What Animates Political Debates? Analyzing Ideological Perspectives in Online Debates between Opposing Parties

Saud Alashri[1], Sultan Alzahrani[1], Lenka Bustikova[2], David Siroky[2], Hasan Davulcu[1]
[1] School of Computing, Informatics and Decision Systems Engineering, Arizona State University
[2] School of Politics and Global Studies, Arizona State University
salashri@asu.edu, ssalzahr@asu.edu, lenka.bustikova@asu.edu,
david.siroky@asu.edu, hdavulcu@asu.edu

Abstract

Internet and social media created a public space for online debate on political and social issues. A **debate** is defined as a formal discussion on a set of related issues in which opposing perspectives and arguments are put forward. In this paper, we aim to develop automated perspective discovery techniques which would contribute to the understanding of frames that drive contention between radical right and liberal political actors in Eastern Europe using Slovakia as a case study. Based on computational analyses of their online corpus, we develop an analytical tool to identify contentious frames between ideological opposites and to predict online escalation. A **frame** is the packaging of an element of rhetoric in such a way as to encourage certain interpretations and to discourage others. Our experimental data comprises nearly 10,000 documents downloaded from the official websites of radical right and liberal Slovak political parties spanning a decade between 2004-2014. Our perspective discovery algorithm not only identifies contentious vs. ignored frames, but it also utilizes these frames to predict online escalation with an accuracy of 82.8% (F-measure). We also present qualitative analysis of the resultant frames.

1 Introduction

Hotly debated issues span all spheres of human activity; from liberal vs. conservative politics, to radical vs. counter-radical religious debate, to climate change debate in scientific community. Many prominent 'camps' have emerged within Internet debate [11]. There are many applications [26, 5, 21] for recognizing politically-oriented sentiment in texts.

A **debate** is defined as a formal discussion on a set of related issues in which opposing arguments are put forward. Initially, we observe that given a certain issue, each camp's web sites mostly discuss their own perspectives related to that issue, and occasionally discuss others' perspectives, relating them back to their own perspectives.

While radical right parties in Western Europe often target immigrants, this is less the case in Eastern Europe, where radical right parties tend to mobilize against domestic ethnic minorities that have been settled for a long period of time. Across Eastern and Western Europe, the radical right party family is the fastest growing party family. Radical actors mobilize mass social movements, push for anti-democratic policies, and produce social unrest and violent conflicts that directly impinge on society's stability and sustainability. Our paper examines how radical right party outlets and their "ideological twins" (the proponents of multiculturalism, social liberalism and minority rights) create a dynamic dyad of "radicals" and "liberals". Each member in the dyad feeds off the other in creating its own symbols and contentious frames. **Framing** is accomplished when a choice of words, phrases, metaphors, images, and other rhetorical devices favor one interpretation of a set of facts, and discourage other interpretations. Framing often presents facts in such a way that implicates a problem that is in need of a solution. A special case is *adversarial framing*, which "is typically competitive, fought between parties or ideological factions, where issues are debated and framed in opposing terms" [10].

This actor-oriented approach offers a more dynamic view than previous studies, which have explained radical right mobilization by reference to relatively stable socio-economic structures and legacies of political engagement. The radical right and their "ideological twins" mobilize in bursts that are often short-term and sporadic as shown in Figure 1. This figure illustrates the volumes of documents generated by radical (red line) and liberal (blue line) political party outlets in Slovakia over a ten-year period, between 2004 and 2014. We operationalize the attitudes of the ideological twin actors, liberals and radicals, in Eastern Eu-

