Readme

"Efficient Detection of Online Communities and Social Bot Activity During Electoral Campaigns" Harvard Dataverse

Ludovic Rheault[†] and Andreea Musulan[‡]

Description

The files contained in this repository can be used to reproduce the results reported in the main text and the supplementary materials (online appendix) of the following paper:

Rheault, Ludovic and Andreea Musulan. 2021. "Efficient Detection of Online Communities and Social Bot Activity During Electoral Campaigns." *Journal of Information Technology & Politics*, doi:10.1080/19331681.2021.1879705.

File Size

The uncompressed files and library installation procedure outlined below will require approximately 5gb of disk space.

Package Requirements

Results were computed using Python 3.8.5. The required libraries are detailed in the file requirements.txt (see procedure below).

[†]Assistant Professor, Department of Political Science and Munk School of Global Affairs and Public Policy, University of Toronto.

[‡]PhD Student, Department of Political Science, University of Toronto.

Reproducing Figures and Tables

We recommend using a virtual environment. Below is a summary of the steps to set up a virtual environment and run the main scripts. These commands are designed to be run from the operating system's command line interface.

Note: On some operating systems, the *python3* and *pip3* commands may be needed below if Python 3 is not the default language associated with the *python* command.

1. Download and extract the Dataverse repository files, then enter the directory:

```
unzip dataverse_files.zip
cd dataverse_files
```

2a. Install the Python virtualenv package (if needed):

```
pip install virtualenv
```

2b. Create and activate a virtual environment:

```
virtualenv bots
source bots/bin/activate
```

3. Install the requirements:

```
pip install -r requirements.txt
```

4. Reproduce the main paper's results (note, figures and tables will be overwritten in the figures/ and tables/ directory):

```
python table1.py
python table2.py
python figure1.py
python main_model.py
```

5. To reproduce the results from the online appendix:

```
python appendix.py
```

6. Deactivate the virtual environment:

```
deactivate
```

To fully remove the virtual environment files, simply delete the directory bots/ that was created during the procedure.

Notes on Virtual Environments

Using a Python virtual environment allows to install specific library requirements to reproduce results while avoiding conflicts with other versions of these libraries that may already be installed on a given machine. This step is not entirely necessary, but it is a good practice to preserve stability when working on multiple projects on the same computer.

License

These materials are released under a Creative Commons Zero ("CC0") license. Third-party software libraries are subject to their respective licenses, and Twitter data is subject to the company's terms of use.

Terms of Use

To comply with Twitter's terms of use, we do not bulk distribute the raw content of the tweets. Following common practice on the Dataverse, researchers may "rehydrate" the data using tweet IDs if they wanted to retrieve the original data in full. With the exception of public personalities, we also hashed the usernames and the unique user ID strings to avoid over-emphasizing the descriptive identifiers of individual users in our dataset. The methods emphasized in the source paper are designed to study aggregate trends, without the need to single out specific users. The authors do not condone usage of these materials for purposes that would violate Twitter's terms of use.