Intro

This document describes the codes and the data used to perform numerical exercises and empirical analysis in the paper "*Endogenous social interactions with unobserved networks*" by Marco Battaglini, Eleonora Patacchini and Edoardo Rainone.[1] First, it describes script by script the codes used to produce all the tables and figures in the paper. Second, it describes the data on U.S. Congress loaded by the codes and used for the empirical analysis. The codes are in the 'codes' folder and the data is in the 'data' folder, both are in the "Replication_material.zip" file.

Codes

Each code used in the paper is reported in blue. Codes are grouped by their main purpose.

Examples

badgood.py: python code that produces level curves of the objective function in (15) and illustrates the value of the objective function in (14), as reported in Figure 2, in the simple setting presented in Example 1.

Numerical Experiments and Empirical Application

All the codes are in Matlab and run with version 2017a. Before launching any code be sure that the unzipped folder is set in the path with the subfolders and the necessary packages are downloaded and saved in that folder. Packages and toolboxes used are reported at the end of this section.

Comparative statics exercises

These codes prepare the scripts for the numerical experiments and produce the boxplots used for the comparative statics. They are in the folder \ABC_algo_application followed by the name of the comparative exercise after the last "_". The "R" codes regard the exercises on the number of observations. The "N" codes regard the exercises on the number of nodes in the networks. The "phi" codes regard the exercises on the elasticity of network formation. The "density" codes regard the exercises on the density of the networks. The "chains" codes regard the exercises on the number of MCMCs used in ABC algorithm. See 4.3.2 of the paper for more details.

main_comparative_TYPE.m: scripts to create the settings for comparative statics simulations then used in Figure 5, 6, A.4, A.5, A.6, A.7, Table A.7 according to the estimation type. TYPE: N (number of observations), R (number of networks), phi (value of the parameter phi), density (density of the networks) and chains (number of MCMCs used for the Bayesian estimation).
box_plots_TYPE.m : scripts to produce the boxplots to compare different simulations' results presented in Figure 5, 6, A.4, A.5, A.6, A.7 according to the estimation type.

Congress data preparation

These codes collect and prepare the data for the empirical analysis of the U.S. Congresses in the paper. The purpose of each function is reported here. More detailed comments can be found in each function.

[1] The authors thank Valerio Leone Sciabolazza and Julien Neves for precious help in collecting data and developing the codes.

A02_econometric_specification_ABC_input.m: uses the input data (see below) to build the econometric specification and the data for the ABC estimation procedure, see Table A.2 for a description.

ABC estimation functions

These codes implement the ABC estimation algorithm, as described in the paper, for both the numerical exercises and empirical analysis. The purpose of each function is reported here. More detailed comments can be found in each script.

