Political attacks in 280 characters or less

A new tool for the automated classification of campaign negativity on social media

Petkevič, V.; Nai, A.

DOI
10.1177/1532673X211055676

Publication date
2022

Document Version
Final published version

Published in
American Politics Research

Article

American Politics Research
2022, Vol. 50(3) 279–302

DOI: 10.1177/1532673X211055676
journals.sagepub.com/home/apr


Political Attacks in 280 Characters or Less: A New Tool for the Automated Classification of Campaign Negativity on Social Media

Vladislav Petkevic[1], and Alessandro Nai[2]

Abstract

Negativity in election campaign matters. To what extent can the content of social media posts provide a reliable indicator of candidates' campaign negativity? We introduce and critically assess an automated classification procedure that we trained to annotate more than 16,000 tweets of candidates competing in the 2018 Senate Midterms. The algorithm is able to identify the presence of political attacks (both in general, and specifically for character and policy attacks) and incivility. Due to the novel nature of the instrument, the article discusses the external and convergent validity of these measures. Results suggest that automated classifications are able to provide reliable measurements of campaign negativity. Triangulations with independent data show that our automatic classification is strongly associated with the experts' perceptions of the candidates' campaign. Furthermore, variations in our measures of negativity can be explained by theoretically relevant factors at the candidate and context levels (e.g., incumbency status and candidate gender); theoretically meaningful trends are also found when replicating the analysis using tweets for the 2020 Senate election, coded using the automated classifier developed for 2018. The implications of such results for the automated coding of campaign negativity in social media are discussed.

Keywords

negative campaigning, US Midterms, machine learning, neural networks, incivility

Introduction

Modern politics is a hard-fought business. The public is increasingly hostile toward those they consider as their rivals (Iyengar et al., 2012; Iyengar & Westwood, 2015), antagonistic and aggressive political figures are on the rise across the globe (Nai & Martinez i Coma, 2019a), and attacks seem at times the very essence of election campaigns (Ansolabehere et al., 1994; Lau & Pomper, 2004). To be sure, negativity in politics matters. Negative information, when compared to equivalent "positive" information, is more likely to be seen, processed, and remembered (e.g., Rozin & Royzman, 2001). Also because of that, some scholars show that negative messages can convey important and useful information to the voters, promote issue knowledge, cue the voters that the election is salient, and thus worth the emotional and cognitive investment, and ultimately stimulate the interest of the public (Finkel & Geer, 1998; Martin, 2004; Geer, 2006). On the other hand, however, strong evidence suggests that negative campaigning can be a detrimental force in modern democracies. Negative and harsh campaigns can reduce turnout and political mobilization, depress civic attitudes such as political efficacy and trust, foster apathy, and generally produce a "gloomier" public mood (Ansolabehere et al., 1994; Thorson et al., 2000; Yoon et al., 2005). On top of this, a case can be made that decreased trust in the political game and increased political cynicism are likely to reinforce the consolidation of antagonistic and disruptive movements, which often "feed" off the public discontent. Whether a positive or detrimental force, few would contest that negativity is a key component of contemporary electoral democracies.

In recent years, the dynamics of electoral campaigning have been reshaped by the emergence of social media (Gainous & Wagner, 2014; Graham et al., 2016; Straus et al., 2013). Online communication, especially via social media, allows political actors to "cut the middlemen"—for instance, journalistic gatekeeping—and communicate directly with their audience, in what is often referred to as "one-step flow of

[1]Faculty of Social and Behavioral Sciences, University of Amsterdam, Amsterdam, Netherlands
[2]Amsterdam School of Communication Research (ASCoR), University of Amsterdam, Amsterdam, Netherlands

Corresponding Author:
Vladislav Petkevic, Faculty of Social and Behavioral Sciences, University of Amsterdam, Nieuwe Achtergracht 166, 1018 WV, Amsterdam, Netherlands.
Email: v.petkevich@uva.nl

communication" (Bennett & Manheim, 2006). Such facilitated access to the people is one of the reasons why online communication is particularly favored by populists (Engesser et al., 2017). In recent years, several studies have assessed the presence of negativity in social media (e.g., Auter & Fine, 2016; Ceron & d'Adda, 2016; Evans et al., 2014; Evans & Clark, 2016; Gainous & Wagner, 2014; Gross & Johnson, 2016). Broadly speaking, these studies find confirmation that the main trends of strategic campaigning found for traditional techniques—for instance, that challengers tend to attack more than incumbents (Gainous & Wagner, 2014)—are also found when looking at campaigning on social media. Those existing studies tended to rely on manual coding of social media posts. For instance, in their analysis of the drivers of negativity of Facebook during the 2010 Midterms, Auter and Fine (2016) manually coded more than 14,000 posts. Similarly, Ceron and d'Adda (2016) hand-coded more than 15,000 Tweets published by competing candidates prior to the 2013 Italian general election. Recent advances of machine learning approaches have made it increasingly affordable to dive into very large amounts of data, which were inaccessible—or required time-intensive coding efforts—up to recently due to insufficient computational power and the preference for manual coding. In this article, we expand the growing literature on automated classification of textual data within the context of political communication. We introduce a neural network classifier that we trained to automatically annotate the tweets of candidates competing during the 2018 US Senate Midterms elections, in terms of the presence of political attacks. The algorithm was run on approximately 16,000 tweets, posted by 63 candidates for the period between September 1st and November 6th, 2018 (the day of the election). After presenting the results of the classification, the article will test the external and convergent validity of the measure; more specifically, we will check whether the results make sense in terms of factors that can be theoretically expected to drive the presence of negativity in the candidates' tweets (e.g., the incumbency status of the candidate), and by triangulating the measure with an independent dataset about the content of candidates' campaigns in the 2018 Midterms, as assessed by expert surveys (Nai & Maier, 2020).

The reason for studying the 2018 midterm elections was two-fold. Firstly, from a conceptual standpoint, the US senate elections provide a unique opportunity to study a series of extremely similar elections with a reduced number of competitors, happening simultaneously within the same broad societal, cultural and, ultimately, political context (Lau & Pomper, 2001, 2004)—while, at the same time, being able to control for the most important differences at the contextual level (e.g., how close the race was). Yet, even if driven by state-level dynamics, Senate Midterms elections all participate to the broader national context and political dynamics. The 2018 Midterms were not an exception in this sense, and the results in each state had fundamental national implications in terms of, for example, the control of the upper house (so central to the recent dynamics of presidential impeachment of Donald J. Trump). In other terms, the Senate Midterms are an ideal research setting, allowing all the benefits of variation—both at the candidate and context levels—while keeping most of the broad cultural and political dynamics, assumed to be shared across all state-level elections at bay, so to speak. Indeed, especially compared with Presidential elections, Senate elections can be seen as "methodologically superior" for the study of campaign dynamics (Lau & Pomper, 2004, p. 6). Secondly, the 2018 Midterm Senate elections provide us with the unique opportunity to test the convergent validity of our data by comparing it with other, independent data about the same elections and the same dynamics (i.e., how negative the candidates in the Midterms went against each other). More specifically, we will triangulate the tone of the candidates' campaign on Twitter with expert ratings provided by independent scholars (Nai & Maier, 2020).

The rest of this article proceeds as follows. In section 2, we describe the empirical procedure employed to develop the algorithm for the automated coding of the negativity in tweets. Section 3 then presents three tests. First, we test the convergent validity of the algorithm, by comparing it with the measure of negativity from independent data using expert surveys. Second, we test its external validity by checking our measurement against some trends that can be theoretically expected (i.e., the fact that challengers should be expected to be more likely to go negative, or that female candidates tend to use gentler campaigns. Finally, third, we investigate whether applying the coding algorithm to a different set of data—the campaign on Twitter during the 2020 Senate election—yields results that are also theoretically valid. As we will see, our algorithm scores well in both external and convergent validity, suggesting that the automated coding of social media posts is an effective alternative to standard measurement of campaign content. The last section concludes the discussion and glimpses over the directions of future research.

Supporting materials for this article are available at the following Open Science Foundation repository: https://osf.io/up826/. The repository includes (i) the Jupyter Notebook (Python) file with the code that was used to pre-process the raw data, build the classifier, and annotate the whole dataset, (ii) the annotated dataset of all tweets, (iii) the archive with the text of all the collected tweets, and (iv) an excel file with the reliability assessment of the initial sample of tweets coded (see below).

Measuring Negative Campaigning in Tweets

Data and Procedure

During the 2018 Midterms, 33 Class 1 Senate seats were up for grabs (plus additional special elections in Minnesota and Mississippi to fill vacancies, but which we will not analyze here); Democrats were holding 26 of these seats, and

Republicans 9. Excluding some scattered small third-party candidates that ran in a handful of elections (e.g., the Libertarian Rusty Hollen in West Virginia), 66 main candidates competed overall to fill these 33 seats, in as many bipartisan races.

The data (tweets) used in this study were collected via vicinitas.io, a website that allows for bulk downloading of tweets retroactively based on Twitter handles (usernames). Prior to it, an online search for Twitter pages of all contemporaneous Senate election candidates was performed to determine which of the candidates used Twitter for their political campaigns and what their Twitter handles were.[1] The handles were then supplied to vicinitas.io to collect the tweets for the period of September 1, 2018–November 6, 2018 (the day of the election), for a total of $N = 16{,}173$ tweets.[2] Three candidates did not, to the best of our knowledge, post any tweets in that period (even though they do have a twitter handle): Chele Chiavacci (R, NY, @CheleNYC), Leah Vukmir (R, WI, @LeahVukmir), and Lawrence Zupan (R, VT, @LawrenceZupan). The analyses discussed in this article thus concern the 63 remaining candidates (see Table A1 in Appendix A). The number of tweets per candidate collected varies considerably, from $N = 24$ for Mitt Romney (R, UT, @MittRomney) to $N = 1028$ for Rick Scott (R, FL, @SenRickScott), with an average of 256.7 tweets per candidate. Figure 1 plots the number of tweets per day, per party. The figure shows a rather marked increase in the number of published tweets as election day nears, for both parties, similarly to what found in Gross & Johnson (2016) for the 2016 Republican primaries.

A codebook was developed to measure the tweets on four dimensions of negativity: negative tone, personal attacks, policy attacks, and incivility. Each of these dimensions was to be coded dichotomously as either present or absent in a given tweet. A random sample of 200 tweets was then coded by four coders independently to assess intercoder reliability. The initial Krippendorff's alpha reliability scores were *.79* for negative tone, *.52* for political attacks, *0.50* for personal attacks, and *0.66* for incivility. Given such suboptimal scores, the tweets where disagreements between coders occurred were analyzed by the researchers and the coders were consulted to establish any systematic differences in interpretation of negativity dimensions between the coders. Based on the resulting observations, the codebook was reworked, and its instructions elaborated upon.

Once the coders were briefed on the new instructions, each of them was given a random sample of 100 tweets to annotate. That proved to be insufficient as the negativity dimensions were coded as absent in the overwhelming majority of tweets (M = 89%). Building a successful machine learning classifier presupposes providing a sufficient number of all possible examples during the algorithm training: if there are only a few examples of any given dimension, the model will not be able to accurately generalize the text features that determine the presence or absence of that dimension. To account for such an imbalance, the coders were provided with new random samples of the data and asked to annotate the tweets on one dimension until at least 200 tweets were coded as "present" for the respective dimension. The annotated tweets were then combined producing a dataset of 1186 tweets.

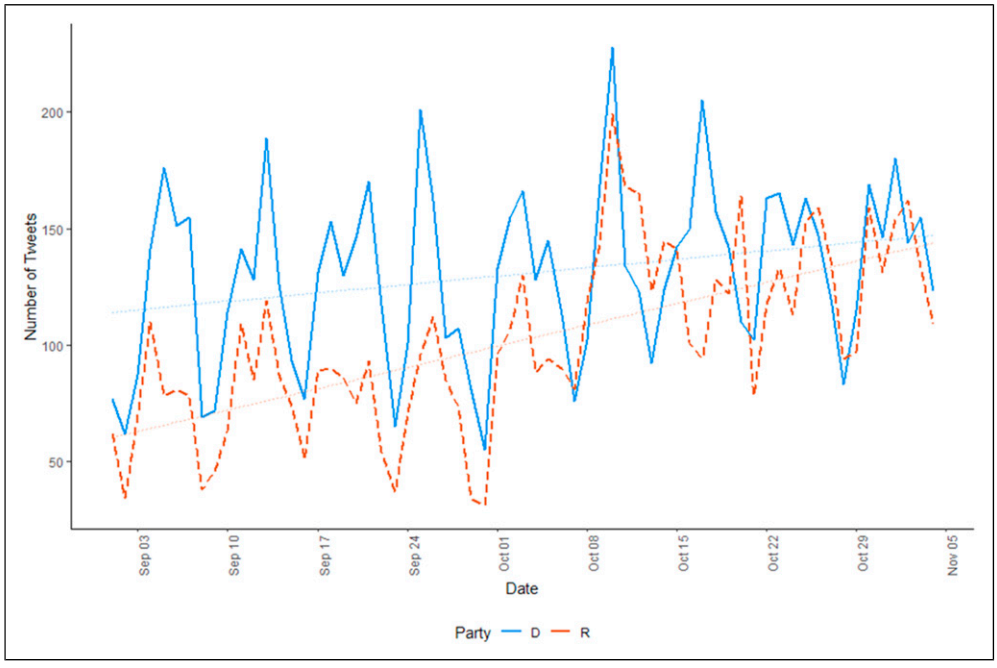

Figure 1. Frequency of tweets per day and party (2018 Senate election).

The Algorithm

A multilayer perceptron neural network (MLP; Pedregosa et al., 2011) classifier was trained to automatically annotate the remainder of the tweets. MLP is a type of feedforward neural network where connections between neurons (nodes) do not form loops but are directed onto subsequent layers only, as opposed to recurrent neural networks (RNNs) in which such loops are an integral part. All nodes, with the exception of the input ones, are activated with a non-linear activation function (a function that determines the node's output given its input), usually a sigmoid or a rectifier, with possible value in the ranges of *[−1:1]* and *[0:1]*, respectively. The number of hidden layers in the model and the number of neurons in them is chosen freely. All input nodes are connected with a certain weight to all nodes of the first hidden layer, all nodes of the hidden layer are connected to the nodes of the following layer, and so on until the output layer is reached. The initial connection weights are chosen randomly and are adjusted via backpropagation during training (Rosenblatt, 1961).

