBERTa shows 2.9 higher F1 than BoW. Our pretrained RoBERTa generates more efficient contextual embedding for text information of bills and stakeholders (e.g., summary), and thus better determines the impact of a bill on its stakeholders. Including the sponsors’ info in the pretrained RoBERTa leads to the best text model. *(2) In the graph-based models*, Deepwalk/GCN exhibits a sharp drop in F1, by ignoring the heterogeneity of relations in the graph and thus producing inefficient representations for them. RGCN overcomes this issue and approaches the best text model with F1 of 63.9. *(3) Our joint text-graph* model combines the strengths of the graph and text models and delivers 3.3 points higher F1.

## 6.4 Exp 2: Impact on Bill Cleavage Prediction

Next, we focus on the performance of different models in determining voting cleavages/survival, with *Unknown*, *Known*, *Predicted* winners-losers in the legislative graph. In Table 7, we report the results for the bill survival and party-based voting cleavages (results for the other cleavages in Appendix, Table 11). We can make a few observations: First, our stakeholder analysis helps all models to better decode state policies, when comparing the same model in the Unknown and Known winners-losers settings: *(1) In the text-based models*, prediction on the textual information of both bills and known winners-losers delivers a higher F1 than only on the text of bills (e.g., Pretrained RoBERTa model gets a 5.4% boost in F1 in predicting party competitive bills). Similarly, *(2) In the graph-based models*: RGCN overcomes



Type	Models		Pass/ Fail	Comp. (Party)	Inverse Comp. (Party)
	Winner/ Loser	Embedding			
Naive	Unknown	Majority Sponsor	47.2 50.6	43.1 52.3	48.3 53.4
		Known	WL Correlation	58.7	51.3
Text-based	Unknown	BoW	48.1	56.8	48.6
		Pre. RoBERTa	49.8	58.2	49.3
	Known	BoW	49.4	61.2	50.1
		Pre. RoBERTa	51.9	63.6	51.8
	Predicted	BoW	48.3	59.1	49.3
		Pre. RoBERTa	50.1	61.4	49.8
Graph-based	Unknown	DeepWalk	59.9	49.1	49.8
		GCN	60.8	50.4	48.2
		RGCN	64.3	52.8	49.8
	Known	DeepWalk	62.6	49.6	50.3
		GCN	63.6	52.2	50.3
		RGCN	66.3	53.7	52.4
	Predicted	DeepWalk	61.3	47.7	50.1
		GCN	62.3	51.2	49.5
		RGCN	65.6	52.3	52.1
Our Joint Text + Graph	Unknown	Pre. RoBERTa + RGCN	69.8	69.9	59.1
	Known		73.7	71.8	59.8
	Predicted		70.7	71.6	59.5

**Table 7:** Effect of winners/losers information on the graph and text-based models in different downstream vote classification tasks. See the results for other demographic voting cleavages in Table 11.

the limitations of Deepwalk in handling heterogeneous relation types (winner-loser vs. roll-call) and delivers consistent gains in the setting *Known*. (3) *Our model* has the best performance due to generating and optimizing a joint graph and text representation for legislators, bills, money donors, and stakeholders in the setting with known winners and losers. Second, by focusing on the models with the *Predicted* winners-losers information, we observe: (5) *Our model* still beats the other baselines, due to our unified model for roll-call and winner-loser training as well as our text-based legislative graph abstraction (Section 5). Of course, there is an expected drop in F1 across different models including ours, when we consume predicted winner-loser relations instead of human-labeled ones. This drop could be tolerable in most cases, thus not hindering the automation of our stakeholder analysis and leveraging its results in downstream vote analysis tasks (ethical considerations in Section 7).

