Universität Bremen

Global Dynamics of Social Policy CRC 1342

Networks and geographies of global social policy diffusion.

Culture, economy and colonial legacies.

Updates for Calculations October 2020

New Colonial Network

As promised we decided and worked a lot on a time-varying colonial network in which countries are linked via colonial dependencies/histories. We provide it in an edgelist format in which the colonized country 'nominates' its colonizer. Therefore, a colonial dependency is denoted by the outdegree and consistent with the other edgelist formats we provide. In short, the network fulfills the following criteria:

- Time-Variant: i.e. countries at yearly intervals from 1880 to 2010
- Weighted: i.e. connections of countries have differing weights based on colonial status in that year
    - existing colonial relationship has the value of 1
    - no colonial relationship has the value of 0
    - time before colonial relationship started has the value of 0
    - past colonial relationship have differing values < 1
- Possibility for different values for past colonial relationship
    - constant value denoting past relationship either as 0.3; 0.5; or 0.7
    - decay parameter as a log function and additionally as an exponential function
- Inclusion of not only European colonial powers, but every relationship in which power was asserted over a space which was and is not still the space of said nation state exerting the power through "colonialism"; e.g. USSR
- Directed: i.e. (analogous to survey data) countries that were colonized "nominate" their colonizers

Data and Definition of "colonial tie"

The raw data is based on the COLDAT Dataset[1] by Bastian Becker in combination with CEPII[2] and own data collection. Relationships that we added were for example that Georgia 'nominates' Russia as colonizer from 1801-1990 and Iran (Persian Empire) as well as Turkey (Ottoman Empire) from 1555-

[1] Becker, Bastian, 2019, "Colonial Dates Dataset (COLDAT)", https://doi.org/10.7910/DVN/T9SDEW
[2] Head, Keith and Thierry Mayer, 2014, "Gravity Equations: Toolkit, Cookbook, Workhorse." Handbook of International Economics, Vol. 4,eds. Gopinath, Helpman, and Rogoff, Elsevier.

1801. In line with CEPII we did not add military occupations of e.g. Poland – Germany during World War II. CEPII follows a somewhat non-transparent definition of colonial links, however, boiling it down would mean:
a colonial relationship should involve long-term, civilian administration that includes significant settlement[3]. Short-term military occupation does not fit within these criteria.

Decay Parameters

We calculated two different decay parameter which both have advantages as well as disadvantages.

The first is a logarithmic parameter calculated as:

$$1 / (1 + \log(\textit{No. years since colony ends +1}))$$

It has a quite rapid fall in values i.e. 0.59 one year after colony ends and 0.29 ten years after. However, the values tend to get small very slowly like for example 0.17 at 100 years after colonization.

For the second way of calculating the decay we used the following exponential function

$$\exp(-(\textit{No. years since colony ends}/40))$$

Here the characteristics are in a sense flipped: The exponential function has a rather steady slow decay of the value with e.g. 0.97 one year after colonization and 0.77 ten years after colonization. On the other hand, the values tend to get smaller a lot faster than in the log function with e.g. 0.08 at 100 years after colonization. An example of how that looks in the dataset can be found on the last page.

We decided on using the exponential decay parameter precisely because the values tend to get quite small, i.e. the influence via that link of colonial history is diminishing. Nevertheless, all of the solutions are implemented in the new dataset. We, however, believe that one standard which is used throughout the edited volume is most beneficial. See below for changes in the script to incorporate the new dataset.

Issues stemming from the September Workshop

During our workshop in September there where two major issues not yet solved. The first one is the way of how *time dependence* should be controlled in our diffusion models. Secondly, it is still an open question, how to deal with countries that were part of other countries, and came into existence after having split from the overarching unit. In other words, the second issue deals with the definition of our units of observation in analyses that assume a *consistent* set of subjects over time.

Time dependence

The problem of *time dependence* of the diffusion process is related to unobserved heterogeneity. If the process strictly followed a logistic growth, the baseline hazard rate would be low at the beginning, high when the share of adopters approximates 50%, and then decrease again. However, it is unlikely that the diffusion process follows this pattern in all projects. The first advantage of the piecewise constant rate model is that it does not *ex ante* determine the functional form of the hazard rate over time.

