Online Supporting Information

Mapping Digital Campaign Strategies: How Political Candidates Use Social Media to Communicate Constituency Connection and Policy Stance

James P Cross[1] Derek Greene[1] Stefan Müller[1] Martijn Schoonvelde[2]

[1] *University College Dublin*; [2] *University of Groningen*

Appendix 1: Training set validation and inter-coder reliability

Coding instructions

Each research assistant was asked to code a set of tweets based on two categories of interest - electioneering and policy content. Electioneering includes evidence of canvassing, event announcements, election material, and other campaign-related activities. Policy content includes screenshots of news articles, policy positions, policy outputs, and other policy-related content. Many tweets contained both images and text. To account for this, each tweet was first classified using only image content where it existed, and then using images and text together. This allowed us to account for the multi-modal nature of Twitter communications. The coding task was organised as follows:

1. For tweets without images:
    (a) Electioneering (text) - 0/1 dummy
    (b) Policy content (text) - 0/1 dummy
2. For tweets with images:
    (a) Electioneering (image only) - 0/1 dummy
    (b) Electioneering (image + text) - 0/1 dummy
    (c) Policy content (image only) - 0/1 dummy
    (d) Policy content (image + text) - 0/1 dummy

1 Inter-coder reliability and validation

Our first set of results pertain to validation of the training sets. We employ Cohen's κ Cohen (1960) to measure inter-coder agreement, where κ is calculated as follows:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{1}$$

In the above equation, p_o is the relative agreement observed between our coders, and p_e is the probability of agreement by chance, using the observed data to calculate the probabilities of each observer observing each category randomly. A κ of 1 represents complete agreement between the coders, while a κ of 0 means no agreement between the coders beyond what would be observed by chance. Generally a κ of 0.7 is seen as a reasonable threshold above which inter-coder reliability can be considered acceptable.

Table A1 provides an overview of inter-coder reliability scores for each of the four possible categories we are interested in identifying. As can be seen, agreement is significantly higher than expected for all categories, ranging from 83%–95% depending on the category in question.

For tweets that contain text only, κ is greater than our required 0.7 threshold for the electioneering coding and just below this threshold for the policy content coding. This is deemed as acceptable given the complexity associated with coding policy in tweets. For the tweets that contain both images and text, κ is greater than our required 0.7 threshold for both indicators of electioneering and for the indicator of policy content based on image and text considered together. As a result, we employ the categories based on image plus text coding in our analyses for tweets that contain both.

Category	N	Agreement	Expected agreement	Kappa	S.E.	k-alpha
Electioneering (text only)	800	95.62%	68.63%	0.8603	0.053	0.861
Electioneering (image only)	750	92.90%	66.66%	0.7872	0.0364	0.787
Electioneering (image + text)	750	94.38%	54.72%	0.8758	0.0365	0.876
Policy content (text only)	800	83.48%	50.88%	0.6637	0.0351	0.663
Policy content (image only)	750	89.02%	64.97%	0.6867	0.0365	0.687
Policy content (image + text)	750	86.48%	51.51%	0.7211	0.0366	0.719

Table A1: Inter-coder reliability scores (Cohen's kappa and Krippendorf's alpha reported)

Once the individual coding task was complete and in order to improve the quality of the final training set, the coders were asked to undertake a second round of coding where they identified cases of disagreement from the first round and reached a consensus coding for the final training set.

Appendix 2: Classifier setup and performance

Experimental setup. To identify a suitable representation and classifier combination to use for the automatic classification of our full dataset of unlabelled tweets, we conduct experiments on our training set of 1550 tweets. Specifically we consider a range of three different experimental configurations for choice of representation:

- *Bag-of-words*: Preprocessing steps involve (i) Twitter-specific tokenization with NLTK (Loper and Bird 2002), (ii) English stopword filtering using a custom list with 422 entries, (iii) stripping of @-mentions and URLs, (iv) term filtering based on term length range $\in [2, 20]$, (v) minimum document frequency 2; (vi) log-based TF-IDF term weighting, (vii) L2 document length normalisation.

- *Document embeddings*: Embeddings are generated using DistilBERT (Sanh et al. 2019) as implemented in the *Transformers* library (Wolf et al. 2020) (model 'distilbert-base-uncased'). Default parameter values were used in all cases.