run_me_TYPE.m: main scripts to set and lunch the ABC estimation, with TYPE: baseline (standard ABC estimation without restrictions); phi_0 (ABC estimation with restriction phi = 0); curlyphi_0 (ABC estimation with restriction lambda = 0); G_alumni_alpha_1 (ABC estimation with restriction phi = 0 and exogenous network equal to the alumni network); Cospo_resid (ABC estimation with restriction on phi = 0 and exogenous network equal to the cosponsorship network with residual from a first step with alumni network as in Battaglini, Leone Sciabolazza and Patacchini, 2020); G_cospo_hat (ABC estimation with restriction on phi = 0 and exogenous network equal to the prediction of the cosponsorship network with residuals from a first step with alumni network); fast_dist_free (ABC estimation without restictions and distribution free unobservables). For the empirical application, these scripts use the input produced by "A02_econometric_specification_ABC_input.m", in which the outcome variable and other features of the estimation are set. Estimates with simulated data take about five days with five networks of 100 nodes each. Estimates with U.S. Congress data take about three weeks for parsimonious specifications with restrictions. The computational burden increases with the introduction of unobservables. The characteristics of the machines used for the estimation are reported at the end of the document. The computation with multiple chains is done using parallel computing.
J00_two_step: script that joins the first and second step of the ABC algorithm.
J01_ABC_vector_TYPE.m: run the ABC algorithm according to the TYPE of estimation of the run_me_TYPE launched.
J02_create_data.m : create the data for Monte Carlo experiments.
J03_simulation_main.m: set the features for the ABC estimation, combining different functions.
J04_simulation_plot.m: plots the output of the algorithm, posterior of parameters and estimated networks included, and print tables and figures. Specifically, it produces the numbers in Table 1, 2, 3, 4, 5, A.3, and Figure A.9, A.10, A.11, A.12, A.13 according to the estimation type.
J05_combine_chains.m: combine different chains when multiple chains are used.
J06_get_params.m: get the parameters for the algorithm setting.
J07_get_prior.m: get the prior distribution.
J08_get_starting.m: get the starting value of the chain.
J09_get_kernel.m: set the kernel for the iteration.
J10_acceptance_prob.m: evaluate the proposal and the actual value of the parameter.
J11_get_posterior.m: compute the posterior distribution.
J12_ksdensity_pdf.m: extract PDF from numerical vectors.
J13_ksdensity_cdf.m: extract CDF from numerical vectors.
J14_clean_estimation.m: clean the output of the algorithm from burn-in period, stuck chains and produce aggregate outputs and plots.
J14_clean_estimation_individual.m: clean the output of the algorithm from burn-in period and produce outputs and plots for each chain.
J15_create_script.m: create the script to perform the specific ABC estimation.
J16_network_analysis.m: performs network analysis on the estimated network and produces network plots. Specifically, Figure 3, 4, 7, A.1, A.2, A.3, A.8, A.14, A.15, A.16, and numbers in Table 6, A.1, A.5, A.6, A.8 according to the estimation type.

Functions used to run the numerical experiments are in the folder: \ABC_algo_experiments.
Functions used for the empirical estimation with U.S. Congress data are in the folder: \ABC_algo_application.

Counterfactual Analysis

These codes perform the counterfactual analysis using the U.S. Congresses data and the output of the estimation algorithm, as described in the paper. The purpose of each function is reported here. More detailed comments can be found in each function.

C03_extremism.m: performs the counterfactual analysis in which the level of extremism is moderated and produces the results in Table 7, A.4, Figure 8.
C02_alumni_link_formation.m: performs the counterfactual in which the influence of alumni connections is removed from the network and produces the results in Table 8.

Auxiliary functions

These codes are used to perform both the numerical exercises and empirical analysis and serve as inputs of the codes reported above. The purpose of each function is reported here. More detailed comments can be found in each function.

A00_medians_and_post_estimation.m: computes medians and post-estimation values
A03_compute_expected_values: computes expected values of the network and outcome and produces fit measures of the model.
A04_simulate_Y: simulates Y based on parameters' posterior distributions and random errors.
A05_expected_Y: computes the expected value of Y based on median parameters' posterior distributions.
X_link_shape.m: defines the functional form g(x_i,x_j) for the link formation process.
X_link_shape_indiv.m: creates the individual level data for the link formation process.
plot_pol_net_G.m : creates figures of networks using the XX algorithm.
net_comparison: produces plots and statistics to compare two different networks.
Wgen.m: generates random circular networks
Wgen_1.m: generates random circular networks (variant)
get_tables.m: computes tables from ABC results
block_net.m: divides a network in sub components
tabellizzare_bayesian.m: produces tables with p-values based on estimated posterior distributions
measures_for_networks_without_plots_bub: computes and produces a series of network indicators and metrics

Matlab packages used

- ComDetTBv090: https://it.mathworks.com/matlabcentral/fileexchange/45867-community-detection-toolbox
- matlab_bgl: https://www.mathworks.com/matlabcentral/fileexchange/10922-matlabbgl
- Spatial Econometrics toolbox: https://www.spatial-econometrics.com/

Download the packages and save them in the 'codes' folder.

Matlab toolboxes used

Optimization toolbox
Parallel computing toolbox

Data

All the data used for the empirical application is in the file "inputs.mat" in Matlab format. Each object in this .mat file is listed in red and described below. Objects are grouped by type of information content.