## 7 Conclusion

We took a new data-driven approach to analyze state legislation in the US. We showed that identifying the winners/losers of state bills can (1) inform the public on the directions of state policies, and (2) build a nationwide context for a better understanding of legislators’ roll-call behaviors. Thus, we proposed a text-based graph abstraction to model the interplay of key players in the state legisla-

tive process, e.g., bills, stakeholders, legislators, and donors to legislators’ campaigns. Next, to automate our analysis, we developed a shared text and graph embedding architecture to jointly predict winners/losers of bills and legislators’ votes on them. We created a new dataset using different data sources and human annotation and evaluated the strength of our architecture against existing models. We hope this work will provide a starting point for further studies examining the impact of policy decisions on individuals and groups, an important step towards making the democratic process more transparent.

## Ethical Considerations

Analyzing state legislation is a sensitive task, where unexpected results of research and deployed ML systems can create misguided beliefs on the government policies on important topics (e.g., health, education). Thus, we would like to discuss some ethical aspects related to our work in terms of data and model (considering potential scenarios suggested by Chandrabose et al., 2021):

**1. Selection of data sources.** While there can be different inherent imbalances in the state legislature (e.g., gender and party distribution), we were not able to identify that our data sources adding systematic political and social biases to our study, e.g., towards demographic populations of legislators. All our data sources (e.g., LegiScan and FollowTheMoney) are publicly available and have been used by the political science community over the years. LegiScan (LegiScan, 2019) is a nonpartisan and impartial legislative tracking and reporting service for state bills. FollowTheMoney (FollowTheMoney, 2019) is a nonpartisan, nonprofit organization revealing the influence of campaign money on state-level elections and public policy in all US states. Finally, Ballotpedia (Ballotpedia, 2019) is a nonpartisan, nonprofit organization providing a brief introduction, biography, committee assignment, and general information on legislators across different years. Our study combined these data sources for analyzing state bills in a broad context, thus contributing to reduced *data bias* for all models evaluated in this paper.

**2. Selection of states.** In addition, to help mitigate the risk of data collection bias or topic preference that can be introduced through the choice of specific state legislatures, we randomly picked a “red”, a “blue”, and a “purple” state (indicating

a significant majority for Republicans, Democrats or more balanced state legislature, respectively). There were some restrictions in terms of collecting the data from the above sources (e.g., FollowTheMoney and Ballotpedia). These data sources and services often limit the number of API calls and queries for retrieving the data for educational institutions. Besides this, annotating the data through Amazon MTurk was expensive for us so we conducted our study on four highly discussed topics in state bills (i.e., health, education, agriculture, and law). We will explore ways of expanding our dataset to more states and topics over time.

**3. Disguised winners and losers.** In theory, the authors of state bills (e.g., interest groups selling fill-in-the-blank bills to legislators) may try to reframe bills (disguise winners or losers) to further their political aims. At the first glance, this could pose a challenge to our bill annotation, dataset, and stakeholder analysis. As described in Section 3, the state legislative process has a multi-stage reviewing process in two chambers (e.g., first reading, second reading, and third reading). Thus, we have observed that it is hard to hide the impact of bills on their relevant stakeholders from our qualified annotators, i.e., the authors and multiple vetted MTurk workers for each example, in practice. In addition, our work on MTurk maps the impact of policies suggested by bills to winners and losers. Thus, it already considers those stakeholders that are not mentioned in the text explicitly (More details in Appendix A.1).

**4. Winners and losers analysis.** The analysis, aligning demographic cleavages with winners and losers preferences, is done at an aggregate level based on the data we annotated. These preferences could be influenced by other factors beyond demographics. Deriving conclusions from this analysis could require longitudinal studies, capturing the change of these patterns over time, for example when analyzing policies intended to help correct inequities towards marginalized groups. Our goal is to provide a tool for domain experts that would point at nuanced, stakeholder specific, legislative preferences that can be studied further in order to determine their significance.

**5. Handling abstain votes.** There are abstain (absent and N/A) votes in our dataset. However, we did not include them in our study due to their extremely low frequency (for our proposed model and other baseline models). We leave this evaluation as

a future work.