- *Sentence models*: For these representations, we make use of the Sentence-BERT sentence transformers proposed by Reimers and Gurevych (2020) and implemented in Python[1]. Here we consider three pre-trained models: (i) 'distilroberta-v1'[2], (ii) 'ms-marco-MiniLM-L-6-v22'[3], (iii) 'all-mpnet-base-v2'[4]. Again default parameter values were used.

In conjunction with these representations, we use three standard classifiers as implemented by *scikit-learn* (Pedregosa et al. 2011). To identify appropriate parameters for each of the classification algorithms for a given representation, we apply a grid search on the training data using the following ranges:

- k-NN: Number of neighbours $k \in [1, 20]$

- Linear SVM with SGD training: regularisation term $\alpha \in [0.0001, 1]$; elastic net mixing parameter L1-ratio $\in [0.05, 1]$

- Logistic regression: inverse of regularisation strength term $C \in [0.001, 1000]$

To produce robust results, we use k-fold cross-validation with 5 folds.

[1]See `https://www.sbert.net`
[2]See `https://huggingface.co/sentence-transformers/all-distilroberta-v1`
[3]See `https://huggingface.co/cross-encoder/ms-marco-MiniLM-L-6-v2`
[4]See `https://huggingface.co/sentence-transformers/all-mpnet-base-v2`

Experiment	Model	Classifier	F1	Accuracy	BAR	Precision	Recall
Bag-of-words	TF-IDF	k-NN	0.1539	0.6529	0.5169	0.5329	0.2267
	TF-IDF	LR	0.6393	0.8594	0.7360	**1.0000**	0.4719
	TF-IDF	SVM	0.7972	0.8994	0.8505	0.8573	0.7459
Document embeddings	DistilBERT	k-NN	0.6024	0.8200	0.7233	0.7285	0.5161
	DistilBERT	LR	0.7624	0.8852	0.8255	0.8428	0.6977
	DistilBERT	SVM	0.7621	0.8690	0.8399	0.7690	0.7775
Sentence models	distilroberta-v1	k-NN	0.7952	0.8903	0.8629	0.7884	**0.8040**
	distilroberta-v1	LR	**0.8429**	**0.9226**	**0.8795**	0.9099	0.7871
	distilroberta-v1	SVM	0.8348	0.9219	0.8659	0.9509	0.7459
	MiniLM-6v2	k-NN	0.7457	0.8742	0.8172	0.8064	0.6950
	MiniLM-6v2	LR	0.7753	0.8903	0.8335	0.8538	0.7120
	MiniLM-6v2	SVM	0.7714	0.8948	0.8227	0.9156	0.6682
	mpnet-base-v2	k-NN	0.7717	0.8819	0.8410	0.7933	0.7532
	mpnet-base-v2	LR	0.8155	0.9097	0.8599	0.8916	0.7532
	mpnet-base-v2	SVM	0.8082	0.9077	0.8532	0.9012	0.7364

Table A2: Comparison of 5-fold cross-validation results for different representation and classifier combinations on the training data for the electioneering task.

Experiment	Model	Classifier	F1	Accuracy	BAR	Precision	Recall
Bag-of-words	TF-IDF	k-NN	0.6125	0.4703	0.5305	0.4476	**0.9701**
	TF-IDF	LR	0.6462	0.7452	0.7218	0.8039	0.5502
	TF-IDF	SVM	0.6549	0.7194	0.7077	0.6986	0.6218
Document embeddings	DistilBERT	k-NN	0.7366	0.7961	0.7823	0.8261	0.6803
	DistilBERT	LR	0.7761	0.8187	**0.8099**	0.8252	0.7445
	DistilBERT	SVM	0.7351	0.7935	0.7803	**0.8451**	0.6832
Sentence models	distilroberta-v1	k-NN	0.7696	0.8090	0.8015	0.8014	0.7460
	distilroberta-v1	LR	0.7706	0.8110	0.8031	0.8109	0.7445
	distilroberta-v1	SVM	0.7578	0.8032	0.7932	0.8074	0.7191
	MiniLM-6v2	k-NN	0.7484	0.7923	0.7839	0.7835	0.7221
	MiniLM-6v2	LR	0.7499	0.7974	0.7879	0.8023	0.7177
	MiniLM-6v2	SVM	0.7265	0.7923	0.7768	0.8321	0.6625
	mpnet-base-v2	k-NN	**0.7775**	0.8103	0.8063	0.7877	0.7759
	mpnet-base-v2	LR	0.7762	**0.8174**	0.8093	0.8180	0.7491
	mpnet-base-v2	SVM	0.7607	0.8090	0.7985	0.8209	0.7207

Table A3: Comparison of 5-fold cross-validation results for different representation and classifier combinations on the training data for the policy content task.

