Whose fingerprint does the news show? Developing machine learning classifiers for automatically identifying Russian state-funded news in Serbia

Denkovski, O.; Trilling, D.

International Journal of Communication 14(2020), 4428–4452 1932–8036/20200005

Whose Fingerprint Does the News Show? Developing Machine Learning Classifiers for Automatically Identifying Russian State-Funded News in Serbia

OGNJAN DENKOVSKI[1]
DAMIAN TRILLING
University of Amsterdam, The Netherlands

Democratic nations around the globe are facing increasing levels of false and misleading information circulating on social media and news websites, propagating alternative sociopolitical realities. One of the most innovative actors in this process has been the Russian state, whose disinformation campaigns have influenced elections and shaped political discourse globally. A key element of these campaigns is the content produced by state-funded outlets like RT and Sputnik, whose articles are republished by underfunded or sympathetic local media, as well as coordinated groups that attempt to shape mainstream political narratives. Using a tailored text-as-data approach, we examine the thematic and linguistic differences in articles produced by U.S. and Russian state-funded and mainstream outlets in Serbia. We use 11 features (frames and in-text characteristics) to construct an article country-source classifier with a high degree of accuracy. The article contributes toward an understanding of the structural characteristics of Russian state-funded news in the Western Balkans, enhances the application of computational text analysis in Serbian, and provides suggestions for the application of text-as-data methods to the study of online disinformation.

Keywords: disinformation, computational text analysis, text classification, automated frame identification, fragmented audience, Western Balkans

The creation and spread of misleading, polarizing, or false information has become an increasingly prominent topic in academia and public debate (e.g., Bradshaw & Howard, 2018; Galante & Ee, 2018). In comparison with traditional media systems, characterized by few channels of information flow in which "effective gatekeeping against wild or dangerous narratives" (Bennett & Livingston, p. 128) is possible, social media platforms and online news have made it increasingly difficult to control the development of local or global narratives (Bennett & Livingston, 2018). In this environment, the creation and spread of news content have become largely horizontal processes, allowing for unverified news to easily reach millions across the globe

Ognjan Denkovski: o.denkovski@uva.nl
Damian Trilling: d.c.trilling@uva.nl
Date submitted: 2019–11–16

[1] We thank the University of Amsterdam Digital Communication Methods Lab for funding this research.

(Nemr & Gangware, 2019). Recently, an increasing number of systematic efforts to make use of these developments have been tied to state actors, generally driven by geopolitical goals with narratives targeting radical segments of foreign populations. Many argue that states such as Russia, China, Iran, and Turkey are among the most innovative in this process (Janda, 2016; Polyakova & Boyer, 2018; Prague Security Studies Institute, 2019). Our study uses the case of Russian state-funded outlets in Serbia, frequently described as disinformation tools for the Russian government, to examine the possibility for automatically detecting foreign state-sponsored content and with that foreign disinformation campaigns ("Disinformation Analysis," 2018).

As long as content published on a website or Facebook group is branded as, for instance, Sputnik, linking this content to (Russian) state funding is trivial. But can this content be traced back to a source through linguistic and content based in-text characteristics when the source is not explicitly mentioned? Answering this question can (1) enhance our understanding of what characterizes potentially malicious state-sponsored content, (2) show whether such content can be distinguished from content from other news sources in a meaningful way, and (3) pave the way for developing automated systems that recognize content from specific (state) actors where the source is obscured. Automatically detecting this content can contribute toward the identification of, for instance, extremist social media communities, such as the Sputnik-linked group recently shut down by Facebook for spreading anti-NATO propaganda (Waterson, 2019), as well as help monitor for the development of new narratives by state-funded outlets associated with disinformation campaigns (Cerulus, 2019; Woolley & Howard, 2016). The relevance of these insights becomes particularly evident when considering the broad set of challenges faced by researchers focused on developing approaches for identifying disinformation while relying on measures of content veracity (Hjorth & Adler-Nissen, 2019).

We propose an alternative approach by combining text-as-data methods, machine learning, and manual content analysis to create and compare profiles of articles from Russian state-funded outlets and U.S. state-funded and mainstream outlets. We consider 11 explanatory variables grouped in two feature sets—namely, issue-specific frames and linguistic properties. These features are used to answer two questions:

1) *Is it possible to create distinct profiles of U.S. and Russian state-funded news?*

2) *Can these differences be used to automatically distinguish the source of a news article?*

Serbia is a particularly adequate case for the study of Russian state-funded news due to the low media freedom rates within the country, the existing pro-Russian sentiment, and the wide reach of Serbian media in the region (Klepo, 2017; Prague Security Studies Institute, 2019). While the study is based in Serbia, the methodology and theoretical framework proposed are applicable to a broader context (Bennett & Livingston, 2018).

Theoretical Framework, Context, and Feature Selection

What Is Disinformation?

The ease of producing and spreading content online, combined with the increasing popular demand for alternative interpretations of political and social events, has led to the development of a wide range of "alternative" news sources that promote ethnic nationalism and antiglobalist conspiracies, as well as for-

profit fake or sensationalist news (Bandeira, Barojan, Braga, Peñarredonda, & Argüello, 2019; Beam, Hutchens, & Hmielowski, 2018; Bennett & Livingston, 2018; Garimella & Weber, 2017). The various actors and diversity of content produced has inspired a wide range of concepts and frameworks for describing information that misleads, deceives, and polarizes, including fake news, hyperpartisan content, junk news, and clickbait, to name a few. Not all of these terms are equally useful.

Two well-defined concepts that offer a theoretical possibility for content evaluation with automated approaches are misinformation and disinformation, the creation of which is distinct from what one traditionally expects from information published in journalistic media (Bennett & Livingston, 2018; Jackson, 2017). Misinformation describes the "inadvertent" process of sharing "false information" and is often tied to reckless journalism, but it is not intended to cause harm, nor is it produced in a coordinated fashion. Disinformation refers to the "purposeful dissemination of false information" (Nemr & Gangware, 2019, p. 4) created with the intention to mislead, harm, or promote political interests (Bennett & Livingston, 2018). This definition of disinformation implies an organized and continuous process based on the development of a legitimate following with political, cultural, and emotionally charged content. Such a process also includes making use of this following during pivotal events, while relying on a diverse range of ideological positions targeting numerous and sometimes ideologically opposed social groups (Bradshaw & Howard, 2018; Galante & Ee, 2018; Hjorth & Adler-Nissen, 2019).

Outlets involved in disinformation campaigns can be assumed to follow distinct news production patterns where the primary goal is not the veracity or objectivity of news items (Altheide, 2004; Bennett & Livingston, 2018; Bradshaw & Howard, 2018). However, past research demonstrates that identifying disinformation-like content through content veracity with automated methods is a challenging task largely due to the subtleties of content manipulation (e.g., Hjorth & Adler-Nissen, 2019). We propose a method that relies on characteristics of content selection and creation for automatically distinguishing news from a particular state-funded source. We demonstrate that such an approach could be further refined for identifying disinformation campaigns from state-funded or private organizations, both in the Western Balkan region and further abroad.

As the large body of literature on automated text analysis consistently shows (see, for instance, the literature review on tweet analysis by Orellana-Rodriguez & Keane, 2018), both linguistic and substantial features can be of importance. Linguistic features, such as the length of a text or its complexity, may be proxies for a more latent concept (such as, for instance, being of Russian origin). These features are indicators, but are not necessarily substantially related to the categories of interest. Substantial features refer to the substance of content—for instance, the applied framing. In contrast to linguistic features like the length of a text, an anti-Western frame for instance, has a direct, substantial link with the classes we try to distinguish. We are interested in both groups of features as well as how important they are relative to each other. Specifically, we examine whether issue-specific frames and linguistic properties of text can be used to distinguish between content from different sources, with a frame selection based on content that we expect to characterize Russian disinformation in Serbia and abroad. Therefore, we ask the following:

RQ1: Can statistically significant differences be observed in news production routines and linguistic habits (issue-specific frames and linguistic features) of Russian and U.S. outlets in Serbia?