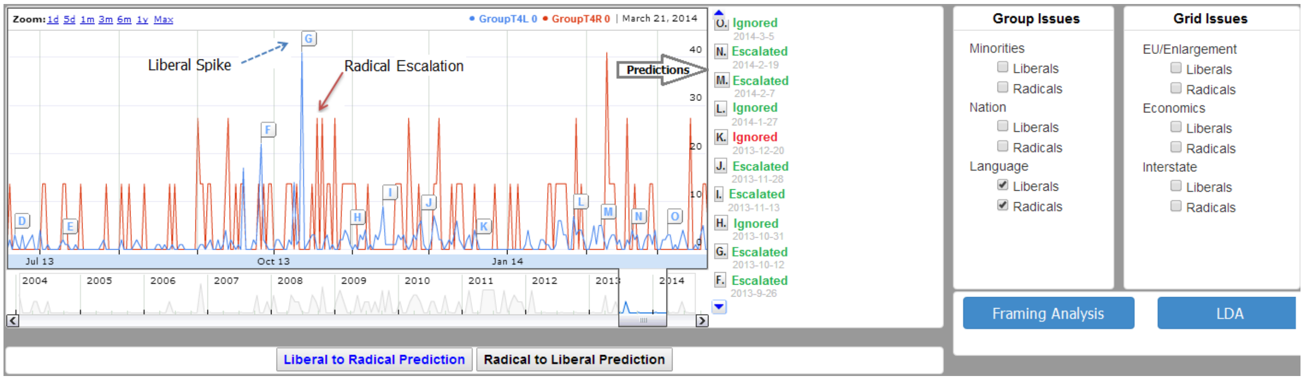

Figure 1: Contentious Frames Analyzer (CFA) Tool

rope using two theoretically derived scales: *grid* and *group*. Grid and group dimensions of the social control theory [13] guided us in selecting political parties to be included in our case study and to identify *key issues* in their corpus. The details of the social control theory are presented in the methods and theory section. Issues such as minorities, nation and language are listed under the 'group' variable and issues like economics, EU/enlargement, and interstate relations are listed under the 'grid' variable.

Not all frames related to an issue resonate within the dyad. The advantage of our approach is that we will be able to parse out seemingly contentious frames that, on the surface, might seem likely to lead to escalation from those frames that might seem innocent at first, but possess a dynamic resonance which serves as a prelude to online escalation. We expect escalation only when frames resonate (i.e., polarize actors). The tool determines which liberal frames resonate with the radicals and which radical frames resonate with the liberals. This classification forms the basis of a predictive model to anticipate when contention leads to escalation. For example, in Figure 1, alphabetic annotations on the timeline correspond to liberal spikes and predictions about their effects; as "escalated" or "ignored" by the radicals. A liberal spike with "contentious" frames is likely to be controversial and cause an escalation, where as a liberal spike with "ignored" frames will likely lead to no response. In the "predictions" panel, green labels indicate a hit and red labels indicate a miss by the classifier.

The rest of the paper is organized as follows. Section 2 presents related work. Problem definition is presented in Section 3. Section 4 presents methodology and theory. Section 5 presents the experimental results and the quantitative analysis. Section 6, presents qualitative analysis of the resultant frames from both camps. The final section concludes the paper and discusses future work.

2 Related Work

For long time, automating and analyzing political perspectives and ideologies by computer algorithms have been quite a focus of research. In 1973, Abelson [1] simulated political ideologies on machines, and in 1978 Carbonell [9] introduced a system to interpret a text relating to a given ideology or political event.

Recently, there has been a lot of work in the area of opinion mining which mostly focuses on mining user reviews about some products such as in [5]. In political context mining, [2] studied topic modeling to predict political orientation of the author of a document. Others used sentiment analysis [14, 27, 4] to predict political orientation of a person (Republican vs. Democratic) or agreement/disagreement on political issues [15, 3, 19].

In this paper, our primary contribution is the development of an automated perspective discovery technique which would contribute to the understanding of frames shared by one side of a debate, and opposed or ignored by the other side. Secondly, we show that, our perspective discovery algorithm not only identifies relevant frames, but it also yields a high accuracy predictor for online escalation.

3 Problem Definition

In this study, we would like to automatically discover the perspectives of opposing political parties on a given set of issues, and identify the underlying contentious frames from one camp that might lead to a debate. Next, we utilize identified frames as features to predict whether a temporal spike (i.e. a relatively higher volume of documents during a fixed period of time) from one camp will trigger a reaction from the other camp. The following general steps describe our methodology:

- Run a simple term frequency - inverse document frequency (TF-IDF) [17] based technique on the entire corpus to generate a large candidate list of keywords for inclusion in grid-group scales. Select the top 100 *n-gram* terms (1 to 3 grams).
- Political scientists on our team scan the list, and select relevant keywords indicating hotly debated *grid-group* issues.
- For each *grid-group* issue we perform the following steps:
  - Run Latent Dirichlet Allocation, LDA [6] over each camp's corpus matching an issue to get their *latent topics*;
  - Detect and label the spikes from each camp as *escalated* or *ignored* (by the other camp) according to the mechanisms described in Section 4.5;
  - Use latent topics and a feature selection algorithm to determine issue-specific discriminative *contentious* vs. *ignored* frames;
  - Use discriminative frames to train a dichotomous classifier as a predictor of online escalation as described in Section 4.6.

4 Methods and Theory

4.1 Overall System Model

We devised an end-to-end pipeline consisting of document collection from political parties' websites, key issues identification and mapping them to *grid-group* scales, issue-specific topic inference for each camp, *escalated/ignored* spikes detection, identification of discriminative frames and a predictive model for *escalation* of online debate.

