An Automated Approach to Measuring Competitive Issue Framing at Scale

Vincent Hopkins
Mark Pickup
Simon Fraser University

Motivation
Political groups frame issues to build support for their policy agenda. Issue framing helps unite diverse coalitions, but it may also invite competition from political opponents. Identifying political “discourse coalitions” – communities of allied interest groups that invoke similar semantic concepts during legislative debate (Hajer 1993) – is valuable but has proven not easily amenable to large-scale quantification. It involves identifying discursive concepts (“frames”), associating each frame with a political actor, and measuring the shared frames and other ties connecting those actors. The biggest scalability hurdle is reliance on human coders. The sheer amount of political debate – thousands of utterances between hundreds of actors over days, months or years – seems to demand a selection of a subset of the data for a small-N case study analysis. Similarly, the complexity of debate would appear to call for qualitative analysis. Indeed, the study of political discourse is perhaps best associated with the qualitative and postmodern traditions of social science (e.g. Hajer 1993, 1995; Schmidt 2008; Fischer 2012).

Here, we propose a scalable, automated approach to identify discourse coalitions and measure the degree to which they compete over framing. Our approach builds off Leifeld and Haunss (2012), and combines web-scraping, topic modeling, and network analysis. First, we collect legislative committee debates and use unidimensional scaling to identify a political actor’s support (or opposition) for a given legislative proposal. Second, we use topic modeling to associate each actor with one or more issue frames. Third, we use network analysis to model the extent to which competing discourse coalitions compete over issue framing during legislative debate. Empirically, we define such coalitions as like-minded interest groups that invoke similar frames when testifying before a parliamentary committee. When coalitions are cohesive, with overlapping frames within them but few overlapping frames between them, it suggests a highly competitive framing environment. The approach is useful because it provides a modular and scalable solution to the problem of measuring issue framing across time or political systems. We provide an empirical example using 131 pieces of legislation across 15 policy areas – over 1.09M utterances.

Issue Frame Competition
We begin by situating our contribution in the framing literature. Our project is about “frames in communication,” which Chong and Druckman (2007a) describe as “the words, images, phrases,
and presentation styles that a speaker (e.g., a politician, a media outlet) uses when relaying information about an issue or event to an audience” (100). We distinguish between *valence* frames and *issue* frames. Valence frames involves substantively similar and logically equivalent concepts. For example, a politician may frame the benefits of a housing subsidy by emphasizing the larger, annual benefit (e.g. “$1,200 a year”) rather than the smaller, monthly benefit (e.g. “$100 a month”). Issue frames juxtapose substantively different concepts. For example, a liberal group may describe a proposed oil pipeline using words related to climate change (e.g. “pollution”, “greenhouse gas”) while a conservative group may describe that same pipeline in relation to economic growth (e.g. “GDP”, “jobs”). In this paper, we focus on issue frames.

Issue framing matters because people’s judgement often depends on how a decision is framed. In the classic “Asian disease” experiment, Kahneman and Tversky (1981) showed a simple valence frame can change people’s risk tolerance. Considerable evidence of issue framing effects also exists, such as the “government spending” experiment of Sniderman and Theriault (2004). Framing effects are one of the best documented phenomena in behavioural science. Given this, one might be excused for thinking voters’ preferences change with each new frame they encounter. Fortunately, experimental evidence suggests framing is only successful when frames are unopposed in political discourse (Chong and Druckman 2007b; Druckman et al. 2010). When opposing groups are allowed to present coherent, compelling frames, they can cancel each other out.

Research on issue frame competition has shed tremendous light on the potential for elites to frame issues and drive the public agenda (e.g. Terkildsen et al. 1998; Glazier and Boydstun 2012), the role of narrative and messaging in holding together coalitions of interest groups (e.g. Stone 1997; Sabatier & Weible 2007), and the importance of social network analysis to quantitatively assess conflict between competing coalitions (e.g. Schneider 2013; Schneider 2017). Surprisingly, however, there is little research on the extent of competitive issue framing in the real world. This is likely because empirical tools are lacking. There is no scalable solution to the problem of measuring issue frame competition at scale.

In this article, we propose a scalable approach to (1) identify political groups that use similar issue frames; (2) distinguish such groups from their opponents; and, (3) measure the extent to which each “side” of an issue is able to coordinate on one or more issue frames. We focus on an important part of the frame building process – the moment a proposed bill moves beyond the “trial” stage and receives serious debate by a permanent legislative committee.

Our project draws inspiration from Leifeld and Haunss (2012), Leifeld (2013) and Schneider (2017), who propose a quantitative approach to identify “discourse coalitions” in political debate. A discourse coalition is a collection of interest groups that invoke similar semantic concepts during legislative debate (Hajer 1993). Coalition members may not agree on everything, but they uphold a common narrative or storyline in support of a shared policy

---

1 In contrast, a “frame in thought” refers to an individual’s perceptions, as in “an economic frame of mind” (Chong et al. 2013).
agenda. This narrative is ultimately comprised of issue frames – phrases, images, metaphors, etc.

Leifeld (2013) and Schneider (2017) propose a four-step approach. First, they manually code newspaper articles to identify relevant interest groups and assess their support for a given piece of legislation. Second, they use those same newspaper articles to identify relevant policy concepts, such as “economic growth and stability.” Third, they construct a bipartite graph with interest groups as one set of nodes and concepts as another. Fourth, they project the bipartite network to a one-mode network that connects groups with concepts. This is an “actor congruence network,” which the authors then use to classify interest groups into various discourse coalitions. A key dependent variable is weighted network density, both within and across each respective “side” of the issue.2

Our Case Study
Before we present our data and approach, we explain our choice of Canada as a case study.

First, experimental evidence confirms Canadian MPs are sensitive to framing effects. Sheffer et al. (2017) successfully replicate the findings of the Asian disease experiment with politicians in Canada, Belgium and Israel. This is an important finding because it suggests political knowledge and savvy may not inoculate people against cognitive errors. At the same time, the Asian disease experiment uses a non-competitive valence frame. It remains an open question whether competitive electoral forces and deliberative institutions may buffer against issue framing effects.