Five outlets are analyzed: N1, Radio Slobodna Evropa (RSE), and Glas Amerike (VOA) representing U.S. mainstream and state-funded outlets, and Sputnik Serbia and Vostok Vesti representing Russian state-funded outlets. N1 is a regional informative partner of CNN with ≈200,000 followers on Facebook, and while not funded by the U.S. government, it does represent news production routines of mainstream U.S. media and adds to the generalizability of the study findings. RSE and VOA are U.S. state-funded outlets with ≈215,000 and ≈125,000 followers on Facebook, respectively (Klepo, 2017). Sputnik Serbia is the official Serbian branch of Sputnik, whereas Vostok, whose funding and ownership are not disclosed, republishes RT and Sputnik international content in Serbian, with ≈125,000 and ≈80,000 followers on Facebook, respectively (Klepo, 2017). These outlets were chosen based on their representativeness of content creation and selection routines by outlets in Serbia from the countries of interest and the availability of data.

Disinformation Across Media Systems: The Case of Serbia

State-funded media such as the BBC World Service and the U.S. Information Agency have long been integral parts of foreign public diplomacy, shaping how these countries and their actions are perceived abroad (Cull, 2009; Yablokov, 2015). Not unlike partisan media, state-funded outlets aim to frame the perception of state-relevant issues among foreign audiences, while also defining what issues should be perceived as relevant, thus setting the agenda of public discussions (de Vreese, 2005; Entman, 1993; Yablokov, 2015). Qatari-owned Al-Jazeera is an excellent example of the substantial reach achieved by some state-funded outlets, despite concerns regarding its independence and political impartiality (Fahmy & Al Emad, 2011; Miles, 2010; Robinson, 2005). However, Russian state-funded outlets have often stood out among their competitors. In two interviews in 2012 and 2013, editor-in-chief Margarita Simonyan discussed RT's role in the "information war" against the "whole Western world," noting how "information weapon[s]" and audiences should be used in "critical time[s]" (as cited in Nimmo, 2018). In practice, this goal has often resulted in disinformation campaigns based on false and misleading information presented with highly charged language. This explicit use of disinformation campaigns as part of a broader operational logic of (Russian) state-funded outlets provides the grounds for the current study.

The repercussions of disinformation manufactured by Russian state-funded outlets are tangible in stable Western democracies, as illustrated by the recent establishment of both EU and NATO task forces for countering Russian disinformation (Janda, 2016). The need for this preparedness became apparent following the Lisa case, as 700 protestors of Russian and German origin massed in front of Chancellor Merkel's chancellery due to a rape incident that was ultimately shown to be fabricated, but which was extensively covered by RT and Sputnik (Baade, 2019). In Western Balkan countries, where "disinformation and fake news is the norm" (Bechev, 2018, p. 4), public opinion is particularly at risk (Denkovski, forthcoming; Klepo, 2017).

In Serbia, critical media are largely underfunded and often under threat of violence, while government sympathetic outlets dominate the news agenda. This situation partially motivated Serbia's latest protests, ongoing since November 2018 (and suspended as of March 2020 because of the coronavirus pandemic)—the most significant since the fall of President Milošević, yet largely neglected by public broadcaster RTS (Fidanovski, 2019; "One in Five," 2019). At the same time, journalists in Serbia are facing many of the issues faced by journalists globally, such as a decrease in advertising revenue, an increased

demand for rapid content production, and news production routines based on for-profit reasoning (Andresen, Hoxha, & Godole, 2017; Schauster, Ferrucci, & Neill, 2016). This media environment, combined with a polarized Western- or Eastern-oriented public opinion, as well as the existing anti-Western attitudes among Serbs, makes Serbia an "ideal" space for the success of geopolitical narratives from foreign actors (Eisentraut & de Leon, 2018; Stronsky & Himes, 2019).

Russian media has been particularly successful in this respect. Research from the U.S. Senate Foreign Affairs Committee shows that Russian popularity among Serbs increased from 47.8% in 2015, the year Sputnik Serbia was launched, to 60% in June 2017 (Prague Security Studies Institute, 2019). President Putin's popularity in Serbia, whom government-friendly media portray as a supporter of Serbia's claim to Kosovo, is second only to that of President Vučič ("Disinformation Analysis," 2018). The success of Russian media can be explained by two factors: the existing (or at least perceived) connection between Russia and Serbia combined with anti-Western attitudes in the population, but perhaps more relevantly, the free-for-all policy of Russian outlets, which do not charge a fee for republishing their content. An increasing number of sympathetic, underfunded or for-profit local outlets have made use of this possibility and either actively republish content from Russian outlets in Serbia or base their own reporting on this content (Stronsky & Himes, 2019). One study suggested that one-third of local outlets publish articles about international actors without noting sources or authors, much of which based on pro-Russian and anti-Western attitudes ("Analiza," 2018). However, U.S. state-funded and mainstream media also maintain a strong presence in Serbia, with outlets such as Radio Slobodna Evropa (RSE), Glas Amerike (VOA), and N1 reaching sizable audiences (Klepo, 2017).

A qualitative analysis of the coverage by N1 and Sputnik Serbia showed "noticeable differences in the selection of topics and interlocutors" (p. 3), as well as ideological stances toward the local government, foreign-policy, and social issues reflecting the existing geopolitical divisions in the country (Klepo, 2017). Several studies have since reached similar conclusions (Bechev, 2018; Eisentraut & de Leon, 2018; Stronsky & Himes, 2019). Building on these findings, we test 11 theoretically relevant features for their potential to distinguish between U.S. and Russian state-funded news.

Which Features Characterize (Russian) State-Funded News?

Recent advances in computational text analysis offer numerous possibilities that can contribute to automated identification of content linked to disinformation (Burscher, Odijk, Vliegenthart, de Rijke, & de Vreese, 2014; Conroy, Rubin, & Chen, 2015; Jain, Sharma, & Kaushal, 2016). The literature suggests two relevant feature types that are applicable to the current study: issue-specific frames and linguistic properties. The latter are extracted automatically, while to obtain the former, a small manually coded data set is used to train a supervised machine learning classifier that predicts frame presence in the entire data set. For this task, we build on the holistic frame identification method proposed by Burscher et al. (2014), which uses human "yes/no" responses to multiple indicator questions for each frame.

Through identifying issue-specific frames, we aim to capture common practices and routines in reporting about the examined topics (de Vreese, 2005, p. 54; Matthes & Koring, 2008). The issue-specific frames examined in the study build on past findings regarding Russian narratives in Serbia and abroad, such

as anti-Western sentiment, Russian opposition to Kosovo's independence, and the promotion of a pan-Slavic/Orthodox connection between Russia and Serbia (Bechev, 2018; Eisentraut & de Leon, 2018; Klepo, 2017). The first issue-specific frame, Serbian victimhood, refers to the presentation of Serbia as a victim of Western geopolitics, through references to the NATO bombing of Serbia or the "loss" of Kosovo (Bechev, 2018; Eisentraut & de Leon, 2018). The anti-West frame refers to the presentation of Western actors as a threat to Serbia or the "global order" (Eisentraut & de Leon, 2018; Klepo, 2017). The pro-Russian frame emphasizes the economic, political, and cultural ties between Russia and Serbia (Stronsky & Himes, 2019). Finally, the Russian might frame emphasizes the relevance of Russia in global geopolitics as well as Russia's zeal to defend Serbian interests (Bechev, 2018; Stronsky & Himes, 2019). We expect these features to be significantly more prevalent in content from Russian state-funded outlets.

The second feature set captures variations in linguistic habits of journalists, building on natural language processing (NLP) research, which demonstrates that linguistic properties can reflect author ideology (Schoonvelde, Brosius, Schumacher, & Bakker, 2019). For instance, Brundidge, Reid, Choi, and Muddiman (2014) demonstrate that conservative political bloggers in the U.S. use simpler language than their liberal counterparts. Also, Horne and Adali (2017) demonstrate that fake news has longer titles and uses simpler, more repetitive content in the text body, whereas Chakraborty, Paranjape, Kakarla, and Ganguly (2016) demonstrate that clickbait content can be distinguished from legitimate news with features such as word length and common audience bait phrases. We use this literature to select several features that can be useful for identifying content designed for virality (e.g., longer titles and simpler language), potentially characterizing disinformation content based on our expectations of the routines behind disinformation manufacturing and the desired reach of the content. Working within the possibilities of automated text analysis in Serbian, we test the utility of seven linguistic properties: title length, article length, ratio of unique words, ratio of substantive (not stop-) words in the article text, ratio of substantive words in the article title, average word length, and named entities in the article texts (e.g., references to actors and institutions).