We apply the complete process above separately for the electioneering and policy content tasks. The results for the former are listed in Table A2, while the results for the latter are given in Table A3. We report scores for five standard evaluation measures: classification accuracy, precision, recall, F1-measure, and balanced accuracy (BAR). Note that precision and recall are calculated with respect to the positive class in each task.

Across both tasks, we see that the sentence embeddings yield the most accurate models, followed by the BERT-based document embeddings. In terms of specific combinations, we identified a sentence-based approach using a distilled RoBERTa model ('distilroberta-v1') combined with an SVM classifier as being the most effective combination when considering both tasks. We see that this combination not only yields high classification accuracy scores for the two tasks, we also see strong performance in relation to the balanced

accuracy (BAR) and F1 measures, which consider differences arising from imbalance between class sizes in the data.

Appendix 3: Classifier verification

To verify our choice of classification combination (distilroberta-v1 + SVM). We use the original full training set of 1550 tweets to train the model, and then use our additional 800 tweet hold-out set for testing. This is performed for both the electioneering and policy content tasks.

Table A4 shows the results of these experiments according to 5 performance measures. We see that the scores for the different measures are in line with our previous classification results reported in Appendix 2, which serves to validate our choice of representation and classifier combination.

Task	Accuracy	BAR	F1	Precision	Recall
Electioneering	0.9225	0.8820	0.8442	0.8984	0.7962
Policy content	0.8213	0.8135	0.7660	0.7500	0.7826

Table A4: Classification results for the sentence model using a pre-trained base model ('distilroberta-v1') and SVM classifier on the 800 tweet hold out set.

Appendix 4: Alternative explanations and robustness checks

Competitiveness (measured through betting odds) as a predictor of electoral success