Using MLP, just like any other neural network, presupposes supplying numeric values, not text, as input values. Consequently, a numeric representation of text (tweets in our case) is required. Such mapping of words (or phrases) to numeric vectors is referred to as word embedding. Multiple techniques have been developed to accomplish the task (c.f. Mikolov et al., 2013; Levy & Goldberg, 2014). For this project, we opted for a pre-trained model from SpaCy owing to its state-of-art performance, compact size, and processing speed. The model[3] includes over 685,000 vectors of 300 dimensions each trained on Common Crawl[4] and OntoNotes 5[5] using convolutional neural networks (Honnibal & Johnson, 2015).

The Procedure

To assess baseline classification performance, a zero-rule algorithm was used, whereby the labels in the test dataset are predicted on the basis of the most common label in the training dataset. This simple classifier provided us with a reference point which could be used to establish if the "sophisticated" classification models showed any meaningful improvements in their classification abilities. The results of the zero-rule classifier are presented in Table 1.

Next, the data was pre-processed before supplying it to the MLP model. The text of the annotated tweets was "cleaned": they were stripped of punctuation and single characters, all words were converted to lowercase, and stop-words were removed. Stop words are those words that do not carry any particular information relevant to the meaning of the text, such as "the," "at," and "to"; in our case, we use a pre-compiled list of stop words from Spacy.[6] The words also were lemmatized: inflectional endings of words were removed keeping only their "base forms" (lemmas) such that the same words occurring in different grammatical forms across tweets would be transformed into identical ones. Below is an example sentence from one of the tweets (1) in its "raw" format and (2) with the pre-processing steps discussed above conducted:

Table 1. Classification Statistics for the Zero-Rule Classifier.

	F1 Score (Absence of Dimension)	F1 Score (Presence of Dimension)	Area under ROC Curve
Negative tone	0.67	0.00	0.50
Personal attack	0.85	0.00	0.50
Political attack	0.84	0.00	0.50
Incivility	0.90	0.00	0.50

(1) Another example of the corruption our current representation is a part of and the false claims that @SenatorCarper is for the environment.
(2) exampl corrupt current represent fals claim senatorcarp environ

Since the stop-word removal and lemmatization are not guaranteed to result in a better classifier performance (sometimes resulting in the opposite effect), four different pre-processed datasets were created with either (1) none, (2 and 3) one, or (4) both of these steps skipped.

As the next step, the words contained in the tweets were vectorized using the pre-trained word-embedding model discussed above and an average vector with 300 dimensions for every tweet was calculated (the number of dimensions is determined by the dimensionality of vectors supplied with the word-embedding model). All out-of-vocabulary words (such as hashtags and mentions) were mapped to zero vectors when computing the average tweet vector. Taking an average of all word embeddings per tweets was done to reduce the processing time needed for the classifier training.[7] As discussed above, vectorizing the input data allowed for (1) representation of words as numeric values required for the next step of model training while (2) preserving the (contextual) meanings of the original words through assignment of similar vector values to words similar in meaning. As the final pre-processing step, the annotated negativity data was reshaped into an array of binary label vectors (e.g., *[1,0,0,1]* for a single tweet) to allow for a convenient parsing of them into a multi-label MLP model. These transformations resulted in two numeric arrays: one with *(1168,300)* dimensions for the textual data, and one with *(1168,4)* dimensions for the negativity dimensions. With both the text and the corresponding label data transformed into the desired formats, we moved onto determining the most fitting hyper-parameters for the data at hand.

To be able to evaluate the model's performance, 20% of the data (N=234) was set aside as a testing dataset to eventually be used to calculate accuracy statistics of the trained model. Using the remaining 80%, the best parameters were estimated for a multi-label prediction model, a model in which all dimensions of negativity are predicted concurrently. Three-fold cross-validation (i.e., splitting the data into three equal parts and predicting each individual part based on the training of the remaining two) was used to minimize the chance of accidental high performance of the model that is not generalizable to the rest of the data. An alternative approach of training four separate single-label models (one for each negativity dimension) and estimating their hyper-parameters was attempted, too. This was done to see whether greater accuracy can be achieved by fine-tuning independent models rather than having one model predicting all labels simultaneously.

The performance of the models was assessed by having the model predict the labels of the testing dataset that we had set aside. The evaluation was made on the basis of F1-scores and the fraction of area under the receiver operating characteristic (ROC) curve. Precision and recall scores were also assessed with preference for "conservative" models. In the end, the best performance was achieved using the multi-label classifier trained on data that had stop-words removed but was not lemmatized, with the average weighted F1 score of 0.84 (Table 2). For all the possible parameter choices used in the grid search and the classification statistics of all classifiers please consult Appendix B.

Figure 2 plots the number of negative tweets per day between September and November 2018, separately for Democratic and Republican candidates. The graph shows that Democrats started rather negative but somehow reduced the share of negativity near the end of the campaign. Republicans, on the other hand, rather consistently went more negative as time came close to election day.

Algorithm Bias

Before moving onto evaluating external and convergent validity of the automatic classification, we assessed the consistency in the algorithm's predictions and attempted to detect potential systematic errors in its classification. To do that, we leveraged the testing dataset by comparing the true negativity values of the hand-coded tweets to those predicted by the algorithm. Figure 3 presents daily time series of the true and predicted values of all four dimensions of negativity (the time series were smoothed using a 7-day moving average). The figure suggests that the algorithm is consistent in its predictions of the negativity dimensions and does not significantly deviate from the true values at any time.

For a more formal analysis, we ran a series of chi-square tests comparing the frequencies of the presence and absence of negativity between the manually annotated and predicted tweets. Separate chi-square tests were run for tweets posted by different groups of candidates to see whether the algorithm performs equally well under all circumstances. The difference between the true and predicted values were assessed for tweets subset by candidates' party, gender, incumbency status, state-level race closeness, and state level of Trump support in the 2016 presidential elections. In total, 48 chi-square tests were performed, with none indicating a significant difference between the true and predicted values. Additionally, we performed topic modeling using density-based clustering (Campello et al., 2013) on the entire dataset (extracting 15 topics such as discussions of economic policy or encouragements to cast a vote) to check whether the algorithm's performance is consistent over different topics. Chi-square tests comparing the true and predicted negativity values were run for each topic with all tests producing statistically insignificant results. Finally, we also performed correlation analyses to determine whether there is a relationship between the candidates' frequency of posting and the candidates' average negativity scores (on all four dimensions). These tests also produced no statistically significant results. Together these analyses suggest that the algorithm's performance is consistent and unbiased.

Table 2. Classification Performance of the Best Model.

	F1 Score (Absence of Dimension)	F1 Score (Presence of Dimension)	Area under ROC Curve
Negative tone	0.81	0.83	0.82
Political attack	0.92	0.75	0.83
Personal attack	0.89	0.77	0.82
Incivility	0.94	0.77	0.85

Testing the Algorithm

We present below three sets of analyses that we implemented to test for the convergent and external validity of our developed algorithm. The first test (convergent validity) checks whether the measurement of campaign negativity from our algorithm yields results that are in line with other, independent measurements of the same phenomena, in our case expert judgments about the campaign content of candidates having competed in the 2018 Senate elections (Nai & Maier, 2020). The second test (external validity) checks whether our measurement is theoretically meaningful, that is, is able to show trends that reflect string theoretical assumption—in our case, regarding the drivers of campaign negativity. Finally, the third test (replication) tests whether applying the coding algorithm to a different set of data—the campaign on Twitter during the 2020 Senate election—yields results that are also

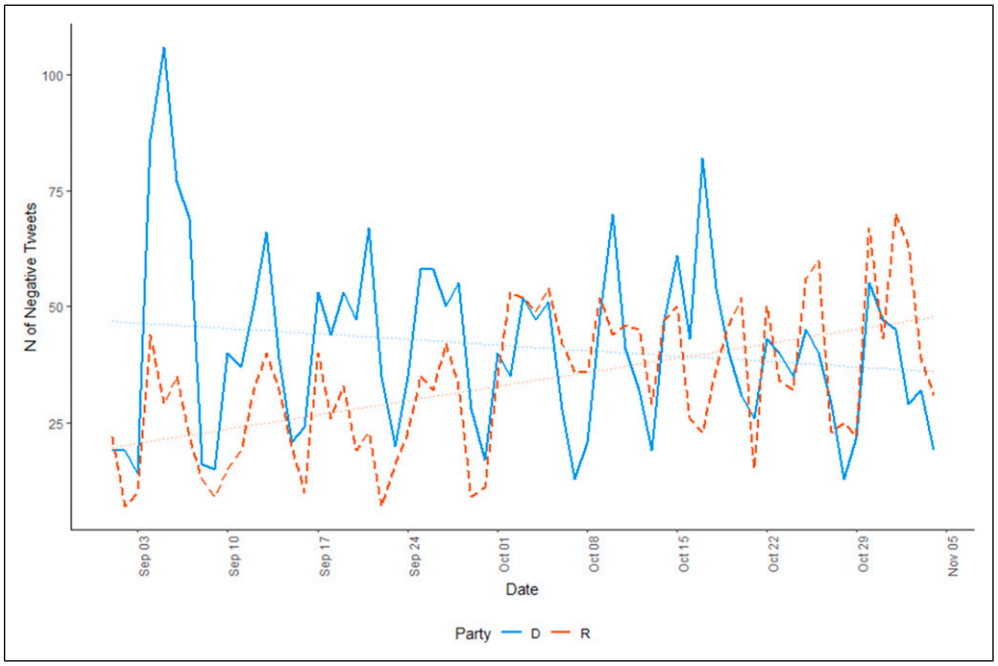

Figure 2. Frequency of negative tweets per day by party (2018 Senate election).

theoretically valid. As we will discuss below, all three sets of tests yield positive results for our algorithm.

Convergent Validity: Negativity in Twitter and Expert Ratings

Convergence. Valid new constructs should ideally be in line with similar existing constructs. We test here convergent validity by comparing our Twitter measures of negativity with completely independent data gathered within the framework of an expert survey (Nai & Maier, 2020). In the direct aftermath of the 2018 Midterms, we distributed a standardized survey to a sample of scholars with expertise in elections, politics, or political communication working for a US higher education institution. Among other things, experts were asked to evaluate the content of campaigns of the two competing candidates for the Senate in that state. We were not able to gather any expert opinions for North Dakota and West Virginia, and only one expert provided ratings for candidates in Hawaii, Nevada, and Wyoming, which we excluded from our analyses for robustness reasons; only candidates for whom at least two scholars provided independent ratings are included in our analyses. Analyses are run for the remaining 49 candidates (see Table A1 in Appendix A).

The number of experts that answered our survey varies between 2 (e.g., for Delaware) and 30 (California), with an average of 8.04 experts per candidate. Table A1 in Appendix A lists all candidates and presents how many experts rated their campaign. On average, experts in the sample lean unsurprisingly to the left (M = 3.22/1-10, SD = 1.43); 66% of them identify as a Democrat, 21% as Independent, and only 4% as a Republican (4% prefer not to say). 27% of them are female. Experts rated themselves as quite familiar with election campaigns in their state (M = 7.81/0-10, SD = 2.05) and estimated that the survey was easy to answer (M = 7.52/0-10, SD = 2.39).

Experts were asked to rate to what extent candidates used "negative campaigning" against their opponent during the election, that is, to what extent they relied on campaigning messages "criticising their opponents' programs, ideas, accomplishments, qualifications, and so forth." For each candidate, they provided a rating between −10 "Very negative" and +10 "Very positive," which we simplified into a 0-10 negativity scale where 10 means "Very negative." Experts also had to evaluate whether candidates attacked mostly on policy issues or on the personal characteristics and character of their opponents, using a scale from 1 "Exclusively policy attacks" to 5 "Exclusively character attacks." For the sake of comparison, we transformed this variable into a 0-10 scale, where 10 means "Exclusively character attacks." Finally, experts were asked to what extent the candidates used "fear appeals," on a 0-10 scale where 10 means "very high use."