Surely, there are also some flexible parametric approaches: the hazard rate can be e.g. a function of log(time) and log($time^2$). Depending on the signs of the respective coefficients, the hazard rate can decrease, increase or follow a sickle-shaped form. The problem of this parametric approach, however, is a strong multicollinearity of this time-variable with those explanatory variables that have

[3] Mayer, Thierry; Head, Keith and John Ries, 2008, "The Erosion of Colonial Trade Linkages after Independence", CEPII, Working Paper No 2008-27. http://www.cepii.fr/PDF_PUB/wp/2008/wp2008-27.pdf.

non-stationary trends. If there is, for instance, a non-stationary trend in GDP and the density of the trade network (steady growth over time), the long, decreasing tail of the sickle-function would negatively correlate with these non-stationary trends. Keep in mind that our data is based on countries' long-term historical trends: they are in many cases non-stationary. Again, this is an advantage of the flexible piecewise constant rate model: it captures time dependence imprecisely by the step function, and thereby does not to introduce such kind of strong multicollinearity into the event history model.

Non-independent observations

The second problem is that there are historical time periods in which several countries did not exist because they were part of a larger, embracing unit. An ideal-type example of this case are countries of former Yugoslavia. If Serbia and Croatia introduced a social policy when they both where part of former Yugoslavia – so it was the overarching unit which actually did this introduction – , Serbia and Croatia are non-independent observations. This has to do with the way the dataset is built for the diffusion analysis: here, the set of nodes in the network is constant over time, which implies that Serbia and Croatia existed *before*, *during* and *after* Yugoslavia existed.

Our approach to address this problem is to regard Serbia and Croatia as a kind of "spatial patches" – well aware of the fact that many countries actually changed their borders over history (this is measurement error). In this view, Serbia and Croatia are spatial patches that were at risk of introducing the social policy *before*, *during* and *after* Yugoslavia existed. Yugoslavia will not be regarded as a subject in our data set, but Serbia and Croatia (and all other countries which formerly belonged to former Yugoslavia) are indeed distinct subjects. These subjects are, however, not statistically independent from each other!

In the event history diffusion model we address this statistical non-independence by correcting the standard errors using the standard Huber-White method. In our case, this method requires a time-dependent indicator-variable of the respective cluster of countries. This indicator-variable defines a cluster ID, in this case all countries that belonged to former Yugoslavia. It is important that countries can enter or leave the respective context over time, while the context ID variable remains constant.

If a country has not yet joined the context or has already left the context, it gets an ID variable which identifies each *country-year* (!) as the context ID. This procedure has the following advantage: it accounts for the statistical non-independence of observations >when< they were part of an overarching cluster by using the corrected standard error, but it does not impose any standard error correction in the hazard model for country-years *not* belonging to the respective cluster or to any other cluster. Otherwise, if independent countries just had the 'iso3' code as a cluster ID, the Huber-White method would account for the clustering of country-years within an independent country. This would be appropriate in some methods of panel analysis, but not in event history analysis. Here, we use country-years to define the (conditional) hazard rate by (conditionally) dividing the number of events by the sum of time at risk.

Changes to the originally provided script

The changes that are described here refer to the original script sent in an email on 09.09.2020 by Fabian Besche-Truthe. The Script had the title: "soc_pol_diffusion_1_4" and is now called "soc_pol_diffusion_1_6"

1. preparation of the trade network

We realized that with only using a log function to smooth out the much skewed trade values we created negative values. That in itself would be not such a huge problem, however since we are dealing with network weights a negative weight does not only make no sense theoretically, it actually distorts the calculation of the exposure variable a lot. We end up with values outside the range of 0 – 1. To circumvent that we changed the script in line 109 in a way that we make sure to not end up with negative log_values:

trade_dyadic$log_value <- ifelse(trade_dyadic$smoothtotrade > 0, log(trade_dyadic$smoothtotrade + 1), 0)

2. Creating the diffnet_colony object

Since we updated the colonial ties network we need to change a little bit in the creation of the respective diffnet object starting in line 456.

1. As the default weight variable we use the log decay parameter. So the line of code when choosing the weights of the network is: "w = colonial_ties_edgelist$weight_decay_exp,"
2. The colonial network is now a directed one, so we changed the code at the respective place (line 467) to: "undirected = FALSE,"

3. Including the Cluster ID

We decided on including the Cluster ID at the stage where we combine the exposure variables to a final dataframe. We do that, because, when creating a diffnet object, every variable is supposed to be either factor or an interval, i.e. the package does not recognize character variables (This seems to be a bug or at least it does not work). However, the cluster_id variable has more factor levels than there are observations, and we do want to keep it intact in the sense that the values are still understandable and not just numbers.

