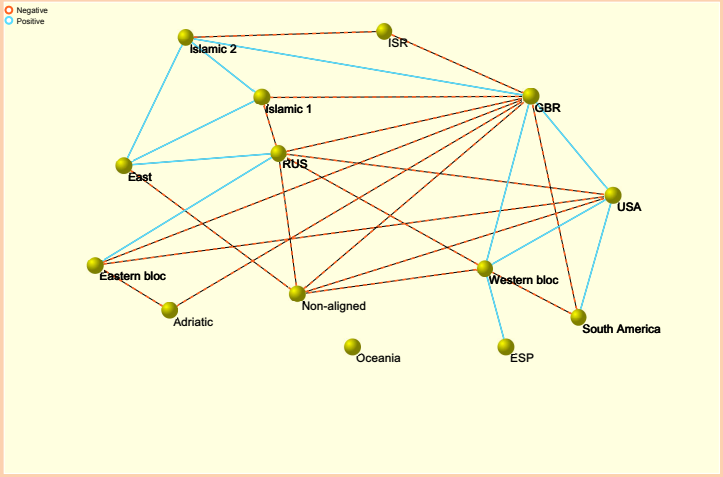
Network macro stability with micro level instability



Patrick Doreian

University of Ljubljana

University of Pittsburgh

ARS15, Anacapri

April 29-30, 2015

version: April 13, 2015 / 20 : 51

Outline

Introduction

- Discerning the underlying structure of a network is a critical issue for understanding networks.
- Tracking this structural feature over time is even more important.
- There are two broad approaches for doing this:
  - Blockmodeling and generalized blockmodeling
  - Community detection
- Are there conditions under which these methods fail?
- If so, what are the conditions for failure?
- What alternatives are there when these conditions exist?

Blockmodeling

Key ingredients:

- Selecting some form of equivalence for partitioning vertices into positions;
- Choosing between indirect fitting or direct fitting of a blockmodel;
- Opting for an inductive or a deductive approach and
- Fitting a blockmodel. But . . .
- The default for most application appears to be indirect fitting of inductive blockmodels.
- My **very strong** preference is for direct fitting of deductive blockmodels.

Induction

- Switching on a blockmodeling algorithm and accepting whatever is returned as a result is a *supreme expression of ignorance*!
- We often know more than what this *mindless* analysis implies:
  - This knowledge can be based on substance.
  - It can be based on knowledge of the specific empirical context that is studied.
  - Knowledge can stem also from an understanding of similar situations.
- If we do have such knowledge it is *remarkably foolish* to act as if we are clueless about what we study.
- This knowledge can be used fruitfully to pre-specify blockmodels.

Pre-specification of blockmodels

The selection of any equivalence type implies a set of **ideal block types**.

Items that can be specified in advance of an analysis include:

- The number of positions (clusters);
- The locations of some or all block types;
- The minimum number of units in a position;
- Units that *must* be together in a position;
- Units that *must not* be together in a position;
- Penalties for specific types of of inconsistencies with ideal block types

There can be partial or complete pre-specification of a blockmodel.

Two empirical examples

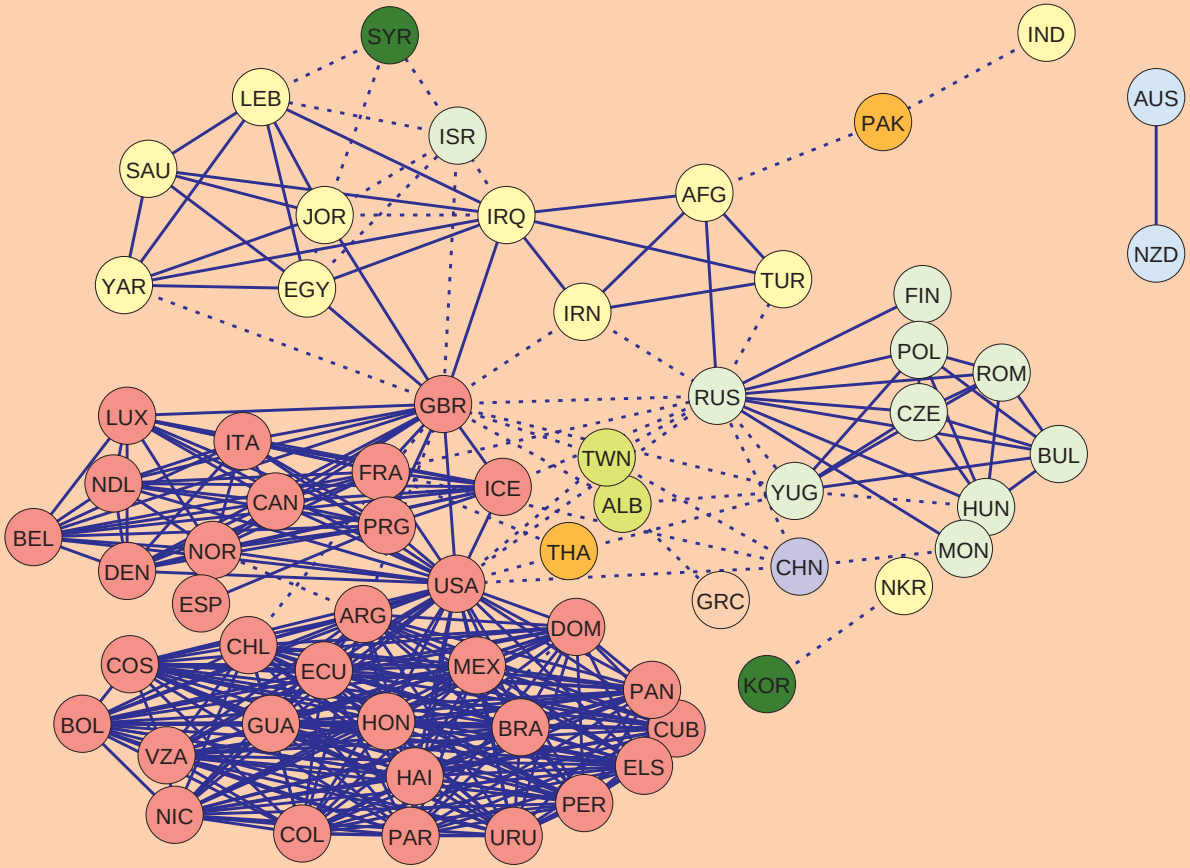
- The level of structural balance in the international system of signed relations (Correlates of War data).
  - Positive ties: joint memberships in alliances, unions and inter-governmental agreements for states.
  - Negative ties: states being at war, in conflict with each other without military involvement, border disputes and sharp ideological or policy disagreements.
- Financial Direct Investment (FDI) is a financial relation involving the acquisition of direct ownership (represented by voting securities) by a single company located in one country (the parent) of a company located in a different country (the foreign affiliate or target). The investments for this example are in electical infrastructure.