Table 1. All Features (N = 11).

Issue-specific frames	Linguistic features
Serbian victimhood	Named entities
Anti-West	Article length (in words)
Pro-Russian	Title length (in characters)
Russian might	Ratio of unique words in article text
	Ratio of substantive words in article text
	Ratio of substantive words in article title
	Average word length in text (in characters)

Table 1 shows all features used in the study, grouped by feature set. Together, these features represent a theoretically supported set of metrics for creating distinct profiles of Russian and U.S. news. The issue-specific frames attempt to capture distinctions in issue-specific institutional practices, while the linguistic features capture stable linguistic habits of journalists from outlets tied to each country. The study tests the utility of both sets and any combination thereof for the automated distinction of article country source.

RQ2: Which feature set (issue-specific frames, linguistic properties, or combination thereof) can best inform the automated distinction of Russian and U.S. news?

Method

We scraped news articles from the sites under study and conducted a manual content analysis (N = 1,108) to get training and test data which we used to develop supervised machine learning (SML) classifiers for frame prediction in the entire data set (N = 10,132). We then attempted to automatically classify the articles on the basis of country source. All data scraping was conducted in Python with Selenium and BeautifulSoup (Nair, 2014; Salunke, 2014). Data processing was conducted with Pandas, NumPy, and NLTK (Bird, Klein, & Loper, 2009; McKinney, 2011; van der Walt, Colbert, & Varoquaux, 2011). Tasks specific to the Serbian language were conducted with Polyglot and Transliterate (Al-Rfou, Perozzi, & Skiena, 2013; Qian, Hollingshead, Yoon, Kim, & Sproat, 2010). We used Scikit-Learn for all machine learning tasks (Pedregosa et al., 2011). All scripts and resources related to the project, including a custom Serbian stop word list and the codebook, are available (https://github.com/OgnjanD/Whose-Fingerprint-does-the-news-show---Online-Appendix).

The data set covers the period between November 1, 2018, and February 1, 2019, with 5,860 (57%) of the articles from U.S. outlets and 4,272 (43%) from Russian outlets. We only considered categories such as world news, politics, economy, and culture, but not "soft" categories such as sports. All articles from these categories are scraped for the time period studied.

Frame Prediction With Supervised Machine Learning and Manual Content Analysis

Frames can be identified top-down or bottom-up: One can either predefine frames or find patterns in the data and interpret them as frames. This is true both for manual content analysis (see Matthes & Kohring, 2008) and for machine learning, where these approaches would correspond to supervised and unsupervised methods. Recently, unsupervised machine learning, in particular so-called topic models, such as the latent Dirichlet allocation (LDA), has become popular to analyze frames (e.g., Heidenreich, Lind, Eberl, & Boomgaarden, 2019; Tsur, Calacci, & Lazer, 2015). However, given that the frames we aim to identify are theoretically motivated (and hence defined a priori), such a bottom-up method is not appropriate and a supervised approach is preferred (Boumans & Trilling, 2016). Besides, it is a debated issue whether topic models actually capture topics or frames (Maier et al., 2018; van Atteveldt, Welbers, Jacobi, & Vliegenthart, 2014).

Burscher et al. (2014) show that the best way to teach supervised classifiers to recognize and predict frames is the holistic approach, where human coders identify the presence or absence of frames through "yes/no" responses to a set of indicator questions associated with each frame. Each indicator question is taken as representative of different but equally important aspects of a frame, and a "yes" response to at least one indicator question indicates the presence of a frame (Burscher et al., 2014).

"Yes/no" responses to these questions for each article in the manually annotated set (*N* = 1,108) are supplied to the SML classifier to replicate human decisions by establishing associations between properties of text and human-indicated frame presence or absence. These classifiers were then used to predict frame presence in the remaining articles (*N* = 9,024). Manual coding was conducted by three recent graduates in communication science and related disciplines, and the first author, all of whom are fluent in Serbian. Coders underwent two training sessions and received feedback thrice, after which each coder was provided with a double stratified sample accounting for both country source and month of publication. Intercoder reliability was assessed with a double-stratified sample (*N* = 31). All variables had satisfactory reliability scores presented in Table 2.

Table 2. Intercoder Reliability Scores.

Frame	Krippendorff's α
Serbian victimhood	.80
Anti-West	.78
Pro-Russian	1
Russian might	.69

Before training the SML classifiers, we removed stop words and punctuation. Then, all remaining words were stemmed and converted to lowercase. We used a customized list of 251 stop words, obtained by combining an existing Serbian stop word list developed for NLTK research and a new list developed on the basis of the most commonly occurring stop words in the study data (Champion, 2019). For stemming, the study relies on a custom Serbian stemmer (Milosevic, 2012).

Evaluating Classifier Performance for Frame and Country Source Prediction

We used scikit-learn's grid search to test 20 model–hyperparameter combinations to determine the optimal classifier for predicting each frame. The performance of the optimal model–hyperparameter combination is tested in combination with both a simple count vectorizer and a tf–idf vectorizer. Using five-fold cross-validation, we evaluated the weighted f1 score of each model–hyperparameter combination, the weighted harmonic mean of precision and recall. Precision indicates the percentage of correctly identified cases (true positives), while recall indicates how many of the cases that should have been identified were actually identified (false negatives). Because of the imbalanced data set resulting from the manual coding, we use a combination of synthetic over- and undersampling with SMOTE+ENN when creating the training set for frame classification. The SMOTE part of the process generates synthetic samples that are used in combination with the real samples, while ENN smoothens the borders of the synthetic cases generated (Lemaître, Nogueira, & Aridas, 2017). This procedure allows for improved recall and precision scores with highly imbalanced data, such as our own.

The weighted f1 score accounts for the proportion of "yes/no" responses when producing the averaged metric, ranging from zero to one. We considered classifiers that perform with a weighted f1 score

greater than 70% as fit for frame prediction on unlabeled data (see also Burscher et al., 2014; Grimmer & Stewart, 2013).

Finally, the manually coded and automatically predicted frames are combined with the automatically extracted linguistic features to predict the country source of an article. First, the variance of each feature is examined to eliminate those features that do not vary significantly across the two country sources and that consequently are not likely to contribute to the country source classification task. The remaining features are used for country-source classification with a set of 119 model–hyperparameter combinations and five-fold cross-validation. To optimize classifier performance, seven feature combinations are considered: all remaining features, each feature set individually, and four feature combinations suggested by selection techniques available in scikit-learn. The feature selection techniques used are univariate feature reduction, recursive feature elimination (RFE), random forest decision trees, and variance analysis (Brownlee, 2016; Khalid, Khalil, & Nasreen, 2014).

Results

We report three distinct analyses. First, we show the distribution of frames per country source based on manually annotated data and the implications of this distribution for the automated frame classification task, after which we evaluate the performance of frame prediction classifiers. In the second step, we test the variance of the features across U.S. and Russian articles, demonstrating the distinctiveness of the profiles generated with the features. We use the features for which variance across U.S. and Russian news is significant for the development of country source classifiers, answering RQ1. At the same time, we evaluate the best performing feature combination and answer RQ2.

Manually Coded Frame Distribution and Supervised Machine Learning Frame Prediction

Figure 1 shows the distribution of manually coded "yes" responses for each frame across the two data sets as a proportion of total articles in each. With the exception of the Serbian victimhood frame, all issue-specific frames are unique to Russian news items, with the anti-West frame present in 17.3% of Russian articles. This distribution suggests that all issue-specific frames can help distinguish between the two data sets.