4.2 Slovak Political Corpus

Our Slovak corpus is comprised of nearly 10,000 news and opinion articles downloaded from the official websites of radical right and liberal political parties spanning a decade of online debate between 2004 and 2014. These organizations were selected by political scientists on our team as representatives from both camps. From the liberal camp, we downloaded Most-Hid (http://www.most-hid.sk/) and SMK - Party of the Hungarian Coalition (http://www.mkp.sk/). From the radical right camp, we downloaded SNS - Slovak National Party (http://www.sns.sk/) and Slovenska Pospolitost - Slovak Brotherhood (https://pospolitost.wordpress.com/). The document corpus, as well as the tools and algorithms presented in the paper will be made publicly available.

4.3 Grid-Group Dimensions

Grid and group dimensions of the social control theory [13] guided us in selecting political parties to be included in our case study and to identify *key issues* in their corpus. The *grid-group* [13] typology characterizes four ideal type modes of social action along a grid-group axis.

Grid dimension captures policy positions of radical actors that are compatible with authoritarianism and social and cultural conservatism. In its pure form, the grid dimension has no ethnic basis. A radical actor scoring high on the social authoritarianism dimension might campaign against accommodating gay and lesbian couples, or against abortion. Similarly, an actor that promotes law and order, along with uncritical obedience to authority, religious or secular, would be classified as a radical grid actor.

Group, the second dimension, captures nationalism and is therefore associated with exclusionary ethnicity-based appeals. It conceptualizes identity in terms of 'the ethnic other' and is grounded in a distinction between the in-group and the out-group. An actor that propagates nationalism on behalf of the titular nationality would qualify as a radical group actor. Core-group ethnocentric, anti-minorities and socially conservative actors are classified as radicals.

The **radical** actors are placed on the extreme ideological poles, which are operationalized as radical nationalism (high group) and radical socio-cultural conservatism (high grid). Radical actors are classified as being high on grid and high on group, or high on one of these two dimensions and 'neutral' on the second dimension. Radical actors are thus either highly nationalistic and/or extremely socially conservative. The **liberal** actors occupy the opposite side of the spectrum, which is characterized by multiculturalism and social-liberalism. This classification logic is consistent with the set of issues that stimulates contentious interactions between the "radicals" and the "liberals". The **issues** involve a wide array of frames ranging from outrages over laws that restrict the language rights of minorities, such as was the case in Ukraine, to outrage over parades of militant paramilitary groups marching in the streets of Hungary. Both of these contentious frames resonate with the ideological twin opposites. In the Ukrainian case, the perception that the rights of the Russian speakers might be curbed, served as one of the catalysts of current support for the separatist movement. In Hungary [7], marches of the Hungarian Guard in September 2012 escalated into attacks on Gypsy/Roma villages and led to deadly clashes.

4.4 LDA Topic Inference for each Camp

Latent Dirichlet Allocation (LDA) [6] is one of the most popular topic inference algorithms that assumes documents are mixture of topics, where a topic is a probability distribution over words, in other words, it uses "bag of words" approach to perform statistical topic modeling, which is a well-established approach to uncover hidden structures in large text corpora. A topic in LDA not only contains the words that describe the topic, but also the words that express the opinions about the topic. We applied LDA separately on both liberal and radical corpora to discover their *latent topics*.

4.5 *Escalated/Ignored* Spikes Detection

We utilize the 68-95-99.7 rule [28] to define a simple method for spike detection. In statistics, the 68-95-99.7 rule, also known as the three-sigma rule or empirical rule, states that in a normal distribution nearly all values lie within three standard deviations (σ) of the mean (μ). We utilize a fixed sized sliding window (experimentally determined as length of 20 weeks), to compute a running average $\mu(20)$ and a standard deviation σ 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 $(\mu(20) + 2\sigma)$. 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.

To categorize each spike as *escalated* or *ignored* we need to define what is a debate. The definition of a **debate** is "a formal discussion on a set of related topics in which opposing perspectives and arguments are put forward" [1]. Based on this definition, spike categorization as (*escalated/ignored*) is based on shared topics between two consecutive spikes from opposing camps. To measure "relatedness" of topics between a pair of consecutive spikes from opposing camps, we utilize Kullback-Leibler (KL) divergence [20] between LDA topics distributions of consecutive spikes. The KL divergence of the probability distributions L, R on a finite set X is defined as:

$$D(L,R) = \sum_{x \in X} L(x) \log \frac{L(x)}{R(x)} \qquad (1)$$

Given two consecutive spikes from opposing camps, liberal spike S_L, and radical spike S_R, we first identify LDA topics of each spike with their distributions within the documents, L is the distributions of S_L topics, and R is the distributions of S_R topics. We then measure the divergence of distributions of topics using the symmetric form of KL-divergence [29] that measures the divergence of the probability distributions L, R on a finite set X of topics as follows:

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

We normalize this measure to be between [0,1] and convert it to a similarity measure as follows:

$$Sim(S_L, S_R) = 1 - D_{normalized}(L,R) \qquad (3)$$

Where $Sim(S_L, S_L)$= 1 means the two distributions of topics between the two spikes are identical.

