

# Predicting Partisan Responsiveness: A Probabilistic Text Mining Time-Series Approach

Lenka Bustikova<sup>1</sup>, David S. Siroky<sup>2</sup>, Saud Alashri<sup>3</sup> and Sultan Alzahrani<sup>3</sup>

<sup>1</sup> Associate Professor, School of Politics and Global Studies, Arizona State University, Tempe, AZ, USA.

Email: [lenka.bustikova@asu.edu](mailto:lenka.bustikova@asu.edu)

<sup>2</sup> Associate Professor, School of Politics and Global Studies, Arizona State University, Tempe, AZ, USA.

Email: [david.siroky@asu.edu](mailto:david.siroky@asu.edu)

<sup>3</sup> Assistant Professor, King Abdulaziz City for Science and Technology, Riyadh, Saudi Arabia.

Email: [salashri@kacst.edu.sa](mailto:salashri@kacst.edu.sa), [szahrani@kacst.edu.sa](mailto:szahrani@kacst.edu.sa)

## Abstract

When do parties respond to their political rivals and when do they ignore them? This article presents a new computational framework to detect, analyze and predict partisan responsiveness by showing when parties on opposite poles of the political spectrum react to each other's agendas and thereby contribute to polarization. Once spikes in responsiveness are detected and categorized using latent Dirichlet allocation, we utilize the terms that comprise the topics, together with a gradient descent solver, to assess the classifier's predictive accuracy. Using 10,597 documents from the official websites of radical right and ethnic political parties in Slovakia (2004–2014), the analysis predicts which political issues will elicit partisan reactions, and which will be ignored, with an accuracy of 83% (F-measure) and outperforms both Random Forest and Naive Bayes classifiers. Subject matter experts validate the approach and interpret the results.

**Keywords:** forecasting, automated content analysis, natural language processing, probabilistic topic modeling, sparse learning, political parties, polarization

One of the most fundamental questions in party politics is whether, and if so when, parties react to other parties. Advances in machine learning and the availability of highly granular textual data make progress on this question possible in a manner than was not previously feasible on a large scale. This paper develops a computational system, building upon recent advances in natural language processing, to analyze partisan debate and responsiveness with a broad potential utility for studying the dynamics of political competition across different scales and contexts. Hotly debated issues span all spheres of human activity, but politics is perhaps the sphere most defined by contentious debates, and much of it is now fully documented online and available for textual analysis. Text mining tools enable researchers to engage in the systematic analysis of text as data in an unprecedented manner, and political scientists have often been at the forefront of developing

**Political Analysis (2020)**

vol. 28:47–64

DOI: 10.1017/pan.2019.18

**Published**

21 June 2019

**Corresponding authors**

Lenka Bustikova, Saud Alashri and Sultan Alzahrani

**Edited by**

R. Michael Alvarez

© The Author(s) 2019. Published by Cambridge University Press on behalf of the Society for Political Methodology.

*Author's note:* We thank Ben Ansell, David Art, Kai Arzheimer, Daniel Berliner, Anita Bodlos, Rebecca Cordell, Stefan Dahlberg, Hasan Davulcu, Pieter Dewilde, Valery Dzutsati, Michael Hechter, Sean Kates, Miki Kittilson, Will Moore, Andrea Pirro, Mark Ramirez, Christian Rauh, Seyedbabak Rezaeedyakenari, Martijn Schoonvelde, Gijs Schumacher, Sarah Shair-Rosenfield, Arthur Spirling, Scott Swagerty, Cameron Thies, Joshua Tucker, Carolyn Warner, Reed Wood, Thorin Wright and two anonymous reviewers for comments. Earlier versions of the paper were presented in Amsterdam at the EU-Engage Automated Text Analysis Conference, hosted by Gijs Schumacher and Martijn Schoonvelde, at the American Political Science Association Conference in 2015 and at the School of Politics and Global Studies Workshop. The project received seed funding from the Center for the Study of Religion and Conflict at ASU. We especially thank Carolyn Forbes for helping to initiate and sustain the project. Supplementary materials for this article are available on the *Political Analysis* website. For Dataverse replication materials, see Alashri *et al.* (2018).

and applying such methods to analyze large-scale data collections of political texts.<sup>1</sup> This paper contributes to these developments by proposing a new sequential computational pipeline to predict action–reaction party dynamics.

## 1 Partisan Responsiveness

Game theoretic approaches to the study of party competition have offered many predictions about how parties should react to the moves of other parties on a given policy dimension, yet the empirical validation of these insights is surprisingly inconclusive. “We know very little,” writes one scholar, “about whether parties respond to policy shifts of rival parties” (Adams 2012, p. 407). With few exceptions (Meguid 2008), most of the empirical studies that investigate partisan responsiveness focus on competition between mainstream parties or between the mainstream parties and smaller parties in their ideological families (Adams and Somer-Topcu 2009).

Do parties on *opposite* poles of the political spectrum react to each other’s agendas and thereby contribute to polarization? Political science as a discipline still knows relatively little about this fundamental issue and lacks tailored methods to analyze the dynamics of polarization originating from the interaction of political parties at the extreme poles of the political spectrum. In order to detect centrifugal tendencies in the party system, we focus directly on the most extreme poles of the party spectrum that drive polarization. The more common approach is to focus on interactions between parties that are ideologically related and spatially proximate (Katz and Mair 1995; Arzheimer and Carter 2009; Abou-Chadi and Krause 2018, cf. Bustikova 2014), whereas here we study parties that are ideologically and spatially opposite. Moreover, whereas responsiveness is typically examined from the perspective of the voter–party linkage (Klüver and Spoon 2016), we analyze responsiveness at the level of political parties, which allows us to directly address the centrifugality of the party system. Voter’s attitudes can contribute to centrifugality, but political parties and politicians are the primary drivers of polarization (Arceneaux and Johnson 2015; Tucker *et al.* 2018, p. 40).

Scholars have found that niche parties (radical right, ethnic, environmental and regionalist) are less responsive to the preferences of the general electorate and to other parties than are mainstream parties (Adams *et al.* 2006; Spoon 2011). Empirical models of party competition that focus on the interactions of ideological “friends”—a mainstream party and its spatially nearby niche party—often ignore movements induced by ideological “foes.” By focusing on ideological foes, on more granular temporal dynamics and on more disaggregated issues, we demonstrate that niche “extreme” parties do react to their political rivals but only on selected topics, which suggests that their reactions may be strategic.

Niche parties are important vehicles of political polarization (Sartori 1976; Ignazi 1992; Evans 2002; Meguid 2008). The dynamic of responsiveness between two rival niche party families, studied in this paper, can enhance our understanding of the dynamics of multiparty systems of polarized pluralism (Sartori 1976). In such party systems, electoral advantage stems from centrifugal competition. If polar opposites are responding to each other’s provocations, polarization ensues because niche parties pull away from the center and stretch the ideological

1 In some of the earliest studies, Abelson (1973) simulated political ideologies on machines and Carbonell (1978) introduced a system to interpret a text relating to a given ideology or political event. Extending these approaches, Salton (1990) proposed an automated document similarity measure to process large data collections in an automated fashion. Grimmer (2009) applied Bayesian Hierarchical Topic Modeling to identify political agendas expressed in the press releases from senators and Monroe *et al.* (2008) developed a probabilistic Bayesian model to identify features (words) that capture partisan dynamics and used it to analyze conflict between Republicans and Democrats in the US senate. More recently, Greene and Cross (2017) developed a new dynamic topic modeling method based on two layers of nonnegative matrix factorization and demonstrated that it can unveil new niche topics and associated vocabularies using a corpus of all English language legislative speeches in the European Parliament plenary. Theocharis *et al.* (2016) utilized automated text analysis and machine-learning methods on tweets from politicians to measure their level of civility prior to an election and to develop an actor-oriented theory of political dialog that is derived from the incentives that Twitter as a forum provides to its users.

spectrum toward its extremes. If niche parties choose to respond to their polar opponent, they can have a harmful impact on the ability of the party system to rally around the center. By weakening centripetal competition, these dynamics contribute to volatility, fragmentation and de-alignment, and thereby undermine the ability of institutionalized mainstream parties to achieve moderation.