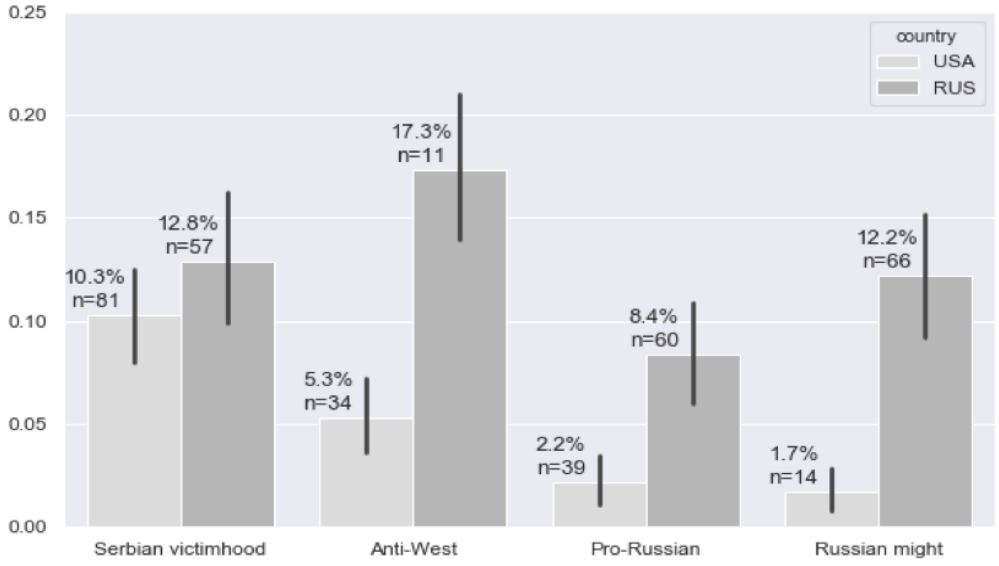

Figure 1. Proportional frame distribution by country source in manually coded data (N = 1,108). Error bars show 95% confidence intervals.

The optimal classifier score for each frame is presented in Table 3, which shows that all frames meet the 70% cutoff for weighted f1 scores. The high f1 scores of the frames are qualified by their precision and recall rates. Though precision varies from 60% to 99%, recall rates for some frames are very poor, with that of the Russian might frame being only 38%. These scores are explained by the distribution of frames in Figure 1, which shows that frame responses are highly unbalanced, with each frame receiving significantly more "no" responses. Thus, while precision rates for "yes" responses for issue-specific frames range between 60% and 73%, indicating that the classifiers can identify the characteristics of frames, the yes recall rates are in the range of 38% to 73%, meaning that the number of false negative responses varies per frame.

Table 3. Optimal Classifier Performance for Frame Prediction.

Frame	Optimal classifier	Weighted f1	"Yes" precision	"Yes" recall	"No" precision	"No" recall
Serbian Vic.	Random forest	89%	67%	48%	93%	96%
Anti-West	Random forest	91%	67%	22%	94%	99%
Pro-Russian	Logistic	97%	73%	73%	99%	99%
Russian might	Random forest	93%	60%	38%	95%	98%

Feature Variance for Country Source Prediction

The four remaining frames are combined with the automatically extracted linguistic features, a total of 11 features characterizing 10,132 articles. To answer RQ1 and to inform the feature selection for country source classification, we first investigate whether there are features that occur equally in both groups and which are thus unlikely to have discriminative power for article source classification (Horne & Adali, 2017). Table 4 shows that the distribution of all frames varies significantly across U.S. and Russian news, as does that of all the linguistic features examined (see Table 5). Based on these findings, all 11 features are considered for country source prediction.

Table 4. Chi Square (χ^2) Scores–Frame Distribution by Country Source.

Frame	$\chi^2(2, N = 10{,}132)$
Serbian victimhood	43.26***
Anti-West	104.08***
Pro-Russian	54.99***
Russian might	177.25***

***$p < .001$.

Table 5. Wilcoxon Rank Sum Test for Linguistic Features.

Feature	z score
Article length	−23.33***
Title length	−14.90***
Ratio substantive words–text	−17.74***
Ratio substantive words–title	−17.73***
Average word length	−27.48***
Ratio unique words	−17.34***

***$p < .001$.

The distributions of these features are shown in Figures 2 and 3. As suggested by the unbalanced distribution of responses in the manual data and by the performance of the classifiers, "yes" response rates are underrepresented in the entire data set, as classifiers conservatively assign "yes" responses, even with the SMOTE + EEN synthetic sample. The standardized values of the linguistic features show that U.S. articles are on average longer, while using simpler and shorter words. Russian articles have longer titles, as well as a higher ratio of unique, longer, and substantive words.

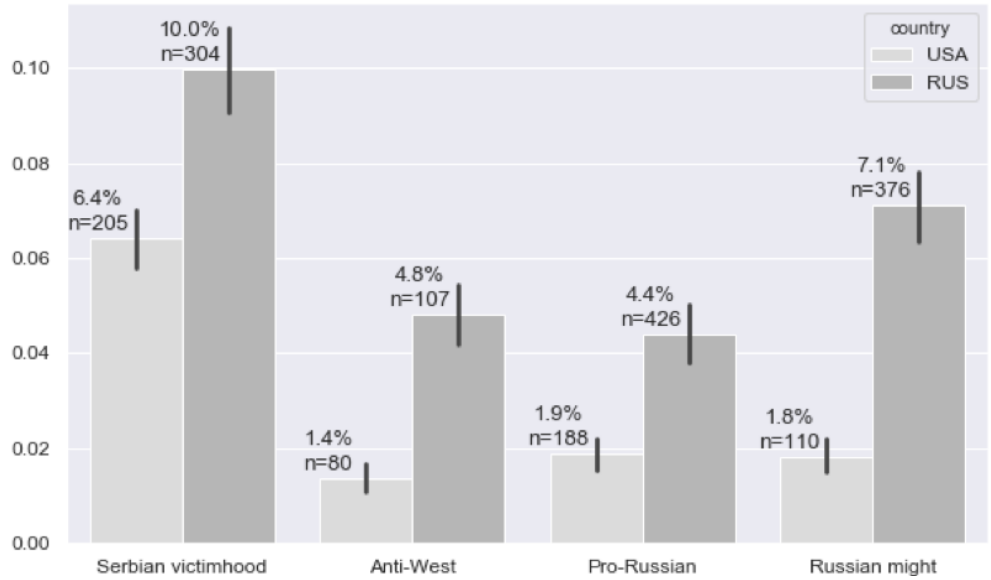

Figure 2. Proportional frame distribution by country source in automatically and manually coded data (N = 10,132). Error bars show 95% confidence intervals.

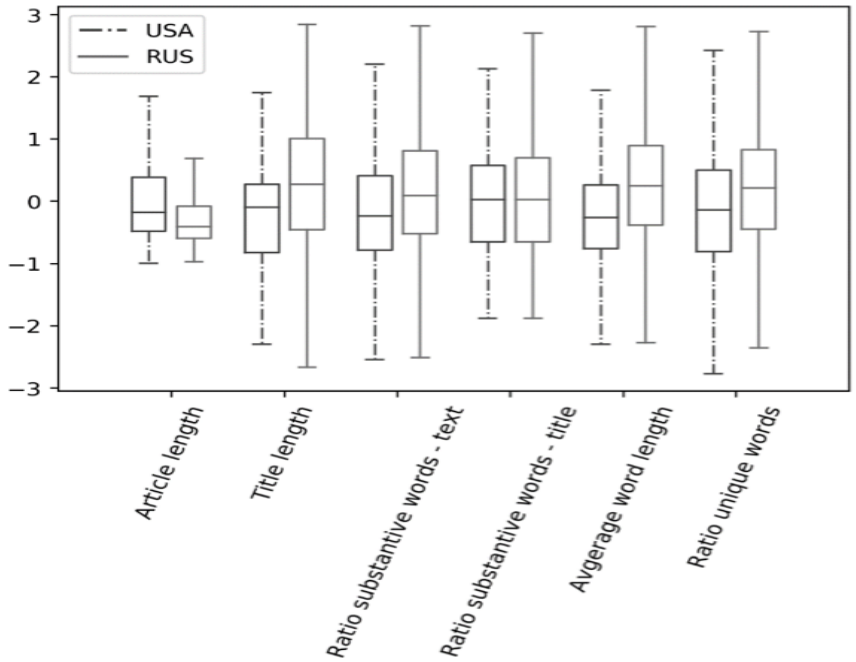

Figure 3. Automated features by country source, z score (N = 10,132). Whiskers show the range of the lower and upper 25% of data.

