CHORUS

Replication Materials

This repository contains the code and data used in the paper "CHORUS: A new dataset of state interest group policy positions in the United States," forthcoming in *State Politics & Policy Quarterly*. If you use our code and/or data, please cite the paper as:

> Hall, Galen, Joshua Basseches, Rebecca Bromley-Trujillo, and Trevor Culhane. 2023. "CHORUS: A new dataset of state interest group policy positions in the United States." *State Politics & Policy Quarterly*. Forthcoming 2023.

Data

The dataset used in the related paper is available online in the SPPQ Dataverse: https://dataverse.unc.edu/dataverse/sppq. To download it locally, run the `code/download.py` file.

Note that we will maintain updated versions of this dataset at a different location. For replicating the results in the paper, please use the version of the dataset available in the SPPQ Dataverse.

File structure

- `data/CO_network_figure_clusters_named.csv` : The network figure data for `Colorado`, with the clusters named by the authors.
- `data/IL_network_figure_clusters_named.csv` : The network figure data for `Illinois`, with the clusters named by the authors.
- `data/MA_network_figure_clusters_named.csv` : The network figure data for `Massachusetts`, with the clusters named by the authors.
- `data/TX_network_figure_clusters_named.csv` : The network figure data for `Texas`, with the clusters named by the authors.
- `data/bills.parquet` : The bills on which positions were recorded, merged with data from LegiScan and the National Conference of State Legislatures for the states in CHORUS.
- `data/block_assignments.parquet` : The block assignments for each organization in each state, from our hierarchical bayesian stochastic block model.
- `data/clients.parquet` : The organizations that recorded positions on bills in CHORUS.
- `data/hbsbm/AZ_testimony_corrected_categorical_blockstate.pkl` : The blockmodel for the `testimony` records in `Arizona`.

- `data/hbsbm/CO_lobbying_corrected_categorical_blockstate.pkl` : The blockmodel for the `lobbying` records in `Colorado` .
- `data/hbsbm/CO_testimony_corrected_categorical_blockstate.pkl` : The blockmodel for the `testimony` records in `Colorado` .
- `data/hbsbm/FL_testimony_corrected_categorical_blockstate.pkl` : The blockmodel for the `testimony` records in `Florida` .
- `data/hbsbm/IA_lobbying_corrected_categorical_blockstate.pkl` : The blockmodel for the `lobbying` records in `Iowa` .
- `data/hbsbm/IL_testimony_corrected_categorical_blockstate.pkl` : The blockmodel for the `testimony` records in `Illinois` .
- `data/hbsbm/KS_testimony_corrected_categorical_blockstate.pkl` : The blockmodel for the `testimony` records in `Kansas` .
- `data/hbsbm/MA_lobbying_corrected_categorical_blockstate.pkl` : The blockmodel for the `lobbying` records in `Massachusetts` .
- `data/hbsbm/MD_testimony_corrected_categorical_blockstate.pkl` : The blockmodel for the `testimony` records in `Maryland` .
- `data/hbsbm/MO_testimony_corrected_categorical_blockstate.pkl` : The blockmodel for the `testimony` records in `Missouri` .
- `data/hbsbm/MT_lobbying_corrected_categorical_blockstate.pkl` : The blockmodel for the `lobbying` records in `Montana` .
- `data/hbsbm/MT_testimony_corrected_categorical_blockstate.pkl` : The blockmodel for the `testimony` records in `Montana` .
- `data/hbsbm/NE_lobbying_corrected_categorical_blockstate.pkl` : The blockmodel for the `lobbying` records in `Nebraska` .
- `data/hbsbm/NJ_lobbying_corrected_categorical_blockstate.pkl` : The blockmodel for the `lobbying` records in `New Jersey` .
- `data/hbsbm/OH_testimony_corrected_categorical_blockstate.pkl` : The blockmodel for the `testimony` records in `Ohio` .
- `data/hbsbm/RI_lobbying_corrected_categorical_blockstate.pkl` : The blockmodel for the `lobbying` records in `Rhode Island` .
- `data/hbsbm/SD_testimony_corrected_categorical_blockstate.pkl` : The blockmodel for the `testimony` records in `South Dakota` .
- `data/hbsbm/TX_testimony_corrected_categorical_blockstate.pkl` : The blockmodel for the `testimony` records in `Texas` .
- `data/hbsbm/WI_lobbying_corrected_categorical_blockstate.pkl` : The blockmodel for the `lobbying` records in `Wisconsin` .
- `data/positions.parquet` : The positions recorded in testimony and lobbying records.

Code

The python code used to generate the data and figures presented in the paper is available in the `code` folder. Code used to create the CHORUS dataset is available for review upon reasonable request.

File structure

- `code/download.py` : Functions to download the data from the SPPQ Dataverse or from Google Drive.
- `code/load.py` : Functions to load the data into memory as pandas dataframes.
- `code/figures.py` : Functions to generate the figures presented in the paper.
- `code/utils.py` : Utility functions for data analysis and plotting.
- `code/hbsbm.py` : Functions to create the hierarchical bayesian stochastic block models. When recreating results from scratch using `run_all_blockmodels_from_scratch()`, note that since the blockmodels are stochastic, the results will not be identical to those presented in the paper.
- `code/main.py` : Main file to run the code, via `main.main()`.

Figures

The figures presented in the paper are available in the `figures` folder. The code used to generate them is available in the `code/figures.py` file. Note that the figures in the paper have been edited for clarity and aesthetics.

Other files

- `requirements.txt` : The required packages to run the code.
- `LICENSE.md` : The license for this repository.
- `CHORUS_logo.png` : The CHORUS logo.
- `CODEBOOK.pdf` : A codebook for the CHORUS dataset.
- `figures/placeholder.txt` : A placeholder file to ensure that the `figures` folder is included in the repository.

Requirements

The code was written in Python 3.10 Most required packages are listed in `requirements.txt`. To install them, run the following command in the terminal:

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

The hbsbm code also requires the `graph-tool` package, which can be installed via Conda but is not available on PyPI. To install it, follow the instructions on the graph-tool website.

Runtime

We ran the code on a premium Google Colab instance with 51 GB of RAM and a Python 3 Google Compute Engine backend. The code took approximately 15 minutes to run. Note that the maximum RAM actually used

was about 13GB, so the code should run on a machine with 16GB of RAM.

License

Contact

For questions or comments, please contact Galen Hall at <galen.p.hall [at] gmail.com>.

Acknowledgments

This project was supported with funding from the Climate Social Science Network (CSSN); see https://cssn.org for more information. We thank the CSSN for their support.

The blockmodeling code is based on the hbsbm package by Tiago Peixoto.

The authors are responsible for all errors.