Small, niche parties are often overlooked because they appear marginal at the macro-electoral level and their supporters are missed by surveys. However, an advantage of text mining is that it allows us to capture the dynamics of responsiveness among small parties that often play an outsized role in party system polarization due to their focus on single issues and ideological purity. In the empirical case analyzed in this paper that focuses on radical right and ethnic parties in Slovakia, all major datasets on political parties, cross-national datasets and public opinion surveys ignore the second, more extreme, radical right party (*Pospolitost'*), which has transformed the landscape of Slovak politics by moving discourse to the extreme with a combative, militant style of politics. Major datasets also ignore the equally important Hungarian splinter movement that formed a new Hungarian ethnic party (*Most-Híd*), a division that has had a significant impact on both radical right mobilization and nation-wide political outcomes. The approach utilized here captures new, small, ascending parties (and factions) that have contributed in important ways to public discourse and to political polarization.

Although ideological opposites seemingly compete on the same cultural dimension, they strategically highlight and suppress their reactions to some topics that their opponents raise. In the long run of an electoral cycle, the dynamics of counter-reactions can wash out, giving the false impression that niche parties are less responsive than mainstream parties, but this is at odds with the microdynamics highlighted in this analysis. The volatile nature of identity politics indicates that polarization is often driven by microbursts that can quickly escalate contestation and, subsequently, recede. Text mining allows us to disaggregate the identity dimension of party competition and, by looking at a multiplicity of topics, to identify with a high level of precision which topics elicit reactions and which are ignored.

This paper contributes to a growing literature using text mining to learn party positions from texts (e.g., treaties, legislative documents and speeches), typically represented as a numerical measure of distance. Building on ideas introduced in Monroe *et al.* (2008, pp. 376–82, 398–99) and Grimmer and Stewart (2013, pp. 3–5), as well as other works on meme diffusion and the temporal characteristics of cascades through social communities (Leskovec *et al.* 2009), we use probabilistic topic modeling features and subject matter expertise to develop and assess a novel end-to-end computational pipeline to predict partisan responsiveness (Alashri *et al.* 2015). Our main methodological contribution is to introduce a new system for detecting, analyzing and predicting partisan responsiveness between political rivals, which we believe has potentially broad application across a variety of contexts and at different levels of analysis.

We first establish that parties from different (antagonistic) party families respond to each other's actions and document the microdynamics of partisan responsiveness that occur within lengthy electoral cycles traditionally studied through the prism of party manifestos and expert surveys of party positions. We simultaneously detect topics that are ignored by the adversarial camps and, using country-specific knowledge, explain the strategic logic that leads party leadership to escalate selectively. We compare the SLEP (Sparse Learning with Effective Projection) classifier used in this paper (Liu *et al.* 2009a) to a Naive Bayes classifier and to a Random Forest classifier. We show that SLEP performs very favorably.<sup>2</sup> We also compare our latent Dirichlet allocation (LDA) approach to a vector space baseline model and a word embedding model. Based on the F-measure, LDA offers the best model.

<sup>2</sup> For an overview of Random Forests, see Breiman (2001) and Siroky (2009). For recent applications in political science to conflict, see Colaresi and Mahmood (2017).

The next section formally defines the research objective and presents a full view under the hood of the methodology, followed by a discussion of the predictive results and substantive findings.

## 2 Predicting Partisan Responsiveness

All adversarial dyads create contentious frames, but not all frames and topics that one political camp raises resonate with the other camp, and sometimes an increase in attention to a topic by one camp is ignored by their political rivals. Given these assumptions, and a set of documents from each side's websites, we first ask: do the documents form a spike around a topic during a particular time period? If so, we then ask whether proximate spikes from an opposing camp are related, and can we use this information to predict partisan responsiveness?

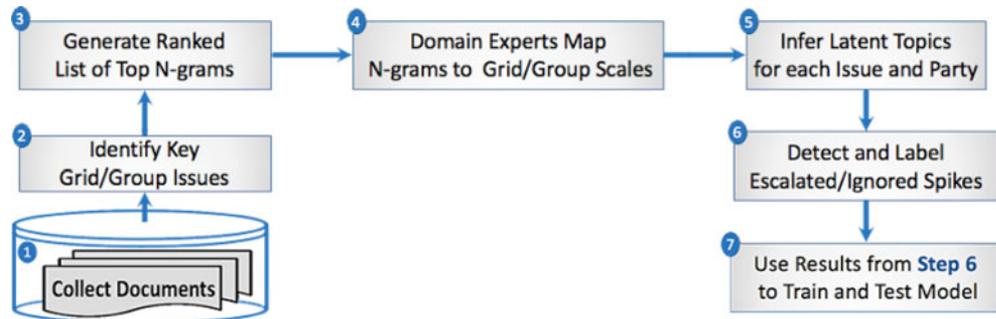
We focus on ideological adversaries: *radical right parties*, which advocate for the sovereign rule of the majority in "their" state, and *ethnic parties*, which stand for minority inclusion and are therefore often at odds with the radical right. In the context of this application to political parties, the proposed computational system is designed to conduct six sequential tasks (without loss of generality): (i) collect documents from political party websites and index them for fast retrieval and processing; (ii) identify key issues using theoretically derived scales and subject matter expertise; (iii) make issue-specific topic inference for each political camp; (iv) detect party-specific spikes that reflect increased attention to a specific issue and measure whether these spikes elicit a response from the opposing political camp (spike relatedness); (v) identify frames that discriminate between ignored/escalated spikes; (vi) train and test predictive models of partisan responsiveness. The result is a framework that can model how political discourse varies over time, detect topics that gain disproportionate attention from each camp and predict which topics solicit reactions from political rivals and which topics are ignored.

Once topical spikes from political opponents are detected and categorized using LDA (Blei *et al.* 2002), we exploit the terms that comprise the topics as features, together with a gradient descent approach known as SLEP (Liu *et al.* 2009a), in order to identify discriminative frames and to predict partisan responsiveness. To assess the framework's predictive accuracy, we use 10-fold cross-validation and 10,597 documents downloaded from the official websites of radical right and ethnic political parties in Slovakia, spanning a decade (2004–2014) of contentious politics. Using the F-measure, the classification accuracy for predicting partisan responsiveness (positive/escalated spikes) ranges from 80% to 89% and the classification accuracy for predicting the negative/ignored spikes (the lack of partisan responsiveness) ranges from 78% to 86%, depending on the topic and political camp. These results compare very favorably to experimentally tuned Naive Bayes and Random Forest classifiers.<sup>3</sup> Subject matter experts then validate and interpret the results.

Figure 1 offers a stylized overview of the system architecture. The numbers on the top left corner of each box represent the order in which these processes are executed. Each of the seven steps in the process is briefly described below, with additional details in the following sections.

- (1) Download documents, label by party, date and store in the database.
- (2) Identify grid/group issues using the theoretical framework.
- (3) Generate a ranked list of the top n-gram terms.
- (4) Subject matter experts map issues onto theoretically informed scales.
- (5) Infer latent topics for each issue and party or camp.
- (6) Detect spikes of documents and label them as escalated or ignored.
- (7) Use the results from step 6 as inputs for the prediction model.

<sup>3</sup> See Figures 1 and 2 in the Supplementary materials.



**Figure 1.** System Architecture.

In step 1, we wrote a set of scripts to download all documents from the websites of radical right parties and ethnic political parties in Slovakia from the beginning of the 2004 calendar year to March 16, 2014. Next, we preprocessed the data to extract text and article dates.<sup>4</sup> Then, we implemented the following methodology:

- Run a simple term frequency–inverse document frequency (TF–IDF) (Hartigan and Wong 1979) on the entire corpus to generate a large candidate list of terms (after removing stopwords) for inclusion. This measure identifies the importance of a word to a document based on its presence in a document (TF) and its rarity at the corpus level (IDF). Select the top  $T$   $n$ -gram terms (1–3 grams).
- Subject matter experts scan the list of  $n$ -grams ranked by frequency and select relevant keywords indicating hotly debated grid/group issues that capture views on group exclusion and state authority.
- For each issue:
  - Select the documents that mention the issue based on given keywords;
  - Run the Mallet algorithm (McCallum 2002) over each political party’s corpus to get their LDA’s *latent topics*, 100 topics each with 20 keywords (Blei *et al.* 2002);
  - Detect and label the spikes from each party as *escalated* or *ignored* by the other camp based on the three-sigma rule (Pukelsheim 1994);
  - Use latent topics and a feature selection algorithm to determine issue-specific discriminative *escalated* versus *ignored* frames;
  - Use discriminative frames to train a sparse-learning classifier (SLEP) to predict partisan responsiveness.