Unsurprisingly, the three measures are strongly correlated, for example, for fear and tone, $r(52) = .75$, $p < .001$, and would load into an additive scale, $\alpha = 0.87$. An exploratory factor analysis (PCA) extracted a single underlying factor explaining 81% of the variance (Eigenvalue = 2.42), with factor loadings between 0.56 and 0.60. This underlying factor varies between −2.91 and +2.81 and can be seen as a broad measure of campaign negativity, as assessed by experts. Amy Klobuchar (D, MN) and David Baria (D, MS) are the two candidates with the lowest scores of campaign negativity

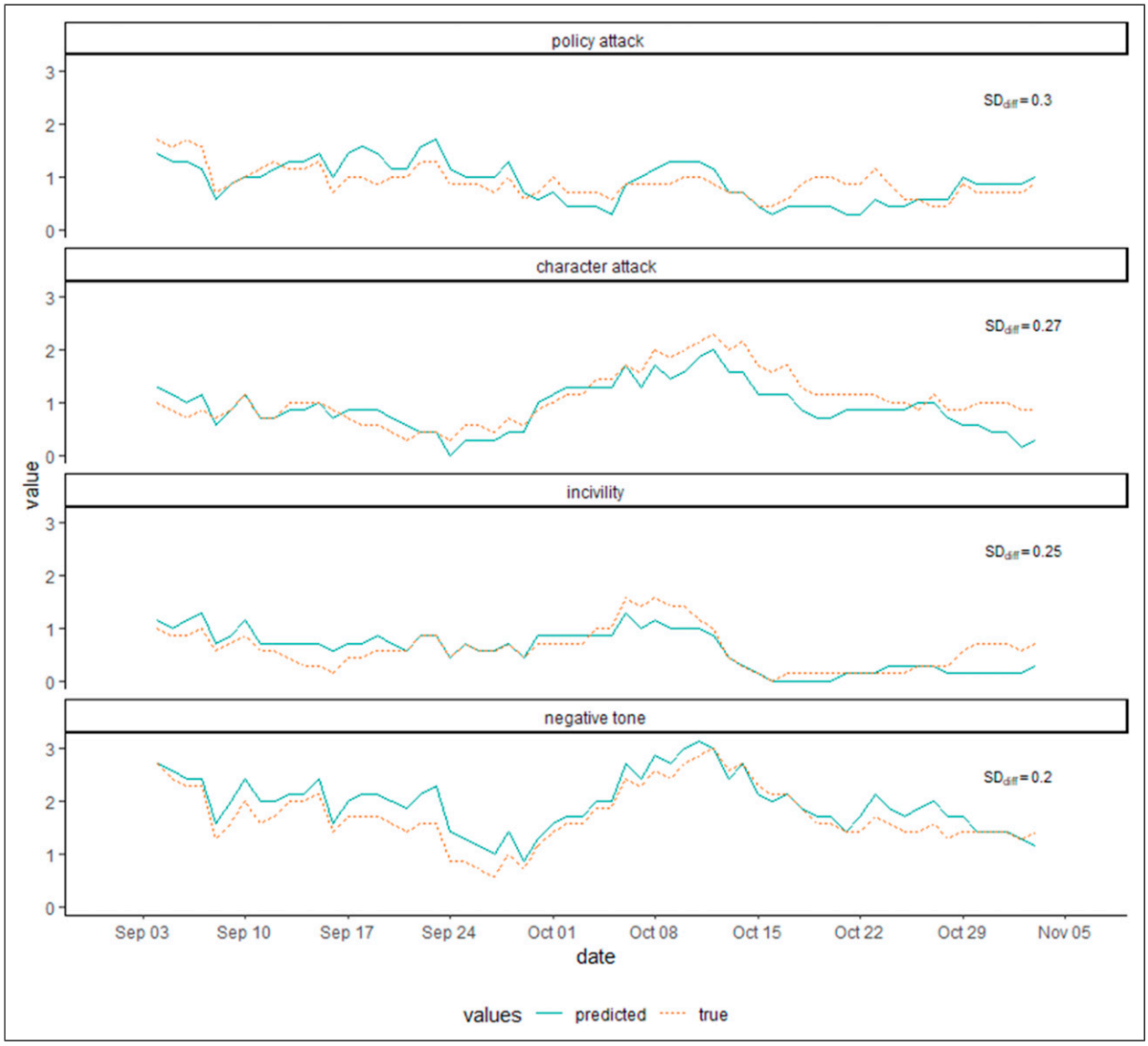

Figure 3. Seven-day moving average of true and predicted values of the four negativity dimensions by day (2018 Senate election). Note. The figure also includes standard deviations of the difference between the true and predicted values for each of the negativity dimensions (top right corner).

according to our experts, whereas Corey Stewart (R, VA) and Matt Rosendale (R, MT) score the highest. Table A1 in the Appendix includes the scores, for each candidate, on both the Twitter and experts' measures of negativity.

To what extent are the two independent measures of campaign negativity—the automated coding of tweets and the broad assessment by election experts—associated? Before answering this question, it is important to stress that the two measures do not necessarily reflect the same phenomenon: the automated measure is specific to the content of campaigns is social media (on Twitter, more precisely), whereas experts were asked to assess the campaign of candidates in general, regardless of the medium. It is known that the use of negativity differs across different communication channels (Walter & Vliegenthart, 2010). Some candidates, for instance, might go very negative in TV ads, and only use Twitter to promote events. In this sense, we should not expect that the two measures are a perfect reflection of each other. Nonetheless, we could expect that both are a proxy of the underlying campaigning style of each candidate, and we thus expect them to be associated somehow.

Table 3 regresses the three measures of campaign negativity via the automated coding of tweets on the expert's assessment of candidate negativity (underlying index), controlling for the usual suspects that drive negativity discussed in the previous section. Perception of campaign negativity is a function of both the content of each ad (or, in this case, tweet) and the total

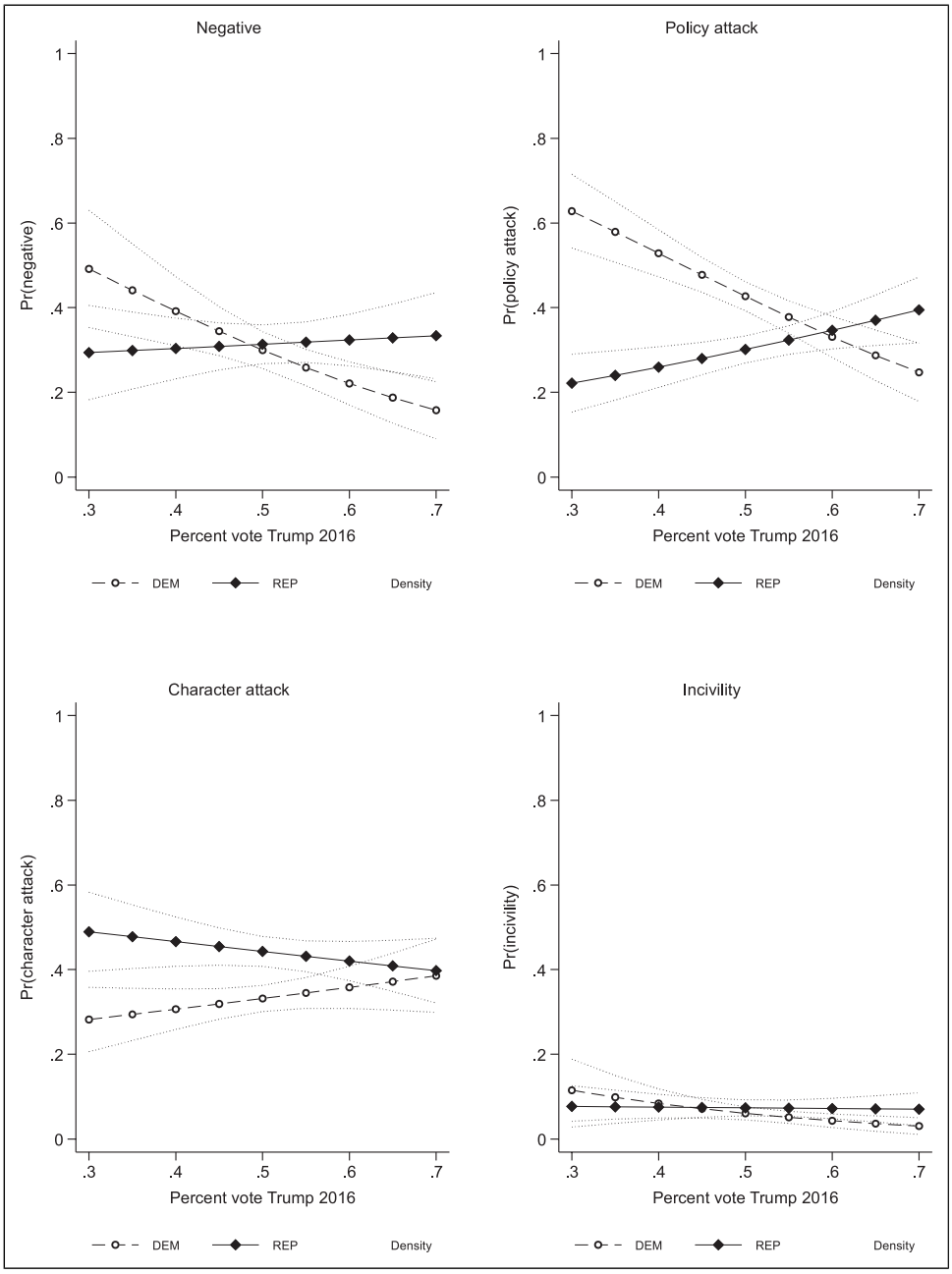

Figure 4. Probability of attack by candidate leaning and percent votes for Trump in 2016; marginal effects (2020 Senate election). Note. Marginal effects, with 95% confidence intervals. Dependent variable is the probability that the tweet was coded as negative (top left-hand panel), as containing a policy attack (within negative tweets only; top right-hand panel), as containing a character attack (within negative tweets only; bottom left-hand panel), and as uncivil (bottom right-hand panel). Full results are in Table C1 (Appendix C).

volume of ads (tweets) people are exposed to (e.g., Stevens, 2009); for instance, a candidate with very negative tweets but having posted only a handful of them would probably be perceived as much less negative than a candidate with fewer negative tweets on average but much more activity on social media. We take this into account by using in our analyses the *volume* of negativity in tweets, that is, the average content of tweets for each candidate (in terms of tone, policy attacks, and character attacks) multiplied by the total number of tweets posted by that candidate. Models in Table 3 are hierarchical linear models, where candidates are nested within states.

Results in Table 3 suggest that the two measures are associated, once controlling for the candidate profile, the closeness of the race, and the percent of votes for Trump in the state in the 2016 Presidential election. Campaigns that are evaluated by the experts as very negative (underlying dimension) have a volume of negativity on Twitter that is twice as high as campaigns that are evaluated as very low in

Table 3. Negativity on Twitter by Expert Assessments (2018 Senate Election).

	Negative Tweets (Volume)			Policy Attacks in Tweets (Volume)			Character Attacks in Tweets (Volume)			Incivility in Tweets (Volume)		
	M1			M2			M3			M4		
	Coef	Se	P	Coef	Se	p	Coef	Se	P	Coef	Se	p
Negativity (experts)[a]	28.90	(13.88)	*	15.04	(14.59)		19.57	(11.42)	†	8.99	(3.94)	*
Republican	−71.12	(41.65)	†	−97.50	(43.78)	*	−76.41	(34.27)	*	−14.36	(11.82)	
Female	−26.56	(29.14)		−10.49	(30.64)		−23.13	(23.98)		−5.52	(8.27)	
Incumbent	−5.68	(32.74)		−47.62	(34.42)		−33.10	(26.95)		4.83	(9.29)	
State tossup[b]	−19.22	(15.69)		−4.13	(16.49)		−14.18	(12.91)		−3.66	(4.45)	
Percent Trump 2016	−46.24	(188.36)		100.55	(198.03)		20.22	(155.01)		−17.41	(53.45)	
Constant	176.12	(85.04)	*	151.73	(89.40)	†	149.50	(69.98)	*	34.84	(24.13)	
N(candidates)	49			49			49			49		
N(states)	27			27			27			27		
R2	0.14			0.12			0.16			0.14		

Note: Dependent variables measure the volume of, respectively, negative tweets, tweets including policy attacks, tweets including character attacks, and tweets including incivility. Volume is computed by multiplying the raw percentage of each quantity of interest (e.g., percentage of negative tweets) for a given candidate by the total number of tweets for that candidate. In all models, observations at the lower level (candidates) are nested into observations at the upper level (states).

[a] Underlying dimension extracted with PCA, from the three dimension of campaign negativity measured by experts: tone, use of character attacks, and use of fear appeals. Varies between −2.9 (lowest negativity) and +2.8 (highest negativity).

[b] Measures the extent to which the state was "safe" (either for Republicans or Democrats) or undecided (tossup) prior to the November 2018 election, based on projections made by POLITICO in the weeks before the vote; the variable varies between 0 "Safe state" and 3 "Tossup."

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, † $p < 0.1$.

negativity (M1). A similar trend exists for the volume of incivility (M4) and character attacks on Twitter (M3), albeit less strongly. We do not find a significant association between experts' negativity measure and the volume of policy attacks on Twitter (M2), which might be due to the fact that the expert measure mostly picked up the harsher components of campaign negativity (harsher personal attacks and fear appeals). All in all, however, the expert general perceptions of candidates' negativity and the volume of tweets coded as negative (and containing character attacks), as coded by our algorithm, go hand in hand quite strongly, and in the expected direction—even controlling for powerful drivers of campaign negativity, to which we turn in section 3.2.

Divergences. To be sure, for some candidates the level of negativity in their campaigns diverges quite considerably from what estimated by experts. In order to uncover any substantial patterns in this sense, we have computed for each candidate the distance between the volume of negativity on Twitter (coming from our algorithm) and the estimated volume of negativity predicted by our inferential model (Table 3, M1). We have computed such residual ($M = 0.0$, $SD = 83.9$) in such a way that positive values indicate that the model underestimated the volume of negativity when compared to the measurement produced by the algorithm, and negative values indicate the opposite (experts overestimated, or the algorithm is more conservative) (see Table D1 (Appendix D) for all divergence scores). By far the most extreme case when it comes to underestimation by the model is represented is Corey Stewart (R, VA). Even accounting for the fact that experts assessed Mr. Steward as having run a very negative campaign, the volume of negative tweets he employed was off the charts, much higher than what the model estimated. This is unlikely to be outlandish. Stewart is known for his "Trumpian" style, affiliations to ultranationalists, and frequent harsh comments against more moderate Republicans (Nwanevu, 2018). As such, it is actually likely that the algorithm correctly picked up the very extreme campaign run by Mr. Stewart, not only in absolute terms but also in comparison to all other candidates—something that experts were of course unable to do. On the other end of the spectrum we find Geoff Diehl (R, MA), Robert Flanders (R, RI), and Jon Tester (D, MT), for whom the algorithm measured less negativity than it was predicted by the model—perhaps indicating that these candidates campaigned more negatively *on other channels*. Perhaps as a confirmation, Mr. Tester is among the candidates having used the higher share of negative TV ads (64%), as attested by the Wesleyan Media Project (WMP; Fowler et al., 2020).[8]

The presence of these outliers is both a cautionary tale and a reassuring finding. On the one hand, it indicates that even if the convergent validity check was successful, extreme cases for which the algorithm is less successful cannot be excluded. On the other hand, the fact that these extreme cases can be explained logically is reassuring and allows for a better understanding of what the algorithm picks up (and not).

External Validity: What Drives Negativity?

Beyond convergent validity, any relevant measure must be able to tell something about the broad phenomenon it captures. Thus, an additional important test for the validity of a measure is to assess to what extent it is useful to predict known dynamics, or it is predicted by known drivers. In our specific case, we test to what extent some known drivers of negativity are associated with the presence or absence of attacks in the tweets. If we expect our algorithm to effectively capture tweet negativity, then we ought to expect that cases in which negativity should be expected to be greater should be more likely to be represented by tweets classified as negative, broadly speaking. It is important to note here that we are not interested in discussing the drivers of campaign negativity on social media. Rather, our substantive point in this section (and the following) is that being able to find theoretically relevant patterns when using our algorithm to code the content of campaigns in social media is likely to indicate, in our opinion, that the algorithm is measuring something that is substantively valid, above and beyond its technical components.

