Replication material for: Modulation of Democracy: Partisan Communication during and after Election Campaigns

—

forthcoming in *British Journal of Political Science*

Bruno Castanho Silva, Lennart Schürmann & Sven-Oliver Proksch

This README provides an overview of the replication materials for the article. The Data section describes the datasets required to reproduce the figures and tables in the paper and in the Online Appendix, and the auxiliary datasets containing the status IDs and the hand-coded tweets for validation. The Analysis section summarises the purpose of the R script to reproduce the analyses in the paper and the Online Appendix.

Note: In compliance with Twitter's Terms of Service (https://developer.twitter.com/en/developer-terms/agreement-and-policy, accessed on February 7, 2023), we cannot publicly share the raw files containing the full text of individual tweets and their associated metadata.We share a list of tweet IDs, which can be used to retrieve the original tweets through the Twitter API. As of February 9, 2023, Twitter has disabled the free access to the Academic API for researchers, but currently claims they are working on a new solution to allow academics access to Twitter data once again. It must be noted that regardless of the API used, it will not give access to tweets that have been deleted since the date when we collected in real time, reason why the exact numbers of tweets downloaded in this way will not match those reported in the paper.

A Main dataset

The main dataset for reproducing tables and figures in the paper is:
sent.campaign_ noncampaign.RData.
It contains 157305 observations, with the following variables:

- `period_seq`: Running time variable (two-week periods)
- `user_id`: MP ID
- `group`: Party name
- `country`: Country
- `sentiment_scaled_by_country`: Sentiment score of the respective period scaled by country (Main dependent variable, is calculated in the script)
- `national_campaign`: Binary variable indicating national campaign period
- `EU_campaign`: Binary variable indicating EU campaign period
- `prime_minister`: Binary variable indicating whether MP belongs to prime minster party. Source: Parlgov
- `junior_partner`: Binary variable indicating whether MP belongs to junior partner party. Source: Parlgov
- `opposition`: Binary variable indicating whether MP belongs to opposition party. Source: Parlgov
- `lrgen`: Left-right placement of party. Source: CHES
- `lrgen_squared`: Left-right placement of party (squared, is calculated in the script). Source: CHES
- `eu_position`: EU position of party. Source: CHES
- `people_vs_elite`: Populism score of party. Source: CHES
- `polling_trend_positive`: Positive polling trend in comparison to last election result. Source: Europe Elects

Additional variables used for analyses in the online appendix:

- `sentiment`: Absolute sentiment score of the respective period (for Online Appendix B)
- `caretaker_Parlgov`: Indicates whether period was a period with a caretaker government; Source: Parlgov (for Online Appendix C)
- `manichean_POPPA`: Some parties see politics as a moral struggle between good and bad. This is often described as a Manichean worldview. Source: POPPA (for Online Appendix E)

- `generalwill_POPPA:` PSome parties consider the ordinary people's interests to be singular (i.e. one can speak of a 'general will'). Source: POPPA (for Online Appendix E)
- `peoplecentrism_POPPA:` Some parties believe that sovereignty should lie exclusively with the ordinary people (i.e. the ordinary people, not the elites, should have the final say in politics). Source: POPPA (for Online Appendix E)
- `antielitism_POPPA:` Some parties can be characterized by their anti-elitism. Source: POPPA (for Online Appendix E)
- `indivisble_POPPA:` Some parties consider the ordinary people to be indivisible (i.e. the people are seen as homogenous). Source: POPPA (for Online Appendix E)
- `populism_POPPA:` Populism general. Source: POPPA (for Online Appendix E)
- `pos:` Number of positive terms
- `neg:` Number of negative terms
- `party_status:` Status of the party (opposition, junior partner or prime minister party)
- `seq_national_election:` Sequence number (two-week periods) relative to the national election
- `seq_eu_election:` Sequence number (two-week periods) relative to the EU election
- `screenname:` Screenname of the MP's Twitter Account
- `period:` Two-week periods time variable consisting of **year_month_first/second half of the month** (e.g., 2019_12_1)
- `doc_id:` Unique identifier combining screenname and period (without _)

B Additional datasets

Additional datasets for validity tests in the online appendix:

1. `sentiment_all_week.RData:` Dataset with sentiment score on a weekly basis (for Online Appendix F).

With the following additional variables:

- `sentiment.w_scaled_by_country:` Sentiment scaled by country on a weekly basis (for Online Appendix F)
- `period_seq.w:` Running time variable (one-week periods) (for Online Appendix F)
- `sentiment.m_scaled_by_country:` Sentiment scaled by country on a monthly basis (for Online Appendix F)

- `period_seq.m`: Running time variable (one-month periods) (for Online Appendix F)

2. `tweets_validation_replication.csv`: Dataset with tweets where sentiment was hand-coded by two student assistants, to conduct the validation in Online Appendix I.