Next, we describe these steps in greater detail.

## 2.1 Text Processing and LDA Topic Inference for Each Camp

We selected parties that were on the opposing poles of the political spectrum: two radical right parties and two ethnic parties. We first collected all the documents, generated a ranked list of  $n$ -gram keywords and placed the most frequent  $n$ -grams on the top of the list.<sup>5</sup> Using the ranked list of  $n$ -gram keywords, experts identified grid/group issues (topics) using a classification scheme derived from general social theory (Douglas 1970; Douglas and Wildavsky 1982).<sup>6</sup>

<sup>4</sup> For preprocessing in text analysis, see Denny and Spirling (2018).

<sup>5</sup> We used a dictionary-based approach to select  $n$ -grams. Our approach builds on Shah *et al.* (2002) and Krippendorff (2004). Two experts, Bustikova and Siroky, coded topics. For validation, the two coders achieved average inter-coder reliability (Krippendorff ratio) of .84, where 1.0 is a perfect match. The Supplementary material (section: Expert Validation) lists discriminative keywords. We also discuss how the experts selected topics and how this approach compares with the Manifesto Data Project and with the Chapel Hill Expert Survey in the Supplementary materials.

<sup>6</sup> More details on grid–group theory and typology are provided in the Supplementary materials.

The grid–group approach to the study of politics has been used to study mass political beliefs (Coughlin and Lockhart 1998), the determinants of dimension dominance (Rehm and Kitschelt 2018), party alignments in Western Europe (Rehm and Kitschelt 2015), complex political orientations beyond the traditional left–right scale (Grendstad 2003) and radical right parties (Bustikova and Kitschelt 2009). As a classification system, it places political orientations into four categories using the two axes of grid and group: hierarchy, egalitarianism, individualism and fatalism. It offers a more nuanced analytical tool for party classification than the left–right placement and is more versatile than the commonly used traditional versus libertarian distinction used in the Chapel Hill Expert Survey. It does not collapse identity onto one dimension and therefore can account for the fact that ethnic inclusion does not necessarily imply social liberalism.<sup>7</sup>

Since our analysis investigates the responsiveness of polar opposites on the so-called “second” axis of party competition (cultural issues as opposed to economic issues), any spike in issues that the party discusses implies both an increase in salience and polarization (Spies and Franzmann 2011). To capture different aspects of polarization and more granular action–reaction dynamics, the identity axis needs to be disaggregated. Grid–group allows the analyst to classify attitudes toward state authority as separate from ethnic issues.

We utilize LDA, one of the most popular topic inference algorithms (Blei *et al.* 2002). It assumes that documents represent a mixture of topics, where a topic is a probability distribution over words. In other words, it uses a “bag of words” approach to perform statistical topic modeling and to uncover hidden structures in large text corpora.<sup>8</sup> In our analysis, LDA outperformed two competitive alternatives: vector space and word embedding models.<sup>9</sup> After identifying the grid–group issues using ranked weighted TF–IDF terms, we applied LDA separately on ethnic and radical right parties’ corpora to discover their party-specific *latent topics*. For each grid–group issue, we determine when an issue is salient for one party (i.e., the issue-specific document volume crosses the threshold and constitutes a spike, as discussed below) and when that leads the other party to respond (with a temporally proximate and topically related spike).

## 2.2 Detecting Spikes, Similarity and Escalation

We utilize the 68–95–99.7 rule for spike detection (Pukelsheim 1994), also known as the three-sigma rule, which states that in a normal distribution nearly all values lie within three standard deviations ( $\sigma$ ) of the mean ( $\mu$ ). We utilize a fixed-sized sliding window (experimentally determined as 20 weeks) to compute a running average  $\mu(20)$  and a standard deviation  $\sigma$  for each issue’s weekly volume distribution from each camp. We designate a weekly volume as a *spike* if the weekly document volume matching an issue exceeds  $(\mu(20) + 2\sigma)$ . We tuned this sliding window to 20 weeks because it showed the best performance. When smaller windows (5 weeks, 10 weeks and 15 weeks) are applied, the resultant spikes are noisy. When larger windows are applied (25 weeks, 30 weeks and 35 weeks), the resultant spikes are sparse.

Spikes are categorized into two categories: (1) “escalated” spikes that trigger a reaction from the other camp or (2) “ignored” spikes that lead to no response—based on the relatedness (goodness of fit) of each topic distribution inferred by LDA in consecutive spikes from opposing camps. Spike categorization (*escalated/ignored*) is a result of shared topics between two consecutive spikes from opposing camps. By matching up consecutive spikes, we capture partisan debates, defined as “formal discussions on a set of related topics in which opposing perspectives and arguments are

7 Traditional–authoritarian–nationalist/green–alternative–libertarian (TAN/GAL) classification scheme of the Chapel Hill Expert Survey collapses identity into one dimension. This implies that, by definition, nationalist parties cannot embrace gender equality and minority ethnic parties are socially liberal. Yet, these “strange” configurations do exist.

8 A topic in LDA contains words that describe the topic and also words that express opinions about the topic.

9 For LDA comparisons to the vector space baseline model and word embedding model, see Figures 3–5 in the Supplementary materials.

put forward.”<sup>10</sup> To measure the “relatedness” of topics between a pair of consecutive spikes from opposing camps, we utilize the Kullback–Leibler (KL) divergence between LDA topic distributions of consecutive spikes (Kullback and Leibler 1951) and then convert it to a similarity measure, scaled between 0 and 1. The KL divergence of the probability distributions  $E, R$  on a finite set  $X$  is defined as<sup>11</sup>

$$D(E, R) = \sum_{x \in X} E(x) \log \frac{E(x)}{R(x)}. \tag{1}$$

Given two consecutive spikes from opposing camps, ethnic party spike  $S_E$  and radical right party spike  $S_R$ , we first identify latent topics of each spike with their distributions within the documents:  $E$  is the distributions of  $S_E$  topics and  $R$  is the distributions of  $S_R$  topics. For example, when comparing two consecutive spikes related to the issue of “Minorities”—one spike with topics from the ethnic camp such as “minority languages, schools, . . . etc” versus the following spike from other camp with topics “gypsy problem, schools, . . . etc”—we measure the distributions of topics in these two spikes with respect to the number of documents matching the “Minorities” topic. We then measure the divergence of topic distributions using the symmetric form of KL divergence (Seghouane and Amari, 2007) that measures the divergence of the probability distributions  $E, R$  on a finite set  $X$  of topics as follows:

$$D(E, R) = \sum_{x \in X} \left( (E(x) - R(x)) \log \frac{E(x)}{R(x)} \right). \tag{2}$$

We normalize this measure with respect to the sum of distributions to be between [0,1] and convert it to a similarity measure as follows:

$$Sim(S_E, S_R) = 1 - D_{normalized}(E, R), \tag{3}$$

where  $Sim(S_E, S_E) = 1$  means the two distributions of topics across the two spikes are identical. If the similarity of topic distributions between the two consecutive spikes from opposing camps exceeds a certain threshold, then we label the first spike as “escalated”; otherwise, it is considered “ignored.”<sup>12</sup>

LDA is used twice. The first time, LDA is applied on the overall corpus (both radical and ethnic corpus) to measure the relatedness of spikes. The second time, LDA is used separately on the radical corpus (radical right parties) and on the ethnic corpus (ethnic parties) to determine issue-specific frames to be exploited as features for the predictive model.

### 2.3 Models

We experimented with three methods: a baseline vector space model, a word embedding model and an LDA model.