Why, and under which circumstances are candidates competing in elections more likely to go negative on their rivals? Research on the drivers of negative campaigning, even if comparatively less developed than research in its effects, provides some cues. First, consistent evidence in the US and internationally suggests that incumbent candidates are less likely to go negative (e.g., Lau & Pomper, 2004; Gainous & Wagner, 2014; Nai, 2020). The rationale here is two-fold. On the one hand, because they previously held an office, incumbents have experience and records to showcase, and have thus greater incentives to go positive; challengers often do not have such a possibility, and are thus more naturally driven to attack the incumbents (Nai, 2020), whose record while in the office is in any case already closely scrutinized by the public at large; because voters tend to rely on retrospective evaluations to make up their mind (Healy & Malhotra, 2013), challengers naturally try to expose bad deeds and inconsistencies in their opponent's record, program, and character. The fact that challengers receive comparatively a weaker media coverage than incumbents (Hopmann et al., 2011) should also act as incentive to go more negative, in the light of evidence showing that negativity is much more likely to attract media attention that positivity (e.g., Geer, 2012). On the other hand, incumbents have much to lose—much more so than challengers—and should thus be more wary of the potentially negative effects of harsh attacks; the public usually tend to dislike harsh campaigns (Fridkin & Kenney, 2011; Johnson-Cartee & Copeland, 1989), and candidates that go excessively negative face the risk of backlash effects where their net favorability in the eyes of the voters drops—instead of the target's (Shapiro & Rieger, 1992; Roese & Sande, 1993). Many scholars have shown that the prospect of electoral failure is a catalyst to adopt a more negative rhetoric (Harrington & Hess, 1996; Walter et al., 2014; Nai & Sciarini, 2018). Negative campaigning aims at reducing support for (and favorability of) the opponents; a candidate that starts with a comparative disadvantage (e.g., lagging behind in the polls) "has not succeeded in attracting undecided voters and, therefore, has to scare off the opponent's voters to stand a better chance" (Elmelund-Praestekaer, 2010, p. 141). Given that incumbents naturally start with a strong comparative advantage over challengers (Cox & Katz, 1996), these latter should face extra incentives to go negative. In their analysis of negative campaigning on Twitter during the 2016 Republican primaries, Gross & Johnson (2016) also find that candidates tended to "punch upwards" (excluding Donald Trump, an exception on many aspects).

Second, some evidence exists that candidates or the right-hand of the political spectrum are more likely to go negative. In the USA, Lau and Pomper (2001) and Gainous and Wagner (2014) show that Republican candidates are more likely to go negative than Democrats, which might perhaps be because GOP strategists have been shown to be more open to the idea of strategic attacks (Theilmann & Wilhite, 1998). On the other hand, some marginal evidence exists that voters identifying with the Democrats are, under some conditions, less sympathetic toward negativity than Republicans or independents (Ansolabehere & Iyengar, 1995; Mattes & Redlawsk, 2015). These trends seem to exist outside of the USA as well; in Switzerland, the most negative party by far during referenda campaigns is the far-right Schweizerische Volkspartei (SVP—Swiss People's Party; Nai & Sciarini, 2018), and an analysis of 172 candidates competing in elections worldwide shows that, indeed, the likelihood of going negative is higher for candidates on the right (Nai, 2020).

Third, we might expect female candidates to be less likely to go negative (but see Maier, 2015; Evans et al., 2014; Evans & Clark, 2016). Even today, reasons still exist to imagine women have a strategic disadvantage, compared to men, when adopting a more negative or harsher rhetoric. Social stereotypes generate shared expectations that female candidates should be passive, kind, and sympathetic (Huddy & Terkildsen, 1993; Fridkin et al., 2009; Krupnikov & Bauer, 2014), and harsh rhetoric clearly contrasts with these stereotypes, generating potentially stronger backlash effects (Kahn, 1996; Trent & Friedenberg, 2008).

The conditions of the race should also shape the strategic considerations leading to the decision to go negative (or not). Some authors suggest that more competitive or "close" races should lead to more negative campaigns because the stakes tend to be higher (Kahn & Kenney, 1999; Lau & Pomper, 2004; Elmelund-Praestekaer, 2008); yet, the opposite is also found (e.g., Francia & Herrnson, 2007). Given what is discussed above for incumbents, we believe that close situations should decrease the use of attacks in a case such as the US Senate elections. In elections with uncertain results, risk aversion should play a particularly important role; on the other hand, if elections run in "safe" states, the risks of

nefarious backlashes should be lower for both challengers and incumbents: the former have probably little to lose, and for the latter any backlash effects are unlikely to put a decisive dent in their chances to secure a victory. We might thus expect negativity to be especially high during elections fought in "safe" states. Finally, evidence exists linking the timing of the campaign with the use of more negative messages, so that little remaining time before the vote increases negativity (Ridout & Holland, 2010; Damore, 2002; Nai & Martinez i Coma, 2019b). According to most existing studies, negative campaigning is more likely to be effective for candidates that are seen as credible on the issues at stake—which is mostly achieved with positive campaigning. In other terms, "by waiting to go negative until after they have established themselves in the mind of voters, candidates may be perceived as more credible, which may increase the veracity of their attacks" (Damore, 2002, p. 673). On the other hand, and following the main rationale discussed above for incumbency and competitiveness of the race, late attacks might be ever riskier, and the time to correct potential backlash effects is limited. In this sense, a rationale could also be developed why negativity should decrease when election day looms—especially when the election is close.

Table 4 tests these assumptions, for the four indicators of negativity measured in the tweets via our algorithm. Model 1 estimates the likelihood that the tweet is coded as "negative," models M2 and M3 test for the presence of, respectively, policy and character attacks (only within tweets that have been coded as negative), and model M4 tests for the presence of incivility. All analyses are hierarchical binary logistic regressions, where tweets are nested within competing candidates.

Table 4 shows several results that are in line with our expectations. Incumbents are less likely to use character attacks (M3), in line with the idea that incumbents have much to lose if they are excessively harsh. Inversely, female candidates are more likely to use policy attacks (M2) and less likely to run character attacks (M3) than their male counterparts, confirming the idea that female candidates have incentives to stay away from harsher attacks that are potentially at odds with social gender stereotypes. No significant results appear for the candidate's partisanship, even if Republican candidates seem marginally to have been less negative overall. Given that the Democrats were generally more active on Twitter than the Republicans during the time period analyzed, the proportion of negative tweets to all tweets posted by a given party was also considered. While there was no significant difference in the average levels of negative tone between the parties, Republican candidates had a significantly larger proportion of tweets containing character attacks ($p < 0.001$, $M = 0.04$) and incivility ($p = 0.003$, $M = 0.02$) while the Democrats had a larger proportion of tweets containing policy attacks ($p = 0.002$, $M = 0.03$). However, considering that (1) perception of campaign negativity is a function of both the content of each ad (or, in this case, tweet) and the total volume of ads (tweets) people are exposed to (e.g., Stevens, 2009), and (2) comparison of the overall volume of negative tweets produces no significant results,

Table 4. Drivers of Negativity on Twitter (2018 Senate Election).

	Negative Tone			Policy Attacks			Character Attacks			Incivility		
	M1			M2			M3			M4		
	Coef	Se	p	Coef	Se	P	Coef	Se	p	Coef	Se	p
Weeks before the vote[a]	0.04	(0.01)	***	0.02	(0.01)	†	0.07	(0.01)	***	0.06	(0.01)	***
Republican	−0.15	(0.24)		−0.14	(0.19)		−0.21	(0.25)		−0.25	(0.38)	
Female	−0.34	(0.22)		0.31	(0.17)	†	−0.41	(0.23)	†	−0.39	(0.34)	
Incumbent	−0.03	(0.23)		0.06	(0.18)		−0.53	(0.24)	*	0.33	(0.36)	
State tossup[b]	−0.13	(0.09)		0.12	(0.07)		−0.22	(0.10)	*	−0.12	(0.15)	
Percent Trump 2016	−1.93	(1.05)	†	−0.21	(0.85)		−0.56	(1.13)		−1.97	(1.66)	
Constant	0.09	(0.52)		−0.40	(0.42)		−0.33	(0.56)		−2.51	(0.81)	**
N(tweets)	15,998			5052			5,052			15,998		
N(candidates)	63			63			63			63		
Log likelihood	−9060.9			−3385.0			−3095.6			−3207.9		

Note: All dependent variables are binary and are measured at the tweet level; they measure, respectively, whether the tweet is classified as negative (M1) the presence of a policy attack (within negative tweets only, M2), the presence of a character attack (within negative tweets only, M3), and the presence of incivility (M4). All models are hierarchical binary logistic regressions. In all models, observations at the lower level (tweets) are nested into observations at the upper level (candidates). The models exclude the tweets published the day of the election.

[a] Varies between 1 "last week before the election" and 10 "10 weeks before the election."

[b] Measures the extent to which the state was "safe" (either for Republicans or Democrats) or undecided (tossup) prior to the November 2018 election, based on projections made by POLITICO in the weeks before the vote; the variable varies between 0 "Safe state" and 3 "Tossup."

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, † $p < 0.1$.

these minor differences in negative dimensions between the parties are likely inconsequential. Turning to the campaign conditions, tweets during elections fought in competitive states—in our case, states whose result was expected to be a likely tossup[9]—are less include harsher character attacks (M3). The effect is relatively substantial; marginal effects show that the estimate share of character attacks tweets goes from about 22% in "tossup" states to about 36% in "safe" states. Finally, tweets published late in the race are less likely to be negative, uncivil, and if negative they are less likely to use harsh character attacks. This effect contradicts usual trends in the literature, suggesting that the end of the race tends to increase in negativity and harshness (e.g., Ridout & Holland, 2010; Nai & Martinez i Coma, 2019b), but could make sense in light of the effects shown for incumbency and closeness of the race: when the stakes are higher (in this case, little remaining time), risk averse behaviors kick in and negativity goes down.

A Replication Check: The 2020 Election

An even more severe test for the external validity of our automated measurement is to assess whether it is also able to predict theoretically meaningful dynamics in *another* election. To do this, we have used the algorithm developed for the 2018 election, discussed in this article so far, to classify the content of Tweets published by candidates having competed in the *2020* Senate election. In the November 2020 election, 33 Class 2 Senate seats were contested (plus additional special elections in Arizona and Georgia to fill vacancies, which we will not analyze here); 12 were previously held by Democrats and 21 by Republicans. Out of all candidates competing in the (mostly) two-party races for these seats, we were able to collect the Twitter posts for 63 of them (Table A2, Appendix A).[10] We collected via vicinitas.io all tweets published by these candidates for the period between August 29, 2020 and November 3, 2020 (the day of the election), matching the length of data collection used for 2018. A total of total of $N = 24{,}762$ tweets were collected, an increase of 53% compared to 2018. The number of tweets per candidate collected varies considerably, from $N = 3$ for Ben Sasse (R, NE, @SenSasse) to $N = 2055$ for John Cornyn (R, TX, @JohnCornyn), with an average of 393.0 tweets per candidate. These tweets were classified using the algorithm developed for the 2018 election (i.e., the same MLP classifier fitted on the manually annotated data from the 2018 election was used).

Table 5 replicates the models discusses in the previous section for 2018, and regresses the four dimensions classified by the algorithm (negative tone, policy attacks, character attacks, and incivility) on the profile of candidates and the nature of the context, that is, whether the state was a tossup or safe for either candidate, and the percentage of support for Trump in the previous presidential election (2016).

As for 2018, and in line with the literature, results for 2020 indicate that the harshest attacks (character attacks) are significantly less likely for incumbents and for female

Table 5. Drivers of Negativity on Twitter (2020 Senate Election).

	Negative Tone			Policy Attacks			Character Attacks			Incivility		
	M1			M2			M3			M4		
	Coef	Se	p	Coef	Se	p	Coef	Se	p	Coef	Se	p
Weeks before the vote[a]	−0.01	(0.01)	*	0.00	(0.01)		−0.01	(0.01)		−0.03	(0.01)	**
Republican	0.12	(0.17)		−0.52	(0.13)	***	0.46	(0.11)	***	0.25	(0.20)	
Female	−0.18	(0.19)		0.05	(0.15)		−0.29	(0.13)	*	−0.19	(0.23)	
Incumbent	−0.08	(0.17)		0.35	(0.13)	**	−0.50	(0.11)	***	−0.31	(0.20)	
State tossup[b]	−0.01	(0.08)		−0.01	(0.06)		0.02	(0.05)		−0.03	(0.10)	
Percent Trump 2016	−1.78	(0.90)	*	−0.81	(0.69)		0.04	(0.60)		−1.81	(1.07)	†
Constant	0.11	(0.47)		−0.08	(0.36)		−0.41	(0.31)		−1.71	(0.56)	**
N(tweets)	23,011			6914			6,914			23,011		
N(candidates)	59			58			58			59		
Log likelihood	−13,389.6			−4446.7			−4491.6			−5502.7		

Note: All dependent variables are binary and are measured at the tweet level; they measure, respectively, whether the tweet is classified as negative (M1) the presence of a policy attack (within negative tweets only, M2), the presence of a character attack (within negative tweets only, M3), and the presence of incivility (M4). All models are hierarchical binary logistic regressions. In all models, observations at the lower level (tweets) are nested into observations at the upper level (candidates). The models exclude the tweets published the day of the election.

[a] Varies between 1 "last week before the election" and 10 "10 weeks before the election."

[b] Measures the extent to which the state was "safe" (either for Republicans or Democrats) or undecided (tossup) prior to the November 2020 election, based on projections made by POLITICO in the weeks before the vote; the variable varies between 0 "Safe state" and 3 "Tossup."

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, † $p < 0.1$.

candidates (M3). Furthermore, incumbents were more likely to prefer policy attacks when going negative (M2; this trend was also present in 2018, but the coefficient was not significant).