Second, Canada’s interest group system is an excellent case of what Lijphart (2012, 3) calls “free for all” pluralism. He argues majoritarian institutions encourage interest groups to compete for access to government, while consensus institutions (e.g. multi-party systems, balanced executive and legislative powers) lead to “coordinated” corporatism. Coordination discourages competition by offering certain groups a seat at the decision-making table, such as formal ties between social democratic parties and labor unions. In other words, we expect interest groups to offer vigorous debate to all parties when testifying before a Canadian parliamentary committee.

Third, there remains little research using large corpuses of committee data. Parliamentary speech is paving new ground for research on topics like political polarization (Peterson & Spirling 2017; Rheault & Cochrane 2019), issue salience (Curran et al. 2018), and the role of emotion in politics (Rheault et al. 2016). Most studies use plenary proceedings, such as Question Period or all-member debates, rather than committee proceedings. The importance of this depends on the question of interest; however, there is good reason to think plenary speech may be different from committee speech. Especially in contemporary majoritarian

2 Weighted density is the number of observed, weighted edges in a network divided by the total possible number of unique ties. Note that unlike unweighted density, this may exceed one.
parliaments, plenary speech by ordinary members may be subject to direction by party leaders. Plenary debates are highly scripted, involve prominent members of the government and opposition, and feature prominently in news coverage. By contrast, committee work is less scripted, mostly involves backbenchers and invited groups, and is largely ignored by media (Docherty 2005). This matters because partisans might be more willing to “break rank” with their respective party when opening their mouths in committee. Committees are also a place where legislators spend much of their time posing questions to citizen groups rather than speechifying. This gives opposition parties an opportunity to invite a diversity of organized interests, creating an environment where lively debate occurs on matters of national importance.

Our Data: 11 Years of Committee Debates
Our data comprise nearly every legislative committee utterance by Canadian parliamentarians and witnesses between January 2006 and June 2017. This includes just over 1.09M utterances – covering 15 committees across six parliaments and one change in governing party. We apply four filters to the full corpus. First, we select only those committee meetings during which a piece of legislation was the subject of debate. Second, we filter out procedural utterances, most of which are tagged by the House of Commons. Third, we remove committee members from the data, leaving just those organized groups that presented before committee. The result is 51,809 utterances for 1,172 unique groups discussing 131 bills. Finally, to facilitate topic modeling, we concatenate all utterances by each interest group during each committee debate, filtering out any concatenation shorter than 25 words.\(^3\) This yields 2,745 “documents”, which we turn into a series of sparse document term matrices.

Below, we plot basic information about the dataset. Figure 1 plots the percentage of all bills debated in each committee. About one in three are discussed in the Justice committee. This is more than the next two most popular committees, Finance and Human Resources, put together. Figure 2 plots the mean number of speakers for each legislative debate. On average, 25-35 speakers testify on each bill in our data. However, there is quite a bit of variation, especially for the Environment, Agriculture, Ethics, and Health committees. Budget bills are debated in the Finance committee, which invites the largest number of interest groups on average (just over 40). Figure 3 plots the mean number of words spoken by each interest group during a committee meeting. This shows quite a lengthy debate, with the average interest group using 2000-3000 words during an appearance.

---

\(^3\) This is necessary because some committee members and witnesses are silent, save for the occasional short interjection (e.g. “Would you repeat that, please?”).
Fig. 1

Figure 1: Percentage of Bills by Committee

Fig. 2

Figure 2: Mean # of Speakers for Each Bill, by Committee
Our Approach

Given the size of our corpus, it is impractical for human coders to read every committee transcript and identify discursive concepts. We use quantitative text analysis to automate the process. Figure 4 presents an overview of the process. To begin, we use unidimensional scaling to identify opposing sides of each legislative proposal. Next, we use an unsupervised topic modeling approach to extract substantive information from committee testimony and allocate each interest group to one or more topics – that is, issue frames. Finally, we create a bipartite network with interest groups and issue frames as nodes. We project this to a unipartite network, with interest groups as nodes and ties between them representing a shared frame. We then calculate sample statistics, such as density and nominal assortativity, as well as inferential statistics using Exponential Random Graph Models.

Running Example

We illustrate each step using a running example: Committee debate surrounding Bill C-36 (“Protection of Communities & Exploited Persons Act”) from the 1st session of the 41st parliament (below). This was a Conservative government bill to restrict sex work. It was also a highly contested bill. The governing Conservatives enjoyed a majority on the committee. Two opposition parties were present (New Democrats and Liberals) and together held a minority of
seats. The committee met several times in July 2014 to consider the bill. In total, 45 organized groups testified before the committee.

Pre-Processing
Before doing any analysis, we use the pretext package in R to assess the robustness of our textual data to pre-processing decisions (Denny & Spirling 2018). The package allows the user to test the sensitivity of their analysis to six decisions: using n-grams; stemming; removing stopwords, punctuation, or numbers; removing infrequent terms; and turning all tokens lowercase.

We show what this looks like using the corpus for Bill C-36. The scale is such that negative coefficients produce more “usual” outcomes while positive coefficients produce unusual outcomes. The results (Figure 5, below) suggest that using n-grams, stemming tokens, removing stopwords, and removing punctuation make for more “usual” conclusions. There is no evidence that other things, like removing numbers, has much of an effect. As a rule, we do not implement any decision with a positive and statistically significant coefficient. Otherwise, we do.

![Figure 5](image.png)

Fig. 5.