If the similarity of topics distributions between the two consecutive spikes from opposing camps exceeds a certain threshold, then we label the first spike as "escalated", otherwise it is labeled as "ignored." In section 5, we show the experimentally determined issue-specific thresholds that we used for this similarity measure.

4.6 Framing Analysis and Predicting Online *Escalation*

During a debate on a particular issue, like minority, both radical and liberal camps discuss different perspectives such as "gypsy problem" vs. "minority rights". 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, however, then relate them back to their own perspective.

Once escalated and ignored spikes from one camp are determined, we aim to identify contentious frames that would trigger a debate from the other camp.

[1] Oxford Online Dictionary

Framing is accomplished when a choice of words, phrases, topics and other rhetorical devices favor one interpretation of a set of facts, and discourage other interpretations. A special case is adversarial framing, which "is typically competitive, fought between parties or ideological factions, and [where issues] are debated and framed in opposing terms" [10]. Normally, framing research requires qualitative analysis of a number of texts by subject matter experts to identify and code a set of frames. This is a time-consuming process that does not scale well.

In order to address the scalability problem we resort to sparse learning framework [22], with the underlined motivation to select a subset of discriminating features that can (a) identify contentious and ignored frames and (b) classify a spike as escalated or ignored. The following steps describe our algorithm:

1. For each *key grid-group* issue, run LDA to get latent topics for one camp

2. Filter the frame $\times$ spike matrix to include only the top 2,000 terms representing frames from one camp (100 topics each with 20 terms)

3. Formulate the problem in a general sparse learning framework [22]. In particular, the logistical regression formulation presented below fits this application, since it is a dichotomous spike classification problem (e.g. those spikes from one camp leading to escalation vs. ignored):

$$\min_x \sum_{i=1}^{n} w_i \log(1 + \exp(1 + y_i(x^t a_i + c))) + \lambda|x| \quad (4)$$

In the formula above, 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), and $A = [a_1, a_2, \ldots, a_m]$ 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), and the 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, $|x|_1 = \sum |x_i|$ is 1-norm of the x vector. We used the SLEP [23] sparse learning package that utilizes gradient descent approach to solve the above convex and non-smooth optimization problem. The frames with non-zero values 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 contentions frames, and those with negative polarity correspond to ignored frames that are typically ignored by the other camp.

5 Experimental Evaluation

5.1 Grid/Group Topics

We run a simple term frequency - inverse document frequency (TF-IDF) [17] based technique on the entire corpus to generate a large candidate list of stopword eliminated *n-gram* keywords for inclusion in grid-group scales. We selected the top 100 *n-gram* terms (1 to 3 grams) then translate them to English. Political scientists on our team examined the list as candidate issues. Guided by social control theory [13] political scientists hand-picked following *key issues* and mapped them to the group and grid dimensions, as shown below in Table 1.

Table 1: Grid and Group Issues

Group	**Grid**
Minorities	EU/Enlargement
Nation	Economics
Language	Interstate

5.2 Frames Identification and Quantitative Evaluation for Predicting Escalation

Our framework categorizes spikes as: 1) escalated/ignored liberal spikes, 2) escalated/ignored radical spikes, and identifies contentious frames driving the escalation from each camp.

We detect escalated/ignored spikes as follows:

- Run Mallet [24] over each camp corpus to get their LDA *latent topics* (default setting is 100 topics each with 20 keywords),

- For each issue of the grid and group dimensions, we identified corresponding spikes based on the three-sigma rule,

- Next, each spike is labeled as escalated/ignored based on the relatedness of its topics distributions to the topics distributions of the following spike from the other camp. We used KL divergence to measure distance between two distributions of LDA topics from consecutive spikes from opposing camps. Then, we covert KL divergence to a normalized similarity measure between [0,1].