**6. Handling other countries and languages.** While our dataset is specific to the US, the problem we studied, stakeholder analysis, can be generalized to legislation from other countries and in different languages. Although we have not evaluated such bills (due to lack of data sources), we expect such legislation to produce winners or losers to provide practical solutions to their local problems. In particular, our framework offers a multi-relational graph abstraction and prediction models to analyze stakeholders of bills (winners/losers) and the voting behavior of legislators. These techniques can support non-US national and state-level legislative processes. To accommodate other languages, one could adopt cross-lingual embedding models, e.g., XLM-R (Conneau et al., 2019) instead of RoBERTa, in our architecture.

## Acknowledgement

We would like to acknowledge the members of the PurdueNLP lab. We also thank the reviewers for their constructive feedback. The funding for the use of mTurk was part of the Purdue University Integrative Data Science Initiative: Data Science for Ethics, Society, and Policy Focus Area. This work was partially supported by an NSF CAREER award IIS-2048001.

## References

- Tanya Bagashka and Jennifer Hayes Clark. 2016. Electoral rules and legislative particularism: Evidence from us state legislatures. *American Political Science Review*, 110(3):441–456.
- Ballotpedia. 2019. State-level political encyclopedia data. <https://ballotpedia.org/>.
- 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.
- GIS Census. 2019. Gis census data. <https://www.nhgis.org/>.
- Aravindan Chandrabose, Bharathi Raja Chakravarthi, et al. 2021. An overview of fairness in data-illuminating the bias in data pipeline. In *Proceedings of the First Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 34–45.
- Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Xlmr. *arXiv preprint arXiv:1911.02116*.

- Maryam Davoodi, Eric Waltenburg, and Dan Goldwasser. 2020. Understanding the language of political agreement and disagreement in legislative texts. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5358–5368.
- 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.
- 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.
- Gerald Gamm and Thad Kousser. 2010. Broad bills or particularistic policy? historical patterns in american state legislatures. *American Political Science Review*, 104(1):151–170.
- 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.
- Suchin Gururangan, Ana Marasović, Swabha Swayamdipta, Kyle Lo, Iz Beltagy, Doug Downey, and Noah A Smith. 2020. Don’t stop pretraining: adapt language models to domains and tasks. *arXiv preprint arXiv:2004.10964*.
- Mikael Henaff, Joan Bruna, and Yann LeCun. 2015. Deep convolutional networks on graph-structured data. *arXiv preprint arXiv:1506.05163*.
- WA State House. 2021. Washington State Health Care & Wellness Committee. <https://leg.wa.gov/House/Committees/HCW/Pages/default.aspx>. [Online; accessed 19-July-2021].
- 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.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Diego Marcheggiani and Ivan Titov. 2017. Encoding sentences with graph convolutional networks for semantic role labeling. *arXiv preprint arXiv:1703.04826*.
- 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.
- 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-follower network, communication networks, and vote agreement of the us members of congress. *Communication Research*, 43(7):996–1024.
- J Roland Pennock. 1979. Another legislative typology. *The Journal of Politics*, 41(4):1206–1213.
- 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*.

- Rajkumar Pujari and Dan Goldwasser. 2021. Understanding politics via contextualized discourse processing. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1353–1367.
- Ramit Sawhney, Arnav Wadhwa, Shivam Agarwal, and Rajiv Shah. 2020. Gpols: A contextual graph-based language model for analyzing parliamentary debates and political cohesion. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4847–4859.
- 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*.
- 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.