Step 1. Extract Policy Position

Denny and Spirling (2018) use a straightforward definition of “unusual” results: “how documents ‘move’ relative to one another when they apply some transformation to the [document term matrix]” (180). They calculate Cosine similarity and for each combination of pre-processing decisions (e.g. stemming + lowercase + etc.), rank each document (e.g. Document 1 is closer to Document 2 than it is to Document 3, etc.). They calculate a “preText” score, which measures the difference in rankings between combinations. The procedure uses linear regression to estimate the sensitivity of this score to a given pre-processing decision. An example of these mean estimates is given in Fig.5 (above).
Next, we use a unidimensional scaling algorithm, *Wordfish*, to determine each speaker’s position on each bill. *Wordfish* is an unsupervised scaling model that estimates unobserved (latent) ideological positions using a “bag-of-words” approach (Slapin & Proksch 2008). It is parametric, assuming each actor $i$ uses word $j$ as drawn from a Poisson distribution:

$$ y_{ij} = \text{Poisson}(\lambda_{ij}) $$
$$ \lambda_{ij} = \exp (\alpha_i + \psi_j + \tau_j \times \theta_i) $$

where $y_{ij}$ is the count of word $j$ for each document/speaker $i$, $\lambda$ is the mean and the variance of the underlying distribution, $\alpha_i$ represents fixed effects for each speaker, $\psi_j$ represents fixed effects for each word, $\tau_j$ is a word’s “discrimination” parameter (i.e. the importance of each word in discriminating between different speakers’ ideologies), and $\theta_i$ is the estimate of each speaker’s ideological position (Slapin & Proksch 2008, 709). Words with large absolute values of $\tau_j$ are helpful in distinguishing political actors with a given value of $\theta_i$. For example, words like “anthropogenic climate change” and “renewable energy” might have a large, positive $\tau_j$ values. These words might help distinguish right-leaning groups with a negative $\theta_i$ from left-leaning groups with a positive $\theta_i$.

*Wordfish* works best when the “documents” (in this case, speakers giving testimony) stick to the main topic of discussion – for example, addressing one piece of legislation rather than several all at once. Monte Carlo simulations also show the algorithm may falter when the number of documents is very small (Proksch & Slapin, 2009). We focus on the subset of committee debates for which a piece of legislation is on the published agenda and the number of witnesses is at least five. This leaves 131 bills, or about two-thirds of all bills debated in our sample.

We run *Wordfish* on each committee debate, generating $\theta_i$ estimates for interest groups giving testimony. We show what this looks like using Bill C-36 (Figure 6). Blue nodes are groups with positive, statistically significant $\theta_i$ values. Red nodes are groups with negative, statistically significant values. Grey nodes are groups with $\theta_i$ values statistically indistinguishable from zero.
As expected, groups that oppose legalized sex work are generally on one end of the scale, with higher $\theta_1$ values. This includes Mothers Against Trafficking Humans, Sextrade 101, and policing organizations. At the other end of the scale, groups that support legalized sex work receive lower $\theta_1$ values. 5 This includes Prostitutes of Ottawa-Gatineau, Canadian Alliance for Sex Work Law Reform, and the British Columbia Civil Liberties Association. Wordfish provides bootstrapped standard errors, meaning the $\theta_1$ estimate for some groups (e.g. the Canadian Association of Elizabeth Fry Societies) has a 95% confidence interval indistinguishable from zero. We code these groups as “0”. Otherwise, groups with a negative, statistically significant score are coded “-1” while those with a positive, statistically significant score are coded “1”. 6

5 Note that our Wordfish models are unidentified, in the sense that positive and negative values have no substantive meaning. An identified Wordfish model generates the same absolute $\theta_1$ values with the same rank ordering. This is fine for our purposes, as we only wish to find out which groups are on opposing sides in legislative debate rather than support or oppose the specific bill.

6 As it happens, a previous study considered this same committee debate in 2017. Johnson et al. (2017) hand-code each interest group’s support or opposition to Bill C-36. We compared their coding to ours using Cronbach’s Alpha. The reliability coefficient is 0.87. Out of the 44 groups in the two samples, there were six discrepancies. First, we coded three groups as “neutral” (a category that does not exist in the Johnson et al. (2017) schema). Second, we allocated three groups (Ratanak International, the Asian Women Coalition Ending Prostitution, and the Evangelical Fellowship of Canada) to the “wrong” category.
We then move on to frame representation. We use a topic model. The first step is to select the number of \( k \) topics (i.e. issue frames). There are several ways to determine \( k \). For example, one might measure perplexity across different values of \( k \) or visually inspect another metric across different values to identify an “elbow,” or inflection point in a curve. One might also minimize (or maximize) an LDA-specific criterion, in the manner of Arun et al. (2010) or Deveaud et al. (2014).\(^7\) Alternatively, one might opt to skip this step entirely and use something like Hierarchical Dirichlet Allocation. This determines the optimal number of topics as part of the topic modeling process, thereby cutting out the need for an initial clustering algorithm.

For computational ease, we use a fast, unsupervised clustering algorithm to select \( k \): affinity propagation. This works by iteratively assigning each cell in a matrix to its nearest data center (called an “exemplar”). The algorithm defines the optimal number of clusters as that which minimizes the distance between all points. Affinity propagation requires a similarity matrix. Following its original use case (Frey & Dueck 2007), we use negative squared Euclidean distance. Below, we present an illustrative example using five tokens and five groups for Bill C-36.

<table>
<thead>
<tr>
<th>Walk With Me Canada Victim Services</th>
<th>section</th>
<th>charter</th>
<th>right</th>
<th>sex</th>
<th>worker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Canadian Alliance for Sex Work Law Reform</td>
<td>10</td>
<td>0</td>
<td>14</td>
<td>32</td>
<td>6</td>
</tr>
<tr>
<td>Criminal Lawyers Association</td>
<td>11</td>
<td>4</td>
<td>4</td>
<td>42</td>
<td>32</td>
</tr>
<tr>
<td>Government of Manitoba</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>Evangelical Fellowship of Canada</td>
<td>1</td>
<td>0</td>
<td>7</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1 Document Term Matrix for Sample from Bill C-36

First, we calculate negative squared Euclidean distance for two groups: the Canadian Alliance for Sex Work Law Reform and the Evangelical Fellowship of Canada. The first group used the word “sex” 42 times, while the second group used the same word just 6 times. We subtract these values and square the difference (i.e. \((42 - 6)^2 = 36^2 = 1296\)). We do this for each of \( N \) word pairs, sum them together, and multiply by \(-1\) such that

\[
d^2 = -\Sigma_j^N(p_j - q_j)^2
\]

where \( d^2 \) is negative squared Euclidean distance, \( p_j \) represents the number of times the Canadian Alliance for Sex Work Law Reform uses word \( j \), and \( q_j \) is the number of times the Evangelical Fellowship of Canada uses that same word. In this case, \( d^2 = -2353 \). Below, we show a heatmap for the example in Table 1 (Figure 7). Darker colors represent smaller values – that is, less distance between these two groups. The results suggest the Evangelical Fellowship of Canada and the Canadian Alliance for Sex Work Law Reform are the “farthest” from one another \((d^2 = -2353)\), while the Evangelical Fellowship of Canada and the Government of Manitoba are the closest \((d^2 = -183)\).