#### 2.3.1 Baseline Vector Space Model

In our baseline model (vector space model), we directly modeled the similarity approach by using the cosine similarity over spikes’ frequent keyword vector representation of  $E$  and  $R$ , without requiring a *lower dimensional space* representation of the data, for example, inferring topic distribution LDA or word embedding document to vector as follows:

$$Sim(S_E, S_R) = Cosine(E, R) = \frac{E \cdot R}{\|E\|_2 \cdot \|R\|_2}. \tag{4}$$

<sup>10</sup> Oxford online dictionary.

<sup>11</sup> Intuitively, this is a divergence (D) between the probability distributions of topics discussed by ethnic parties (E) and by radical parties (R) on a finite set of topics (X).

<sup>12</sup> In Section 3.3, we show the experimentally determined issue-specific thresholds that we used.

The similarity measure is a sparse vector representation of frequent keywords. We computed all distances between consecutive spikes using Equation (4) and, by thresholding, determined the labels of spikes whose measure is larger than or equal to the mean, indicating an “escalated” spike from the opposing camp; otherwise, they are labeled as “ignored.”

### 2.3.2 Word Embedding Model

Word embedding utilizes neural networks to encode the context into a denser, lower dimensional space. It is a highly effective method of capturing semantic relations where each document is represented by a real number vector such that similar documents are closer to one another than dissimilar documents in a geometric space. We employed the Paragraph Vector Distributed Bag of Words (PV-DBOW) proposed by Mikolov *et al.* (2013) and Le and Mikolov (2014) to infer the real number vector of a document (a.k.a. doc2vector). After training the PV-DBOW model over our corpus, we infer vectors for each spike and computed all distances using Equation (4).<sup>13</sup>

### 2.3.3 LDA Model

The LDA model can be viewed as a three-level Bayesian probabilistic model to learn distributions of topics over documents and words. After training an LDA model, we infer topic distribution for each spike’s topics. Then, we determine the labels of spikes based on the KL measure, which captures the divergence of distributions between two consecutive spikes (Equation (3)).

## 2.4 Framing Analysis and Predicting Escalation

During a debate on a particular topic, both radical right parties and ethnic parties discuss different perspectives.<sup>14</sup> Once escalated and ignored spikes from one camp are determined, we use a sparse-learning framework (Liu *et al.*, 2009a), with the aim of selecting a subset of discriminating features that can identify and classify contentious (escalatory) spikes as opposed to ignored ones. The following steps describe our algorithm:

- (1) For each *key grid-group* issue, run LDA to get latent topics for one camp.<sup>15</sup>
- (2) Filter the frame  $\times$  spike matrix to include only the top 2,000 terms representing frames from one camp (100 topics (or topic dimensions) each with top 20 terms inferred).
- (3) Formulate the problem as a logit model streamlined by the SLEP framework (Liu *et al.*, 2009a) to predict escalated versus ignored spikes. Formally,

$$\min_x \sum_{i=1}^m w_i \log(1 + \exp(-y_i(x^t a_i + c))) + \lambda \|x\|_1, \quad (5)$$

where  $a_i$  is the vector representation of the  $i$ th spike,  $w_i$  is the weight assigned to the  $i$ th spike ( $w_i = 1/m$  by default),  $A = [a_1, a_2, \dots, a_m]$  is the frame  $\times$  spike matrix,  $y_i$  is the polarity of each spike (+1 for an escalated spike and -1 for an ignored spike),  $x_j$ , the  $j$ th element of  $x$ , is the unknown weight for each frame, ( $\lambda > 0$ ) is a regularization parameter that controls the sparsity of the solution and  $\|x\|_1 = \sum |x_j|$  is the 1-norm of the  $x$  vector.

The sparse-learning approach (SLEP) relies on a gradient descent algorithm to solve the above convex and nonsmooth optimization problem (Liu *et al.*, 2009b). The frames with nonzero values

13 Cosine fits better as each document is represented by a point in a geometry space. Through thresholding, we determined the labels for the spikes.

14 During the design of an automated perspective detection algorithm, we made the following simplifying assumptions: (1) each camp will mostly discuss their own perspective in a debate; (2) each camp will occasionally mention others’ perspectives but relate them back to their own perspective.

15 Determining the number of topics can be done using various methods (e.g., elbow curves, AIC, BIC, etc.). Among these approaches, LDA tends to be most resilient when the number of topics,  $k$ , increases (Blei *et al.* 2010). However, larger  $k$  imposes additional computational costs and makes convergence of the posterior probability estimate more difficult. Finding the right  $k$  also requires qualitative validation by experts. After multiple trials, we determined that the most applicable  $k$  was 100. Later, we determined that the results are robust to minor changes to  $k$  (e.g.,  $[+/- \frac{k}{10}]$ ).

on the sparse  $x$  vector yield the discriminant factors for classifying a spike as escalated or ignored based on their polarity (positive or negative). Frames with positive polarity correspond to escalated frames and those with negative polarity to ignored frames.

### 3 Analysis

#### 3.1 The Data Corpus

The corpus comprises 10,597 news and opinion articles downloaded from the official websites of radical right and ethnic political parties in Slovakia from 2004 and 2014. From the ethnic camp, we downloaded all documents from *Most-Híd* (<http://www.most-hid.sk>) and from *SMK—Party of the Hungarian Coalition* (<http://www.mkp.sk>). From the radical right camp, we downloaded all documents from the *SNS—Slovak National Party* (<http://www.sns.sk>) and *Slovenská Pospolitosť—The Slovak Brotherhood* (<https://pospolitost.wordpress.com>). The document volume is roughly equal between the two camps.

While the method has broad potential applications for studying a diverse set of cases and topics, a word is in order about why this party system serves as an interesting and important case study to introduce this approach to predicting partisan responsiveness. First, the radical right parties and the ethnic parties in Slovakia are both relatively large compared to some “niche parties” in other countries; so they are politically relevant for coalition formation.

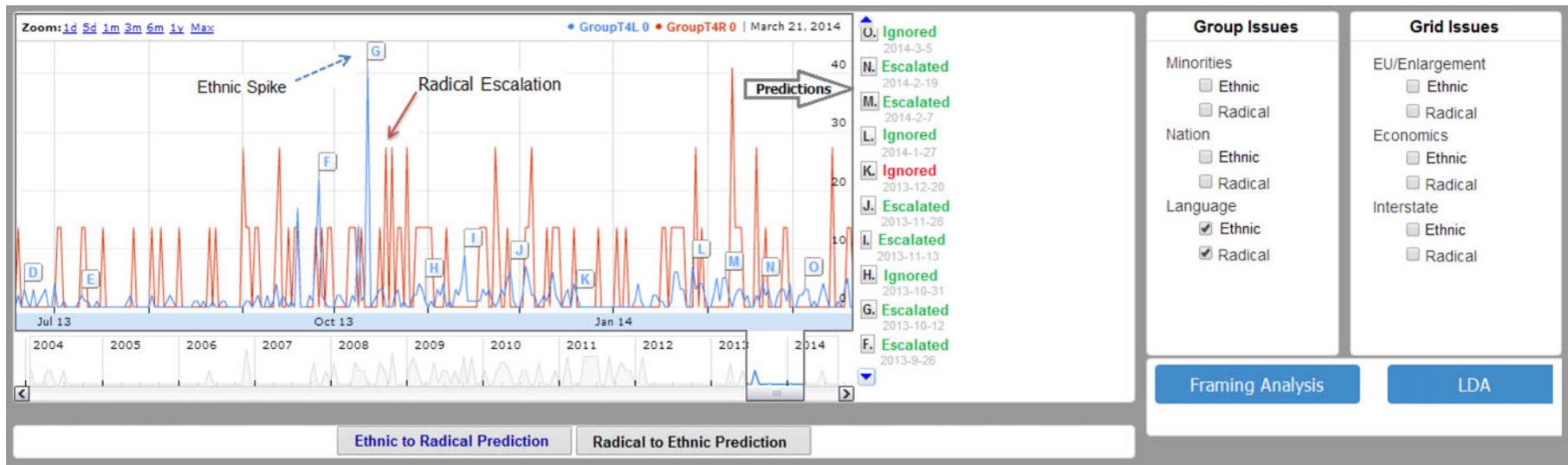
Second, the political space in Slovakia has been characterized by a high degree of variation in the extent to which it is polarized on issues of national identity, and this variability allows us to track a truly dynamic process of contestation. Finally, over the past two decades, the political scene in Slovakia has been quite stable in terms of the actors that anchor both political poles. This provides consistency over time in the analysis since the actors are identifiable with transparent profiles and reputations that have been established over a relatively long time period (Gyárfášová *et al.* 2015; Baboš, Világi, and Oravcová 2016; Kluknavská and Smolík 2016; Guasti and Mansfeldová 2018).