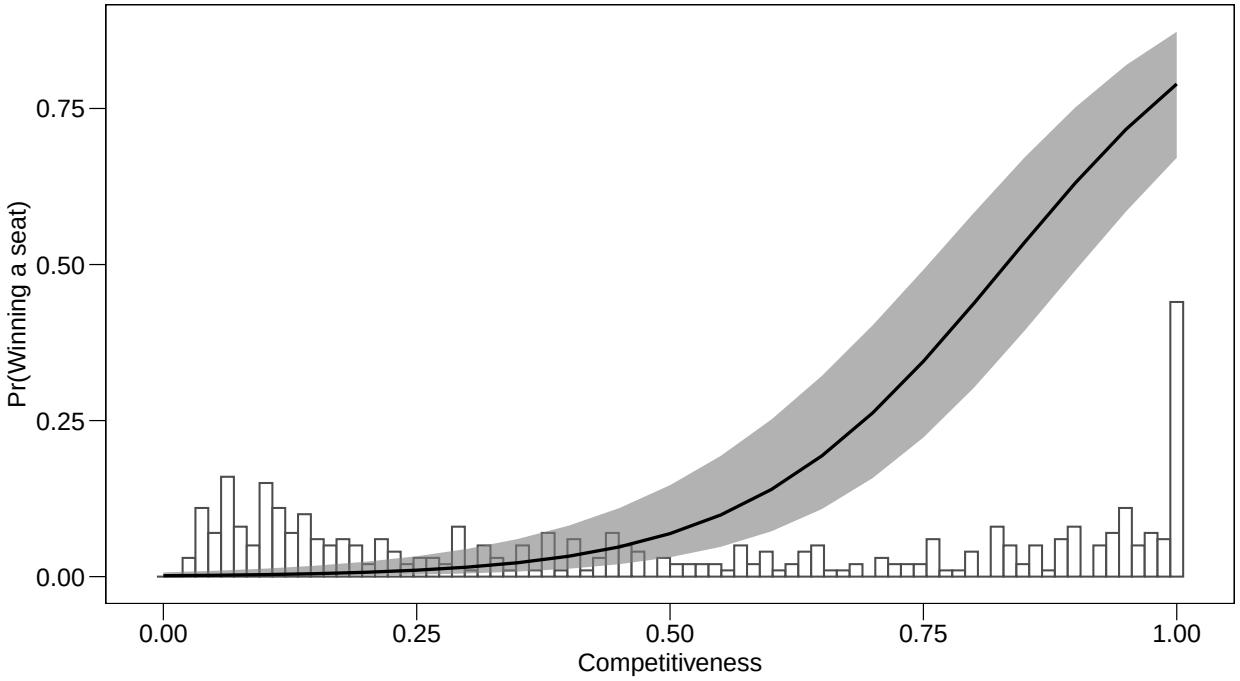

Figure A1: Predicted probabilities of winning a seat, conditional on a candidate's competitiveness derived from betting odds. Predicted probabilities based on Model 2 of Table A5. The control variable *Party* is held constant at 'Fianna Fáil', the party with the most candidates running for a seat in 2020. The histogram on the x-axis shows the distribution of the competitiveness scores.

Alternative model specifications

In our pre-analysis plan we specify a baseline alternative model specification which is not presented in the main text for the sake of brevity. This model just includes the variables of interest and a set of controls, excluding interaction terms. The vector of covariates associated with the model presented below are specified as follows:

$$\boldsymbol{\beta_i^s X} = \beta_1\, Competitiveness + \beta_2\, Experience + \beta_3\, Gender + \beta\, Controls + \epsilon \tag{2}$$

Table A6 presented the results of this analysis. As can be seen the results are somewhat different to what

Table A5: Predicting whether a candidate obtains a seat, conditional on competitiveness and a candidate's party affiliation. The models include all candidates with an active Twitter account.

	Model 1	Model 2
(Intercept)	-3.16^{***}	-4.00^{*}
	(0.31)	(1.85)
Competitiveness	5.05^{***}	7.85^{***}
	(0.47)	(0.84)
Party: Fianna Fáil (ref.: Aontú)		−2.53
		(1.90)
Party: Fine Gael		−1.47
		(1.88)
Party: Green Party		−0.33
		(1.90)
Party: Labour		−1.23
		(1.99)
Party: Other/Independent Candidates		−0.56
		(1.88)
Party: Sinn Féin		3.42
		(1.93)
Party: Social Democrats		0.75
		(1.95)
Party: Solidarity-PBP		1.01
		(1.96)
AIC	326.36	267.31
BIC	334.21	306.58
Log Likelihood	−161.18	−123.65
Deviance	322.36	247.31
Num. obs.	375	375

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

is presented in the main body of the text, but this is a result of not interacting our variables of interest with the campaign period variable.

	(1)					
	Both		Electioneering		Policy	
Competitiveness	0.720	(-1.28)	1.022	(0.13)	1.124	(0.55)
Previous campaign experience	1.133**	(2.95)	1.058	(1.93)	1.178***	(4.99)
Non-Campaign Period	0.169***	(-15.56)	0.238***	(-18.39)	1.079	(0.87)
Male	1.188	(1.24)	1.153	(1.61)	0.998	(-0.01)
Time until election	0.994***	(-8.98)	0.992***	(-16.56)	1.001	(1.61)
Fianna Fáil	0.448**	(-2.82)	1.086	(0.38)	0.509***	(-3.96)
Fine Gael	0.262***	(-4.43)	0.968	(-0.14)	0.428***	(-5.81)
Green Party	0.908	(-0.32)	1.003	(0.02)	0.870	(-0.79)
Labour	0.537	(-1.82)	0.910	(-0.38)	0.446***	(-3.80)
Other/Ind	0.599	(-1.48)	0.902	(-0.44)	0.638	(-1.92)
Sinn Féin	0.600	(-1.62)	1.178	(0.72)	0.495**	(-3.00)
Social Democrats	0.694	(-1.07)	0.987	(-0.05)	0.749	(-1.57)
Solidarity-PBP	1.960*	(2.10)	1.901**	(2.62)	1.363	(1.58)
District magnitude	0.726**	(-2.72)	0.940	(-0.75)	0.885	(-1.31)
Candidates running	1.044	(1.35)	1.012	(0.61)	1.008	(0.27)
Urban	1.024	(0.18)	1.025	(0.29)	1.123	(1.02)
Twitter length	1.000	(-1.30)	1.000**	(-3.18)	1.000	(-1.64)
Followers count	1.000	(1.11)	1.000	(0.04)	1.000*	(2.02)
Rolling average of no. of tweets (7 days)	0.963***	(-3.50)	0.956***	(-6.19)	0.979**	(-2.66)
Sum of likes on previous 10 tweets	1.000*	(-2.09)	1.000	(-1.56)	1.000**	(-2.62)
Sum of retweets on previous 10 tweets	1.001*	(2.06)	1.000	(1.03)	1.001*	(2.54)
Rolling average of retweets (7 days)	1.006	(1.79)	1.002	(0.94)	1.004	(1.42)
Observations	101864					

z statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A6: Multinomial logit regression analysis

Table A7 replicates the regression found in Table 2 but without clustering standard errors. As can be seen, many of the variables of interest appear significant in this model, in contrast to Model 1 presented in the paper. To further probe these differences, we conducted an LR test to assess the added value of including interaction effects. The results suggest that Model 1 with the interaction effects provides a better fit to the data, with the interaction terms explaining significant additional variance in the data. Model 1 reveals conditional relationships that Model A1 assumes are constant.

	(1)					
	Both		Electioneering		Policy	
Competitiveness	0.940	(-0.66)	0.848**	(-2.63)	1.145***	(4.64)
Previous campaign experience	1.079***	(4.13)	1.047***	(3.81)	1.182***	(31.91)
Campaign Period	5.780***	(23.06)	3.460***	(22.39)	0.811***	(-5.06)
Campaign Period × Competitiveness	0.660***	(-3.62)	1.354***	(3.68)	0.907	(-1.43)
Campaign Period × Previous campaign experience	1.076**	(3.26)	1.016	(0.93)	0.979	(-1.60)
Male	1.127*	(2.00)	1.172***	(3.99)	0.945**	(-3.04)
Male × Campaign Period	1.171*	(2.16)	1.056	(1.03)	1.394***	(7.71)
Time until election	0.994***	(-12.99)	0.993***	(-24.13)	1.001***	(5.22)
Fianna Fáil	0.446***	(-6.58)	1.078	(0.66)	0.507***	(-11.34)
Fine Gael	0.261***	(-10.87)	0.965	(-0.31)	0.427***	(-14.40)
Green Party	0.900	(-0.88)	0.999	(-0.01)	0.867*	(-2.38)
Labour	0.538***	(-4.91)	0.911	(-0.79)	0.442***	(-12.93)
Other/Ind	0.594***	(-4.27)	0.892	(-0.99)	0.632***	(-7.52)
Sinn Féin	0.600***	(-4.12)	1.177	(1.40)	0.494***	(-11.39)
Social Democrats	0.694**	(-2.94)	0.993	(-0.06)	0.755***	(-4.39)
Solidarity-PBP	1.942***	(5.20)	1.894***	(5.20)	1.361***	(4.59)
District magnitude	0.727***	(-10.46)	0.941**	(-2.77)	0.884***	(-8.65)
Candidates running	1.044***	(5.37)	1.013*	(2.10)	1.008*	(2.00)
Urban	1.021	(0.61)	1.019	(0.75)	1.120***	(7.38)
Twitter length	1.000***	(-5.21)	1.000***	(-10.85)	1.000***	(-10.80)
Followers count	1.000*	(2.54)	1.000	(0.18)	1.000***	(6.59)
Rolling average of no. of tweets (7 days)	0.964***	(-13.43)	0.958***	(-17.77)	0.980***	(-12.78)
Rolling average of retweets (7 days)	1.006***	(3.91)	1.002	(1.18)	1.004***	(4.71)
Sum of likes on previous 10 tweets	1.000***	(-5.09)	1.000**	(-2.96)	1.000***	(-9.13)
Sum of retweets on previous 10 tweets	1.001***	(4.77)	1.000	(1.76)	1.001***	(9.13)
Observations	101864					

z statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A7: Multinomial logistic regression analysis without clustered standard errors

2 Differences in campaign strategy across parties

Figure A2 compares the averages of electioneering and policy tweets across parties. Each dot marks one candidate. Policy content is more prevalent across all parties than electioneering content. Note that the year of birth is only available for politicians who were elected to Dáil Éireann.