Finally, Figures 4a and 4b show word cloud renderings of the named entities used in article texts, demonstrating significant differences across articles from both countries. The renderings show the most discussed entities by outlets from each country, with entity size corresponding to the frequency of mention.

Figure 4a. Word cloud representation of named entities in U.S. articles (N = 10,132).

Figure 4b. Word cloud representation of named entities in Russian articles (N = 10,132).

The entities show that U.S. news in Serbia has a strong focus on Serbia and Serbian affairs as indicated by references to Serbia ["srbija/srbiji/srbije"], Belgrade ["beograd"], and President Vučić. In comparison, whereas Russian news also reports on Serbia ["srbije/srbiji"], their articles extensively discuss the U.S. ["sad"], the EU["eu"], Russia ["rusije/rusija"], and Kosovo["kosova"]. Figures 4a and 4b show the substantive differences in actor focus, an observation confirmed by an analysis of the five most frequently named entities in articles from both countries, shown in Table 6. The Online Appendix contains a listing of the 15 most common entities in articles (https://github.com/OgnjanD/Whose-Fingerprint-does-the-news-show---Online-Appendix).

Table 6. Top Five Named Entities per Country Source.

Five most referenced entities–U.S. news	Frequency	Five most referenced entities–Russian news	Frequency
Serbia ["srbija"]	1,698	USA ["sad"]	1,109
USA ["sad"]	1,241	Serbia [srbija"]	1061
Serbia ["srbiji"]	1,017	Russia ["rusije"]	869
Serbia ["srbije"]	839	Russia ["rusija"]	607
EU ["eu"]	815	Kosovo ["kosova"]	594

The analysis of human coded data, frame prediction on the basis of this data, and feature variance all indicate that the 11 features can be used to develop distinct Russian and U.S. profiles. The analyses suggest that the linguistic features article length, average word length, and named entities can be particularly informative, along with the pro-Russian, anti-West, and Russian might frames.

Country Source Classification

In the development of a country source classifier, the study tests the classifying ability of seven combinations of features—namely, all 11 remaining features, each feature set individually, and four feature combinations suggested by feature selection techniques available in the scikit-learn framework (Khalid et al., 2014). The output of the feature selection analyses is presented in Table 7.

Table 7. Features Selected by Feature Selection Analyses.

Feature selection output			
Univariate (*N* = 3)	Recursive (*N* = 3)	Random forest (*N* = 5)	Variance threshold (*N* = 7)
Article length	Article length	Avg. word length	Article length
Anti-West frame	Avg. word length	Ratio substantive words	Avg. word length
Russian might frame	Title length	Article length	Title length
		Ratio unique words	All frames
		Title length	

The selected features from each feature selection technique consistently indicate linguistic features such as article length, title length, and average word length as the most relevant features. Optimal model–hyperparameter scores for each feature set are reported in Figure 5, along with the corresponding weighted f1 score of a most frequent dummy classifier, which uses simple rules for classification and provides a baseline classification rate against which model performance can be evaluated (Pedregosa et al., 2011).

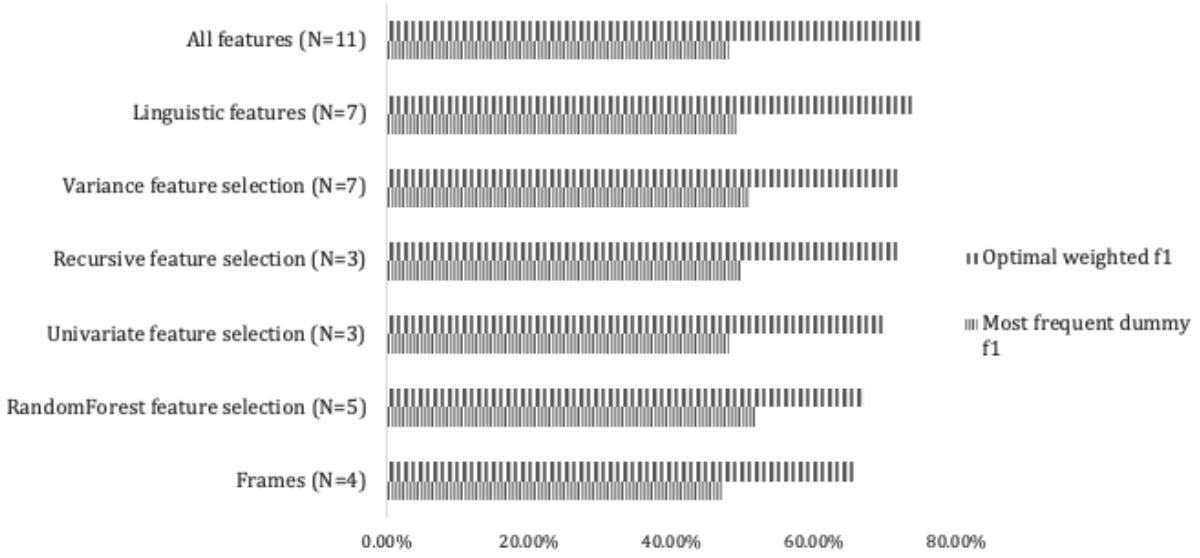

Figure 5. Country source classification in descending order of feature performance.

The classifier performance rates show that the best performing classifier is based on all of the remaining features (*N* = 11), with a weighted f1 score of 75%, closely followed by the linguistic features (*N* = 7), with a weighted f1 score of 73%. Three more feature combinations achieve scores around 70%, all of which significantly outperform their corresponding dummy classifiers. The analysis shows that the automated distinction of Russian and U.S. state-funded news is possible with a relatively high degree of accuracy. These findings also answer RQ2, demonstrating that the most useful feature set is the linguistic feature set, in particular, the feature's article length, title length, average word length, and ratio of unique words. The worst performing classifier is based exclusively on the issue-specific frames, with a weighted f1 score of 66%. However, all features combined slightly outperform only the linguistic features, indicating that the frames contribute to the country source classification. Figure 6 shows the precision and recall rates of each feature combination.

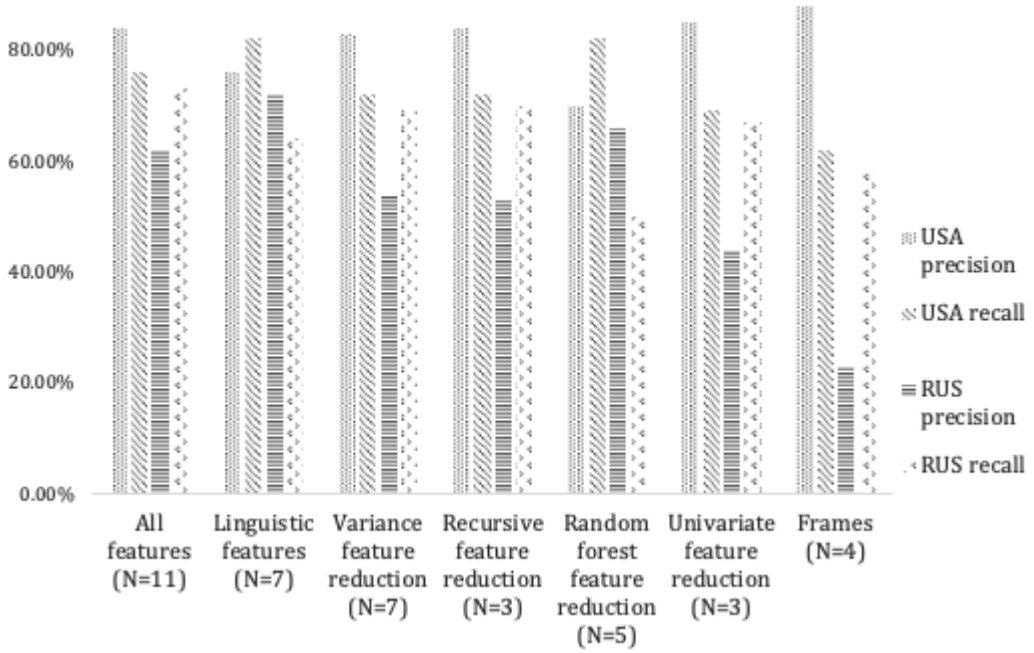

Figure 6. Recall and precision rates per country source for each model.