Figure 2 displays spikes of attention over time to one topic (in this case, “Language”), by both camps (radical spikes are red and ethnic spikes are blue), and shows that the adversaries mobilize in bursts.<sup>16</sup> It also shows whether a spike from one camp is ignored or reacted to by the other camp in the form of a new spike on the same topic. The bottom panel in Figure 2 shows (in gray) the overall volume of contentious frames (2008–2011) during a period of intense debate over a very restrictive language law, adopted in September 2009 with the help of the SNS, and the ensuing efforts of ethnic parties to soften its negative impact on Hungarians. Finally, Figure 2 illustrates the volume of documents generated by radical (red line) and ethnic (blue line) political party outlets in Slovakia over the entire 10-year period (between 2004 and 2014) that we analyze.

Not all topics resonate within the dyad. The “predictions” panel on the right of Figure 2 shows a timeline (using alphabetic annotations) that corresponds to ethnic party spikes and predictions about whether they will lead to “escalation” as a result of the radical right parties responding. Green labels indicate a “hit” (correct prediction) and red labels indicate a “miss” (incorrect prediction) by the classifier. In the screenshot displayed, which covers from July 2013 to March 2014, the classifier correctly hits 9 of out 10 spikes.

Subject matter experts selected the *key issues* shown in Table 2 and mapped them onto the group (nationalism) and grid (state authority) dimensions. Focusing on these six topics, the framework categorizes spikes as either escalated or ignored for both radical right parties and for ethnic parties and then uses this information to predict partisan responsiveness. To determine whether a pair of consecutive spikes on one of these six topics is related, the mean similarity for each grid/group issue was used as the threshold. If the similarity between a spike (from one camp)

<sup>16</sup> For “burstiness,” see Eggers and Spirling (2018).



**Figure 2.** Contentious Frames Analyzer Tool. Zoom on the period from the middle of July 2013 to the middle of March 2014. Below the main plot is a secondary plot that shows the entire time period from 2004 to 2014. Users can simply slide the window to the time period of interest and the main plot will zoom in on this period and identify ignored and escalated spikes for a given topic, which can be selected on the right (Minorities, Nation, Language, Interstate (Relations), Economics, EU/Enlargement), and for a given actor in the dyad (ethnic or radical right party), which can be selected below the main plot: Ethnic eliciting Radical reaction prediction (left) or Radical eliciting Ethnic reaction prediction (right).

**Table 1.** Number of Analyzed Documents.

Ethnic Parties	#of (Docs)	Radical Right Parties	#of (Docs)
SMK (Party of the Hungarian Coalition)	4143	SNS (Slovak National Party)	4800
Most-Híd—Bridge	640	Slovenská Pospolitosť	1014
<b>Total</b>	<b>4783</b>	<b>Total</b>	<b>5814</b>

*Ethnic parties* are: SMK—Strana maďarskej koalície, Party of the Hungarian Coalition (now Strana maďarskej komunity) and Most-Híd—Bridge).

*Radical right parties* are: SNS—Slovenská národná strana, Slovak National Party and Slovenská Pospolitosť, Slovak Brotherhood.

**Table 2.** Group (Nationalism) and Grid (State Authority) Issues.

Group	Grid
Minorities	EU/Enlargement
Nation	Economics
Language	Interstate Relations

and the following spike (from the other camp) exceeds the mean similarity for an issue, then the first spike is labeled *escalated* and otherwise it is labeled *ignored*.

To illustrate, Figure 3 shows the similarity measure ( $y$ -axis) for ethnic party spikes, for each of the six grid/group issues ( $x$ -axis). Dots represent spikes and boxes show the means (which vary between 0.35 and 0.45), along with the first and third quartiles, with whiskers for the 95% confidence intervals on the similarity measure. Once spikes from each camp are detected and categorized, the terms that comprise the topics are used as features for the SLEP classifier (Liu, 2009b) to identify discriminative frames. Using 10-fold cross-validation (McLachlan, 2004), we calculated the precision, recall and F-measure (Perry, 1955).

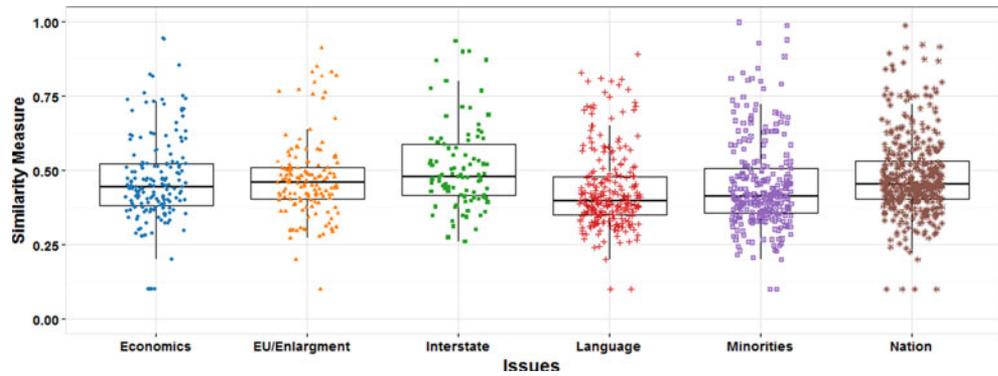
### 3.2 Model Performance and Frame Detections Based on the LDA Model

Table 3 displays the performance of the vector space model, the word embedding model and LDA. In this table, the LDA model outperformed the other two models (F-measure).<sup>17</sup> Using the same set of LDA-based features, the SLEP classifier outperforms both the Naive Bayes and Random Forest classifiers in terms of the overall F-measure, which is consistently higher for SLEP than for either alternative classifier across *all* escalated and ignored topics.<sup>18</sup>

Figure 4 shows the “language” topics over which ethnic parties and radical parties fight (e.g., languag(e) Slovak, Hungarian school, minor(ity) nation, human right, educ(ational) minist(ry)). The Venn diagram shows two intersected circles, where the first circle belongs to the first spikes with top terms and the second circle represents the spikes from the other party. The intersecting area represents the common terms. Language education emerges as an intrinsic focal point spurring debate, in this case, prompted by ethnic parties and followed by a spike from radical right parties.

<sup>17</sup> Tables 1–4 in the Supplementary materials present the results of the LDA model for each party. Table 1, in the Supplementary materials, shows the accuracy for predicting ethnic party spikes that lead to radical right party responses. The accuracy varies between 81% and 89% for different issues (F-measure). Table 2 shows the accuracy for predicting ethnic spikes that the radical right parties ignore, which varies between 78% and 84% (F-measure). The average F-measure for predicting outcomes of ethnic spikes is therefore 82.9%. Similarly, Table 3 shows the accuracy for predicting ethnic party responsiveness to radical right party spikes, which varies between 80% and 86% depending on the issue (F-measure). Table 4 shows the accuracy for predicting radical spikes that are ignored by the ethnic parties, which varies between 78% and 86% (F-measure). The average F-measure for predicting outcomes of radical spikes is about the same: 82.7%.

<sup>18</sup> Classification results for Naive Bayes and Random Forest are provided in Figure 1 and Figure 2 in Supplementary materials. We also found that Random Forest mostly outperforms Naive Bayes in terms of F-measure.