3 Alternative specifications of competitiveness

The measure of competitiveness specified in the paper is based on the betting odds that a given candidate will win a seat. As a robustness test, we also measure competitiveness as the first-preference vote share at the previous election. First preference votes are regarded as a suitable measure of candidate support in Irish

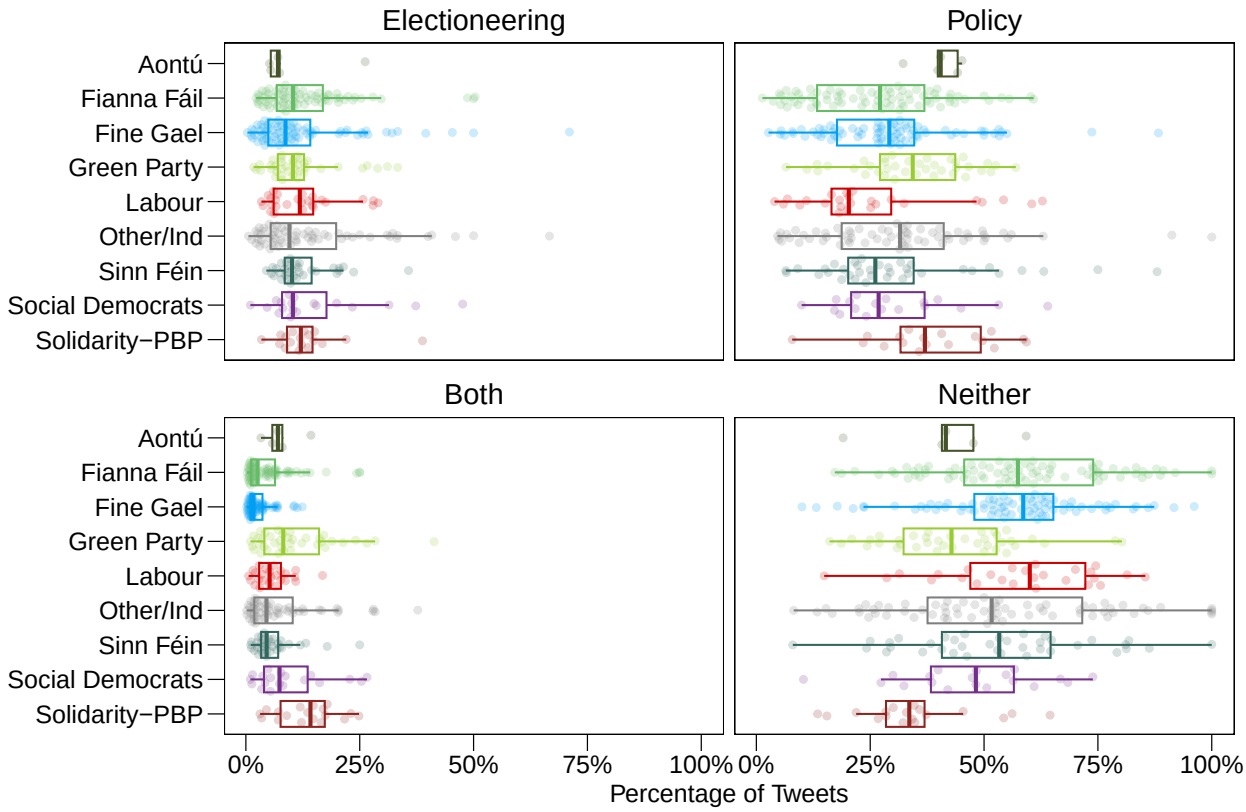

Figure A2: The prevalence of electioneering and policy tweets across parties.

elections conducted under PR-STV (Benoit and Marsh 2008). We rerun our models with this alternative and present the results in Table A8. Our results are robust to this alternative specification.

4 Age as a confounder of experience

A candidate's age is a potential confounder for experience. Older candidates may differ from younger candidates in their Twitter activity for many reasons (e.g. cohort effects, education, familiarity with technology). We collected the year of birth of all candidates who ever obtained a seat in the parliament and matched their dates of births with our Twitter dataset. Figure A3 shows that older do not differ substantively from younger candidates. The correlations between the year of birth and proportions of electioneering ($r = -0.06$) and policy tweets ($r = -0.1$) are very low.

Besides significantly reducing the sample size due to missing data on the age variable, the inclusion of age in our regression analysis also makes little differences to the substantive results as demonstrated in Table A9.