The recall and precision rates shown on Figure 6 indicate that the classifiers are suitable for recognizing both U.S. and Russian articles, with precision scores of 84% and 72%, respectively, in the best models. However, recall rates for U.S. articles are slightly higher, with an optimal score of 82%, whereas Russian article recall scores range between 58% and 73% in the optimal models. The higher recall scores for U.S. outlets may be a result of the slightly imbalanced data, with U.S. articles representing 57% of the total articles considered, or they may demonstrate that U.S. content has unique characteristics that allow for higher recall.

Conclusion and Discussion

This study demonstrated the potential of computational text analysis for the distinction of U.S. and Russian news in Serbia. The proposed methodology and theoretical framework are applicable to other contexts with adaptations of the issue-specific frames used, while all linguistic features can be directly applied and expanded on with data from other contexts. The findings support past literature which suggests that news production values and linguistic habits are significantly varied across media systems, ideology, and authors suggesting that large-scale profiling of outlet news production values, as well as linguistic habits of journalists, has extensive potential for political communication and text classification research (Horne & Adali, 2017; Schoonvelde et al., 2019). Moreover, we demonstrate that the distinctions between U.S. and Russian state-funded news in Serbia are best captured by these simple linguistic properties of text. This

finding suggests that a large-scale cross-country analysis of the differences between (Russian) state-funded outlets and known, reputable local or foreign media could be conducted regardless of researcher language barriers, allowing for robust approaches for detecting and countering disinformation campaigns from specific actors.

The study demonstrates that the theoretical feature selection was effective, as all suggested features show differences across the reporting of outlets from both countries, with U.S. outlets characterized by longer and simpler news, while Russian outlets use longer titles, complex language, and promote anti-Western narratives or Russian interests. Conversely, U.S. media are primarily focused on content related to Serbia, with comparatively minimal discussion of Russia. In line with past findings, this study answers RQ1 and demonstrates observable differences in both the style and content of reporting from U.S. and Russian outlets (Bechev, 2018; Klepo, 2017).

However, the study also shows that although Russian state-funded news contains anti-Western and divisive elements, these are present in no more than 20% of articles—a finding supported by the high intercoder reliability scores for these frames (all above 69%). This finding qualifies past research suggesting that RT, Sputnik, and related outlets are far more concentrated on promoting Russian geopolitical interests than what traditional notions of journalistic routines and codes of conduct would imply (Bechev, 2018; Prague Security Studies Institute, 2019). This study cannot make claims about the objectiveness or quality of the reporting of Russian state-funded outlets, but it demonstrates that a majority of Russian content does not contain the divisive elements suggested by past qualitative research, even though these narratives are an integral part of Russian content (Galante & Ee, 2018; Richter, 2017; Stronsky & Himes, 2019).

Additionally, we show that the differences in content can be used to automatically predict the source of an article with a weighted f1 score of 75%, a relatively high-performance score. This score is in line with that of past automated text classification research, with reliable scores typically ranging between 70% and 90% (Burscher et al., 2014; Chakraborty et al., 2016; Horne & Adali, 2017). Practically, these findings demonstrate that further quantitative research about the distinctions between Russian state-funded and Western media has merit and is called for. A machine learning classifier such as the one developed in this study can (in principle) be applied to the discourse of social media groups to identify networks that exchange and discuss Russian state-funded news, including automated accounts (Bradshaw & Howard, 2018; Hanlon, 2018; Helmus et al., 2018; Woolley & Howard, 2016). Such work would require additional robustness checks, as well as analysis of potential concept drift, i.e., the change in meaning of a frame, in a given context or period (Gama, Žliobaitė, Bifet, Pechenizkiy, & Bouchachia, 2014).

From a theoretical perspective, the issue-specific frames examined are closely aligned with common elements of nationalist discourse elsewhere. These frames, capturing anti-Western attitudes, ethnic grievances, and revisionist nationalist themes may belong to a single, broader frame. Researchers interested in examining this broader narrative may benefit from the simple yet effective approach offered by the holistic frame identification method when combined with custom, issue-specific frames and well-defined indicators. Helmus et al. (2018) use a similar approach to identify pro-Russian communities in Eastern European social media networks based on the language used and the linguistic properties of discourse. This study offers an additional

approach to this field of study by suggesting that suspicious users and groups can be identified based on the content shared and potentially on the corresponding linguistic properties in the discourse about it.

Finally, in line with past findings, the study demonstrates that simple and easily obtained linguistic properties of text, such as article length or language complexity, can be particularly informative for automatically distinguishing between U.S. and Russian news (Brundidge et al., 2014; Schoonvelde et al., 2019). This finding has significant implications for text classification tasks in widely spoken languages for which numerous automated approaches are available for detailed linguistic feature extraction, which could allow for extremely precise text classification in a wide range of fields (Grimmer & Stewart, 2013). However, the study also demonstrates that text classification research is possible in languages with very limited standardized tools and only basic text-processing capabilities, such as Serbian.

Limitations

We need to highlight that our study does not identify Russian disinformation, nor does it inform us about how to counter it. This task would require an entirely different methodological and theoretical approach, which would primarily attempt to quantitatively identify what disinformation is. Given the complexities and subtleties of disinformation content, such a task may go beyond the capacities of automated text analysis. However, these challenges could be overcome by developing a manually annotated data set of Russian disinformation content, the profile of which could then be compared with "normal" Russian news, so as to determine if disinformation can automatically be distinguished from normal news content. Furthermore, the binary framework that guides the analysis prevents a nuanced representation of the media systems considered. U.S. media is diverse, and the outlets used in this study do not reflect U.S. journalism as a whole, nor can U.S. journalism be taken as entirely representative of Western journalism. Adding European state-funded outlets to the analysis would provide a more robust and convincing analysis. However, other foreign media are underrepresented in the Serbian context, while the Serbian branch of the BBC, an excellent additional set of data does not permit access to its archives. This study also does not consider news from Serbian local outlets. Examining whether a classifier could distinguish this source of news as well would provide further support for the findings.

Secondly, frame classifier scores were not optimal for all frames, placing the validity of the final frame distribution in question. In the automatically determined distribution of the frames, "yes" responses for all frames were proportionally underrepresented when compared with the manually coded data. Though the SMOTE + EEN synthetic sample did improve recall and reliability scores, further improved classifiers are called for in future studies, as the predicted frames in the current study likely underrepresented the presence of all frames considered. Ultimately, this issue is due to the small manually annotated data set used (N = 1,108). Burscher et al. (2014), whose holistic method the current study is based on, achieve high frame prediction rates with supervised classifiers trained with a data set nearly tenfold the one used in this study. Regardless, our findings show that automatically detecting issue-specific frames is possible with a relatively high degree of accuracy even when the sample size is small, provided well-delineated frames and properly trained coders. The approach was shown to be effective for issue-specific frames, which are present in at most 20% of Russian articles and no more than 10% of the total data set (Burscher et al, 2014).

Finally, the linguistic features used in this study are relatively simple, despite their effectiveness. With the exception of the named entities, all features can be obtained with simple text processing, and they do not represent the significantly broader range of possibilities for automated linguistic feature extraction. A key variable for inclusion would have been named-entity-associated sentiment; however, an examination of existing sentiment analysis packages for Serbian, such as those provided by the Polyglot library, demonstrated that the reliability of performance was not suitable for the study (Al-Rfou et al., 2013). Researchers based in languages for which a broader set of text analysis tools are available are encouraged to also consider the sentiment of article titles; the sentiment associated with named entities; formal language complexity scores, such as the Flesch reading score; topic taxonomies provided by tools such as Textrazor; and various other NLP tools available.

