Description of Dataset and Replication Materials

Replication materials for: Bruno Castanho Silva and Sven-Oliver Proksch (2021). "Fake It Til You Make It: A Natural Experiment to Identify European Politicians' Benefit from Twitter Bots". *American Political Science Review* 115(1): 316–322: https://doi.org/10.1017/S0003055420000817

Dataset

data_castanhosilva_proksch_apsr.rds: aggregated dataset where each row is one member of parliament.

Description of Variables

The following variables are included in *data_castanhosilva_proksch_apsr.rds.*
Note: not all of these variables are needed to solve the homework.

- *user_id_str*: Twitter user ID.
- *male*: whether the MP is male (1) or female (0). Obtained from everypolitician.org.
- *terms*: number of terms served by the MP in parliament. Obtained from MPs' Wikidata page.
- *min*: whether the MP ever held a cabinet position. Obtained from MPs' Wikidata page.
- *seat_share*: Seat share of the MP's party in parliament from the CHES dataset.
- *cabinet*: whether the MP's party was in government in July 2018, obtained from the ParlGov dataset.
- *pt_country*: combined party name to which the user belongs (party names following the Chapel Hill Expert Survey denominations) + country code (also from CHES).
- *country*: country where the MP comes from.
- *family*: party family according to the CHES.
- *eu_pos*: EU position of the MP's party according to the EU position question in the CHES dataset.
- *fol.b*: last available number of followers up to July 10, 2018.
- *log_fol.b*: natural logarithm of fol.b.
- *fr.b*: last available number of friends up to July 10, 2018.

- *day.b*: day from which fol.b and fr.b are counted. Date in the format of the daymonth2 variable in df.daily.csv.
- *fol.a*: first available follower count after July 13, 2018.
- *fr.a*: first available friends count after July 13, 2018.
- *day.a*: day from which fol.a and fr.a are counted. Date in the format of the daymonth2 variable in df.daily.csv.
- *folrat*: ratio of followers after the purge in relation to before. Calculated as (fol.a/fol.b)*100 - foldiff: difference in the number of followers after the purge in relation to before. Calculated as fol.a - fol.b.
- *frrat*: ratio of friends after the purge in relation to before. Calculated as (fr.a/fr.b)*100
- *frdiff*: difference in the number of friends after the purge in relation to before. Calculated as fr.a - fr.b.
- *sent.overall*: estimated sentiment of all tweets posted by the MP from February to the end of June 2018, using the Lexicoder Sentiment Dictionary.
- *sent.eu*: estimated sentiment of the EU-related tweets posted by the MP from February to the end of June 2018, using the Lexicoder Sentiment Dictionary.
- *radright*: dummy for whether the MP's party belongs to the radical right family according to the CHES.
- *radleft*: dummy for whether the MP's party belongs to the radical left family according to CHES.
- *ches_pid*: party ID from the CHES dataset.

Note by authors: In compliance with Twitter's Terms of Service (https://developer.twitter.com/en/developer-terms/agreement-and-policy, accessed on August 12, 2020), we cannot publicly share the raw files containing the full text of the tweets or their metadata. We therefore share datasets with aggregated information from them (daily follower counts and sentiment estimates), as well as a list of tweet ID's, which can be used to retrieve the original tweets through the Twitter API. One function that does that in R is explained in the following link: https://rdrr.io/github/ashoksiri/rtweet/man/lookup_statuses.html. Nevertheless, if one downloads the tweets from this list, the follower counts that appear are those of the time of downloading, not of when the tweet was posted, making it impossible to reproduce the original dataset as is and the time series of follower counts. Therefore, we provide the list of statuses from February to June, which were used to estimate MPs' sentiment, and which can be reproduced.