\(^7\) These tools are now easily available in R via the ldatuning package.
Using this matrix, affinity propagation identifies the number of clusters that minimize distance. Using the full corpus for Bill C-36, affinity propagation suggests 11 clusters.

Selecting the number of topics is akin to climbing down what Sartori (1970, 1040) describes as the “ladder of abstraction.” Increasing the number of topics maximizes a concept’s dimensions (i.e. intension/connotation) but minimizes the class of cases to which it applies (i.e. extension/denotation) (Sartori, 1970, 1040-1044). In our experience, affinity propagation is a “conservative” algorithm that suggests a small number of clusters (< 10). Other clustering algorithms, such as k-means, or topic-model specific procedures (e.g. picking the number of topics based on perplexity) tend to suggest far more topics. We use affinity propagation because it has some nice properties – it is fast and appears to suggest a “reasonable” number of topics – but we invite other ideas for this important step.

After selecting the number of issue frames evoked during committee debate, we associate each interest group with each frame using a Latent Dirichlet Allocation (LDA) topic model. This consists of categorizing each “bag-of-words” according to an estimated statistical distribution of underlying themes (Grimmer and Stewart, 2013, 18-19). Depending on the technique, topic models assume each document contains multiple topics, and permit a more fine-grained categorization than topic dictionaries. There are a variety of topic modeling techniques, such as the expressed agenda model (Grimmer, 2010), the dynamic multitopic model (Quinn et al., 2010) and the Latent Dirichlet Allocation (LDA) model (Blei et al. 2003).

We use the simplest topic model, LDA. It estimates how different political actors talk about different issues. It first examines correlations of words across documents, then groups words

---

8 We recognize the importance of other techniques, such as the Structural Topic Model or the Correlated Topic Model, and encourage the user to explore other topic modeling techniques to improve on the approach.
into thematic categories. It assumes each document contains a mixture of topics, which may be represented as a proportion.

Formally, Blei et al. (2003) define a “topic” as “a distribution over a fixed vocabulary” (78). This is useful for our purposes because it is consistent with the concept of an issue frame. For example, when left-leaning groups speak about a proposed carbon tax, they may frame the issue in terms of climate change. This would constitute a topic, the fixed vocabulary of which might yield a high probability of words like “greenhouse gas,” “emissions,” and “polluters.” Conversely, right-leaning groups might frame the same bill in terms of economic impacts. This also constitutes a topic, the fixed vocabulary of which could yield a high probability of words like “jobs,” “GDP,” and “tax burden.”

In this paper, we treat topics as issue frames.

Our LDA model infers a posterior, conditional probability that a given interest group $i$ invokes issue frame $k$ based on four things (Blei 2012, 80-81):

$$p(\beta_k, \gamma_i, z_j | w_i)$$

The first three terms are unobserved: $\beta_k$ is the distribution of words in each topic; $\gamma_i$ is the probability a given document is “about” issue frame $k$; and, $z_j$ is the corresponding issue frame for each word $j$. The remaining term is the only one observed: $w_i$ represents the actual words used in document $i$. As Blei (2012, 81) describes it, the conditional probability can be expressed as the joint probability distribution divided by $w_i$:

$$\frac{p(\beta_k, \gamma_i, z_j, w_i)}{p(w_i)}$$

The model proceeds by assigning each word in document $i$ to one of $k$ issue frames. The model assumes most words do not distinguish between frames. As such, it adopts a Dirichlet distribution with a sparse $\alpha$ parameter. This yields an initial set of parameter estimates based on the posterior. We use Gibbs sampling to iteratively improve and estimate the posterior distribution (with 2000 iterations and 1000 burn-in samples).

The final model returns two sets of probability estimates: one for the probability that each document that is “about” issue frame $k$ ($\hat{\gamma}_i$) as well as the probability that a given word comes from that frame ($\hat{\beta}_k$). For our purposes, $\hat{\gamma}_i$ is an estimate of the probability that a given interest group $i$ invokes issue frame $k$.

---

9 The researcher is responsible for setting the value of $k$. As discussed, we use affinity propagation for this step.
We implement an LDA model using our running example of Bill C-36. In Figure 8a, we show the most distinguishing terms for each of 11 issue frames during Bill C-36. For example, the top tokens in frame four include “tool,” “offenc,” and “law_enforc” among others. These terms distinguish frame four from the other 10 frames. Some frames are more distinguishing than others. This is evident in Figure 8b. It shows, for example, that frames six and seven are really only used by one group. This is an extreme case.

(a) Top 5 Tokens for Each Issue Frame, Bill C-36

(b) Topic Probabilities for Each Group, Bill C-36

Fig. 8
In Figure 9, we compare the list of groups that were more likely to use frames two and eight. We also color each group by its Wordfish coding. Whereas frame two emphasizes professional dimensions of sex work, frame five emphasizes gendered dimensions of sex trafficking (e.g. using tokens such as “survivor” and “women_girl”). Bill C-36 was controversial, so it is not surprising to see such sharp division in the use of these frames. Groups most strongly associated with frame two include several sex workers’ groups (e.g. Prostitutes of Ottawa-Gatineau) as well as groups that support the legal rights of sex workers (e.g. Canadian Alliance for Sex Work Law Reform). In contrast, groups most strongly associated with frame eight include anti-human trafficking groups (e.g. Mothers Against Trafficking Humans), and those explicitly opposed to sex work (e.g. Sextrade 101).