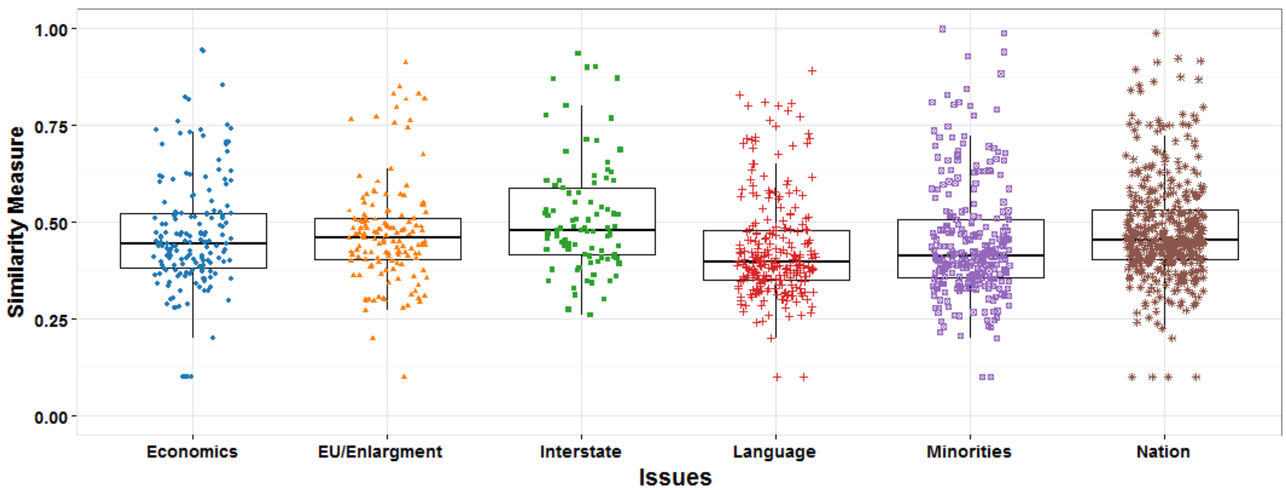

Figure 2: Distributions of Similarity Measures for Liberal Spikes

Figure 2 shows the box-and-whisker plots of similarity measures (y-axis) of liberal spikes, for each grid/group issue (x-axis). Dots represent spikes and boxes represent mean (which varies between 0.35 and 0.45), first and third quartile, and whiskers represent the 95% confidence interval of similarity measures. We use the mean similarity for each grid/group issue as a threshold to determine whether a pair of consecutive spikes is related or not. If similarity between a spike (from one camp) and the following spike (from the other camp) exceeds the mean similarity for an issue, then the first spike is labeled as *escalated*, otherwise it is labeled as *ignored*.

Once spikes from each camp are detected and categorized, we utilize the terms that comprise the topics as features and the sparse logistical regression classifier SLEP [23] to identify discriminative frames. Next, we perform 10-fold cross validation for measuring the classifier's predictive accuracy between escalated/ignored spikes. Qualitative analysis of the identified contentious/ignored frames is also presented in the next section.

Table 2: Predictions for *escalated* liberal spikes

Issue	Dimension	TP	FP	Precision	Recall	F-measure
Minorities	Group	0.79	0.15	0.82	0.79	0.81
Nation	Group	0.95	0.29	0.82	0.95	0.88
Language	Group	0.88	0.19	0.83	0.88	0.86
EU/Enlargement	Grid	0.77	0.12	0.87	0.77	0.82
Economics	Grid	0.86	0.26	0.78	0.86	0.82
Interstate	Grid	**0.92**	**0.25**	**0.86**	**0.92**	**0.89**

Table 2 shows the accuracies for predicting liberal spikes that lead to escalation by the radicals. These accuracies vary between 81% and 89% (F-measure)

Table 3: Predictions for *ignored* liberal spikes

Issue	Dimension	TP	FP	Precision	Recall	F-measure
Minorities	Group	**0.85**	**0.21**	**0.82**	**0.85**	**0.84**
Nation	Group	0.71	0.042	0.92	0.71	0.80
Language	Group	0.81	0.12	0.86	0.81	0.83
EU/Enlargement	Grid	0.87	0.23	0.78	0.87	0.82
Economics	Grid	0.74	0.14	0.82	0.74	0.78
Interstate	Grid	0.75	0.08	0.85	0.75	0.80

Table 4: Predictions for *escalated* radical spikes

Issue	Dimension	TP	FP	Precision	Recall	F-measure
Minorities	Group	0.83	0.13	0.85	0.83	0.84
Nation	Group	0.79	0.19	0.81	0.79	0.80
Language	Group	0.80	0.16	0.84	0.80	0.82
EU/Enlargement	Grid	0.84	0.22	0.83	0.84	0.84
Economics	Grid	0.82	0.15	0.86	0.82	0.84
Interstate	Grid	**0.85**	**0.15**	**0.87**	**0.85**	**0.86**

Table 5: Predictions for *ignored* radical spikes

Issue	Dimension	TP	FP	Precision	Recall	F-measure
Minorities	Group	**0.87**	**0.17**	**0.85**	**0.87**	**0.86**
Nation	Group	0.81	0.20	0.79	0.81	0.80
Language	Group	0.83	0.20	0.80	0.83	0.82
EU/Enlargement	Grid	0.78	0.16	0.79	0.78	0.78
Economics	Grid	0.85	0.18	0.80	0.85	0.82
Interstate	Grid	0.85	0.15	0.83	0.85	0.84

for different issues. Table 3 shows the accuracies for predicting liberal spikes that are ignored by the radicals. These vary between 78% and 84% (F-measure). Average F-measure for predicting outcomes of liberal spikes (as escalated vs. ignored) is 82.9%.