References

Al-Rfou, R., Perozzi, B., & Skiena, S. (2013). *Polyglot: Distributed word representations for multilingual NLP*. ArXiv:1307.1662.

Altheide, D. L. (2004). The control narrative of the Internet. *Symbolic Interaction*, *27*(2), 223–245. doi:10.1525/si.2004.27.2.223

Analiza medijskog izveštavanja o međunarodnim akterima: Slučaj Srbije, BiH, Crne Gore i Makedonije [An analysis of media reporting about foreign actors: The case of Serbia, B&H, Montenegro & North Macedonia]. (2018). Retrieved from https://crta.rs/wp-content/uploads/2018/06/Regionala-analiza-medijskog-izvestavanja.pdf

Andresen, K., Hoxha, A., & Godole, J. (2017). New roles for media in the Western Balkans. *Journalism Studies, 18*(5), 614–628. doi:10.1080/1461670X.2016.1268928

Baade, B. (2019). Fake news and international law. *The European Journal of International Law, 29*(4), 1357–1376. doi:10.1093/ejil/chy071

Bandeira, L., Barojan, D., Braga, R., Peñarredonda, L. J., & Argüello, P. F. M. (2019). *Disinformation in democracies: Strengthening digital resilience in Latin America.* Retrieved from https://www.atlanticcouncil.org/publications/reports/disinformation-democracies-strengthening-digital-resilience-latin-america

Beam, M. A., Hutchens, M. J., & Hmielowski, J. D. (2018). Facebook news and (de) polarization: Reinforcing spirals in the 2016 U.S. election. *Information, Communication & Society*, *21*(7), 940–958. doi:10.1080/1369118X.2018.1444783

Bechev, D. (2018). *Understanding Russia's influence in the Western Balkans.* Retrieved from https://www.hybridcoe.fi/wp-content/uploads/2018/10/Strategic-Analysis-2018-9-Beshev-.pdf

Bennett, W. L., & Livingston, S. (2018). The disinformation order: Disruptive communication and the decline of democratic institutions. *European Journal of Communication*, *33*(2), 122–139. doi:10.1177/0267323118760317

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: Analyzing text with the natural language toolkit*. Sebastopol, CA: O'Reilly.

Boumans, J. W., & Trilling, D. (2016). Taking stock of the toolkit: An overview of relevant automated content analysis approaches and techniques for digital journalism scholars. *Digital Journalism*, *4*(1), 8–23. doi:10.1080/21670811.2015.1096598

Bradshaw, S., & Howard, P. N. (2018). *Challenging Truth and Trust: A Global Inventory of Organized Social Media Manipulation* (Computational Propaganda Project Research Report 2018.1). Oxford, UK: Oxford Internet Institute, Oxford University.

Brownlee, J. (2016, May 20). *Feature selection for machine learning in Python.* Retrieved from https://machinelearningmastery.com/feature-selection-machine-learning-python/

Brundidge, J., Reid, S. A., Choi, S., & Muddiman, A. (2014). The "deliberative digital divide": Opinion leadership and integrative complexity in the U.S. political blogosphere. *Political Psychology, 35*(6)*,* 741–755. doi:10.1111/pops.12201

Burscher, B., Odijk, D., Vliegenthart, R., de Rijke, M., & de Vreese, C. H. (2014). Teaching the computer to code frames in news: Comparing two supervised machine learning approaches to frame analysis. *Communication Methods and Measures, 8*(3), 190–206. doi:10.1080/19312458.2014.937527

Cerulus, L. (2019, January 17). *Facebook takes down two Russian disinformation networks in Eastern Europe.* Retrieved from https://www.politico.eu/article/facebook-takes-down-two-russian-disinformation-networks-in-eastern-europe/

Chakraborty, A., Paranjape, B., Kakarla, S., & Ganguly, N. (2016). Stop clickbait: Detecting and preventing clickbaits in online news media. *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 9–16). San Francisco, CA: IEEE Press. doi:10.1109/ASONAM.2016.7752207

Champion, J. (2019). *WbSrch Engine, extra stop words* [Serbian stop word list]. Retrieved from https://github.com/Xangis/extra-stopwords

Conroy, N. J., Rubin, V. L., & Chen, Y. (2015). Automatic deception detection: Methods for finding fake news. *Proceedings of the Association for Information Science and Technology* (Vol. 52, pp. 1–4). Silver Spring, MD: American Society for Information Sciences. doi:10.1002/pra2.2015.145052010082

Cull, N. J. (2009). *The Cold War and the United States Information Agency: American propaganda and public diplomacy, 1945–1989.* New York, NY: Cambridge University Press. doi:10.1017/CBO9780511817151

de Vreese, C. H. (2005). News framing: Theory and typology. *Information Design Journal, 13*(1), 51–62. doi:10.1075/idjdd.13.1.06vre

Denkovski, O. (forthcoming). *Infodemics, a snap election, and a (lukewarm) Western welcome: North Macedonia's identity at stake on Twitter.* Prague, Czech Republic: Prague Security Studies Institute.

Disinformation analysis on the Western Balkans: Lack of sources indicate potential disinformation. (2018, August 3). Retrieved from https://euvsdisinfo.eu/disinformation-analysis-on-the-western-balkans-lack-of-sources-indicates-potential-disinformation/

Eisentraut, S., & de Leon, S. (2018). *Propaganda and disinformation in the Western Balkans: How the EU can counter Russia's information war.* Berlin, Germany: Konrad Adenauer Stiftung. Retrieved from https://www.kas.de/documents/252038/253252/7_dokument_dok_pdf_51729_2.pdf/33dbbc29-eb30-e4ec-39c8-a4c976319449?version=1.0&t=1539647810617

Entman, R. M. (1993). Framing: Towards a clarification of a fractured paradigm. *Journal of Communication, 43*(4), 51–58. doi:10.1111/j.1460-2466.1993.tb01304.x

Fahmy, S. S., & Al Emad, M. (2011). Al-Jazeera vs Al-Jazeera: A comparison of the network's English and Arabic online coverage of the U.S./Al Qaeda conflict. *International Communication Gazette, 73*(3), 216–232. doi:10.1177/1748048510393656

Fidanovski, K. (2019, March 20). *"Macedonian scenario" unraveling in Serbia, but not the one Vucic has in mind*. Retrieved from https://europeanwesternbalkans.com/2019/03/20/macedonian-scenario-unraveling-serbia-not-one-vucic-mind/

Galante, L., & Ee, S. (2018). *Defining Russian election interference: An analysis of select 2014 to 2018 cyber enabled incidents.* Retrieved from https://www.atlanticcouncil.org/in-depth-research-reports/issue-brief/defining-russian-election-interference-an-analysis-of-select-2014-to-2018-cyber-enabled-incidents/

Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, *46*(4), 1–37.

Garimella, V. R. K., & Weber, I. (2017). A long-term analysis of polarization on Twitter. *Proceedings of the Eleventh International AAAI Conference on Web and Social Media* (pp. 528–531). Palo Alto, CA: AAAI Press.