Interestingly, Table 5 presents two instances of results that match the theoretical expectations discussed above but were not present in 2018. First, general negativity and incivility increase as the election day nears, as shown in the literature (Ridout & Holland, 2010; Damore, 2002; Nai & Martinez i Coma, 2019b). Second, Republican candidates were significantly more likely to use character attacks and less likely to use policy attacks (the effect of being a Republican on the presence of attacks in general and the use of uncivil messages is positive, but not significant), also in line with American and comparative research showing a greater propensity for right-wing candidates to go negative (Lau & Pomper, 2001; Gainous & Wagner, 2014; Nai, 2020). Additional models (Appendix C) show furthermore that Democrats and Republicans use different attacks depending on the political leaning of the context in which they campaign. As substantiated in Figure 4 with marginal effects, Democrats increasingly use character attacks in Republican states (high percentage of Trump votes in 2016) and policy attacks in Democrat strongholds (low percentage of Trump votes in 2016). Inversely, Republicans use more character attacks in Democratic states, and policy attacks in Republican states. Taken together, these results indicate that in more unfavorable contexts both Democrats and Republicans tend to prefer a more confrontational approach via harsher attacks, but rely on constructive, policy attacks in their strongholds—reflecting the general idea that conflict spawns conflict (Maier & Nai, 2021).

All in all, trends for 2020 are theoretically meaningful, suggesting that our automated measurement comes with a strong potential for replication in different contexts, on top of performing well in convergent and external validity.

Applications and Limitations

All in all, the results discussed in the previous sections seem to indicate that automated classifications are able to provide reliable measurements of campaign negativity. Triangulations with independent data show that our automatic classification is strongly associated with the experts' perceptions of the candidates' campaign. Furthermore, variations in our measures of negativity can be explained by theoretically relevant factors at the candidate and context levels (e.g., incumbency status and candidate gender); theoretically meaningful trends are also found when replicating the analysis on tweets published by candidates during the 2020 Senate election, coded using the automated classifier developed for 2018.

These results face some caveats. First, if the dataset at the tweet level was relatively consequential, it represented a restricted number of candidates overall. Studying the campaign behavior of candidates in the US senate elections offers multiple advantages (Lau & Pomper, 2004), but due to the mechanism of electoral competition in the US only about 60 candidates are actually competing for open seats in any given Midterm election. To be sure, the data at hand was large enough to test for the driving effect of characteristics at the candidate and context levels. Yet, the relatively small N prevented us for more nuanced analyses, especially in terms of composition effects—for instance, it could be argued that the gender plays a different role for Democrats and Republicans, but testing for such moderating effects would stretch the data too thinly. Second, the dynamics discussed in this article reflect a very narrow electoral setting, which only rarely finds a match outside of the US. At odds with the imperative of expanding our current knowledge of negative campaigning outside of the abundantly studied US case (e.g., Nai and Walter, 2015), our article plays the card of simplicity versus innovation when it comes to case selection. To be sure, there are several excellent reasons for selecting the US Senate Midterms for our study; yet, further research is direly needed to confirm the exportability of US results in different settings. Third, and relatedly, the jury is out regarding the exportability of our algorithm to textual data in other languages (but see, e.g., Huang et al., 2013).

Finally, some critical questions about the process behind algorithmic classification of negativity still remain to be answered. Although our tests suggest that the new automatic classification method exhibits satisfactory performance and is generalizable to other contexts, a lot remains unknown about what features the algorithm considers relevant and how exactly it decides on its classification of negativity. Yet, transparency in the algorithmic classification process is important: not only to be able to definitively assess the features' theoretical relevance and justifiability, but also to better detect possible systematic errors in the algorithm's predictions. Recent developments in explainable artificial intelligence (XAI) have sparked the creation of a number of frameworks aimed at opening the "black box" of machine learning algorithms. These frameworks rely on various mathematical models that attempt to approximate the black box model's predictions (globally or locally) while still allowing for interpretability of the results. Some of the widely used frameworks include LIME (Ribeiro et al., 2016), which uses local linear approximations to mimic the black box model's behavior, and SHAP (Lundberg & Lee, 2017) which borrows from game theory and utilizes Shapley values for model approximation. Regardless of the employed method, all of these frameworks aim to discern which features, and to what extent, contribute to the classification result. Employing them in future research on automated classification of negativity can shed some light on the algorithmic decision-making process and elucidate how robust and context-independent it is. Doing so can further strengthen the case for usability of automated measures of negativity and help discern in which scenarios such measures are most effective and in which less so.

With these limitations in mind, it is important to explicitly state what we believe the best usages for our algorithm are. First, from a practical standpoint, our algorithm can be used to assess the level

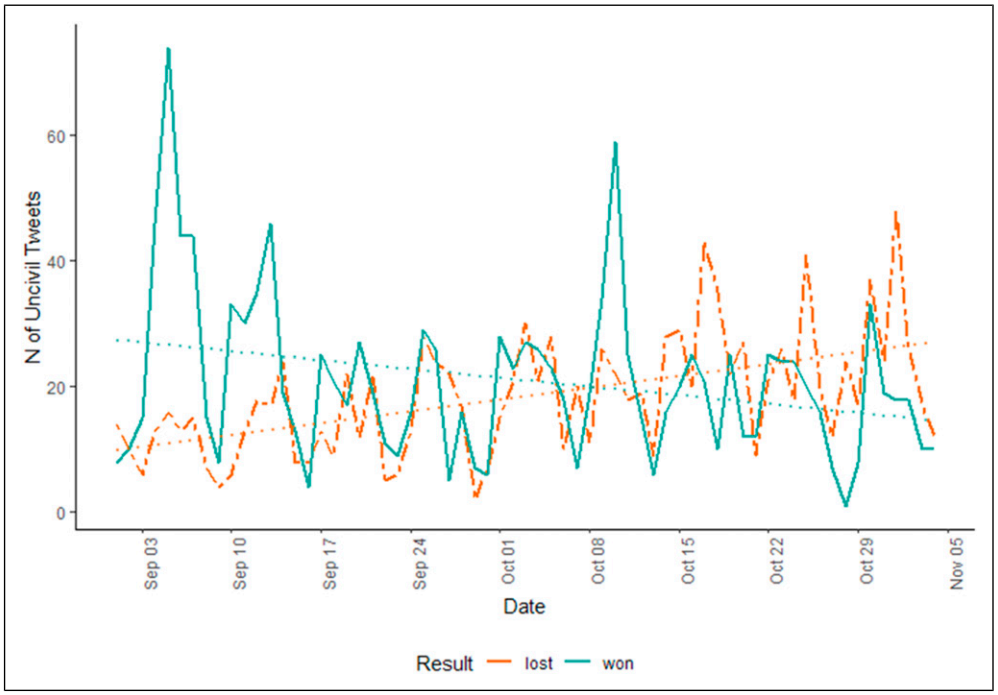

Figure 5. Frequency of policy attacks per day and election outcome (2018 Senate election).

of negativity in upcoming elections that reflect the structural and institutional patters of elections investigated here—that is, single-seat elections with a small number of competing candidates elected via plurality voting. This can be either done post-hoc once the election is over, or be scaled up as an interactive tool for real-time campaign assessment. But, given the materials on which the algorithm was developed, we strongly advocate against the use of the algorithm for the classification of elections under, for example, a plurality formula, where the absence of a zero-sum game makes it much more complicated to assess automatically who the target of attacks and incivility might be (and who might benefit from it). Second, from a methodological standpoint, our algorithm can however serve as a starting point to develop expanded classifications of campaign negativity that are less contingent on the institutional circumstances we faced here. Finally, and within the parameters discussed above, our algorithm can be used to provide independent external validity checks for all scholars engaged in content analysis of election campaigns in traditional communication channels, such as TV ads. Being largely automated, quite cost-effective, and scalable for large datasets, the algorithm ticks all the boxes for validity checks.

Discussion and Conclusion

The study of campaign messages has only recently turned toward the candidates' communication in social media (Gainous & Wagner, 2014; Graham et al., 2016; Straus et al., 2013). Several studies assessed the presence of negativity in social media (e.g., Auter & Fine, 2016; Ceron & d'Adda, 2016; Evans et al., 2014; Evans & Clark, 2016; Gainous & Wagner, 2014; Gross & Johnson, 2016), and broadly confirmed that the main trends of strategic campaigning found for traditional techniques—for instance, that challengers tend to attack more than incumbents (Gainous & Wagner, 2014)—are also found when looking at campaigning on social media.

Within this framework, our article introduces a neural network classifier that we trained to automatically annotate the tweets of candidates competing during the 2018 US Senate Midterms elections. The algorithm, trained on approx. 1000 tweets, was run on over 16,000 tweets, posted by 63 candidates for the period between September 1st and November 6th, 2018 (the day of the election) to classify the presence of political attacks (both in general and separately for policy and character attacks) and the use of incivility. Safe for the description of the algorithm construction, the bulk of the article was dedicated to a series of tests, showing the performance of our algorithm in terms of external and convergent validity of the measure. The first series of tests assessed the convergent validity of our measure. To do so, we compared it with independent data gathered within the framework of an expert survey (Nai & Maier, 2020). Controlling for several determinants at the candidate and state levels, our results show that campaigns that are evaluated by the experts as very negative (underlying dimension) have a volume of negativity on Twitter that is twice as high as campaigns that are evaluated as very low in negativity; a similar trend exists for the volume of incivility and character attacks on Twitter, albeit less strongly. All in all, this suggests high convergent validity of our automated measure of negativity on Twitter.

The second series of tests checked whether our automated measure "makes sense" in terms of factors that can be theoretically expected to drive the presence of negativity in the candidates' tweets. Our results show that consistently with evidence in the US and internationally (e.g., Lau & Pomper, 2004; Nai, 2020), incumbents are significantly less likely to go negative on their rivals. Also, in line with the theory (e.g., Kahn, 1996; Trent & Friedenberg, 2008), female candidates are less likely than their male counterparts to run harsher character attacks and more likely to attack on policy. Our results also show that tweets in competitive states are less likely to be negative and uncivil, and that tweets published *late* in the race are less likely to be negative, uncivil. This last effect is at odds with results found elsewhere (e.g., Ridout & Holland, 2010; Nai & Martinez i Coma, 2019b), but makes sense in light of the effects shown for incumbency and closeness of the race: when the stakes are higher (little remaining time), risk averse behaviors are probably triggered, reducing the chances that candidates go negative. The third series of tests took the external validity check a step further, and assessed whether the classifier developed for the 2018 election is able to provide a theoretically meaningful measurement of the tone of campaigns on Twitter for a different election altogether—the 2020 Senate election. Results show that as for 2018 (and in line with the literature), results for 2020 indicate that character attacks are significantly less likely for incumbents and for female candidates, and that incumbents were more likely to prefer policy attacks when going negative. In 2020, we also find several results that perfectly match the theoretical expectations, but were not present in 2018: negativity increases as the election day draws near, and Republican candidates were significantly more likely to use character attacks and less likely to use policy attacks. An additional series of models showed furthermore that both Democrats and Republicans tend to use harsher attacks in states controlled by the other party, in pine with the general idea that more conflictive contexts provide incentives for harsher campaigns (Maier & Nai, 2021). The fact that results are at times different from the trends showed for the 2018 election suggest that contextual determinants are quite likely to play an important role as well. The 2020 election was a rather different political event than the 2018 one, for many reasons—it included a presidential contest, whereas the 2018 only featured elections for the House and the Senate, and of course it took place in a unique political context marred by extreme polarization and the unfolding drama of the COVID-19 pandemic. These contextual drivers are likely to alter the dynamics at play and affect the incentives for candidates to go negative. Yet, the fact that for both the 2018 and 2020 elections the trends shown are broadly in line with the theoretical expectations reinforces our main claim that our automated classifier is able to yield an externally valid measurement of campaign negativity.

Our article contributes by expanding the state of the art in computational research for political communication. From a practical point of view, it demonstrates that automatic classification is a viable and efficient method for content analysis of campaign negativity. Researchers can rely on this method in the future to investigate dynamics of negative campaigning across parties and elections, the relationship between negativity and electoral outcomes, and drivers of campaign negativity, among other things. Furthermore, from a theoretical perspective, by enriching the tools at our disposal to measure negativity, our article indirectly contributes to a broader understanding of negative campaigning dynamics. As discussed above, our measure confirmed some established trends in the literature when it comes to the drivers of negative campaigning in elections. The mammoth task at hand now is to deepen our understanding of its consequences. Does negativity work as intended, by increasing the electoral prospect of the attackers while at the same time reducing support for the target? Does negativity mobilize or demobilize? And, on the long term, is negativity nefarious for our current standards of liberal democracy? It is not our goal here to develop on these fundamental, normative questions. We conclude by simply suggesting that fine-grained measurements of candidate negativity in social media have potentially an important role to play for these questions. Figure 5 plots the frequency of attacks per day, and depending on whether the candidate ultimately won or lost the race for the seat.

The figure seems to indicate a trend where an increase of negativity over time is associated with electoral defeat, in line with the idea that negativity is a risky business. To be sure, this simple association is not enough to conclude that (an increase in) negativity has detrimental effects on electoral results. The relationship between the use of attack politics and electoral success could very well be reversed so that candidates lagging behind in the polls and facing an electoral defeat could face enhanced incentives to attack (e.g., Skaperdas & Grofman, 1995; Harrington & Hess, 1996; Nai & Martinez i Coma, 2019b). The dynamics of attack politics and electoral support are complex, and likely to be marred by two-sided causality—negativity can drive more (un)favorable electoral standings in polls, which in turn set up (dis)incentives to go negative subsequently, and do forth (e.g., Blackwell, 2013). Such fine-grained analysis that also requires intermediate assessments of electoral standings (e.g., poll data) is beyond the scope of this article. But the data at hand is particularly appropriate to disentangle such nuanced dynamics, and further research in this sense is under way.

Appendix

Appendix A. Candidates

Appendix B. Classification Performance

Appendix C. Additional results

Appendix D. Divergences between real and predicted values for negative volume in tweets

Appendix A

Candidates

Table A1. Candidate Scores, by State (2018 Senate Election).