Finally, we turn the LDA results into network data. For each interest group, the model assigns a probability that the group’s testimony relates to one of $k$ issue frames. Each interest group may invoke multiple discursive concepts during debate. Accordingly, we assign topics with to each group if their $\hat{p}_i$ estimate is above a baseline threshold (i.e. greater than $1/k$ topics). This yields a column of groups and a column of frames.\textsuperscript{10} We represent this as an affiliation network, with interest groups as one set and frames as another. We then follow Leifeld and Haunss (2012), projecting the bipartite network to a one-mode “actor congruence network” that connects groups with frames.

In Figure 10, we plot the actor congruence network for Bill C-36. Ties represent an issue frame (i.e. LDA topic category). A tie between each node means both groups used the same frame during their committee testimony. Ties are weighted, such that darker ties represent multiple connections. For example, we highlight two red nodes for the Prostitutes of Ottawa-Gatineau group and the Canadian Alliance for Sex Work Law Reform. We also show illustrative tokens of one issue frame connecting them: “client,” “sex_work,” and “advertis”. We also highlight two

\textsuperscript{10} There are three bills where $k = 1$. After removing these as well as instances where the Wordfish algorithm fails to converge, the total number of bills drops from 151 to 131.
blue nodes for the York Regional Police and the Canadian Women’s Foundation. We show representative tokens from one issue frame connecting them: “survivor,” “women_girl,” and “sex_traffick”.

The figure shows strong evidence of competitive issue framing. There are many more ties within each colored node set than between them.

**Actor Congruence Network for Bill C-36**

The intuition is straightforward: When coalitions are cohesive with overlapping frames within them but few between, it suggests a highly competitive framing environment. As blue nodes become more compact and move away from the red nodes, it means the opposing groups use very different issue frames when talking about the same bill. As the two clouds merge into one, it means opposing groups invoke similar frames. The weighted density for the blue nodes is 0.77. Weighted density for red nodes is 0.95.\(^{11}\) This is a preliminary indication of competitive issue framing, as it suggests the blue and red nodes are, respectively, more tightly connected to one another than to the rest of the network.

The dynamic is closely related to the network concept of homophily, according to which individuals who share one attribute are more likely to share ties than those who do not (McPherson 2001, 416).\(^ {12}\) Strong homophily means groups who are on the same side of an issue are more likely to invoke the same issue frame, relative to groups on opposing sides.

We calculate two network homophily statistics. First, we calculate nominal assortativity. A positive assortativity coefficient means nodes that share an attribute (in this case, positive vs. negative \(\theta_i\) values) share many connections. For Bill C-36, the assortativity coefficient is moderate: 0.28 on a -1 to +1 scale. Second, we estimate an Exponential Random Graph Count Model (ERGCM) to estimate the number of shared connections between groups. This lets us control for global and local network properties. We estimate an ERGCM model, controlling for

---

\(^{11}\) Unweighted density for the full network is 0.46, for the blue nodes is 0.60, and for the red nodes is 0.76.

\(^{12}\) For simplicity, we measure “uniform homophily” assuming each side of an issue has the same propensity for within-group ties (Krivitsky 2015).
some individual heterogeneity. The results (Table 2) suggest competitive issue framing in the debate over Bill C-36. The odds of two opposing interest groups invoking the same issue frame are $\exp(\beta_1) = \exp(-0.66) = 0.51$. The odds of two groups on the same side invoking the same frame are ($\exp(\beta_2) = \exp(1.49)$) = 4.43 times higher.

Table 2. ERGCM of the Actor Congruence Network for Bill C-36

<table>
<thead>
<tr>
<th></th>
<th>Log Odds (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum (intercept)</td>
<td>-0.66* (0.11)</td>
</tr>
<tr>
<td>Individual Heterogeneity</td>
<td>3.45* (0.26)</td>
</tr>
<tr>
<td>Wordfish Homophily</td>
<td>1.49* (0.16)</td>
</tr>
<tr>
<td>N speakers</td>
<td>45</td>
</tr>
<tr>
<td>N ties (unique ties)</td>
<td>529</td>
</tr>
<tr>
<td>N ties (sum of weighted edges)</td>
<td>745</td>
</tr>
<tr>
<td>AIC</td>
<td>-252.4</td>
</tr>
<tr>
<td>BIC</td>
<td>-237.7</td>
</tr>
</tbody>
</table>

$\text{P}^* < 0.05$

Next, we provide a validation check of our approach.

**Validation Using Null Models**

We begin by creating 100 “null” models with lorem ipsum text randomly distributed across each witness for Bill C-36. The idea is to create a corpus that, by design, has no competitive issue framing. We retain the same *Wordfish* estimates for each group as that estimated from the actual data.

In terms of the topic model, we should expect all speakers are equally likely to invoke each topic. The predicted probability that each group $i$ invokes topic $k$ should be equal to $1/k$. We can think of this in terms of bias:

$$Bias(\hat{p}_i) = E[\hat{p}_i] - \gamma_i$$

where $\hat{p}_i$ represents the predicted probability each group $i$ invokes topic $k$ and $\gamma_i$ is the known probability under the null model, equal to $1/k$ topics.

In terms of the network model, we would expect network density to be at or near 1, with assortativity and homophily coefficients around 0.

---

13 This is the covariance of the square roots of dyad values that fall around each actor (Krivitsky 2012, 2015).
Below, we present our measure of bias for each group across 100 simulations. The results are encouraging. For the topic model, we find the mean bias in topic probability is around zero for all 46 groups (below, Figure 11). In other words, there was no systematic variation in topic usage between groups. The standard errors are given as the standard deviation of $\text{Bias}(\hat{\gamma}_i)$ divided by the square root of the number of samples ($N=100$). There also appears to be little relationship between the length of each utterance and the magnitude of the bias (Figure 12).