Similarly, Table 4 shows the accuracies for predict-

Table 6: Contentious Frames

Contentious Frames of Liberals	Issues	Contentious Frames of Radicals
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 Rydlo, introduction of euro, Slovak's entry, Carphathian basin
Language law, Czechoslovak republic, World War, collective guilt, Slovak-Hungarian relations, post-war	INTERSTATE	World War II, death penalty, Jozef Rydlo, emergence (origin) of Slovak state, Project "Studnica", anti-Zionism

ing radical spikes that lead to escalation by the liberals. These accuracies vary between 80% and 86% (F-measure) for different issues. And, Table 5 shows the accuracies for predicting radical spikes that are ignored by the liberals. These vary between 78% and 86% (F-measure). Average F-measure for predicting outcomes of radical spikes (as escalated vs. ignored) is 82.7%. The overall accuracy (F-measure) for prediction of online escalation is 82.8%.

6 Qualitative Analysis of Identified Frames

We supplement quantitative evaluations with a qualitative analysis of the identified "contentious frames" in Table 6, which lists issue-specific frames used by each camp that elicit a reaction from the other camp. We compare these with "ignored frames" of each camp, shown in Table 7, and draw general conclusions.

In Slovakia, a key political cleavage between radical and liberal parties is the status of the Hungarian minority, particularly its language rights, residing mostly in southern Slovakia [8, 12, 25]. The left column of Table 6 depicts the Hungarian-Slovak political cleavage over language and Hungarian minority rights clearly. Radicals respond strongly when liberals talk about the "language law", "minority language", and "mother tongue". This finding is consistent with scholarship on Slovak politics and ethnic relations [16, 18].

Slovakia is a home to two ethnic minorities: Hungarians and Roma (Gypsy). Hungarian minority is wealthier, politically organized and spatially concentrated. Roma population is often portrayed as "welfare parasites". It is economically marginalized, socially ostracized, and, contrary to the Hungarian minority, it has never formed a political party that would represent its demands in the Parliament.

The Roma population is both an economic and an ethnic minority. The computational results are remarkable at capturing the differential responses of radicals and liberals to different ethnic and social groups. Radicals, as we would expect [7], respond to frames that push the rights of Hungarian speakers by ethnic pro-Hungarian liberal parties. Liberals however do not respond to the rhetoric of radicals challenging their right to be politically accommodated but to more inflammatory frames linked to Roma, historical memories, legacy of fascism and racism. Scholars of Slovak politics know that radicals attack both Hungarians and Roma but they do not know that liberals, as shown by results here, are more likely to respond to radical frames that are not related to the rights of Hungarians. These computational results thus unmask the strategies of both liberal and radical camps in unexpected and counter-intuitive ways.

This, most likely a strategic response, allows the liberals to frame radicals as fascists and xenophobes, not as political adversaries. Liberals respond when the radicals discuss "protection of the republic," "white race" and the "Gypsy problem". The right column in Table 6 repeatedly shows that liberals re-

Table 7: Ignored Frames

Ignored Frames of Liberals	Issues	Ignored Frames of Radicals
Public Education Act, Lisbon strategy, member states (EU), conference of Bishops	**NATION**	Turkey's accession (to the EU), healthcare, unified SNS, SNS council, Defense Ministry, Department of Transportation, average wage
Ministry of Environment, subsidies, annual budget, judicial council, EU grant	**LANGUAGE**	Serb republic, discussion by SNS, Turkey's accession (to the EU), church provinces, pope Benedict, Jan Hus
Musical institution, EU funds, cultural minorities, theatre, national music, cultural activities, subsidies	**MINORITY**	President of Slovakia, SNS Council, Turkey's accession (to the EU), church provinces, Jan 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 Transporation
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
Archibishop 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

spond when radicals use xenophobic frames, particularly against Gypsies/Roma. A major historical cleavage in Slovakia is also reflected in these results. Liberals respond when radicals attempt to whitewash World War II, during which Slovakia was for the first time an independent state, and to romanticize leaders of the period (such as Jozef Tiso) who collaborated with the Nazis and implemented the Nuremberg laws in Slovakia.