So we load the Cluster ID in line 621ff and include it into the diff_data dataset.

4. Estimation of Clustered Standard errors (Huber/White) and implementation in stargazer3

To estimate the corrected standard errors we first need to load two new packages: "sandwich" and "lmtest". These were added right at the beginning of the script where all the other packages are loaded (lines 3-10), in addition, in line 11, source (i.e. load) the function stargazer3 (stargazer2 will be loaded automatically as well, this can be used in case you do not want to display corrected standard errors but odds ratios).

Starting line 764 the coeftest function calculates the correct standard errors and corresponding statistical significance for the model output. The standard errors are corrected after the model calculation; this is not a mistake.

1. Save the output from coeftest in a new object
2. You need to specify the model twice in the function, please be aware.
3. Give that output over to stargazer3 (as first argument), along with the original glm output (the model used to calculate the correct standard errors, with the argument origin_model = …).
4. If you have multiple models for stargazer, wrap them in list()
5. The output is a html table with odds ratios but does not display standard errors.
6. The stargazer3 function is custom fit for our needs, it can be modified like the normal stargazer function, but if you wish to include standard errors (it is not common to display odds ratios with standard errors) or confidence intervals, let us know, they need to be calculated separately and inserted manually due to the standard error correction.

Appendix

Example of new colonial ties edgelist

ego_id	alter_id	year	weight_constant_0.3	weight_constant_0.5	weight_constant_0.7	weight_decay_log	weight_decay_exp
…..	……	……	……	……	……	……	……
AGO	PRT	1974	1	1	1	1	1
AGO	PRT	1975	1	1	1	1	1
AGO	PRT	1976	0.3	0.5	0.7	0.59062	0.97531
AGO	PRT	1977	0.3	0.5	0.7	0.47651	0.95123
AGO	PRT	1978	0.3	0.5	0.7	0.41906	0.92774
AGO	PRT	1979	0.3	0.5	0.7	0.38322	0.90484
AGO	PRT	1980	0.3	0.5	0.7	0.3582	0.8825
AGO	PRT	1981	0.3	0.5	0.7	0.33945	0.86071
AGO	PRT	1982	0.3	0.5	0.7	0.32473	0.83946
AGO	PRT	1983	0.3	0.5	0.7	0.31277	0.81873
AGO	PRT	1984	0.3	0.5	0.7	0.30279	0.79852
AGO	PRT	1985	0.3	0.5	0.7	0.2943	0.7788
AGO	PRT	1986	0.3	0.5	0.7	0.28695	0.75957
AGO	PRT	1987	0.3	0.5	0.7	0.28051	0.74082
AGO	PRT	1988	0.3	0.5	0.7	0.2748	0.72253
AGO	PRT	1989	0.3	0.5	0.7	0.26968	0.70469
AGO	PRT	1990	0.3	0.5	0.7	0.26507	0.68729
AGO	PRT	1991	0.3	0.5	0.7	0.26088	0.67032
AGO	PRT	1992	0.3	0.5	0.7	0.25704	0.65377
AGO	PRT	1993	0.3	0.5	0.7	0.25352	0.63763
AGO	PRT	1994	0.3	0.5	0.7	0.25027	0.62189
AGO	PRT	1995	0.3	0.5	0.7	0.24725	0.60653
AGO	PRT	1996	0.3	0.5	0.7	0.24444	0.59156
AGO	PRT	1997	0.3	0.5	0.7	0.24181	0.57695
AGO	PRT	1998	0.3	0.5	0.7	0.23935	0.5627
AGO	PRT	1999	0.3	0.5	0.7	0.23703	0.54881
AGO	PRT	2000	0.3	0.5	0.7	0.23485	0.53526
AGO	PRT	2001	0.3	0.5	0.7	0.23278	0.52205
AGO	PRT	2002	0.3	0.5	0.7	0.23083	0.50916
AGO	PRT	2003	0.3	0.5	0.7	0.22897	0.49659
AGO	PRT	2004	0.3	0.5	0.7	0.22721	0.48432
AGO	PRT	2005	0.3	0.5	0.7	0.22553	0.47237
AGO	PRT	2006	0.3	0.5	0.7	0.22393	0.4607
AGO	PRT	2007	0.3	0.5	0.7	0.22239	0.44933
AGO	PRT	2008	0.3	0.5	0.7	0.22093	0.43823
AGO	PRT	2009	0.3	0.5	0.7	0.21952	0.42741
AGO	PRT	2010	0.3	0.5	0.7	0.21817	0.41686