![Bias in Gamma Probability for a Null Model of Bill C-36, by Group](image1)

**Fig. 11**

![Bias in Gamma Probability for a Null Model of Bill C-36, by Word Count of Each Utterance](image2)

**Fig. 12**
To give a visual sense of what a committee debate looks like without competitive issue framing, Figure 13 shows a bipartite projection from one of the null models. It shows a well-connected network with no appearance of homophily. Network density equals 1. Nominal assortativity equal -0.017. The ERGCM homophily coefficient is 0.000 with undefined standard errors, likely owing to convergence difficulties given the large number of ties between nodes.

Results

Below, we present the results of our approach across the full sample of 131 bills. We focus on the ERGCM results, as they provide a measure of uncertainty and are correlated with nominal assortativity (Pearson’s correlation = 0.42).

Figure 15 (next page) plots ERGCM homophily coefficients (with 95% confidence interval) for each bill, ranked by magnitude. We also separate the results by number of speakers. Green estimates are statistically significant. Red are not. There is evidence of competitive issue framing in approximately 57% of legislative debates. In these cases, two interest groups on the same side of an issue were more likely to invoke the same frame than were two opposing interest groups. In the remaining 43%, there is no evidence of competitive issue framing. In these cases, opposing interest groups were equally likely to invoke the same frame as were interest groups on the same side of an issue.

14 The 0.05 significance level is used for hypothesis testing throughout.
ERGCM Coefficients by Number of Witnesses

- 5-10 witnesses
- 10-20 witnesses
- 20+ witnesses

Fig. 15
There also appears to be a strong relationship between the number of interest groups giving testimony and the degree of competitive issue framing. Among those legislative debates with fewer than 10 testifying groups, just 22% of bills show evidence of homophily.\textsuperscript{15} When the number of witnesses increases to 10-20, the proportion of bills exhibiting homophily increases to 68%. And roughly 86% of bills with more than 20 witnesses show evidence of competitive issue framing. We regard this as a “face validity” check. Interest groups are most successful when they mobilize on issues that resonate with voters (Klüver 2018; Hopkins, Pickup & Klüver 2018). Similarly, committee activity fluctuates with changes in issue salience (e.g. May, Sapotichne & Workman 2006; Burstein and Hirsh 2007). On this basis, we would expect to see more competition over issue framing as the number of interest group witnesses increases.

We also note there is no evidence of statistically significant negative homophily. This is also encouraging. Negative homophily would imply two opposing interest groups are more likely to share the same frame than are two groups on the same side of an issue. We can think of no theoretical reason to expect negative homophily and would regard such a result as strong evidence against our approach.

Across all statistically significant debates, we note that our running example of Bill C-36 is in the upper-mid-range of the pack. The smallest effect is $\exp(0.26, \text{SE}=0.11) = 1.30$. The largest is $\exp(4.25) = 70.11$. With a log-odds coefficient of 1.49, Bill C-36 lies around the 60\textsuperscript{th} percentile.

Next, we provide an intuitive sense of these results by discussing committee debate over two bills. The first exhibits competitive issue framing. The second does not. In the Appendix, we review two additional bills and plot the actor congruence networks for all four.

\textit{Bill C-69}

The Impact Assessment Act (Bill C-69) was a Liberal government proposal from February 2018. The bill sought to restructure the federal approval process for large natural resources projects. The Liberals are a moderate-centrist party, and Bill C-69 was part of a grand environmental compromise. On the one hand, the government would support – and even pay for – an oil pipeline in western Canada. On the other hand, Bill C-69 would tighten the regulatory system and improve consultation with affected parties, including indigenous communities. In the end, Bill C-69 was criticized on the left and the right. Left-leaning politicians argued the bill did not go far enough. One member of the New Democratic Party argued C-69 was evidence of the government’s hypocrisy: “climate leaders do not try to ram through massive bitumen pipelines” (Julian 2019). Conservative politicians claimed the bill would harm the economy by “killing Canadian innovation and killing Canadian jobs” (Stubbs 2018). Bill C-69 was debated at the Environment and Sustainable Development Committee, comprising seven members of the governing Liberal Party and four opposition members.

\textsuperscript{15} This may also have something to do with the relatively small size of these networks, which increases standard errors and makes it harder to reject the null hypothesis of no homophily.
During committee debate, 97 groups testified. Several indigenous groups appeared, including the Mikisew Cree First Nation and the Metis National Council. Many of these groups were supportive and pushed for specific amendments to improve the bill. In their testimony, some sought to frame the bill as a chance to address unresolved treaty rights. For example, Mr. Bill Namagoose of the Grand Council of the Crees (Eeyou Istchee) testified:

“The key message of our submission today is that Bill C-69 should provide for a carve-out, or distinct regime, to address specifically the [James Bay and Northern Quebec Agreement, JBNQA] territory. In so doing, Bill C-69 must guarantee the treaty rights of the Cree of the Eeyou Istchee under the JBNQA, as recognized in the Moses decision [ed. note: a 2010 Supreme Court case], to be active and mandatory participants in the assessment of development projects in Eeyou Istchee carried out under federal legislation.” (Canada 2018a).

Environmental groups also testified. These groups often used different language than indigenous groups. For example, Mr. Joshua Ginsberg of Ecojustice Canada framed the bill in terms of the broader relationship between environmental impacts and liberal progressivism:

“The bill also does not recognize that in Canada, vulnerable populations such as low-income populations, indigenous communities, and socially marginalized groups are disproportionately exposed to environmental hazards while also disproportionately lacking access to environmental benefits. In other words, environmental approvals often lack environmental justice.” (Canada 2018a)

Business associations representing the extractive industries also testified. These groups were usually opposed to the bill. Rather than stress specific amendments, many framed the bill as a drag on the country’s economic wellbeing. For example, Mr. Chris Bloomer of the Canadian Energy Pipeline Association framed the bill as a direct – possibly even intentional – threat to the oil and gas industry:

“With built-in climate change tests covering upstream and downstream emissions, it is preposterous to expect that a pipeline proponent would spend upwards of a billion dollars only to be denied approval at the end, because the project must account for emissions from production of the product to consumption in another part of the world. If the goal is to curtail oil and gas production and to have no more pipelines built, this legislation may have hit the mark.” (Canada 2018b)

In the final vote, Bill C-69 passed the House of Commons on sharp party lines. Members of the Liberal Party voted in favour, while the New Democrats, Conservatives, and Bloc Québécois voted against.