Grimmer, J., & Stewart, B. M. (2013). Text as data: The promise and pitfalls of automatic content analysis methods for political texts. *Political Analysis*, *21*(3), 267–297. doi:10.1093/pan/mps028

Hanlon, B. (2018, December 19). *A long way to go: Analyzing Facebook, Twitter, and Google's efforts to combat foreign interference.* Retrieved from https://securingdemocracy.gmfus.org/a-long-way-to-go-analyzing-facebook-twitter-and-googles-efforts-to-combat-foreign-interference/

Heidenreich, T., Lind, F., Eberl, J. M., & Boomgaarden, H. G. (2019). Media framing dynamics of the "European refugee crisis": A comparative topic modelling approach. *Journal of Refugee Studies, 32*, i172–i182. doi:10.1093/jrs/fez025

Helmus, T. C., Bodine-Baron, E., Radin, A., Magnuson, M., Mendelson, J., Marcellino, W., . . . & Winkelman, Z. (2018). *Russian social media influence: Understanding Russian propaganda in Eastern Europe.* Retrieved from https://www.rand.org/content/dam/rand/pubs/research_reports/RR2200/RR2237/RAND_RR2237.pdf

Hjorth, F., & Adler-Nissen, R. (2019). Ideological asymmetry in the reach of pro-Russian digital disinformation to United States audiences. *Journal of Communication*, *69*(2) 168–192. doi:10.1093/joc/jqz006

Horne, B. D., & Adali, S. (2017). This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. *Eleventh International AAAI Conference on Web and Social Media* (pp. 759–766). Palo Alto, CA: AAAI Press. Retrieved from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM17/paper/view/15772/14898

Jackson, D. (2017). *Issue brief: Distinguishing disinformation from propaganda, misinformation, and "fake news."* Retrieved from https://www.ned.org/wp-content/uploads/2018/06/Distinguishing-Disinformation-from-Propaganda.pdf

Jain, S., Sharma, V., & Kaushal, R. (2016, September). Towards automated real-time detection of misinformation on Twitter. *2016 International Conference on Advances in Computing, Communications and Informatics* (pp. 2015–2020). Piscataway, NJ: IEEE Press. doi:10.1109/ICACCI.2016.773234

Janda, J. (2016). *The Lisa Case: STRATCOM Lessons for European States* (Security Policy Working Paper Publication No. 11/2016). Berlin, Germany: *Federal Academy for Security Policy.* https://www.baks.bund.de/sites/baks010/files/working_paper_2016_11.pdf

Khalid, S., Khalil, T., & Nasreen, S. (2014). A survey of feature selection and feature extraction techniques in machine learning. In K. Arai & A. Mellouk (Eds.), *Proceedings of the 2014 Science and Information Conference* (pp. 372–378). London, UK: IEEE Xplore. doi:10.1109/SAI.2014.6918213

Klepo, N. (2017). *Geopolitical influence on media and media freedom in the Western Balkans.* Retrieved from https://davastrat.files.wordpress.com/2017/09/dava-analytic-brief-no-3-20171.pdf

Lemaître, G., Nogueira, F., & Aridas, C. K. (2017). Imbalanced-learn: A Python toolbox to tackle the curse of imbalanced data sets in machine learning. *The Journal of Machine Learning Research, 18*(1), 559–563.

Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., . . . & Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures, 12*(2/3), 93–118. doi:10.1080/19312458.2018.1430754

Matthes, J., & Koring, M. (2008). The content analysis of media frames: Toward improving reliability and validity. *Journal of Communication, 58*(2)*,* 258–279. doi:10.1111/j.1460-2466.2008.00384.x

McKinney, W. (2011). Pandas: A foundational Python library for data analysis and statistics. *Python for High Performance and Scientific Computing, 14*(9), 1–9.

Miles, H. (2010). *Al Jazeera: How Arab TV news challenged the world*. London, UK: Hachette.

Milosevic, N. (2012). *Stemmer for Serbian language*. arXiv preprint arXiv:1209.4471.

Nair, V. G. (2014). *Getting started with Beautiful Soup.* Birmingham, UK: Packt.

Nemr, C., & Gangware, W. (2019). *Weapons of mass distraction: Foreign state-sponsored disinformation in a digital age.* Retrieved from https://static1.squarespace.com/static/5714561a01dbae16 1fa3cad1/t/5c9cb93724a694b834f23878/1553774904750/PA_WMD_Report_2019.pdf

Nimmo, A. (2018, January 8). *Question that:* RT*'s military mission*. Retrieved from https://medium.com/dfrlab/question-that-rts-military-mission-4c4bd9f72c88

"One in five million": Protesting Serbia's muzzled media. (2019, February 3). *Al Jazeera.* Retrieved from https://www.aljazeera.com/programmes/listeningpost/2019/02/million-protesting-serbia-muzzled-media-190202102248693.html

Orellana-Rodriguez, C., & Keane, M. T. (2018). Attention to news and its dissemination on Twitter: A survey. *Computer Science Review*, 29, 74–94. doi:10.1016/j.cosrev.2018.07.001

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . & Duschensay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Polyakova, A., & Boyer, S. (2018). *The future of political warfare: Russia, the West, and the coming age of global digital competition.* Retrieved from https://www.brookings.edu/wp-content/uploads/2018/03/fp_20180316_future_political_warfare.pdf

Prague Security Studies Institute. (2019). *Western Balkans at the crossroads: Assessing non-democratic external influence activities*. Retrieved from

https://docs.wixstatic.com/ugd/2fb84c_68a12132ec2d4e16b5a699ca586eaca6.pdf?fbclid=IwAR1fCbJCfFcVgNiHWF_OEXOArB9w86O88aj6Fx6oy2p3Tw16RL5J1fqhUao%20(WB,%202019)%20https://davastrat.files.wordpress.com/2017/09/dava-analytic-brief-no-3-20171.pdf

Qian, T., Hollingshead, K., Yoon, S. Y., Kim, K. Y., & Sproat, R. (2010). A Python toolkit for universal transliteration. In D. Tapias, D. I. Russo, O. Hamon, S. Piperidis, N. Calzolari, K. Choukri, . . . & M. Rosner (Eds.), *Proceedings of the 7th International Conference on Language Resources and Evaluation* (pp. 2897–2901). Paris, France: European Language Resources Association.

Richter, M. (2017). *What we know about* RT *(*Russia Today*)* (Kremlin Watch Report 10.09.2017). Retrieved from https://www.europeanvalues.net/wp-content/uploads/2017/09/What-We-Know-about-RT-Russia-Today-1.pdf

Robinson, P. (2005). *The CNN effect: The myth of news, foreign policy and intervention*. Abingdon, UK: Routledge. doi:10.4324/9780203995037

Salunke, S. (2014). *Selenium webdriver in Python: Learn with examples.* Scotts Valley, CA: CreateSpace.

Schauster, E. E., Ferrucci, P., & Neill, M. S. (2016). Native advertising is the new journalism: How deception affects social responsibility. *American Behavioral Scientist*, *60*(12), 1408–1424. doi:10.1177/0002764216660135

Schoonvelde, M., Brosius, A., Schumacher, G., & Bakker, B. N. (2019). Liberals lecture, conservatives communicate: Analyzing complexity and ideology in 381,609 political speeches. *PLOS ONE*, *14*(2), e0208450. doi:10.1371/journal.pone.0208450

Stronsky, P., & Himes, A. (2019). *Russia's game in the Balkans*. Retrieved from https://carnegieendowment.org/2019/02/06/russia-s-game-in-balkans-pub-78235

Tsur, O., Calacci, D., & Lazer, D. (2015). A frame of mind: Using statistical models for detection of framing and agenda setting campaigns. In C. Zong & M. Strube (Eds.), *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing: Vol. 1. Long Papers.* (pp. 1629–1638). Stroudsburg, PA: Association for Computational Linguistics. doi:10.3115/v1/P15-1157

van Atteveldt, W., Welbers, K., Jacobi, C., & Vliegenthart, R. (2014). *LDA models topics . . . But what are "topics"?* Retrieved from http://vanatteveldt.com/wp-content/uploads/2014_vanatteveldt_glasgowbigdata_topics.pdf

van der Walt, S., Colbert, S.C., & Varoquaux, G. (2011). The NumPy array: A structure for efficient numerical computation. *Computing in Science & Engineering, 13*(2), 22–30. doi:10.1109/MCSE.2011.37

Waterson, J. (2019, Jan 17). *Facebook removes hundreds of pages 'linked to Russian site'.* Retrieved from https://www.theguardian.com/technology/2019/jan/17/facebook-removes-hundred-of-pages-allegedly-linked-to-russian-site-sputnik

Woolley, S. C., & Howard, P. N. (2016). Political communication, computational propaganda, and autonomous agents: Introduction. *International Journal of Communication*, *10*, 4882–4890.

Yablokov, I. (2015). Conspiracy theories as a Russian public diplomacy tool: The case of *Russia Today* (*RT*). *Politics*, *35*(3/4), 301–315. doi:10.1111/1467-9256.12097