	(1)					
	Both		Electioneering		Policy	
Previous campaign experience	1.067	(1.13)	1.019	(0.44)	1.190***	(5.31)
First preference share	1.002	(0.16)	1.013	(1.17)	1.002	(0.16)
Campaign Period	6.082***	(9.86)	3.736***	(9.78)	0.861	(-1.09)
Campaign Period × First preference share	0.970*	(-2.34)	0.996	(-0.36)	0.988	(-1.44)
Campaign Period × Previous campaign experience	1.092	(1.50)	1.049	(1.12)	0.990	(-0.30)
Male	1.115	(0.67)	1.098	(0.91)	0.972	(-0.23)
Male × Campaign Period	1.130	(0.63)	1.111	(0.71)	1.359*	(2.25)
Time until election	0.994***	(-9.49)	0.992***	(-16.43)	1.001	(1.59)
Fianna Fáil	0.402**	(-3.10)	1.062	(0.31)	0.513***	(-4.25)
Fine Gael	0.247***	(-4.47)	0.939	(-0.32)	0.421***	(-5.83)
Green Party	0.849	(-0.51)	0.979	(-0.10)	0.849	(-0.90)
Labour	0.497	(-1.94)	0.920	(-0.35)	0.426***	(-4.01)
Other/Ind	0.581	(-1.54)	0.935	(-0.30)	0.638*	(-1.99)
Sinn Féin	0.641	(-1.14)	0.954	(-0.19)	0.491*	(-2.41)
Social Democrats	0.639	(-1.24)	0.972	(-0.12)	0.724	(-1.72)
Solidarity-PBP	1.357	(0.84)	1.560	(1.88)	1.112	(0.52)
District magnitude	0.763*	(-2.23)	0.985	(-0.18)	0.898	(-1.22)
Candidates running	1.036	(1.09)	1.012	(0.63)	1.011	(0.40)
Urban	1.092	(0.65)	1.033	(0.39)	1.124	(1.04)
Twitter length	1.000	(-1.38)	1.000**	(-3.15)	1.000	(-1.67)
Followers count	1.000	(0.84)	1.000	(-0.27)	1.000*	(2.06)
Rolling average of no. of tweets (7 days)	0.969**	(-2.88)	0.961***	(-5.19)	0.982*	(-2.21)
Rolling average of retweets (7 days)	1.006	(1.94)	1.002	(0.76)	1.004	(1.43)
Sum of likes on previous 10 tweets	1.000	(-1.88)	1.000	(-1.57)	1.000*	(-2.52)
Sum of retweets on previous 10 tweets	1.001	(1.87)	1.000	(1.13)	1.001*	(2.44)
Observations	104420					

z statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A8: Multinomial logit regression analysis with clustered standard errors