						Tweets					Experts				
Name	Party	State	Gender	Incumbent	Twitter handle	Negative[a]	Policy attacks[b]	Char. attacks[d]	Unciv[d]	N	Tone[e]	Char. attacks[f]	Fear[g]	NEG[h]	N
Martha McSally	R	Arizona	F		@SenMcSallyAZ	0.11	0.75	0.00	0.00	71	9.00	5.00	0.00	0.06	2
Kyrsten Sinema [i]	D	Arizona	F		@SenatorSinema	0.15	0.60	0.25	0.00	131	8.00	2.50			2
Kevin de Leon	D	California	M		@kdeleon	0.29	0.48	0.37	0.05	218	4.95	3.75	3.10	−0.53	30
Dianne Feinstein	D	California	F	yes	@SenFeinstein	0.59	0.51	0.46	0.09	349	2.85	2.50	2.30	−1.56	30
Matthew Corey	R	Connecticut	M		@MattCoreyCT	0.34	0.16	0.52	0.05	186	4.25	1.25	1.50	−1.73	8
Chris Murphy	D	Connecticut	M	yes	@ChrisMurphyCT	0.32	0.45	0.23	0.07	625	1.75	2.50	1.50	−2.01	8
Rob Arlett	R	Delaware	M		@RobArlett	0.57	0.40	0.57	0.02	154	7.50	7.50	5.00	1.51	2
Tom Carper	D	Delaware	M	yes	@SenatorCarper	0.42	0.60	0.19	0.09	150	2.00	1.25	2.00	−2.16	2
Bill Nelson	D	Florida	M	yes	@SenBillNelson	0.15	0.61	0.11	0.02	118	3.85	4.00	3.20	−0.71	6
Rick Scott	R	Florida	M		@SenRickScott	0.02	0.15	0.00	0.00	1028	8.10	7.50	5.20	1.71	6
Ron Curtis [i]	R	Hawaii	M		@rcurtis808	0.28	0.54	0.29	0.03	99	1.50		3.00		1
Mazie Hirono [i]	D	Hawaii	F	yes	@maziehirono	0.52	0.48	0.39	0.09	44	1.50		2.00		1
Mike Braun	R	Indiana	M		@braun4indiana	0.48	0.49	0.50	0.15	258	9.05	5.75	8.00	2.13	7
Joe Donnelly	D	Indiana	M	yes	@SenDonnelly	0.13	0.56	0.22	0.02	210	6.00	3.50	4.70	0.03	7
Eric Brakey	R	Maine	M		@SenatorBrakey	0.46	0.45	0.29	0.05	360	7.65	5.75	4.00	0.86	5
Angus King	D	Maine	M	yes	@SenAngusKing	0.23	0.38	0.46	0.07	57	0.60	1.25	1.50	−2.62	5
Tony Campbell [i]	R	Maryland	M		@Campbell4MD	0.26	0.45	0.45	0.02	43					8
Ben Cardin [i]	D	Maryland	M	yes	@SenatorCardin	0.47	0.48	0.36	0.14	186	0.60		0.60		8
Geoff Diehl	R	Massachusetts	M		@RepGeoffDiehl	0.25	0.18	0.45	0.02	87	7.30	5.75	7.00	1.47	17
Elizabeth Warren	D	Massachusetts	F	yes	@ewarren	0.35	0.34	0.32	0.06	445	4.60	2.75	5.20	−0.39	17
John James	R	Michigan	M		@JohnJamesMI	0.17	0.40	0.31	0.01	496	4.80	4.50	6.20	0.35	9
Debbie Stabenow	D	Michigan	F	yes	@SenStabenow	0.19	0.57	0.07	0.01	72	2.20	2.50	2.20	−1.74	9
Amy Klobuchar	D	Minnesota	F	yes	@amyklobuchar	0.24	0.27	0.29	0.02	324	1.10	0.75	0.30	−2.91	8
Jim Newberger	R	Minnesota	M		@NewbergerJim	0.31	0.29	0.38	0.04	68	6.80	5.00	6.80	1.11	8
David Baria	D	Mississippi	M		@dbaria	0.26	0.41	0.26	0.02	674	2.15	0.00	0.70	−2.76	6
Roger Wicker	R	Mississippi	M	yes	@SenatorWicker	0.17	0.34	0.31	0.02	173	3.85	2.50	5.00	−0.69	6
Josh Hawley	R	Missouri	M		@HawleyMO	0.53	0.44	0.52	0.21	339	7.10	5.75	6.70	1.35	7
Claire McCaskill	R	Missouri	F	yes	@clairecmc	0.36	0.48	0.24	0.07	249	8.30	6.75	6.70	1.91	7
Matt Rosendale	R	Montana	M		@MattForMontana	0.34	0.45	0.31	0.10	439	9.50	7.50	8.50	2.81	2
Jon Tester	D	Montana	M	yes	@SenatorTester	0.29	0.52	0.07	0.05	101	8.50	2.50	6.00	0.68	2
Deb Fischer	R	Nebraska	F	yes	@SenatorFischer	0.04	0.25	0.00	0.00	104	5.25	2.50	2.00	−1.04	2
Jane Raybould	D	Nebraska	F		@JaneRaybould	0.28	0.46	0.35	0.04	164	5.25	1.25	4.00	−0.90	2
Dean Heller [i]	R	Nevada	M	yes	@SenDeanHeller	0.14	0.42	0.05	0.00	140	8.50	7.50	8.00	2.45	1
Jacky Rosen [i]	D	Nevada	F		@SenJackyRosen	0.17	0.57	0.10	0.02	127	3.00	2.50	3.00	−1.36	1
Bob Hugin	R	New Jersey	M	yes	@BobHugin	0.57	0.65	0.25	0.10	184	8.60	8.75	3.30	1.71	6
Bob Menendez	D	New Jersey	M		@SenatorMenendez	0.46	0.15	0.64	0.13	445	6.35	5.75	5.80	0.96	6
Martin Heinrich	D	New Mexico	M	yes	@MartinHeinrich	0.43	0.46	0.41	0.06	90	1.85	0.00	1.30	−2.69	3
Mick Rich	R	New Mexico	M		@MickRich4Senate	0.40	0.20	0.61	0.03	103	6.00	5.00	7.00	0.96	3
Kirsten Gillibrand	D	New York	F	yes	@SenGillibrand	0.30	0.56	0.30	0.06	403	2.60	2.50	3.00	−1.46	12
Kevin Cramer	R	North Dakota	M		@SenKevinCramer	0.14	0.62	0.52	0.00	57					
Heidi Heitkamp	D	North Dakota	F	yes	@HeidiHeitkamp	0.29	0.65	0.15	0.03	414					
Sherrod Brown	D	Ohio	M	yes	@SherrodBrown	0.30	0.50	0.23	0.03	520	6.15	4.75	5.50	0.58	9
Jim Renacci	R	Ohio	M		@JimRenacci	0.26	0.35	0.35	0.00	274	8.30	8.25	6.30	2.21	9
Lou Barletta	R	Pennsylvania	M		@RepLouBarletta	0.29	0.58	0.17	0.00	82	7.15	6.00	7.50	1.62	18
Bob Casey Jr.	D	Pennsylvania	M	yes	@SenBobCasey	0.44	0.33	0.44	0.13	82	4.50	3.75	4.10	−0.41	18
Robert Flanders	R	Rhode Island	M		@flanders4senate	0.27	0.23	0.37	0.05	111	7.50	5.75	7.50	1.64	3
Sheldon Whitehouse	D	Rhode Island	M	yes	@SenWhitehouse	0.50	0.34	0.44	0.13	412	3.50	2.50	3.00	−1.24	3
Marsha Blackburn	R	Tennessee	F		@MarshaBlackburn	0.24	0.40	0.30	0.02	42	8.90	5.00	8.50	2.01	4
Phil Bredesen	D	Tennessee	M		@PhilBredesen	0.22	0.38	0.33	0.01	673	3.75	3.75	4.00	−0.62	4
Ted Cruz	R	Texas	M	yes	@SenTedCruz	0.29	0.45	0.17	0.03	363	7.40	5.50	7.60	1.57	13
Beto O'Rourke	D	Texas	M		@BetoORourke	0.17	0.49	0.16	0.01	687	2.85	3.00	3.00	−1.26	13
Mitt Romney [i]	R	Utah	M		@MittRomney	0.29	0.29	0.43	0.04	24	1.00		0.80		4
Jenny Wilson	D	Utah	F		@JennyWilsonUT	0.15	0.34	0.28	0.01	471	5.35	4.25	2.00	−0.56	4
Bernie Sanders	D	Vermont	M	yes	@SenSanders	0.61	0.58	0.30	0.07	261	3.75	0.00	6.00	−1.13	2
Tim Kaine	D	Virginia	M	yes	@timkaine	0.36	0.37	0.45	0.13	107	2.90	2.50	3.20	−1.34	8
Corey Stewart	R	Virginia	M		@CoreyStewartVA	0.57	0.47	0.38	0.16	958	9.25	6.25	9.50	2.66	8
Maria Cantwell	D	Washington	F	yes	@SenatorCantwell	0.24	0.61	0.21	0.03	318	1.25	5.00	2.30	−1.29	3
Susan Hutchison	R	Washington	F		@Susan4Senate	0.33	0.56	0.35	0.08	202	7.00	6.25	6.50	1.41	3
Joe Manchin	D	West Virginia	M	yes	@Sen_JoeManchin	0.11	0.57	0.14	0.02	64					
Patrick Morrisey	R	West Virginia	M		@MorriseyWV	0.56	0.39	0.60	0.10	265					
Tammy Baldwin	D	Wisconsin	F	yes	@SenatorBaldwin	0.32	0.76	0.13	0.13	167	4.70	4.00	3.60	−0.41	9
John Barrasso [i]	R	Wyoming	M	yes	@SenJohnBarrasso	0.20	0.40	0.00	0.01	76	4.00	2.50	3.00	−1.11	1
Gary Trauner [i]	D	Wyoming	M		@TraunerforWY	0.28	0.38	0.88	0.03	29	6.00	2.50	4.00	−0.39	1

[a] Percent of tweets classified as "negative" by the algorithm.
[b] Percent of tweets classified as "policy attacks" by the algorithm (only among negative tweets).
[c] Percent of tweets classified as "character attacks" by the algorithm (only among negative tweets).
[d] Percent of tweets classified as containing "incivility."
[e] Tone of the campaign of the candidate, as evaluated by experts; varies between 0 "very positive" and 10 "very negative."
[f] Use of character attacks by the candidate, as evaluated by experts; varies between 0 "policy attacks exclusively" and 10 "character attacks exclusively."
[g] Use of fear appeals by the candidate, as evaluated by experts; varies between 0 "very low" and 10 "very high."
[h] Underlying dimension extracted with PCA, from the three dimension of campaign negativity measured by experts: tone, use of character attacks, and use of fear appeals. Varies between −2.9 (lowest negativity) and +2.8 (highest negativity).
[i] Candidate excluded for comparisons between twitter and expert data, due to missing values on key variables or fewer than 2 expert scores.

Table A2. Candidate Scores, by State (2020 Senate Election).

						Tweets				
Name	Party	State	Gender	Incumbent	Twitter handle	Negative[a]	Policy attacks[b]	Char. attacks[d]	Unciv[d]	N
Doug Jones	D	Alabama	M	Yes	@SenDougJones	0.17	0.32	0.23	0.00	133
Thomas Tuberville	R	Alabama	M		@TTuberville	0.40	0.28	0.47	0.10	231
Al Gross	I	Alaska	M		@DrAlGrossAK	0.57	0.28	0.45	0.19	309
Dan Sullivan	R	Alaska	M	Yes	@SenDanSullivan	0.13	0.00	0.60	0.05	76
Ricky Harrington	I	Arkansas	M		@RickDHarrington	0.27	0.30	0.43	0.07	460
Thomas Cotton	R	Arkansas	M	Yes	@SenTomCotton	0.48	0.35	0.36	0.12	168
Cory Gardner	R	Colorado	M	Yes	@SenCoryGardner	0.17	0.19	0.19	0.02	124
Christopher Coons	D	Delaware	M	Yes	@ChrisCoons	0.45	0.55	0.28	0.10	410
Jon Ossoff	D	Georgia	M		@ossoff	0.30	0.26	0.58	0.09	631
David Perdue	R	Georgia	M	Yes	@sendavidperdue	0.26	0.33	0.41	0.03	105
Paulette Jordan	D	Idaho	F		@PauletteEJordan	0.16	0.26	0.38	0.03	1056
James Risch	R	Idaho	M	Yes	@SenatorRisch	0.13	0.21	0.53	0.01	141
Mark Curran	R	Illinois	M		@ElectMarkCurran	0.43	0.28	0.57	0.16	518
Richard Durbin	D	Illinois	M	Yes	@SenatorDurbin	0.60	0.58	0.30	0.13	458
Theresa Greenfield	D	Iowa	F		@GreenfieldIowa	0.38	0.51	0.24	0.06	452
Joni Ernst	R	Iowa	F	Yes	@SenJoniErnst	0.21	0.35	0.35	0.04	349
Barbara Bollier	R	Kansas	F		@BarbaraBollier	0.30	0.41	0.30	0.09	469
Roger Marshall	R	Kansas	M		@RogerMarshallMD	0.37	0.34	0.38	0.09	378
Amy McGrath	D	Kentucky	F		@AmyMcGrathKY	0.26	0.41	0.23	0.05	465
Mitch McConnell	R	Kentucky	M	Yes	@LeaderMcConnell	0.69	0.48	0.44	0.20	89
Adrian Perkins	D	Louisiana	M		@PerkinsforLA	0.24	0.40	0.40	0.06	676
Bill Cassidy	R	Louisiana	M	Yes	@SenBillCassidy	0.25	0.44	0.23	0.07	306
Sara Gideon	D	Maine	F		@SaraGideon	0.36	0.56	0.21	0.03	608
Susan Collins	R	Maine	F	Yes	@SenatorCollins	0.17	0.30	0.09	0.01	136
Ed Markey	D	Massachusetts	M	Yes	@EdMarkey	0.29	0.55	0.19	0.04	743
Kevin O'Connor	R	Massachusetts	M		@KOCforSenate	0.31	0.21	0.61	0.13	364
John James	R	Michigan	M		@JohnJamesMI	0.32	0.17	0.46	0.07	109
Gary Peters	D	Michigan	M	Yes	@SenGaryPeters	0.27	0.60	0.13	0.02	167
Jason Lewis	R	Minnesota	M		@LewisForMN	0.52	0.37	0.50	0.15	664
Tina Smith	D	Minnesota	F	Yes	@SenTinaSmith	0.36	0.44	0.36	0.09	99
Mike Espy	D	Mississippi	M		@MikeEspyMS	0.21	0.30	0.38	0.04	869
Cindy Hyde-Smith	R	Mississippi	F	Yes	@SenHydeSmith	0.24	0.33	0.35	0.01	170
Steve Bullock	D	Montana	M		@GovernorBullock	0.12	0.47	0.11	0.00	164
Steve Daines	R	Montana	M	Yes	@SteveDaines	0.15	0.33	0.33	0.02	144
Chris Janicek	D	Nebraska	M		@CJSenate2020	0.39	0.26	0.54	0.10	100
Ben Sasse	R	Nebraska	M	Yes	@SenSasse	0.00	0.00	0.00	0.00	3
Corky Messner	R	New Hampshire	M		@CorkyForNH	0.35	0.23	0.50	0.07	607
Jeanne Shaheen	D	New Hampshire	F	Yes	@SenatorShaheen	0.31	0.58	0.21	0.05	312
Cory Booker	D	New Jersey	M	Yes	@CoryBooker	0.14	0.32	0.42	0.03	797
Rik Mehta	R	New Jersey	M		@RikMehta_NJ	0.38	0.27	0.58	0.11	159
Mark Ronchetti	R	New Mexico	M		@MarkRonchettiNM	0.23	0.07	0.41	0.01	117
Ben Ray Luján	D	New Mexico	M		@SenatorLujan	0.19	0.55	0.21	0.02	177
Cal Cunningham	D	North Carolina	M		@CalforNC	0.44	0.52	0.38	0.06	377
Thom Tillis	R	North Carolina	M	Yes	@SenThomTillis	0.30	0.28	0.42	0.03	168
Abby Broyles	D	Oklahoma	F		@abbybroyles	0.29	0.34	0.36	0.05	201
Jim Inhofe	R	Oklahoma	M	Yes	@JimInhofe	0.35	0.45	0.27	0.04	220
Jo Rae Perkins	R	Oregon	F		@PerkinsForUSSen	0.23	0.18	0.65	0.08	173
Jeff Merkley	D	Oregon	M	Yes	@SenJeffMerkley	0.55	0.63	0.25	0.12	373

