Replication Material for 'Fixed-effect regressions on network data'

Koen Jochmans and Martin Weidner

April 4, 2019

This note describes how to replicate the results reported in Section 4 of the paper.

Teacher value-added Our analysis of the estimation of teacher value-added uses the data of Verdier (2018). The source data can be applied for at the North Carolina Education Research Data Center. Full details on how to obtain the data and how to arrive at the teacher graph we use in our work are available in the supplementary material of Verdier (2018). It can be downloaded from the online data archive of *The Review of Economics and Statistics* at the following permant link:

`https://doi.org/10.7910/DVN/ZQ5YCL`

The graph so obtained can be imported in Matlab as four vectors. `edgeid` contains a unique label for each of the 53,741 edges; `sender` and `receiver` identify the nodes involved in each of these edges; `weight` contains the weight assigned to each edge. Call this data set `teachers_read.mat`.

The m-file `TeacherValueAdded.m` opens `teachers_read.mat` and does the following:

1. It relabels teacher identifiers;
2. It identifies the largest connected component in the graph;
3. It constructs the adjacency and degree matrices $\boldsymbol{A}$ and $\boldsymbol{D}$ as well as the Laplacian $\boldsymbol{L}$ and normalized Laplacian $\boldsymbol{S}$.
4. It computes the network statistics of interest, i.e., λ_2 and the various harmonic means;
5. It saves the statistics as `teachers_read_graph.m`.

Next, the m-file `TeacherValueAdded_Statistics.m` first opens `teachers_read_graph.m` and then computes descriptive statistics of the harmonic means as reported in Table 1. It also provides histograms of these distributions.

Occupational network Our illustration of the occupational network uses the full data set of the British Household Panel Survey (BHPS). These data can be requested from the U.K. Data Service at

`https://www.iser.essex.ac.uk/bhps/acquiring-the-data`

We combine all waves of the BHPS and store the person identifier `ahid`, his/her occuptional classification `occupation`, and wave number `year` as an array. Call the data file containing this array `bhps_data.mat`.

The m-file `Occupations.m` opens `bhps_data.mat` and does the following:

1. It asks at what level the occupational classification needs to be considered. In the paper we report results for the most detailed level (type '3' when prompted by Matlab).
2. It relabels person and occupation identifiers;
3. It constructs the weighted adjacency matrix of the one-mode projection on occupations from these bipartite data.
4. It identifies the largest connected component in the graph;
5. It constructs the adjacency and degree matrices $\boldsymbol{A}$ and $\boldsymbol{D}$ as well as the Laplacian $\boldsymbol{L}$ and normalized Laplacian $\boldsymbol{S}$.
6. It computes the network statistics of interest, i.e., λ_2 and the various harmonic means;
7. It saves the statistics as `bhps_data_graph_3D.m`.

Next, the m-file `Occuptations_Statistics.m` first opens `bhps_data_graph_3D.m` and then computes descriptive statistics of the harmonic means as reported in Table 2. It also provides histograms of these distributions.