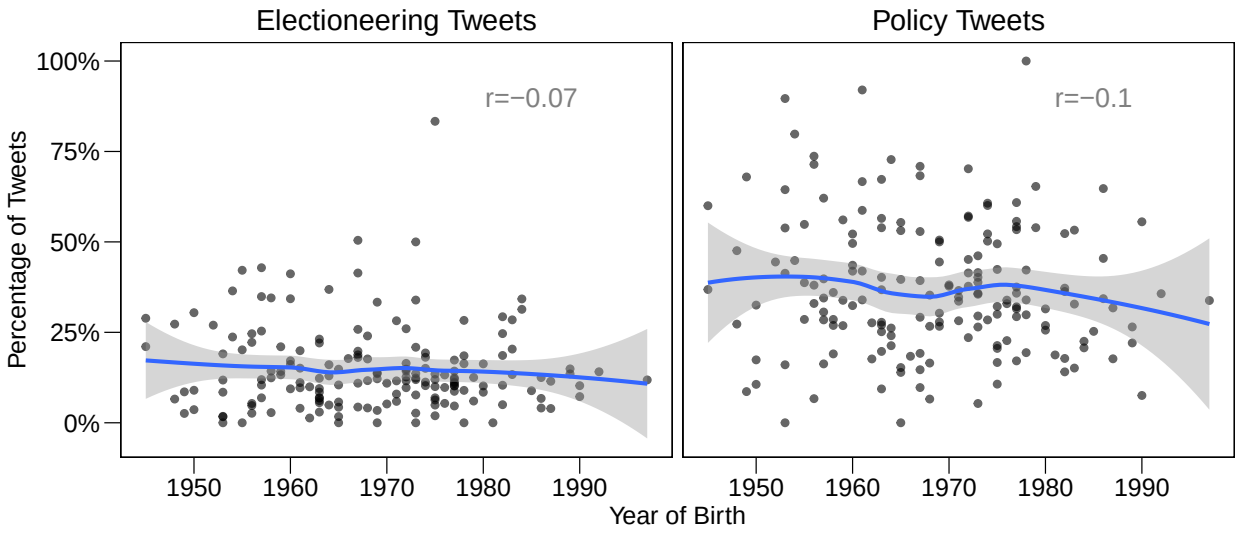

Figure A3: The relationship between age and Twitter activity.

	Both		Electioneering		Policy	
Competitiveness	1.308	(0.59)	0.690	(-1.44)	1.800	(1.65)
Previous campaign experience	1.138	(1.56)	1.053	(0.96)	1.195***	(3.42)
Campaign Period	3.771***	(3.82)	3.100***	(4.61)	0.566*	(-2.29)
Campaign Period × Competitiveness	1.484	(0.95)	1.892*	(2.14)	1.662	(1.65)
Campaign Period × Previous campaign experience	1.060	(0.90)	1.002	(0.05)	0.995	(-0.11)
Male	0.961	(-0.15)	1.081	(0.47)	0.867	(-0.75)
Male × Campaign Period	1.224	(0.65)	0.977	(-0.10)	1.169	(0.78)
Age	1.006	(0.52)	1.006	(0.75)	0.997	(-0.37)
Time until election	0.996***	(-5.31)	0.992***	(-13.31)	1.001*	(2.21)
Fianna Fáil	0.228***	(-6.25)	0.671*	(-2.23)	0.391***	(-5.06)
Fine Gael	0.151***	(-7.67)	0.636*	(-2.40)	0.402***	(-4.95)
Green Party	0.566	(-1.93)	0.925	(-0.48)	0.615	(-1.93)
Labour	0.191***	(-4.31)	0.467**	(-2.60)	0.365***	(-3.37)
Other/Ind	0.127***	(-3.40)	0.584	(-1.68)	0.305***	(-4.50)
Sinn Féin	0.296***	(-3.62)	0.629*	(-2.10)	0.423**	(-2.66)
Social Democrats	0.283*	(-2.46)	0.429***	(-3.46)	0.515*	(-2.17)
Solidarity-PBP	1.228	(0.51)	0.970	(-0.11)	1.455	(1.34)
District magnitude	0.799	(-1.38)	1.057	(0.51)	1.016	(0.16)
Candidates running	1.030	(0.67)	0.996	(-0.16)	0.985	(-0.43)
Urban	1.260	(1.23)	1.278*	(2.07)	1.245	(1.46)
Twitter length	1.000	(-1.63)	1.000***	(-3.52)	1.000	(-1.21)
Followers count	1.000	(1.83)	1.000	(0.92)	1.000	(1.38)
Rolling average of no. of tweets (7 days)	0.990	(-0.64)	0.954***	(-3.57)	0.991	(-0.68)
Rolling average of retweets (7 days)	1.006*	(2.02)	1.003	(1.94)	1.002	(1.42)
Sum of likes on previous 10 tweets	1.000*	(-2.11)	1.000	(-1.08)	1.000**	(-2.80)
Sum of retweets on previous 10 tweets	1.001	(1.52)	1.000	(0.49)	1.001**	(2.90)
Observations	50334					

z statistics in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A9: Multinomial logit regression analysis with clustered standard errors

References

Benoit, K. and Marsh, M. (2008). The campaign value of incumbency: A new solution to the puzzle of less effective incumbent spending. *American Journal of Political Science*, 52(4):874–890.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1):37–46.

Loper, E. and Bird, S. (2002). Nltk: The natural language toolkit. *arXiv preprint cs/0205028*.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Vuncent, D., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

Reimers, N. and Gurevych, I. (2020). Making monolingual sentence embeddings multilingual using knowledge distillation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*. ACL.

Sanh, V., Debut, L., Chaumond, J., and Wolf, T. (2019). Distilbert, a distilled version of bert: Smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. L., Gugger, S., Drame, M., Lhoest, Q., and Rush, A. M. (2020). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45. ACL.