(continued)

Table A2. (continued)

						Tweets				
Jack Reed	D	Rhode Island	M	Yes	@SenJackReed	0.43	0.49	0.39	0.14	113
Jaime Harrison	D	South Carolina	M		@harrisonjaime	0.19	0.26	0.43	0.03	604
Lindsey Graham	R	South Carolina	M	Yes	@LindseyGrahamSC	0.39	0.21	0.50	0.10	259
Dan Ahlers	D	South Dakota	M		@ahlers_dan	0.06	0.20	0.40	0.01	85
Mike Rounds	R	South Dakota	M	Yes	@SenatorRounds	0.10	0.50	0.13	0.01	82
Bill Hagerty	R	Tennessee	M		@BillHagertyTN	0.42	0.40	0.54	0.08	894
Marquita Bradshaw	D	Tennessee	F		@BradshawforTN	0.13	0.28	0.38	0.05	1042
John Cornyn	R	Texas	M	Yes	@JohnCornyn	0.34	0.34	0.46	0.10	2055
Mary Jennings Hegar	D	Texas	F		@mjhegar	0.26	0.37	0.40	0.06	640
Daniel Gade	R	Virginia	M		@gadeforvirginia	0.49	0.39	0.45	0.09	328
Mark Warner	D	Virginia	M	Yes	@MarkWarner	0.34	0.49	0.29	0.08	293
Paula Jean Swearengin	D	West Virginia	F		@paulajean2020	0.23	0.26	0.38	0.04	1126
Shelley Moore Capito	R	West Virginia	F	Yes	@SenCapito	0.17	0.25	0.32	0.02	319
Cynthia Lummis	R	Wyoming	F		@CynthiaMLummis	0.28	0.36	0.53	0.08	160
Merav Ben-David	D	Wyoming	F		@MBenDavid2020	0.20	0.38	0.35	0.06	737

[a]Percent of tweets classified as "negative" by the algorithm.
[b]Percent of tweets classified as "policy attacks" by the algorithm (only among negative tweets).
[c]Percent of tweets classified as "character attacks" by the algorithm (only among negative tweets).
[d] Percent of tweets classified as containing "incivility."

Appendix B

Classification Performance

Table B1. Classification Statistics with Single-Label Classifiers.

	F1 Score (Absence of Dimension)	F1 Score (Presence of Dimension)	Area under ROC Curve
No lemmas, stop-words kept			
Negative tone	0.82	0.83	0.82
Personal attack	0.89	0.74	0.82
Political attack	0.87	0.60	0.73
Incivility	0.96	0.78	0.87
No lemmas, stop-words removed			
Negative tone	0.81	0.83	0.82
Personal attack	0.92	0.75	0.83
Political attack	0.89	0.77	0.82
Incivility	0.95	0.78	0.86
Lemmas, stop-words kept			
Negative tone	0.79	0.80	0.80
Personal attack	0.90	0.72	0.80
Political attack	0.92	0.77	0.84
Incivility	0.96	0.78	0.87
Lemmas, stop-words removed			
Negative tone	0.81	0.81	0.81
Personal attack	0.80	0.72	0.80
Political attack	0.88	0.69	0.79
Incivility	0.94	0.73	0.83

Table B2. Classification Statistics with Multi-Label Classifiers.

	F1 Score (Absence of Dimension)	F1 Score (Presence of Dimension)	Area under ROC Curve
No lemmas, stop-words kept			
Negative tone	0.84	0.83	0.83
Personal attack	0.92	0.72	0.79
Political attack	0.89	0.88	0.75
Incivility	0.91	0.71	0.83
No lemmas, stop-words removed			
Negative tone	0.81	0.83	0.82
Personal attack	0.92	0.75	0.82
Political attack	0.89	0.77	0.78
Incivility	0.92	0.76	0.83
Lemmas, stop-words kept			
Negative tone	0.79	0.80	0.80
Personal attack	0.90	0.72	0.80
Political attack	0.92	0.77	0.84
Incivility	0.96	0.78	0.87
Lemmas, stop-words removed			
Negative tone	0.81	0.81	0.81
Personal attack	0.90	0.72	0.80
Political attack	0.86	0.69	0.79
Incivility	0.94	0.73	0.83

Table B3. Parameter Space Used in Model Selection.

Parameter	Options				
Hidden later sizes	475,237	475	700	300	100
Activation function	tanh	relu			
Solver	sgd	adam	lbfgs		
Alpha	0.0001	0.001	0.05	0.1	
Learning rate	Constant	Adaptive			

Appendix C

Additional results

Table C1. Drivers of Negativity on Twitter (2020 Senate Election); Interaction Republican * Trump 2016.

	Negative Tone			Policy Attacks			Character Attacks			Incivility		
	M1			M2			M3			M4		
	Coef	Se	p	Coef	Se	p	Coef	Se	p	Coef	Se	p
Weeks before the vote[a]	−0.01	(0.01)	*	0.00	(0.01)		−0.01	(0.01)		−0.03	(0.01)	**
Republican (REP)	−2.36	(0.98)	*	−3.70	(0.67)	***	1.55	(0.66)	*	−1.48	(1.24)	
Female	−0.13	(0.18)		0.14	(0.13)		−0.33	(0.12)	**	−0.15	(0.23)	
Incumbent	−0.32	(0.19)	†	0.02	(0.13)		−0.39	(0.13)	**	−0.49	(0.24)	*
State tossup[b]	−0.01	(0.08)		−0.01	(0.05)		0.02	(0.05)		−0.03	(0.10)	
Percent Trump 2016	−4.36	(1.32)	***	−4.17	(0.89)	***	1.19	(0.89)		−3.64	(1.69)	*
REP * Prc Trump 2016	4.85	(1.89)	*	6.28	(1.29)	***	−2.14	(1.27)	†	3.39	(2.40)	

(continued)

Table C1. (continued)

	Negative Tone			Policy Attacks			Character Attacks			Incivility		
	M1			M2			M3			M4		
	Coef	Se	p	Coef	Se	p	Coef	Se	p	Coef	Se	p
Constant	1.51	(0.71)	*	1.73	(0.47)	***	−1.03	(0.47)	*	−0.72	(0.89)	
N(tweets)	23,011			6914			6,914			23,011		
N(candidates)	59			58			58			59		
Log likelihood	−13,386.5			−4436.6			−4490.6			−5501.7		

Note: All dependent variables are binary and are measured at the tweet level; they measure, respectively, whether the tweet is classified as negative (M1) the presence of a policy attack (within negative tweets only, M2), the presence of a character attack (within negative tweets only, M3), and the presence of incivility (M4). All models are hierarchical binary logistic regressions. In all models, observations at the lower level (tweets) are nested into observations at the upper level (candidates). The models exclude the tweets published the day of the election.
[a]Varies between 1 "last week before the election" and 10 "10 weeks before the election."

[b]Measures the extent to which the state was "safe" (either for Republicans or Democrats) or undecided (tossup) prior to the November 2020 election, based on projections made by POLITICO in the weeks before the vote; the variable varies between 0 "Safe state" and 3 "Tossup."
*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, † $p < 0.1$.

Appendix D

Divergences between real and predicted values for negative volume in tweets

Table D1. Divergences between Real and Predicted Values for Negative Volume in Tweets.

Candidate	Party	State	Difference Real-Predicted
Martha McSally	R	AZ	7.73783
Kevin de Leon	D	CA	−83.10718
Dianne Feinstein	D	CA	121.7645
Matthew Corey	R	CT	27.92056
Chris Murphy	D	CT	104.6332
Rob Arlett	R	DE	−41.41886
Tom Carper	D	DE	−25.67026
Bill Nelson	D	FL	−51.52438
Rick Scott	R	FL	−53.9429
Mike Braun	R	IN	40.37614
Joe Donnelly	D	IN	−60.36561
Eric Brakey	R	ME	54.9453
Angus King	D	ME	−60.99203
Geoff Diehl	R	MA	−110.3468
Elizabeth Warren	D	MA	37.59306
John James	R	MI	11.1099
Debbie Stabenow	D	MI	−38.41954
Amy Klobuchar	D	MN	38.92776
Jim Newberger	R	MN	−95.20815
David Baria	D	MS	106.2888

(continued)

Table D1. (continued)

Candidate	Party	State	Difference Real-Predicted
Roger Wicker	R	MS	−23.71232
Josh Hawley	R	MO	119.8291
Claire McCaskill	R	MO	45.06267
Matt Rosendale	R	MT	28.08265
Jon Tester	D	MT	−96.67476
Deb Fischer	R	NE	−11.45242
Jane Raybould	D	NE	−50.25604
Bob Hugin	R	NJ	8.092003
Bob Menendez	D	NJ	65.37453
Martin Heinrich	D	NM	−35.22695
Mick Rich	R	NM	−73.17657
Kirsten Gillibrand	D	NY	37.08179
Sherrod Brown	D	OH	10.92229
Jim Renacci	R	OH	−54.63655
Lou Barletta	R	PA	−86.21456
Bob Casey Jr.	D	PA	−81.1003
Robert Flanders	R	RI	−104.3026
Sheldon Whitehouse	D	RI	90.28143
Marsha Blackburn	R	TN	−60.11504
Phil Bredesen	D	TN	59.19117
Ted Cruz	R	TX	21.91499
Beto O'Rourke	D	TX	41.00295
Jenny Wilson	D	UT	−41.30682
Bernie Sanders	D	VT	35.26049
Tim Kaine	D	VA	−73.29123
Corey Stewart	R	VA	385.6865
Maria Cantwell	D	WA	−13.47005
Susan Hutchison	R	WA	−36.23086
Tammy Baldwin	D	WI	−36.91682

We have computed for each candidate the distance between the volume of negativity on Twitter (coming from our algorithm) and the estimated volume of negativity predicted by our inferential model (Table 3, M1). We have computed such residual (M = 0.0, *SD* = 83.9) in such a way that positive values indicate that the model (i.e., experts) underestimated the volume of negativity when compared to the measurement produced by the algorithm, and negative values indicate the opposite (experts overestimated, or the algorithm is more conservative). Full results are in Table D1.

Acknowledgements

We are very grateful to the journal editor and anonymous reviewers for their careful reading and constructive suggestions, and to Sebastian Stier and Jürgen Maier for precious inputs and feedback. All remaining mistakes are of course our own. Lieke Bos, Camilla Frericks, and Nilou Yekta were instrumental for data collection and coding – thank you! Finally, a sincere thanks to all experts that have donated their time to participate in our expert survey in the aftermath of the 2018 Midterms – and to those who are currently participating in the parallel expert survey covering elections worldwide. We will pay it forward.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

Vladislav Petkevic https://orcid.org/0000-0001-6830-4199
Alessandro Nai https://orcid.org/0000-0001-7303-2693

Notes

1. When multiple handles existed for a given candidate, data was collected for their official campaign account (vs., e.g., their personal one); for instance, for Elizabeth Warren (D, MA), we

used the handle @ewarren (official campaign account) and not the handle @SenWarren (official Senate account).
2. While vicinitas.io provides access only to the latest 3000 tweets posted by the given user, this was sufficient to acquire an exhaustive dataset of all the candidates' tweets as no single candidate exceeded the 3000 limit in the period studied.
3. https://spacy.io/models/en#en_core_web_lg
4. https://commoncrawl.org/
5. https://catalog.ldc.upenn.edu/LDC2013T19
6. https://github.com/explosion/spaCy/blob/master/spacy/lang/en/stop_words.py
7. Training the model on individual word vectors was also performed, but the results produced were worse than those of the average vector model.
8. See their "Supplementary material 1," available on the publisher's website.
9. Election closeness at the state level is measured via a scale ranging between 0 "Safely Democrat/Republican" and 3 "Tossup," using the projections made by POLITICO in the weeks before the vote; see: https://www.politico.com/election-results/2018/house-senate-race-ratings-and-predictions/
10. We did not collect any tweets for John Hickenlooper (D, CO), Lauren Witzke (R, DE), Derrick Edwards (D, LA), Kevin O'Connor (Legal Marijuana Now, MN), Gene Siadek (Libertarian, NE), and Allen Waters (R, RI).

References

Ansolabehere, S., & Iyengar, S. (1995). *Going negative: How attack ads shrink and polarize the electorate*. Free Press.