The ERGCM results suggest Bill C-69 was a highly competitive issue framing environment. The odds of two opposing interest groups invoking the same issue frame – such as Ecojustice Canada and the Canadian Energy Pipeline Association – are \( \exp(\beta_1) = \exp(-0.60, \text{SE}=0.08) = 0.38 \). The odds of two groups on the same side invoking the same frame are \( \exp(\beta_2) = \exp(0.83, \text{SE}=0.12) \) = 2.30 times higher. Both coefficients are statistically significant.
Bill C-41
The Canada-Korea Economic Growth and Prosperity Act (Bill C-41) was a Conservative government proposal from September 2014. The bill was a free trade proposal. It dropped tariffs with South Korea and improved bilateral relations. In the House of Commons, C-41 was so popular that opposition parties competed to shower it with the most praise. One member of the New Democratic Party stated:

“It is my great pleasure and honour to support this bill and this free trade agreement, the crux of which is tariff lines between Canada and South Korea. The NDP believes that this free trade agreement will benefit Canadian industries and that it can produce plenty of positive economic spinoffs for Canadian industries, such as aerospace.” (Liu 2014)

One Liberal legislator felt the need to clarify that his party was even more supportive of the bill, and had been so long before the New Democrats:

“The Liberal Party has consistently, from Korea's initial interest back in 2003, wanted to see a free trade agreement. We have supported the bill in second reading. To that extent, I think it is noteworthy to recognize that the NDP has taken a different road, a road to support free trade agreements. This is something that is very new here in Ottawa. It is a new policy shift for the New Democrats.” (Lamoureux 2014)

Bill C-41 was debated at the International Trade Committee, comprising six members of the governing Conservative Party and four opposition members.

During committee debate, 19 groups testified. Several trade and industry groups appeared, including the Canadian Pork Council and the Business Council of Canada. Many of these groups were supportive of Bill C-41. In their testimony, some framed the bill as an important solution to the problem of decreased Canadian exports to South Korea following its trade agreements with the US and EU. For example, Ms. Claire Citeau of the Canadian Agri-Food Trade Alliance testified:

“It is essential that the Canada-Korea free trade agreement be ratified and implemented by January 1, 2015. South Korea is a lucrative market of 50 million consumers, and a key hub of Asian supply chains. South Korea imports over 70% of its food, and until a few years ago, Canada was a preferred supplier for many agrifood products.” (Canada 2014a)

Labour unions also testified, including the United Food and Commercial Workers Union. These groups were more critical of Bill C-41. Yet some of them used similar issue framing as the business and export groups. For example, Mr. Jim Stanford of Unifor (CBA) also framed the bill in terms of increased demand for Canadian exports. He acknowledged the potential benefits for the meat industry, but also suggested it would increase the trade deficit:

“I acknowledge that some sectors will win, including the export of meat and meat products. Every free trade agreement has winners and losers. The task for policy-makers is to make sure the net impact is going to be positive, and I cannot foresee—perhaps we can talk about this
more in the questions—any scenario in which the increase in meat through pork and beef exports to Korea offsets anything but a tiny fraction of the growth in the bilateral trade deficit, which is clearly going to occur under this deal.” (Canada 2014b)

The debate over Bill C-41 shows an important feature of our approach. A non-significant homophily coefficient does not mean every group agrees on one single issue frame. Rather, it means the degree of competition in the issue framing environment is weak. In this case, all 19 groups invoked the trade-export-benefit frame while just two proposed an opposing frame: trade deficits.

The ERGCM results suggest Bill C-41 was not a competitive issue framing environment. The odds of two opposing interest groups invoking the same issue frame are \( \exp(\beta_1) = \exp(-0.45, SE=0.26) = 0.64 \). The odds of two groups on the same side invoking the same frame are roughly the same (\( \exp(\beta_2) = \exp(0.31, SE=0.46) = 1.37, P\text{-value} > 0.05 \)).

In the final vote, Bill C-41 passed the House of Commons with full support from the Conservative, New Democratic, Bloc Québécois, and Liberal parties.

**Conclusion**

In this paper, we propose a scalable, automated approach to identify discourse coalitions and measure the degree to which they compete over framing. Our approach builds off Leifeld and Haunss (2012) and Leifeld (2017). It combines web-scraping, topic modeling, and network analysis. We believe a virtue of this procedure is that it is modular – that is, any single component may be improved upon and it only strengthens the final product. For example, an improved clustering algorithm at the front end will yield more accurate topic model predictions. Similarly, an improved frame representation strategy will increase the accuracy of the network modeling component. We are not wedded to any single component presented in this paper. Quite the contrary; we invite improvements that remove unnecessary steps or replace them with better tools.

More robustness checks are needed. For example, we would like to try different frame representation tools, such as a Hierarchical Dirichlet Allocation process, or the new “Anchored Correlation Explanation” topic model (Gallagher et al. 2017), designed explicitly to extract frames from news media. We would also like human coders to validate the results of our ERGCM results. This would give us greater confidence that our procedure is actually measuring what we think it is.

One possible limitation to our procedure is its focus on overlapping frames, not overlapping words. For example, one might create a symmetric \( N \times N \) matrix with cells representing some other token-based measure (e.g. number of shared n-grams, cosine similarity). This may yield something similar to our results – after all, our topic models account for the co-occurrence of tokens in document term matrices – but the results might well be different.
Finally, more thought should be given to the role of politicians. As mentioned earlier, committees are a place where elected officials ask questions of witnesses. By including politicians, we might create difficulties for our procedure. Question wording is naturally different from answer giving. One danger is that Wordfish simply picks up on the back-and-forth nature of committee debate rather than estimating true policy positions. Still, this seems worth doing. Committee debates are less partisan than plenary proceedings. Using this approach to study elected officials’ committee speech could yield new insight into political polarization. This would be particularly helpful in political systems where party leaders dominate not just legislative votes but possibly also plenary proceedings.
Bibliography


website:
s=3&DocId=4332869


Appendix: Additional Review of Competitive and Non-Competitive Bills

Bill C-24
The Strengthening Canadian Citizenship Act (Bill C-24) was a Conservative government proposal from June 2014. The bill was controversial. Section 10(2) gave the federal government the power to revoke Canadian citizenship from dual citizens convicted of terrorism, high treason or spying. In the House of Commons, opposition parties called the bill “unconstitutional” and criticized it for establishing “two-tiered citizenship”. Bill C-24 was debated at the Citizenship and Immigration Committee, which comprised six members of the governing Conservative Party and four opposition members.