These frames are most clearly evident in the "Nation", "Language" and "Minority" issues, but they also show up under "Interstate (Relations)", "Economy" and "EU". Interstate relations frames point to current tensions between the Hungarian state and the Slovak state over the rights and status of Hungarians in Slovakia along with the historical tensions, since Slovakia once belonged to Hungary before it joined with the Czech lands after World War I to form Czechoslovakia. Radicals respond when liberals discuss and question the uniformly positive version of these issues that the radicals prefer. The "Economy" issue also reflects these aforementioned cleavages between Slovaks and Hungarians in terms of how public finances pensions, healthcare and social insurance - are allocated to minorities, such as Hungarians and Gypsies/Roma. Even when the "EU" issue is discussed, we can see that issues related to these cleavages arise, since the EU applies pressure on member states to conform to certain protections for minorities. The ethnic polarization is further reflected in the liberal response to radicals mentioning the "Carpathian basin", which is a historical reference to the "Greater Hungary" before Slovakia has been formed as a state. The radical responds when liberals discuss "Hungarians in Slovakia" and "ethnically mixed areas" in the context of the EU, which often means that the EU has requested that Slovakia improve its treatment of the Hungarian minority.

Turning to Table 7, we examine the issue-specific frames that were largely ignored by the other side of the political spectrum. Radicals did not respond when these same cleavages were discussed in cultural terms. Under the issue "Minority", for example, radicals did not respond to the frames "cultural minorities", "theater" and "cultural activities." Although these frames sometimes might 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, liberals ignored radical frames under the issues of "Nation", "Language" and "Minority" that focused on religion ("Pope Benedict", "Jan Hus", "church provinces", "Slovak Church"), since religion does not divide Slovaks and Hungarians (both are Catholic) or Roma. Both radicals and liberals also fail to respond in a significant manner when larger EU-related frames are broached, such as the "Lisbon Strategy", "EU funds", "Turkey's accession to the EU", "adoption of the Euro", "the European People's Party" and "the European rail".

What the radicals and liberals ignore is in some cases just as informative as those frames to which they respond. In this case, the results displayed in Table 6 and Table 7 clearly point to the fact that the key points of contention between radical and liberal political parties are domestic rather than European or international. The issues and frames that animate the political debate in Slovakia between these polarized parties are almost exclusively related to three main cleavages: a political cleavage between Slovaks and Hungarians, a social cleavage focused on the Roma and a historical cleavage that concerns the relationship between Nazism and the first independent Slovak state during World War II.

At the domestic level, however, the computational method allows us to expose differential frames of contention between the liberals and the radicals. The analysis shows that radicals polarize Slovak politics by attacking demands for the expansion of minority rights for Hungarians, such as language rights. The liberals, on the contrary, polarize Slovakia by responding to narratives linked to historical legacies of Nazism and contemporary racism but not to narratives that challenge political successes of liberals in expanding minority protection in Slovakia. These findings challenge a simplistic treatment of dynamics of polarization and contention in political science that often assumes that escalation of ethnic tensions can be collapsed into one-dimensional frames.

7 Conclusion and Future Work

Based on a computational analysis of their websites' content, this paper presents a system and an analytical tool to identify contentious frames and predict online escalation between radical right and liberal political parties. We model the type and volume of spikes by these ideological opposites in Eastern Europe using Slovak parties as a case study. We train a predictive classifier to discover contentious and ignored frames of each camp and predict online escalation. We show that the classifier achieves an average predictive accuracy (F-measure) of 82.8%. We also present a qualitative analysis of the identified frames. Contrary to most studies of political extremism that focus on largely static and structural factors, this study derives the dynamic of contention from the online interactions of ideological opposites. Since extremist activities are erratic and highly variable, static approaches that can account for long-term trends often fail at predicting sudden bursts. In our future work, by matching diachronic online media to real-time events data, we plan to build predictive analytics to determine which types of spikes in which contexts precede which types of real world outcomes.

References

[1] R. P. Abelson. Computer models of thought and language, chapter the structure of belief systems. page 287339. W. H. Freeman and Company, 1973.

[2] A. Ahmed and E. P. Xing. Staying informed: Supervised and semi-supervised multi-view topical analysis of ideological perspective. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, EMNLP '10, pages 1140–1150, PA, USA, 2010. Association for Computational Linguistics.

[3] R. Awadallah, M. Ramanath, and G. Weikum. Opinionetit: Understanding the opinions-people network for politically controversial topics. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, pages 2481–2484, NY, USA, 2011. ACM.

[4] R. Balasubramanyan, W. W. Cohen, D. Pierce, and D. P. Redlawsk. What pushes their buttons?: predicting comment polarity from the content of political blog posts. In *Proceedings of the Workshop on Languages in Social Media*, pages 12–19. Association for Computational Linguistics, 2011.