With the following variables:

- `id_str`: Status ID, can be used to retrieve the original tweet
- `language`: Language of the tweet, from the Twitter API
- `sentiment_coder1`: whether coder 1 assigned **positive**, **negative**, or **neutral** to this tweet
- `sentiment_coder2`: whether coder 2 assigned **positive**, **negative**, or **neutral** to this tweet
- `sentiment_coder1_numeric`: numeric transfomation of `sentiment_coder1` where positive = 1, neutral = 0, and negative = -1.
- `sentiment_coder2_numeric`: numeric transfomation of `sentiment_coder2` where positive = 1, neutral = 0, and negative = -1.
- `sentiment_distance_coders`: distance in the numeric sentiment between the two coders.
- `mean_sentiment_coder`: average numeric sentiment assigned by the coders.
- `campaign`: whether the tweet was posted during an electoral campaign (1) or not (0)
- `user_id_str`: User ID of the MP who posted that status.
- `country`: MP's country
- `period_seq`: Running time variable (two-week periods)
- `sentiment`: The two-week period sentiment for the MP who posted that week in that period.
- `sentiment_R`: The tweet sentiment calculated with the sentiment dictionary
- `sent_cat`: categorical version of `sentiment_R` where values lower than 0 = **negative**, equal to 0 = **neutral**, and higher than 0 = **positive**.
- `sent_hand_cat`: Categorical sentiment obtained from hand-coding, for cases where both coders agreed (**positive, neutral, negative**).

3. `elections_hashtags.xlsx`: Dataset with election dates to create Figure 1.

With the following variables:

- `Country`: Country Name
- `Legislative turn`: Number of legislative period in the country
- `Election date`: Date of the election
- `Call for election`: Binary variable whether there was a call for an election
- `Call for election date`: Date of the call for an election (if there was a call)
- `Extra information - call for election`: Extra information regarding the call for election (for instance if the election is always on a specific date)
- `dictionary`: Twitter-specific hashtags used to refer to the election
- `dictionary_comments`: Extra information regarding the Twitter-specific hashtags used to refer to the election

4. `twitter_IDs_campaign_sentiment.csv`: Dataset with the tweet IDs and sceennames of the politicians to reproduce Online Appendix H: Number of MPs and number of tweets

With the following variables:

- `X1`: Tweet ID, can be used to retrieve the original tweet
- `X2`: Screenname of the of the MP who posted that tweet

C Analysis

The Scripts folder contains one Rmd script to reproduce the figures and tables (Analysis_replication.Rmd). The codes will run as long as the working directory is set to the "Scripts" folder, and the folder structure is kept as it is uploaded here. The Analysis_replication.Rmd script reproduces all Tables and Figures from both the main manuscript and the Online Appendix.

All analyses were conducted with R 4.1.2 running on macOS Big Sur Version 11.6.6. Packages attached (in version):

irr_0.84.1, lpSolve_5.6.18, MLmetrics_1.1.1, utf8_1.2.2, quanteda_3.2.1, reshape2_1.4.4, readxl_1.3.1, nlme_3.1-153, ggpubr_0.4.0, sjPlot_2.8.10, texreg_1.38.5, lme4_1.1-28, Matrix_1.4-1, forcats_0.5.1, stringr_1.4.0, dplyr_1.0.8, purrr_0.3.4, readr_2.1.2, tidyr_1.2.0, tibble_3.1.6, ggplot2_3.4.0, tidyverse_1.3.1

Packages loaded via a namespace (and not attached):

fs_1.5.2, lubridate_1.8.0, insight_0.16.0, httr_1.4.2, tools_4.1.2, backports_1.4.1, R6_2.5.1, sjlabelled_1.1.8, DBI_1.1.2, colorspace_2.0-3, withr_2.5.0, tidyselect_1.1.2, emmeans_1.7.2, compiler_4.1.2, performance_0.8.0, cli_3.4.1, rvest_1.0.2, xml2_1.3.3, bayestestR_0.11.5, scales_1.2.1, mvtnorm_1.1-3, digest_0.6.29, minqa_1.2.4, pkgconfig_2.0.3, htmltools_0.5.2, dbplyr_2.1.1, fastmap_1.1.0, rlang_1.0.6, rstudioapi_0.13, generics_0.1.2, jsonlite_1.8.0, car_3.0-12, magrittr_2.0.2, parameters_0.16.0, Rcpp_1.0.8, munsell_0.5.0, fansi_1.0.2, abind_1.4-5, lifecycle_1.0.3, stringi_1.7.6, carData_3.0-5,

MASS_7.3-55, plyr_1.8.8, grid_4.1.2, sjmisc_2.8.9, crayon_1.5.0, lattice_0.20-45, ggeffects_1.1.1, haven_2.4.3, stargazer_5.2.3, splines_4.1.2, sjstats_0.18.1, hms_1.1.1, knitr_1.37, pillar_1.7.0, boot_1.3-28, estimability_1.3, ggsignif_0.6.3, effectsize_0.6.0.1, stopwords_2.3, fastmatch_1.1-3, reprex_2.0.1, glue_1.6.2, RcppParallel_5.1.5, modelr_0.1.8, standardize_0.2.2, vctrs_0.5.1, nloptr_2.0.0, tzdb_0.2.0, cellranger_1.1.0, gtable_0.3.0, assertthat_0.2.1, datawizard_0.3.0, xfun_0.30, xtable_1.8-4, broom_0.7.12, coda_0.19-4, rstatix_0.7.0, ellipsis_0.3.2,