

Todor Parushev's Chilean data

Inside the data_prep folder

Data Preparation

Original Data

Todor has collected his data in file **ChileNetwork no commas.csv**

This is a .csv format file ("comma separated variables") (Except for the comma formatting it is an ordinary text file.)

Each line of the file shows a connection of a person (director) to a corporation. The line includes id numbers as well as textual names for everything. This number is what identifies a unique person or corporation. Your file may show the same person (or corporation) with slightly different spellings. The program knows they are the same person (or corporation) because they have the same id numbers. The process of editing and cleaning the data is a matter of getting these id numbers right.

Pre-processing the data

First a note on ".csv file". .csv files use the suffix .csv. They can be opened by Excel and edited in Excel. Then save as a .csv file (using "SAVE AS" – unless Excel automatically uses a .csv format for you.)

Now: The data in original data file can be reorganized (or reshaped) in several forms. The mapping (that will come later) requires a corporation-to-corporation table, that is also a .csv file. Caution: Do not use any commas or less-than signs, or greater-than signs within the text of your original data file — no commas that are visible when the file is viewed in Excel. That's why it has "no commas" in its name. If a file has such things in it, edit-replace them in Excel. Replace them with something else.)

Program **Network_Data_Reshape05.py** reorganizes the data into several files including the corporation-to-corporation table that you need. Run the program. It is pre-programmed to seek data in Todor's original data file, using the name for that file that I gave that file.

The several files that the program writes includes two files for which there is no immediate use and a "...cc..." file. The program also puts valuable information on the screen: For each connected set within the full data set, it prints the names of the entities in that connected set (both person names and corporation names).

At the present time the ...cc... file is a corporation to corporation table for *all* the corporations in the original data file.

Before you proceed any further, it is wise to hand-edit this file: Usually there is one large connected set of corporations. Hand-edit the file (in Excel) to remove corporations that are not in this connected set. Here this reduced version of the cc-file is named **Chile.csv**

Inside the *Example Parushev01_2dim_p* Folder

Processing the data with DwV

Now you are prepared for **Example_Parushev01_2dim_p**.

The **Example_Parushev01_2dim_p** folder includes files that you use directly as well as files that the program handles by itself.

In the name of this folder, I have reminded myself that the visualization will be in two dimensions and will use equation “p” (the power-normal).

Inside the folder:

Chile.csv is a copy of the file created in the data preparation phase of the work.

Chile_SETUP_2dimscanDwV.txt is a setup file containing commands that will be received by the program. It is “plain text”. It is not a **Word** file, although **word**, if asked politely can SAVE AS a .txt file.

generic_csv_pscriptspan14.py is the program you need to run

Starting the Program

If you are going to run this Python code from Python’s **IDLE** , then simply double click **generic_csv_pscriptspan14.py** and it will ask you questions.

If you are going to run from a “Terminal” (for Mac OSX or Ubuntu unix) or from a “Command line” (PC), it will be considerably faster.

For Mac, open a “Terminal”.

For Mac and Unix, Navigate to **Example_Parushev01_2dim_p**

For Mac and unix, type **python2.6 generic_csv_pscriptspan14.py**

Whichever path you have used to run the program, the program should start up and ask you a question and then request that you navigate to certain files .

First: A question “Continuation or Startup?”

First answer (always): Startup

Second: Navigate to the SETUP file (named **Chile_SETUP_2dimscanDwV.txt**)

Second answer: Occasionally instead of asking you to navigate, it just sits there and stares at you. This usually means that it has opened a navigation window, but left that window “under” some other file that is open on your desktop. Look around for it, closing unnecessary windows. Then navigate to this SETUP file inside this **Example_Parushev01_2dim_p** folder. Double click on it.

Third: Navigate to the data file (named **Chile.csv**)

Running the Program

Sit back. Enjoy the pictures pictures — somewhere on your desktop. This is going to take several hours. It is working. Give it 3 or 4 hours (at least).

On the Mac, command-option-8 will give you a closeup. (If it doesn’t, then go into Mac Icon that leads to System Preferences and deal with the Universal Access Preferences.)

Output from the Program