## A Appendices

### A.1 Data Annotation Pipeline

Our analysis on MTurk maps the policy described in the bills to potential winners and losers, i.e., stakeholders that would be positively or negatively impacted if the bill passes. The analysis is for the proposed policy, regardless of the legislative outcome (pass a vote or not). Due to lack of space, we did not mention certain aspects of our bill annotation task on MTurk in Section 4. We referred to it as a pipeline (in Section 4) because we fully automated the whole process (e.g., selection of MTurkers, publishing bills, collecting and analyzing results, and etc.) in Python, based on MTurk APIs and other open-source libraries. Annotation of political bills, particularly our winners/losers analysis, turned out to be a challenging task for typical MTurk workers. Thus, we developed an automated quality assurance scheme to ensure high-quality annotations for our study. In particular, we built the following components in our pipeline:

Questions/Tasks
How many women currently serve on the US Supreme Court?
Which party currently has the majority of seats in the US Senate?
What is the topic of the following legislation? <i>Prevention and control of, emergency and involuntary commitment for, and treatment programs and services for drug dependence.</i>
Select the entities that lose benefits from this bill? <i>Requires Oregon Health Authority to commission independent study of costs and impacts of operating basic health program in Oregon. Specifies parameters of study. Requires a report to Legislative Assembly by November 30, 2014. Appropriates money from General Fund to authority for contract costs to conduct study. Declares emergency, effective on passage.</i>

**Table 8:** Some of the questions used in our Political Science qualification test.

1. We developed a Political Science Qualification test on MTurk to evaluate candidate MTurk workers. Our test consists of 20 questions (e.g., Sentiment analysis on the US political text, identifying winners and losers of US bills, basic political knowledge questions). Table 8 shows the first four questions in the test.
2. Our pipeline selected 20 qualified English-speaker annotators who successfully completed 80% of the tasks, assigned them our

qualification label on MTurk, and added them to our pool. We designed the test such that it must be completed by candidates in 30 minutes and those who failed the test were not allowed to take it again. While location was not a determining metric for us in selector annotators (instead of focused on evaluating their knowledge of the US policies and politics), most of our qualified annotators were located in the US.

3. Next, for annotating each bill in our dataset, our pipeline randomly chose 3 annotators from the pool to determine the effect of the bill on the relevant stakeholders (generated based on the topic of the bill).
4. After collecting the result, for each bill, it computed the final winners and losers based on the majority rule. For 5% of bills with no agreement among annotators (each annotator selected different winners/losers), we automatically assigned these bills to two additional MTurkers, and then recalculated the final results/labels based on the majority rule.
5. For a small fraction of bills (around 1%) adding new annotators was not sufficient to reach an agreement, and thus we automatically rejected all the results and restarted the process from Step 3 with a new group of annotators. Finally, for a handful of bills, the authors performed the annotation manually.
6. To monitor the accuracy of our annotation, our pipeline sampled labeled bills from each batch of bills and we (authors) performed winners/losers analysis on them to validate the results. We ended up observing that our pipeline generated fully correct labels (all winners and losers) for 90%+ of bills. Figure 5 shows the distribution of winners and losers associated with the bills in our crowd-sourced pipeline.

### A.2 Ablation Study: Effect of Nationwide Context

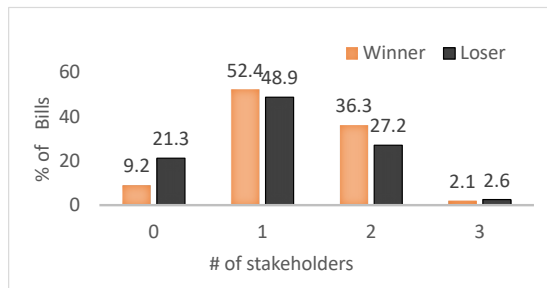
As discussed in Section 3, proposed policies and legislative outcomes at the state level are influenced by the nationwide context. Corporations and lobbying groups coordinate their efforts across multiple states to influence legislators in a similar way. We capture this fact in our graph representation by interconnecting states through common/shared nodes

Model	Embedding	State (Test: IN)	State (Test: OR)	Time (Test: 20%)
Naive	Sponsors	52.3	52.0	53.2
Text-based	Pretrained RoBERTa	65.1	64.8	65.8
Graph-based	RGCN	58.1	58.2	59.5
Our Joint Text + Graph	Pretrained RoBERTa + RGCN	<b>67.5</b>	<b>66.8</b>	<b>67.7</b>

**Table 9:** Performance of best model in each category (defined in Section 6 and Table 6) in predicting winner/loser relations between bills and stakeholders for time-based and state-based splits.