Individual level data

Our analysis considers all of the legislator characteristics provided by the Center for Effective Lawmaking (2018). This data was last accessed on April 9, 2018. It includes the number of years spent in Congress and its squared term, margin of victory and its squared term, the state in which the Congress member was elected, the size of the state congressional delegation, party, chairmanship, majority and minority party leadership, whether the representative is a member of the most powerful committees, previous legislative experience and professionalism, gender, and race. We also include DW ideology of the legislator, as registered by Voteview (2019). This data was last accessed on March 13, 2019. In addition to these factors, we include the age of the legislators as a demographic characteristic. The age of the legislators is derived from the Biographical Directory of the U.S. Congress (2020).

take_1, take_2, take_3, take_4, take_5, take_all:
datasets with individual characteristics for Congress members in the 109th, 110th,111th, 112th, 113th Congress respectively and all pooled with dimension equal to the number of congress members for which all the information is available. A codebook with precise definitions of all variables can be found below.

state_FE_1, state_FE_2, state_FE_3, state_FE_4, state_FE_5; STA:
matrices of state fixed effects for politicians in the 109th, 110th,111th, 112th, 113th Congresses respectively with dimensions equal to the number of congress members (state_FE_N) and the number of states; a structure object that contains 5 adjacency matrices, one for each Congress, with dimension equal to the number of congress members and elements equal to 1 if two politicians are elected in the same state and 0 otherwise (STA).

Topics

This data contains the policy areas in which each member sponsored or cosponsored in the original text the highest number of bills. The policy topic coding system is provided by the Policy Agendas Project (2018). Policy areas are classified according to codebooks created by Baumgartner and Jones. Each entry is coded into one of 21 major topics and 220 subtopics. We use the subtopics to classify exactly the areas of interest. This data was last accessed on August 8, 2018.

dummy_topic_1, dummy_topic_2, dummy_topic_3, dummy_topic_4, dummy_topic_5, dummy_topic; TOP:
matrices of topic fixed effects for politicians in the 109th, 110th,111th, 112th, 113th Congress respectively and all pooled with dimensions equal to the number of congress members and the number of topics in which members have the highest number of bills (dummy_topic_N); a structure object that contains 5 adjacency matrices, one for each Congress, with dimension equal to the number of congress members and elements equal to 1 if two politicians have the same topics and 0 otherwise (TOP).

Networks

Cosponsorship
This data contains information on bills cosponsored by congress members. The ij entry is equal to the number of bills by j that i cosponsored in that congress. Data is provided by Govtrack.us (2017). This data was last accessed on November 28, 2017.

CO; cos_net_all:
a structure object that contains 5 adjacency matrices, one for each Congress, with dimension equal to the number of congress members and element ij equal to the number of bills by j that i cosponsored in that congress (CO); a block-diagonal adjacency matrix with all Congresses pooled together (cos_net_all).

Alumni
This data contains the information on alumni connections among congress members derived from Biographical Directory of the U.S. Congress (2020). The data provided by this source is updated periodically and the website is redesigned over time. We downloaded the version available on April 2020 (last access: April 9, 2020).

HH; alu_net_all:
a structure object that contains 5 adjacency matrices, one for each Congress, with dimension equal to the number of congress members and element ij equal to 1 if i and j graduated from the same institution within four years of each other (HH); a block diagonal adjacency matrix with all Congresses pooled together (alu_net_all).

Committees
This data contains the information on connections among Congress members through Committees, derived from Charles Stewart's Congressional Data Page (2018)**.** We downloaded the data on April 9, 2018.

COM:
a structure object that contains 5 adjacency matrices, one for each Congress, with dimension equal to the number of congress members and element ij equal to 1 if two legislators belong to the same committee.

Data Availability Statement
For all the data sources cited there is no required registration, membership, application procedure, monetary cost, or other qualifications, beyond the URL for download. The data needed to reproduce all the results is freely available in the replication material of this manuscript.