[5] M. Bansal, C. Cardie, and L. Lee. The power of negative thinking: Exploiting label disagreement in the min-cut classification framework. In *Coling 2008: Companion volume: Posters*, pages 15–18. Coling 2008 Organizing Committee, 2008.

[6] D. M. Blei, A. Y. Ng, and M. I. Jordan. Latent dirichlet allocation. *The Journal of machine Learning research*, 3:993–1022, 2003.

[7] L. Bustikova. Revenge of the radical right. *Comparative Political Studies Journal*, 47:1738–1765, 2014.

[8] M. Butora. Nightmares from the past, dreams of the future. *Journal of Democracy*, 18:47–55, 2007.

[9] J. G. Carbonell. Politics: Automated ideological reasoning. *Cognitive Science*, 2(1):27–51, 1978.

[10] D. Chong and J. N. Druckman. A theory of framing and opinion formation in competitive elite environments. *Journal of Communication*, 57(1):99–118, 2007.

[11] L. Dahlberg. The internet and democratic discourse: Exploring the prospects of online deliberative forums extending the public sphere. *Information, Communication & Society*, 4(4):615–633, 2001.

[12] K. Deegan-Krause and T. Haughton. Toward a more useful conceptualization of populism: Types and degrees of populist appeals in the case of slovakia. *Politics & Policy*, 37:821–841, 2009.

[13] M. Douglas. *Natural symbols: Explorations in cosmology*. Routledge, 2004.

[14] M. Efron. The liberal media and right-wing conspiracies: Using cocitation information to estimate political orientation in web documents. In *Proceedings of the 13th ACM International Conference on Information and Knowledge Management*, pages 390–398, NY, USA, 2004. ACM.

[15] V. Etter, J. Herzen, M. Grossglauser, and P. Thiran. Mining democracy. In *Proceedings of the Second ACM Conference on Online Social Networks*, COSN '14, pages 1–12, NY, USA, 2014. ACM.

[16] O. Gyarfasova and G. Meseznikov. Actors, agenda, and appeal of the radical nationalist right in slovakia. In M. Minkenberg, editor, *Transforming the Transformation?*, pages 224–249. 2015.

[17] J. A. Hartigan and M. A. Wong. Algorithm AS 136: A k-means clustering algorithm. *Applied Statistics*, 28(1):100–108, 1979.

[18] T. Haughton and M. Ryba. A change in direction: The 2006 parliamentary elections and party politics in slovakia. *Journal of Communist Studies and Transition Politics*, 24:232–255, 2008.

[19] M. Kaschesky, P. Sobkowicz, and G. Bouchard. Opinion mining in social media: Modeling, simulating, and visualizing political opinion formation in the web. In *Proceedings of the 12th Annual International Digital Government Research Conference*, pages 317–326, NY, USA, 2011. ACM.

[20] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 03 1951.

[21] W.-H. Lin and A. Hauptmann. Are these documents written from different perspectives?: A test of different perspectives based on statistical distribution divergence. In *Proceedings of the 21st International Conference on Computational Linguistics*, ACL-44, pages 1057–1064, PA, USA, 2006. Association for Computational Linguistics.

[22] J. Liu, J. Chen, and J. Ye. Large-scale sparse logistic regression. In *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '09, pages 547–556, NY, USA, 2009. ACM.

[23] J. Liu, S. Ji, and J. Ye. *SLEP: Sparse Learning with Efficient Projections*. Arizona State University, 2009.

[24] A. K. McCallum. Mallet: A machine learning for language toolkit. http://mallet.cs.umass.edu, 2002.

[25] G. Meseznikov, O. Gyarfasova, and D. Smilov. *Populist Politics and Liberal Democracy in Central and Eastern Europe*. IVO (IPA) Working Paper Series, Bratislava, 2008.

[26] T. Mullen and R. Malouf. A preliminary investigation into sentiment analysis of informal political discourse. In *Proceedings of AAAI-2006 Spring Symposium on Computational Approaches to Analyzing Weblogs*, 2006.

[27] S. Park, M. Ko, J. Kim, Y. Liu, and J. Song. The politics of comments: Predicting political orientation of news stories with commenters' sentiment patterns. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, pages 113–122, NY, USA, 2011. ACM.

[28] F. Pukelsheim. The three sigma rule. *The American Statistician*, 48(2):88–91, 1994.

[29] A.-K. Seghouane and S.-I. Amari. The aic criterion and symmetrizing the kullback ndash;leibler divergence. *Neural Networks, IEEE Transactions on*, 18(1):97–106, Jan 2007.