Models			State (Test: IN)	State (Test: OR)	Time (Test: 20%)
Type	Embedding	Winner/Loser			
Naive	WL Correlation	Known	49.5	49.3	50.1
Text-based	Pretrained RoBERTa	Unknown	57.3	57.2	57.4
		Known	59.4	58.8	60.1
		Predicted	58.3	58.0	58.8
Graph-based	RGCN	Unknown	49.5	49.0	51.1
		Known	51.5	50.1	52.1
		Predicted	50.7	49.3	51.5
Our Joint Text + Graph	Pretrained RoBERTa + RGCN	Unknown	62.9	61.3	62.8
		Known	<b>64.1</b>	<b>63.2</b>	<b>65.6</b>
		Predicted	63.8	62.4	64.2

**Table 10:** Effect of winners/losers information on the graph and text-based models in different vote classification tasks, for time-based and state-based data splits.



**Figure 5:** Distribution of winner and loser relations between bills and their stakeholders. Most of the annotated bills had at least one winner or loser. A small portion of the bills (only 3.5%) had no loser and winner. A bill does not necessarily have both a winner and a loser.

in Section 3. We conducted an ablation study to show the benefit of building a nationwide legislative graph. We split common nodes in the legislative graph that were shared across states (e.g., stakeholders, money donors) into state-specific nodes. Then, we repeated a handful of experiments in Section 6. In our classification tasks (both winners/losers prediction and bill cleavage/survival), we observed up to 4.3 points drop in the macro F1. This indicated interconnecting states through common nodes (e.g., stakeholders, and money donors) leads to better contextualized textual+graph embedding. In addition, in another ablation study, we measured the effect of different relation types and

textual attributes of nodes in the legislative graph. For example, our evaluation showed the donors’ information (and relations with legislators) improves the F1 score by at least 2.1 points across different tasks (for graph models listed in Table 6).

### A.3 Additional Results: State and Time-based Data Splits

In Section 6, we evaluated all models using a random split based on bill nodes. Here, We further evaluate the best model in each category (from Table 6, Table 7) with two different data splits: (1) Time-based where test bills are selected from 20% of most recently introduced bills. (2) State-based where we choose the test bills from one specific state and train bills from the other two states. In Table 9, we look at the winners/losers prediction task and the performance of the best model in each category (i.e., naive, text-based, graph-based that we discussed in Section 6). Similarly, in Table 10, we study the best models (from Table 7) for classifying party-based competitive bills.

Overall, we make multiple observations. First, our results of the time- and state-based splits still show that our pretrained textual+graph model outperforms other models in both of the tasks (i.e., winners/losers prediction, and voting cleavage classification). Second, we can see a rather sharp drop across all the models in terms of F1 score for these