REFERENCES

Biographical Directory of the U.S. Congress (2020), "*Biographical Information on Members of the U.S. Congress*", accessed April 9, 2020, http://bioguide.congress.gov/

Center for Effective Lawmaking (2018), "*Legislative Effectiveness Data from 1973-2020*", accessed April 9, 2018, http://www.thelawmakers.org

Charles Stewart's Congressional Data Page (2018), "*Data on Connections among Congress Members through Committees*", accessed April 9, 2018, http://web.mit.edu/17.251/www/data_page.html#2

Congressional Bills Project (2018), "*Congressional Bills Project Data*", accessed April 9, 2018, http://congressionalbills.org

Govtrack.us (2017), "*Data on Bills Cosponsored by Congress Members*", accessed November 28, 2017, http://jhfowler.ucsd.edu/cosponsorship.htm

Policy Agendas Project (2018), "*U.S. Policy Agendas Dataset*", accessed August 8, 2018, https://www.comparativeagendas.net/pages/master-codebook

Voteview (2019), "*DW Nominate Scores Data*", accessed March 13, 2019, https://www.voteview.com/static/data/out/members/HSall_members.csv

Codebook

Variable Name	Description	Source
bill_ss	Number of *substantively significant* bills that each member sponsored.	https://thelawmakers.org/
aic_ss	Number of those bills that received any action in committee.	https://thelawmakers.org/
abc_ss	Number of those bills that received any action beyond committee on the floor of the House.	https://thelawmakers.org/
pass_ss	Number of bills that passed the House.	https://thelawmakers.org/
law_ss	Number of bills that became law.	https://thelawmakers.org/
bill_s	Number of *substantive* bills that each member sponsored.	https://thelawmakers.org/
aic_s	Number of those bills that received any action in committee.	https://thelawmakers.org/
abc_s	Number of those bills that received any action beyond committee on the floor of the House.	https://thelawmakers.org/
pass_s	Number of bills that passed the House.	https://thelawmakers.org/
law_s	Number of bills that became law.	https://thelawmakers.org/
bill_c	Number of *commemorative/symbolic* bills that each member sponsored.	https://thelawmakers.org/
aic_c	Number of those bills that received any action in committee.	https://thelawmakers.org/
abc_c	Number of those bills that received any action beyond committee on the floor of the House.	https://thelawmakers.org/
pass_c	Number of those bills that passed the House.	https://thelawmakers.org/
law_c	Number of those bills that became law.	https://thelawmakers.org/
bill_tot	Total number of bills that each member sponsored.	https://thelawmakers.org/
aic_tot	Number of those bills that received any action in committee.	https://thelawmakers.org/
abc_tot	Number of those bills that received any action beyond committee on the floor of the House.	https://thelawmakers.org/
pass_tot	Number of those bills that passed the House.	https://thelawmakers.org/
law_tot	Number of those bills that became law.	https://thelawmakers.org/
les	Weighted average of the number of a member of Congress's sponsored bill that were introduced, received any action in committee and beyond committee, passed the House, and became law. It differentially weights commemorative, substantive and significant legislation. Created by Volden C. and Wiseman A. E. (2014), Legislative Effectiveness in the	https://thelawmakers.org/