**Figure 3.** Similarity Measures for Ethnic Spikes.

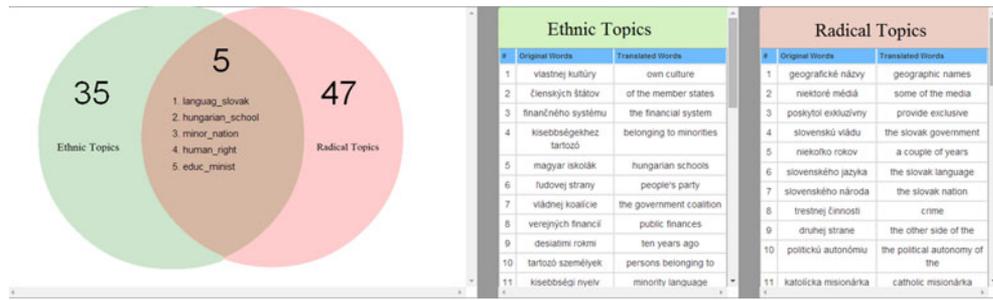
**Table 3.** F-measures of Three Models.

Dimension	Vector Space	Word Embedding	LDA
Escalated Ethnic—Minorities	0.71	0.78	0.81
Escalated Ethnic—Nation	0.80	0.72	0.88
Escalated Ethnic—Language	0.66	0.77	0.86
Escalated Ethnic—EU/Enlargement	0.77	0.86	0.82
Escalated Ethnic—Economic	0.75	0.72	0.82
Escalated Ethnic—Interstate	0.68	0.82	0.89
Ignored Ethnic—Minorities	0.67	0.80	0.84
Ignored Ethnic—Nation	0.76	0.75	0.80
Ignored Ethnic—Language	0.64	0.79	0.83
Ignored Ethnic—EU/Enlargement	0.76	0.81	0.82
Ignored Ethnic—Economic	0.43	0.75	0.78
Ignored Ethnic—Interstate	0.67	0.77	0.80
Escalated Radical—Minorities	0.59	0.82	0.84
Escalated Radical—Nation	0.61	0.80	0.80
Escalated Radical—Language	0.58	0.79	0.82
Escalated Radical—EU/Enlargement	0.57	0.81	0.84
Escalated Radical—Economic	0.95	0.81	0.84
Escalated Radical—Interstate	0.64	0.80	0.86
Ignored Radical—Minorities	0.48	0.79	0.86
Ignored Radical—Nation	0.59	0.79	0.80
Ignored Radical—Language	0.56	0.80	0.82
Ignored Radical—EU/Enlargement	0.52	0.82	0.78
Ignored Radical—Economic	0.50	0.82	0.82
Ignored Radical—Interstate	0.50	0.82	0.84
<b>Mean</b>	<b>0.64</b>	<b>0.79</b>	<b>0.83</b>
<b>Standard deviation</b>	<b>0.12</b>	<b>0.03</b>	<b>0.03</b>

### 3.3 Contentious Frames and Polarization in Slovakia

To assess which topics are contentious and polarize public discourse and which topics are ignored, we turn to the intersecting topics and “contentious frames” identified in Figure 4 and Table 4, which lists issue-specific frames used by each camp that tended to elicit reactions from the other camp. We then compare these with the “ignored frames” in Table 5.

The left column of Table 4 depicts the Hungarian–Slovak political cleavage over language and Hungarian minority rights rather clearly. This is consistent with decade-old fights over the status of



**Figure 4.** Intersecting Topics.

**Table 4.** Contentious Frames.

Contentious Frames of the Ethnic Minority (Radicals Respond)	Issues	Contentious Frames of Radicals (Ethnic Minority Responds)
Against the citizenship law, language law, Slovak citizenship, human rights, Hungarian nationality	Nation	Against racism, Jozef Tiso, Jozef Rydlo, protection of the republic, immunity of the MPs
Minority language, ethnic minorities, Vojvodina Hungarians, Slovak citizenship	Language	Carpathian basin, agent SIS, Jozef Rydlo, immunity of the MPs, Matica Slovenska, MPs of the Carpathian basin, Andrej Hlinka
Ethnic minorities, minority languages, language law, southern Slovakia, citizenship, Ministry of Education, dual citizenship	Minority	White race, gypsy problem, Project “Studnica,” agent of SIS against racism, protection of the republic
Self-governing region, public finances, mother tongue, social insurance, southern Slovakia, European Parliament	Economy	Second pillar (pension reform), healthcare, Jozef Rydlo, MPs of the Carpathian basin
Hungarians in Slovakia, parliamentary elections, ruling coalition, ethnically mixed areas, petitions committee, European Parliament	EU/Enlargement	Death penalty, Jozef Rydilo, introduction of euro, Slovak’s entry, Carpathian basin
Language law, Czechoslovak republic, World War, collective guilt, Slovak–Hungarian relations, postwar	Interstate	World War II death penalty, Jozef Rydlo emergence (origin) of Slovak state, Project “Studnica,” anti-Zionism

the Hungarian minority, particularly its language rights (Bútora 2007; Haughton and Ryba 2008; Mesežnikov, Gyárfášová, and Smilov 2008; Deegan Krause and Haughton 2009). Radical parties tend to escalate on language policies: they respond strongly when ethnic parties talk about the “language law,” “minority language” and “mother tongue.”

Slovakia is home to two ethnic minorities: politically mobilized Hungarians and demobilized, impoverished Roma. The computational results show that radical parties and (Hungarian) ethnic parties react differently to Roma issues. While radicals respond to frames that advance the rights of the Hungarian speakers, the reverse is not true: if radicals challenge the right of Hungarians to be politically accommodated, ethnic parties do not escalate. Instead, ethnic parties escalate if radical right parties launch attacks on Roma: a different ethnic group and when they invoke the legacy of interwar fascism associated with an independent Slovak state. Ethnic parties also respond when the radical right parties discuss “protection of the republic,” “white race” and the “Gypsy (Roma)

**Table 5.** Ignored Frames.

Ignored Frames of the Ethnic Minority (Radicals No Response)	Issues	Ignored Frames of Radicals (Ethnic Minority No Response)
Public Education Act, Lisbon Strategy, member states (EU), conference of Bishops	Nation	Turkey’s accession (to the EU), healthcare, unified SNS, SNS council, Defence Ministry, Department of Transportation, average wage
Ministry of Environment, subsidies, annual budget, judicial council, EU grant	Language	Serb republic, discussion of SNS, Turkey’s accession (to the EU), church provinces, pope Benedict, Jan Hus
Musical institution, EU funds, cultural minorities, theater, national music, cultural activities, subsidies	Minority	President of Slovakia, SNS Council, Turkey’s accession (to the EU), church provinces, Han Hus, united SNS
Jobs, social market economy, Minister of Justice, protect children, European Union, new jobs, unemployment	Economy	SNS Cabinet, Turkey’s accession (to the EU), President of Slovakia, Slovak Parliament, adoption of the euro, Department of Transportation
EU aid, development plan, European People’s Party, Lisbon strategy, member states, Warsaw Pact	EU/Enlargement	Slota said, SNS Plenary Meeting, SNS discussion, Department of Transportation, Serb republic, adoption of Euro, against Zionism
Archbishop Arpad, European Union, European Parliament, Benes Decrees, European People’s Party, European rail	Interstate	SNS club, reformed church, prices growth, boycott of Israel, freedom of speech, Slovak church, church provinces

problem.” The historical dimension associated with attempts to whitewash the fascist legacy of Jozef Tiso, who collaborated with the Nazis, is most evident in the frames “Nation,” “Language” and “Minority.”

Scholars of Slovak politics know that radical right parties attack both Hungarians and Roma, but the computational results also reveal that ethnic Hungarian parties are more likely to respond to radical frames that are *not* related to the rights of Hungarians but rather to Roma, and to historical frames. If ethnic parties escalate on issues of Roma and interwar legacies, they frame radicals as fascists and xenophobes, and thereby diminish radicals as credible adversaries that can be engaged to debate policy. This may explain why ethnic parties stay quiet when radical right parties question policies that expand their (language) rights.

Turning to the frames that were largely *ignored* by the other side of the political spectrum in Table 5, we see that radical right parties did not respond when topics were discussed in cultural terms but rather did so when these topics were discussed in policy terms. Under the issue “Minority,” for example, radicals did not respond to the frames “cultural minorities,” “theater” and “cultural activities.” Although these frames may suggest concessions to minorities, they fall short of recognition as a “national minority,” which implies language rights as well as political and economic power-sharing. Similarly, ethnic parties ignored radical frames under the issues of “Nation,” “Language” and “Minority” that focused on religion (“Pope Benedict,” “Jan Hus,” “church provinces” and “Slovak Church”).