Ansolabehere, S., Iyengar, S., Simon, A., & Valentino, N. (1994). Does attack advertising demobilize the electorate? *American Political Science Review*, *88*(4), 829−838. http://doi.org/10.2307/2082710

Auter, Z. J., & Fine, J. A. (2016). Negative campaigning in the social media age: Attack advertising on Facebook. *Political Behavior*, *38*(4), 999−1020. https://doi.org/10.1007/s11109-016-9346-8

Bennett, W. L., & Manheim, J. B. (2006). The one-step flow of communication. *The ANNALS of the American Academy of Political and Social Science*, *608*(1), 213−232. https://doi.org/10.1177/0002716206292266

Blackwell, M. (2013). A framework for dynamic causal inference in political science. *American Journal of Political Science*, *57*(2), 504−520. https://doi.org/10.1111/j.1540-5907.2012.00626.x

Campello, R. J., Moulavi, D., & Sander, J. (2013, April). Density-based clustering based on hierarchical density estimates. In Lecture Notes in Computer Science (7819) Pacific-Asia conference on knowledge discovery and data mining (pp. 160−172). Gold Coast, QLD, Australia, 14-17 April 2013. Springer.

Ceron, A., & d'Adda, G. (2016). E-campaigning on Twitter: The effectiveness of distributive promises and negative campaign in the 2013 Italian election. *New Media & Society*, *18*(9), 1935−1955. https://doi.org/10.1177/1461444815571915

Cox, G. W., & Katz, J. N. (1996). Why did the incumbency advantage in US House elections grow? *American Journal of Political Science*, *40*(2), 478−497. https://doi.org/10.2307/440290

Damore, D. F. (2002). Candidate strategy and the decision to go negative. *Political Research Quarterly*, *55*(3), 669−685. https://doi.org/10.1177/106591290205500309

Elmelund-Præstekær, C. (2008). Negative campaigning in a multiparty system. *Representation*, *44*(1), 27−39. https://doi.org/10.1080/00344890701869082

Elmelund-Præstekær, C. (2010). Beyond American negativity: Toward a general understanding of the determinants of negative campaigning. *European Political Science Review*, *2*(1), 137−156. http://doi.org/10.1017/S1755773909990269

Engesser, S., Fawzi, N., & Larsson, A. O. (2017). Populist online communication: Introduction to the special issue. *Information, Communication & Society*, *20*(9), 1279−1292. https://doi.org/10.1080/1369118X.2017.1328525

Evans, H. K., & Clark, J. H. (2016). "You tweet like a girl!": How female candidates campaign on Twitter. *American Politics Research*, *44*(2), 326−352. https://doi.org/10.1177/1532673X15597747

Evans, H. K., Cordova, V., & Sipole, S. (2014). Twitter style: An analysis of how house candidates used Twitter in their 2012 campaigns. *PS: Political Science & Politics*, *47*(2), 454−462. http://doi.org/10.1017/S1049096514000389

Finkel, S. E., & Geer, J. G. (1998). A spot check: Casting doubt on the demobilizing effect of attack advertising. *American journal of political science*, *42*(2), 573−595. https://doi.org/10.2307/2991771

Fowler, E. F., Franz, M. M., & Ridout, T. N. (2020). The blue wave: Assessing political advertising trends and democratic advantages in 2018. *PS: Political Science & Politics*, *53*(1), 57−63. http://doi.org/10.1017/S1049096519001240

Francia, P. L., & Herrnson, P. S. (2007). Keeping it professional: The influence of political consultants on candidate attitudes toward negative campaigning. *Politics & Policy*, *35*(2), 246−272. https://doi.org/10.1111/j.1747-1346.2007.00059.x

Fridkin, K. L., & Kenney, P. J. (2011). Variability in citizens' reactions to different types of negative Campaigns. *American Journal of Political Science*, *55*(2), 307−325. http://www.jstor.org/stable/23025053

Fridkin, K. L., Kenney, P. J., & Woodall, G. S. (2009). Bad for men, better for women: The impact of stereotypes during negative campaigns. *Political Behavior*, *31*(1), 53−77. https://doi.org/10.1007/s11109-008-9065-x

Gainous, J., & Wagner, K. M. (2014). *Tweeting to power: The social media revolution in American politics*. Oxford University Press.

Geer, J. G (2006). *Defense of negativity: Attack ads in presidential campaigns*. University of Chicago Press.

Geer, J. G (2012). The news media and the rise of negativity in presidential campaigns. *PS: Political Science & Politics*, *45*(03), 422−427. http://doi.org/10.1017/S1049096512000492

Graham, T., Jackson, D., & Broersma, M. (2016). New platform, old habits? Candidates' use of Twitter during the 2010 British and Dutch general election campaigns. *New Media & Society*, (*18*(5), 765−783. https://doi.org/10.1177/1461444814546728

Gross, J. H., & Johnson, K. T. (2016). Twitter taunts and tirades: Negative campaigning in the age of Trump. *PS: Political*

Science & Politics, *49*(4), 748−754. http://doi.org/10.1017/S1049096516001700

Harrington, J., & Hess, G. (1996). A spatial theory of positive and negative campaigning. *Games and Economic Behavior*, *17*(2), 209−229. https://doi.org/10.1006/game.1996.0103

Healy, A., & Malhotra, N. (2013). Retrospective voting reconsidered. *Annual Review of Political Science*, *16*(1), 285−306. https://doi.org/10.1146/annurev-polisci-032211-212920

Honnibal, M., & and Johnson, M. (2015, September 17-21). An improved non-monotonic transition system for dependency parsing. In *Proceedings of the 2015 conference on empirical methods in natural language processing*, Lisbon, Portugal (pp. 1373−1378). Association for Computational Linguistics.

Hopmann, D. N., de Vreese, C. H., & Albæk, E. (2011). Incumbency bonus in election news coverage explained: The logics of political power and the media market. *Journal of Communication*, *61*(2), 264−282. https://doi.org/10.1111/j.1460-2466.2011.01540.x

Huang, J. T., Li, J., Yu, D., Deng, L., & Gong, Y. (2013). Cross-language knowledge transfer using multilingual deep neural network with shared hidden layers. In IEEE international conference on acoustics, speech and signal processing, Vancouver, BC, Canada, 26-31 May 2013. (pp. 7304−7308).

Huddy, L., & Terkildsen, N. (1993). Gender stereotypes and the perception of male and female candidates. *American Journal of Political Science*, *37*(1), 119−147. https://doi.org/10.2307/2111526

Iyengar, S., Sood, G., & Lelkes, Y. (2012). Affect, not ideology. A social identity perspective on polarization. *Public Opinion Quarterly*, *76*(3), 405−431. https://doi.org/10.1093/poq/nfs038

Iyengar, S., & Westwood, S. J. (2015). Fear and loathing across party lines: New evidence on group polarization. *American Journal of Political Science*, *59*(3), 690−707. https://doi.org/10.1111/ajps.12152

Johnson-Cartee, K.S., & Copeland, G. (1989). Southern voters' reaction to negative political ads in 1986 election. *Journalism Quarterly*, *66*(4), 888−893. https://doi.org/10.1177/107769908906600417

Kahn, K. F. (1996). *The political consequences of being a woman*. Columbia University Press.

Kahn, K. F., & Kenney, P. J. (1999). Do negative campaigns mobilize or suppress turnout? Clarifying the relationship between negativity and participation. *American Political Science Review*, *93*(04), 877−889. http://doi.org/10.2307/2586118

Krupnikov, Y., & Bauer, N. M. (2014). The relationship between campaign negativity, gender and campaign context. *Political Behavior*, *36*(1), 167−188. https://doi.org/10.1007/s11109-013-9221-9

Lau, R. R., & Pomper, G. M. (2001). Negative campaigning by US senate candidates. *Party Politics*, *7*(1), 69−87. https://doi.org/10.1177/1354068801007001004

Lau, R. R., & Pomper, G. M. (2004). *Negative campaigning: An analysis of U.S. Senate elections*. Rowman and Littlefield.

Levy, O., & Goldberg, Y. (2014). Neural word embedding as implicit matrix factorization. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (pp. 2177−2185). Curran Associates, Inc.

Lundberg, S., & Lee, S. I. (2017). A unified approach to interpreting model predictions. arXiv preprint arXiv:1705.07874.

Maier, J. (2015). Do female candidates feel compelled to meet sex-role expectations or are they as tough as men? A content analysis on the gender-specific use of attacks in German televised debates. In A. Nai & A. S. Walter (Eds), *New perspectives on negative campaigning: why attack politics matters* (pp. 129−146). ECPR Press.

Maier, J., & Nai, A. (2021). When conflict fuels negativity. A comparative analysis of the tone of electoral campaigns worldwide using expert ratings. *The Leadership Quarterly*. https://doi.org/10.1016/j.leaqua.2021.101564

Martin, P. S. (2004). Inside the black box of negative campaign effects: Three reasons why negative campaigns mobilize. *Political Psychology*, *25*(4), 545−562. https://doi.org/10.1111/j.1467-9221.2004.00386.x

Mattes, K., & Redlawsk, D. P. (2015). *The positive case for negative campaigning*. University of Chicago Press.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. arXiv preprint arXiv:1310.4546.

Nai, A. (2020). Going negative, worldwide. Towards a general understanding of determinants and targets of negative campaigning. *Government & Opposition*, *55*(3), 430−455. http://doi.org/10.1017/gov.2018.32

Nai, A., & Maier, J. (2020). Dark necessities: Candidates' aversive personality traits and negative campaigning in the 2018 American Midterms. *Electoral Studies*, *68*(2), 102233. https://doi.org/10.1016/j.electstud.2020.102233

Nai, A., & Martinez i Coma, F. (2019a). The personality of populists: Provocateurs, charismatic leaders, or drunken dinner guests? *West European Politics*, *42*(7), 1337−1367. https://doi.org/10.1080/01402382.2019.1599570

Nai, A., & Martinez i Coma, F. (2019b). Losing in the polls, time pressure, and the decision to go negative in referendum campaigns. *Politics & Governance*, *7*(2), 278−296. https://doi.org/10.17645/pag.v7i2.1940

Nai, A., & Sciarini, P. (2018). Why 'going negative'? Strategic and situational determinants of personal attacks in Swiss direct democratic votes. *Journal of Political Marketing*, *17*(4), 382−417. https://doi.org/10.1080/15377857.2015.1058310

Nwanevu, O. (2018, November 6). Trumpism lives on in Corey Stewart's campaign for the Virginia senate. The New Yorker. https://www.newyorker.com/news/dispatch/trumpism-lives-on-in-corey-stewarts-campaign-for-the-virginia-senate

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine

learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825−2830. https://dl.acm.org/doi/10.5555/1953048.2078195

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August 13-17). "Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, San Francisco, CA, United States (pp. 1135−1144). Association for Computing Machinery.

Ridout, T. N., & Holland, J. L. (2010). Candidate Strategies in the Presidential Nomination Campaign. *Presidential Studies Quarterly*, *40*(4), 611−630. http://www.jstor.org/stable/23044843

Roese, N. J., & Sande, G. N. (1993). Backlash effects in attack politics. *Journal of Applied Social Psychology*, *23*(8), 632−653. https://doi.org/10.1111/j.1559-1816.1993.tb01106.x

Rosenblatt, F. (1961). *Principles of neurodynamics: Perceptrons and the theory of brain mechanisms*. Spartan Books.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296−320. https://doi.org/10.1207/S15327957PSPR0504_2

Shapiro, M. A., & Rieger, R. H. (1992). Comparing positive and negative political advertising on radio. *Journalism Quarterly*, *69*(1), 135−145. https://doi.org/10.1177/107769909206900111

Skaperdas, S., & Grofman, B. (1995). Modeling negative campaigning. *American Political Science Review*, *89*(1), 49−61. https://doi.org/10.2307/2083074

Stevens, D. (2009). Elements of negativity: Volume and proportion in exposure to negative advertising. *Political Behavior*, *31*(3), 429−454. https://doi.org/10.1007/s11109-008-9082-9

Straus, J. R., Glassman, M. E., Shogan, C. J., & Smelcer, S. N. (2013). Communicating in 140 characters or less: Congressional adoption of twitter in the 111th Congress. *PS Political Science & Politics*, *46*(1), 60−66. http://doi.org/10.1017/S1049096512001242

Theilmann, J., & Wilhite, A. (1998). Campaign tactics and the decision to attack. *The Journal of Politics*, *60*(4), 1050−1062. https://doi.org/10.2307/2647730

Thorson, E., Ognianova, E., Coyle, J., & Denton, F. (2000). Negative political ads and negative citizen orientations toward politics. *Journal of Current Issues & Research in Advertising*, *22*(1), 13−40. https://doi.org/10.1080/10641734.2000.10505099

Trent, J. S., & Friedenberg, R. V. (2008). *Political campaign communication: Principles and practices*. Rowman & Littlefield.

Walter, A. S., van der Brug, W., & van Praag, P. (2014). When the stakes are high: Party competition and negative campaigning. *Comparative Political Studies*, *47*(4), 550−573. https://doi.org/10.1177/0010414013488543

Walter, A. S., & Vliegenthart, R. (2010). Negative campaigning across different communication channels: Different ballgames? *The Harvard International Journal of Press/Politics*, *15*(4), 441−461. https://doi.org/10.1177/1940161210374122

Yoon, K., Pinkleton, B. E., & Ko, W. (2005). Effects of negative political advertising on voting intention: An exploration of the roles of involvement and source credibility in the development of voter cynicism. *Journal of Marketing Communications*, *11*(2), 95−112. https://doi.org/10.1080/1352726042000315423

Author Biographies

Vladislav Petkevic is a Junior Lecturer at the department of Communication Science at the University of Amsterdam. His research focuses on investigating negativity in electoral campaigns using computational methods. Specifically, his research employs supervised machine learning (SML) and natural language processing (NLP) techniques.

Alessandro Nai is Assistant Professor of Political Communication at the University of Amsterdam. His research focuses on the drivers and consequences of election campaigning, political communication, and the psychology of voting behavior. His recent work deals more specifically with the dark sides of politics, the use of negativity and incivility in election campaigns in a comparative perspective, and the (dark) personality traits of political leaders.