	United States Congress: The Lawmakers, Cambridge University Press.	
les_bill	Sum of bills that each member sponsored weighted by Volden and Wiseman weights.	https://thelawmakers.org/
les_aic_abc	Sum of those bills that received any action in committee or became law weighted by Volden and Wiseman weights.	https://thelawmakers.org/
les_pass_law	Sum of those bills that passed the House or received any action beyond committee on the floor of the House weighted by Volden and Wiseman weights: 10 for *substantively significant,* 5 for *substantive* and 1 for *commemorative/symbolic*.	https://thelawmakers.org/
les_abs	Sum of the last three indicators.	https://thelawmakers.org/
time	A number that identifies the Congress: 1 = 109th , 2 = 110^{th,} 3 = 111th, 4 = 112th , 5 = 113th .	
gender	Dummy variable taking value of 1 if the member of Congress is female, and 0 otherwise.	https://thelawmakers.org/
nowhite	Dummy variable taking value of 1 if the member of Congress is Afro-American or Latino, and 0 otherwise.	https://thelawmakers.org/
party	Dummy variable taking value 1 if the member of Congress is a Democrat, and 0 if the member of Congress is a Republican, and 0 otherwise.	http://web.mit.edu/17.251/www/data_page.html#2
seniority	Number of consecutive years in the House of Representatives.	https://thelawmakers.org/
seniority2	Square of seniority.	
margin	Election Margin of Victory of the member of Congress.	https://thelawmakers.org/
dw	Distance to the center in terms of ideology measured using the absolute value of the first dimension of the dw-nominate score created by McCarty et al. (1997).	https://www.voteview.com/static/data/out/members/HSall_members.csv
deleg_size	Number of seats assigned to member of Congress's State of election.	https://thelawmakers.org/
nchair	Dummy variable taking value of 1 if the member of Congress is a chair of at least one committee.	http://web.mit.edu/17.251/www/data_page.html#2
power	Dummy variable taking value of 1 if the member of Congress is a member of Appropriation, Rules, or Way and Mean.	http://web.mit.edu/17.251/www/data_page.html#2
maj_leader	Dummy variable taking value of 1 if the member of Congress is member of the majority	https://thelawmakers.org/

	party leadership, as reported by the Almanac of American Politics, and 0 otherwise.	
min_leader	Dummy variable taking value of 1 if the member of Congress is member of the minority party leadership, as reported by the Almanac of American Politics, and 0 otherwise.	https://thelawmakers.org/
time	Categorical variable. It takes the value of 1 if the record refers to the 109th Congress, 2 if the record refers to the 110th Congress, 3 if the record refers to the 111th Congress, 4 if the record refers to the 112th Congress, 5 if the record refers to the 113th Congress.	Authors' elaboration
state_leg	Dummy variable taking value of 1 if the member of Congress served in state legislature.	https://thelawmakers.org/
state_leg_prof	Squire's (1992) index of state professionalism relative to Congress.	https://thelawmakers.org/
state	A number that identifies the state in which the Congress member was elected: 1 = Alabama 2 = Alaska 3 = American Samoa 4 = Arizona 5 = Arkansas 6 = California 7 = Colorado 8 = Connecticut 9 = Delaware 10 = District of Columbia 11 = Florida 12 = Georgia 13 = Guam 14 = Hawaii 15 = Idaho	https://thelawmakers.org/

	16 = Illinois 17 = Indiana 18 = Iowa 19 = Kansas 20 = Kentucky 21 = Louisiana 22 = Maine 23 = Maryland 24 = Massachusetts 25 = Michigan 26 = Minnesota 27 = Mississippi 28 = Missouri 29 = Montana 30 = Nebraska 31 = Nevada 32 = New Hampshire 33 = New Jersey 34 = New Mexico 35 = New York 36 = North Carolina 37 = North Dakota 38 = Northern Mariana Islands 39 = Ohio	

	40 = Oklahoma 41 = Oregon 42 = Pennsylvania 43 = Puerto Rico 44 = Rhode Island 45 = South Carolina 46 = South Dakota 47 = Tennessee 48 = Texas 49 = U.S. Virgin Islands 50 = Utah 51 = Vermont 52 = Virginia 53 = Washington 54 = West Virginia 55 = Wisconsin 56 = Wyoming	

Computer characteristics

CPU

```
processor       : 39
vendor_id       : GenuineIntel
cpu family      : 6
model           : 79
model name      : Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz
stepping        : 1
microcode       : 0xb000021
cpu MHz         : 2200.000
cache size      : 25600 KB
physical id     : 1
siblings        : 20
core id         : 12
cpu cores       : 10
apicid          : 57
initial apicid  : 57
fpu             : yes
fpu_exception   : yes
cpuid level     : 20
wp              : yes
```

RAM

```
              total        used        free      shared  buff/cache   available
Mem:      263843204    24835504   229534492     4211676     9473208   234006172
Swap:      94371836        1544    94370292
```