Type	Models		Pass/ Fail	Competitive				Inverse-Competitive			
	Winner/Loser	Embedding		Party	Gender	Ideology	Geography	Party	Gender	Ideology	Geography
Naive	Unknown	Majority	47.2	43.1	47.4	42.6	48.1	48.3	45.7	46.2	45.8
		Sponsor	50.6	52.3	55.7	51.1	48.7	53.4	45.2	46.7	53.2
	Known	WL Correlation	58.7	51.3	50.6	50.4	50.1	54.8	52.6	55.3	52.2
Text-based	Unknown	BoW	48.1	56.8	57.2	57.7	48.3	48.6	49.1	46.3	52.9
		Pretrained RoBERTa	49.8	58.2	58.8	59.1	49.7	49.3	50.7	46.9	56.2
	Known	BoW	49.4	61.2	59.4	60.2	51.7	50.1	53.2	47.1	56.5
		Pretrained RoBERTa	51.9	63.6	60.2	61.8	52.9	51.8	54.7	48.3	58.7
	Predicted	BoW	48.3	59.1	58.8	59.5	50.2	49.3	52.4	46.9	53.6
		Pretrained RoBERTa	50.1	61.4	59.7	61.3	51.3	49.8	52.9	47.1	57.8
Graph-based	Unknown	DeepWalk	59.9	49.1	50.5	47.2	53.7	49.8	59.2	48.7	56.2
		GCN	60.8	50.4	49.7	48.3	54.2	48.2	60.8	49.8	55.6
		RGCN	64.3	52.8	53.1	50.6	55.4	49.8	62.3	51.3	59.4
	Known	DeepWalk	62.2	49.6	52.3	49.8	54.5	50.3	60.7	51.3	58.6
		GCN	63.6	52.2	51.4	50.8	55.3	50.3	62.1	50.4	57.7
		RGCN	66.3	53.7	55.2	51.9	56.5	52.4	64.2	53.7	60.1
	Predicted	DeepWalk	61.3	47.7	52.2	47.5	54.3	50.1	59.6	50.8	57.1
		GCN	62.3	51.2	50.8	48.5	54.3	49.5	61.1	50.1	56.7
		RGCN	65.6	52.3	54.1	51.4	56.2	52.1	62.3	52.6	59.2
Our Joint Text + Graph	Unknown	Pretrained	69.8	69.9	63.7	60.9	61.6	59.1	65.8	60.0	60.7
	Known	RoBERTa	<b>73.7</b>	<b>71.8</b>	<b>67.9</b>	<b>63.5</b>	<b>65.1</b>	<b>59.8</b>	<b>66.2</b>	<b>62.5</b>	<b>63.9</b>
	Predicted	+ RGCN	70.7	71.6	67.2	62.6	63.8	59.5	65.9	60.6	62.2

**Table 11:** Effect of winners/losers information on the graph and text-based models in different vote classification tasks. Extending the results based on the random split of bills in Table 7.

two new splits. The reason is that time-based and state-based splits of bills lead to more unseen nodes (e.g., legislators, money donors), challenging graph models more than text models. For the time-based data split, the performance degradation is slightly less as the number of unseen nodes was fewer in the test dataset. Third, when we use Oregon for testing, we observe there is a higher drop in the performance of models, compared to using Indiana for testing; One potential reason is that Wisconsin and Indiana tend to be Republican states, while OR has been a Democratic one. Forth, our graph abstraction and stakeholders analysis (relations) help even baseline models to better decode state policies, when we compare the performance of models in the bill cleavage/survival tasks, with *Unknown*, *Known*, *Predicted* winners and losers in the legislative graph.

#### A.4 Measuring Disagreement and Labeling Competitive Roll-Calls

As discussed in Section 5, roll-call votes occur when a state-level legislator votes “yea” or “nay” on a bill. In Sections 1 and 5, we defined two types of bill classification tasks to characterize voting cleavages or disagreement across and within different ideological and demographic groups of legislators. Here, we discuss how we measure the disagreement and label bills in each of these tasks: **(1) Inverse-Competitive bills:** Consider a bill where 55% of Men voting Yea and 45% of them voting Nay. We define the disagreement as the percentage of minority votes or 45%. When

the disagreement within a group of legislators (e.g, men) is between 40-49%, we consider the bill as an inverse-competitive bill with a partial tie in votes. The disagreement of 50% is a complete tie. **(2) Competitive bills:** Next, consider a bill with 60% of Women voting Yea and 80% Men voting Nay on a roll call. This bill is competitive because the majority of female legislators voted differently than the majority of male legislators (the cross-group disagreement is 20% = 80%-60% in this case.) Conceptually, we do not need to compute the cross-group disagreement to identify competitive bills.