During committee debate, 30 groups testified. Legal groups appeared, including the Canadian Association of Refugee Lawyers and the David Asper Centre for Constitutional Rights. These groups mostly opposed C-24, and sought to frame it in terms of potential violations of Canada’s Charter of Rights and Freedoms. For example, Ms. Barbara Jackson of the Canadian Bar Association testified:

“We think that [C-24] could raise serious human rights concerns. It does raise serious human rights concerns. It may well contravene the Charter. The Supreme Court of Canada has already ruled in the past that we can’t exile Canadians. By redefining who a Canadian is, you achieve exile. That’s not right. It’s against the Charter. It appears to be against the Charter, and I expect there will be significant litigation.”

Other groups supported the bill, including the Air India 182 Victims Families Association. These groups tended to use different language than groups who opposed it. For example, Ms. Maureen Basnicki of the Canadian Coalition Against Terror made no mention the legal dimensions of Bill C-24. Instead, she framed the bill in terms of Canada’s moral obligation to oppose terrorism:

“Most immigrants do adjust and become productive members of Canadian society, in actions, if not in spirit, accepting Canadian values. Terrorist acts are the exact antithesis of such values. Terrorists, in executing innocent people, denigrate and violate every tenet of the values that make up Canada. Therefore, if Canada allows a convicted terrorist to retain Canadian citizenship, Canada is in effect saying that we accept the terrorist act as part of the fabric of life in Canada.”

In the final vote, Bill C-24 passed the House of Commons on sharp party lines. Members of the Conservative Party voted in favour, while the Liberals, New Democrats, and Bloc Québécois voted against.

Our ERGCM results suggest Bill C-24 was an intensely competitive issue framing environment. The odds of two opposing interest groups invoking the same issue frame are \( \exp(\beta_1) = \exp(-0.97, \text{SE}=0.20) = 0.38 \). The odds of two groups on the same side invoking the same frame are \( \exp(\beta_2) = \exp(1.49, \text{SE}=0.33) = 4.71 \) times higher. Both coefficients are statistically significant.
Bill C-32
The Victims Bill of Rights Act (Bill C-32) was a Conservative government proposal from April 2013. The bill sought to empower victims of crime, for example by giving them the right to information about their perpetrator’s pardon status. In the House of Commons, opposition parties spent much of the debate suggesting specific amendments to improve Bill C-24. Still, many were supportive. One member of the New Democratic Party who scrutinized the bill during committee noted the bill’s broad support among interest group representatives who appeared before the committee:

“I cannot say that the witnesses were on one side or the other. What mattered most to all of the witnesses was putting victims at the centre of the debate. I think that is the most positive thing that stood out about the victims bill of rights. That was the most common remark I heard.”

Bill C-32 was debated at the Justice and Human Rights Committee, which comprised six members of the governing Conservative Party and four opposition members.

During committee debate, 67 groups testified. Several victims groups appeared, including Canadian Parents of Murdered Children and Survivors of Homicide Victims and l’Association Québécoise Plaidoyer-Victimes. Many of these groups were supportive of Bill C-32. In their testimony, some sought to frame the bill as a measured improvement to the criminal justice system. For example, Mr. Joseph Wamback of the Canadian Crime Victim Foundation testified:

“The humane treatment of victims is of absolute, paramount importance to any civilized society. I have been working on and waiting for this for 15 years, and I’m here today to congratulate this government for initiating Bill C-32 and for recognizing the importance of providing protection to Canadian crime victims. Today I am not going to deal with the minutiae of the bill, because I am so pleased that victims’ rights are being considered, and my focus is to recognize its fundamental importance in Canadian society.”

Similarly, Yvonne Lindfield of Canadian Parents of Murdered Children and Survivors of Homicide Victims emphasized the need to balance offender and victims rights, and appealed for cross party unity:

“the implementation of the Canadian victims bill of rights must over time—because it will take time—ensure that all parties operating within the criminal justice system shift their mindset to one of equality for both the offender and the victim. Bill C-32 will have a profound impact on how the criminal justice system, and other government departments and agencies, treat victims. I appeal to all political parties and all levels to work cooperatively to ensure its effective implementation.”

Several other types of groups also testified, including legal associations and indigenous organizations. Some of these groups opposed sections of Bill C-32. Yet many of them used similar language as the victims’ rights groups. For example, Mr. Joshua Ginsberg of the Canadian Bar Association (CBA) also framed the bill in terms of the need to balance victims and offenders’ rights, and suggested improvements to specific clauses:
“The CBA recognizes that an effective criminal justice system must balance the interests of victims of crime, the procedural rights of those accused of crimes, and the public interest in seeing the efficient administration of justice... On the whole, the CBA section believes that this bill is an important step forward, improving the way the criminal justice system responds to victims of crime; however, some of the proposed amendments fail to strike the appropriate balance, leading to fundamental unfairness and some inefficiency.”

In the final vote, Bill C-32 passed the House of Commons with full support from all parties.

The ERGCM results suggest Bill C-32 was not a competitive issue framing environment. The odds of two opposing interest groups invoking the same issue frame are \( \exp(\beta_1) = \exp(-0.67, SE=0.11) = 0.51 \). The odds of two groups on the same side invoking the same frame are roughly the same \( \exp(\beta_2) = \exp(0.27, SE=0.17) = 1.31, P\text{-value} > 0.05 \).

**Actor Congruence Networks and Competitive Issue Framing**

![Competitive Issue Framing](image1.png)

![Non-Competitive Issue Framing](image2.png)

Fig. 1A