Tables 4 and 5 underscore the fact that the key issues of contention between radical right parties and ethnic political parties are almost exclusively related to three main issues: (1) rights of Hungarians as a national minority, (2) hostility toward Roma, an ethnic underclass and (3) an interwar fascist legacy: a historical cleavage that concerns the Nazi collaboration of the first

independent Slovak state during World War II. The computational and qualitative text analysis advances understanding of party politics with a new approach to highlighting the issue-specific causes of political mobilization and polarization. It reveals that radical right parties mobilize when their adversaries discuss minority rights, whereas the ethnic parties respond to adversarial frames that evolve around racism and historical autocratic regime legacies. In sum, parties on the opposing poles of the political spectrum respond to each other and selectively react to each other's polarizing frames.

#### 4 Conclusion

The computation approach introduced in this paper has a broad potential applicability for studying ideological positioning and partisan debates in political science, at different scales and contexts of political competition. Using this framework, scholars can parse, analyze and generate predictions about practically any interesting “debate” between “camps” that produces a large corpus of time-stamped text. While any form of documented debate is a fair game, political debates are particularly ripe for this type of approach because they are frequently both contentious and consequential.

With the growth of online content, political scientists now have more information at their disposal than they can humanly process and understand. Manual processing of such information is time-consuming, costly and does not scale well. This article enriches computational political science by harnessing unstructured data into temporal and topical dimensions for automated analysis to better understand and predict partisan responsiveness. It develops and assesses a new computational tool to discover and predict contentious and ignored frames for each political camp. Using radical right party and ethnic party website content from 2004 to 2014 in Slovakia, the model has an average accuracy (F-measure) for escalated ethnic spikes of 84.7% and an average accuracy for escalated radical spikes of 83.3%. This approach outperforms Random Forest and Naive Bayes classifiers. Using LDA boosts performance over vector space and word embedding models. A qualitative analysis of the contentious and ignored frames yields additional substantive insights and shows that ethnic parties respond more to xenophobic and historical frames, whereas radical right parties react more to frames about minority accommodation. We have also shown that parties on the very opposite poles of the ideological spectrum react to each other's frames and thereby contribute to political polarization.

Although considerable progress has been made in automating content analysis, scholars have also become increasingly aware of its limitations. Grimmer and Stewart (2013) present several issues that scholars applying content analysis models should recognize and engage. First, scholars should acknowledge the complexity of the language and that many quantitative models are incapable of handling language complexity as humans do. Automated content analysis methods will not replace humans, but these methods can magnify our abilities. Here, the automated analysis of partisan responsiveness serves as a complement, rather than a substitute, to subject matter expertise. All the quantitative results in the paper are validated qualitatively by subject matter experts. Second, since there is no global method for automated content analysis, each research problem, along with its data, has to have its own methodology. Although there are general principles and algorithms, there is no “Plug and Play” solution to various research questions. As a result, validating the outputs of content analysis models is a core requirement. One venerable validation approach entails having subject area experts examine the results.

We believe this study represents an important contribution to political science, yet we also wish to highlight several limitations and directions for future research. First, we analyzed the content of official party websites but not other outlets, such as newspapers and social media. Incorporating data from these sources could expand the range of actors and frames, leading to a more comprehensive understanding of partisan dynamics and higher predictive accuracy.

Moreover, these sources could identify emerging trends in real time. Second, more work is needed examining the role of external events, such as elections and protests. Despite these limits, we are hopeful that this approach to understanding partisan responsiveness and polarization will facilitate and inspire additional usage, research and insights into how topic modeling can improve our understanding of party politics and our ability to predict party dynamics.

## Supplementary material

For supplementary material accompanying this paper, please visit <https://doi.org/10.1017/pan.2019.18>.

## References

- Abelson, R. P. 1973. "The Structure of Belief Systems." In *Computer Models of Thought and Language*, edited by R. Schank and K. K. M. Colby, 287–339. San Francisco: W. H. Freeman and Company.
- Abou-Chadi, T., and W. Krause. 2018. "The Causal Effect of Radical Right Success on Mainstream Parties' Policy Positions: A Regression Discontinuity Approach." *British Journal of Political Science*, doi:10.1017/S0007123418000029.
- Adams, J. 2012. "Causes and Electoral Consequences of Party Policy Shifts in Multiparty Elections: Theoretical Results and Empirical Evidence." *The Annual Review of Political Science* 15(1):401–419.
- Adams, J., M. Clark, L. Ezrow, and G. Glasgow. 2006. "Are Niche Parties fundamentally Different from Mainstream Parties? The Causes and the Electoral Consequences of Western European Parties' Policy Shifts." *American Journal of Political Science* 50(3):513–529.
- Adams, J., and Z. Somer-Topcu. 2009. "Policy Adjustment by Parties in Response to Rival Parties' Policy Shifts: Spatial Theory and the Dynamics of Party Competition in Twenty-Five Post-War Democracies." *British Journal of Political Science* 39(4):825–846.
- Alashri, S., S. Alzahrani, L. Bustikova, and D. Siroky. 2018. "Replication Data for: Predicting Partisan Responsiveness: A Probabilistic Text Mining Time-Series Approach." <https://doi.org/10.7910/DVN/TBW4Y4>, Harvard Dataverse, V1.
- Alashri, S., S. Alzahrani, L. Bustikova, D. Siroky, and H. Davulcu. 2015. "What Animates Political Debates? Analyzing Ideological Perspectives in Online Debates Between Opposing Parties." In *Proceedings of the ASE/IEEE International Conference on Social Computing (SocialCom-15)*. Stanford, CA: Academy of Science and Engineering.
- Arceaux, K., and M. Johnson. 2015. "More a Symptom Than a Cause: Polarization and Partisan News Media in America." In *American Gridlock: The Sources, Character, and Impact of Political Polarization*, edited by J. A. Thurber and A. Yoshinaka, 309–336. New York: Cambridge University Press.
- Arzheimer, K., and E. Carter. 2009. "Christian Religiosity and Voting for West European Radical Right Parties." *West European Politics* 32(5):985–1011.
- Baboš, P., A. Világi, and V. Oravcová. 2016. *Spoločenské problémy a politické (ne)riešenia: Vol'by 2016*. Bratislava: STIMUL.
- Blei, D. M., T. L. Griffiths, and M. I. Jordan. 2010. "The Nested Chinese Restaurant Process and Bayesian Nonparametric Inference of Topic Hierarchies." *Journal of the ACM* 57(2):1–30.
- Blei, D. M., A. Y. Ng, and M. I. Jordan. 2002. "Latent Dirichlet Allocation." *Advances in Neural Information Processing Systems (NIPS) 14* 3:601–608.
- Boyce, B. R. 1990. "Concepts of Information Retrieval and Automatic Text Processing: The Transformation Analysis, and Retrieval of Information by Computer." *Journal of the American Society for Information Science* 41(2):150–151.
- Breiman, L. 2001. "Random Forests." *Mach. Learn.* 45(1):5–32.
- Bustikova, L. 2014. "Revenge of the Radical Right." *Comparative Political Studies* 47(12):1738–1765.
- Bustikova, L., and H. Kitschelt. 2009. "The Radical Right in Post-Communist Europe. Comparative Perspectives on Legacies and Party Competition." *Communist and Post-Communist Studies* 42(4):459–483.
- Bútorá, M. 2007. "Nightmares from the Past, Dreams of the Future." *Journal of Democracy* 18(4):47–55.
- Carbonell, J. G. 1978. "POLITICS: Automated Ideological Reasoning." *Cognitive Science* 2(1):27–51.
- Colaresi, M., and Z. Mahmood. 2017. "Do the Robot: Lessons from Machine Learning to Improve Conflict Forecasting." *Journal of Peace Research* 54(2):193–214.
- Coughlin, R. M., and C. Lockhart. 1998. "Grid-Group Theory and Political Ideology: A Consideration of Their Relative Strengths and Weaknesses for Explaining the Structure of Mass Belief Systems." *Journal of Theoretical Politics* 10(1):33–58.
- Deegan Krause, K., and T. Houghton. 2009. "Toward a More Useful Conceptualization of Populism." *Politics and Policy* 37(4):821–841.

- Denny, M., and A. Spirling. 2018. "Text Preprocessing for Unsupervised Learning: Why It Matters, When It Misleads, And What To Do About It." *Political Analysis* 26(2):168–189.
- Douglas, M. 1970. *Natural Symbols. Explorations in Cosmology*. London: Routledge.
- Douglas, M., and A. Wildavsky. 1982. *Risk and Culture: An Essay on the Selection of Technical and Environmental Dangers*, Berkeley, CA: University of California Press.
- Eggers, A. C., and A. Spirling. 2018. "The Shadow Cabinet in Westminster Systems: Modeling Opposition Agenda Setting in the House of Commons." *British Journal of Political Science* 48(2):343–367.
- Evans, J. A. 2002. "In Defence of Sartori: Party System Change, Voter Preference Distributions and Other Competitive Incentives." *Party Politics* 8(2):155–174.
- Greene, D., and J. P. Cross. 2017. "Exploring the Political Agenda of the European Parliament Using a Dynamic Topic Modeling Approach." *Political Analysis* 25(1):77–94.
- Grendstad, G. 2003. "Comparing Political Orientations: Grid-group Theory Versus the Left-Right Dimension in the Five Nordic Countries." *European Journal of Political Research* 42(1):1–21.
- Grimmer, J. 2009. "A Bayesian Hierarchical Topic Model for Political Texts: Measuring Expressed Agendas in State Press Releases." *Political Analysis* 18(1):1–51.
- Grimmer, J., and B. M. Stewart. 2013. "Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts." *Political Analysis* 21(3):267–297.
- Guastrri, P., and Z. Mansfeldová. 2018. *Democracy Under Stress*. Prague, Czech Republic: Institute of Sociology, Czech Academy of Sciences.
- Gyárfášová, O. 2015. "To Sladké Slovo Demokracia. . . . Spokojnosť s Demokraciou a Politické Odcudzenie na Slovensku." *Sociológia-Slovak Sociological Review* 47(4):365–389.
- Hartigan, J. A., and M. A. Wong. 1979. "Algorithm AS 136: A K-Means Clustering Algorithm." *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28(1):100–108. doi:10.2307/2346830.
- Haughton, T., and M. Ryba. 2008. "A Change in Direction: the 2006 Parliamentary Elections and Party Politics in Slovakia." *Journal of Communist Studies and Transition Politics* 24:232–255.
- Ignazi, P. 1992. "The Silent Counter-Revolution." *European Journal of Political Research* 22(1):3–34.
- Katz, R. S., and P. Mair. 1995. "Changing Models of Party Organization and Party Democracy: The Emergence of the Cartel Party." *Party Politics* 1(1):5–28.
- Kluknavská, A., and J. Smolík. 2016. "We Hate Them All? Issue Adaptation of Extreme Right Parties in Slovakia." *Communist and Post-Communist Studies* 49(4):335–344.
- Klüver, H., and J.-J. Spoon. 2016. "Who Responds? Voters, Parties and Issue Attention." *British Journal of Political Science* 46(3):633–654.
- Krippendorff, K. 2004. *Content Analysis: An Introduction to Its Methodology*, 2nd edn. Thousand Oaks, CA: Sage Publications.
- Kullback, S., and R. A. Leibler. 1951. "On Information and Sufficiency." *The Annals of Mathematical Statistics* 22(1):79–86.
- Le, Q., and T. Mikolov. 2014. "Distributed Representations of Sentences and Documents." In *Proceedings of the 31st International Conference on Machine Learning (ICML-14)*, 1188–1196. <http://proceedings.mlr.press/v32/le14.pdf>.
- Leskovec, J., L. Backstrom, and J. Kleinberg. 2009. "Meme-tracking and the Dynamics of the News Cycle." In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 497–506. New York: ACM. <http://doi.acm.org/10.1145/1557019.1557077>.
- Liu, J., J. Chen, and J. Ye. 2009a. "Large-scale Sparse Logistic Regression." In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '09*, 547–555. New York: ACM. <http://www.scopus.com/inward/record.url?eid=2-s2.0-70350663114&partnerID=tZ0tx3y1>.
- Liu, J., S. Ji, and J. Ye. 2009b. "{SLEP}: Sparse Learning with Efficient Projections." <http://yelab.net/software/SLEP/>.
- McCallum, A. K. 2002. "MALLET: A Machine Learning for Language Toolkit." <http://mallet.cs.umass.edu>.
- Mclachlan, G. J., K.-A. Do, and C. Ambrose. 2004. *Analyzing Microarray Gene Expression Data*, vol. 422. Hoboken, NJ: John Wiley and Sons.
- Meguid, B. 2008. *Party Competition Between Unequals*. New York: Cambridge University Press.
- Mesežnikov, G., O. Gyárfášová, and D. Smilov. 2008. "Populist Politics and Liberal Democracy in Central and Eastern Europe." IVO (IPA) Working Paper Series, Bratislava.
- Mikolov, T., K. Chen, G. Corrado, and J. Dean. 2013. "Efficient Estimation of Word Representations in Vector Space." Preprint, [arXiv:1301.3781](https://arxiv.org/abs/1301.3781).
- Monroe, B. L., M. P. Colaresi, and K. M. Quinn. 2008. "Fightin' Words: Lexical Feature Selection and Evaluation for Identifying the Content of Political Conflict." *Political Analysis* 16(4 SPEC. ISS):372–403.
- Perry, J. W., A. Kent, and M. M. Berry. 1955. "Machine Literature Searching X. Machine Language; Factors Underlying Its Design and Development." *American Documentation* 6(4):242–254.
- Pukelsheim, F. 1994. "The Three Sigma Rule." *The American Statistician* 48(2):88–91.

- Rehm, P., and H. Kitschelt. 2015. "Party Alignment: Change and Continuity." In *The Politics of Advanced Capitalism*, edited by P. Beramendi, S. Hausermann, H. Kitschelt, and H. Kriesi, 179–201. New York: Cambridge University Press.
- Rehm, P., and H. Kitschelt. 2018. "Determinants of Dimension Dominance." In *Welfare Democracies and Party Politics: Explaining Electoral Dynamics in Times of Changing Welfare Capitalism*, edited by P. Manow, B. Palier, and H. Schwander, 61–88. Oxford: Oxford University Press.
- Sartori, G. 1976. *Party and Party Systems*. New York: Cambridge University Press.
- Seghouane, A. K., and S. I. Amari. 2007. "The AIC Criterion and Symmetrizing the Kullback-Leibler Divergence." *IEEE Transactions on Neural Networks* 18(1):97–106. doi:10.1109/tnn.2006.882813.
- Shah, D. V., M. D. Watts, D. Domke, and D. P. Fan. 2002. "News Framing and Cueing of Issue Regimes: Explaining Clinton's Public Approval in Spite of Scandal." *Public Opinion Quarterly* 66(3):339–370.
- Siroky, D. S. 2009. "Navigating Random Forests and Related Advances in Algorithmic Modeling." *Statist. Surv.* 3:147–163.
- Spies, D., and S. T. Franzmann. 2011. "A Two-Dimensional Approach to the Political Opportunity Structure of Extreme Right Parties in Western Europe." *West European Politics* 34(5):1044–1069.
- Spoon, J.-J. 2011. *Political Survival of Small Parties in Europe*. Ann Arbor: University of Michigan Press.
- Theocharis, Y., P. Barber, Z. Fazekas, S. A. Popa, and O. Parnet. 2016. "A Bad Workman Blames his Tweets: the Consequences of Citizens' Uncivil Twitter use When Interacting with Party Candidates." *Journal of Communication* 66(6):1007–1031.
- Tucker, J. A., A. Guess, P. Barbera, C. Vaccari, A. Siegel, S. Sanovich, D. Stukal, and B. Nyhan. 2018. *Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature*. Hewlett Foundation. Available at <https://www.hewlett.org/wp-content/uploads/2018/03/Social-Media-Political-Polarization-and-Political-Disinformation-Literature-Review.pdf>.