For both examples, the units are nations.

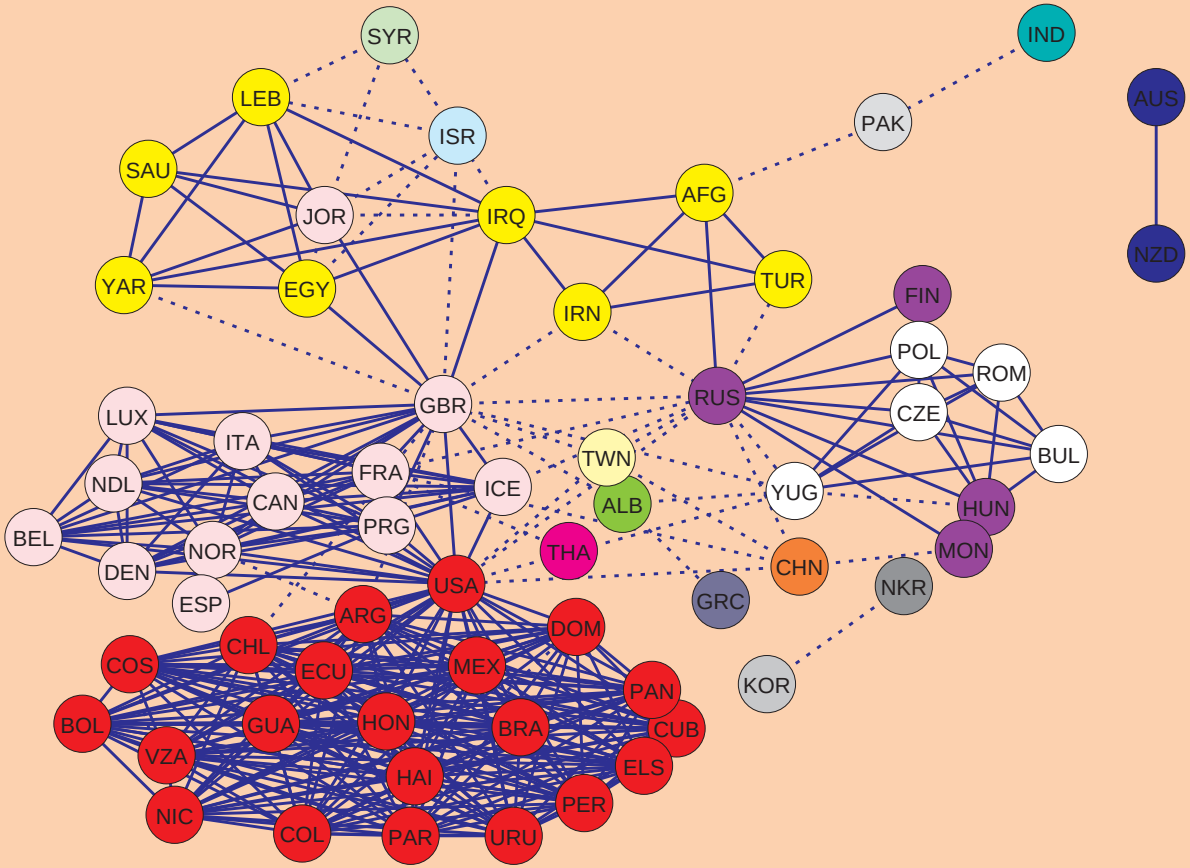
Featues of these networks

- The CoW data are signed with undirected ties. The FDI data are unsigned with directed ties.
- Great instability in both networks:
- The size of the network changes at each time point with units entering and/or departing.
- Many ties are created and/or dropped at each time point.
- For the Cow data: there was a general expansion of the system plus new conflicts and new alliances over time.
- For the FDI investments: there is no general need to keep making the same investments in the same infrastructure in the same places.
- Inductive blockmodeling was a *complete failure* for both networks.

Signed Data: The structural balance failed partition



Signed Data: Community detection's failed partition



The nature of the failures: CoW data

Structural balance

This partition fails because it clusters all states in the Western Bloc with nations from South America due to their positive links to the USA. The resulting positive block is far too large and is not reflective of a real divide between these two clusters of states. This holds, albeit to a lesser extent, for nations in the Middle East.

Communitiy detection

This CD fails in its handling of negative ties by producing a large number of singleton positions that make little structural sense. Its focus on modularity handles the SB problem regarding positive ties well but it cannot handle negative ties in the same way.

Both problems persist over all 51 windows (1946-1999) - neither set of partitions is acceptable.

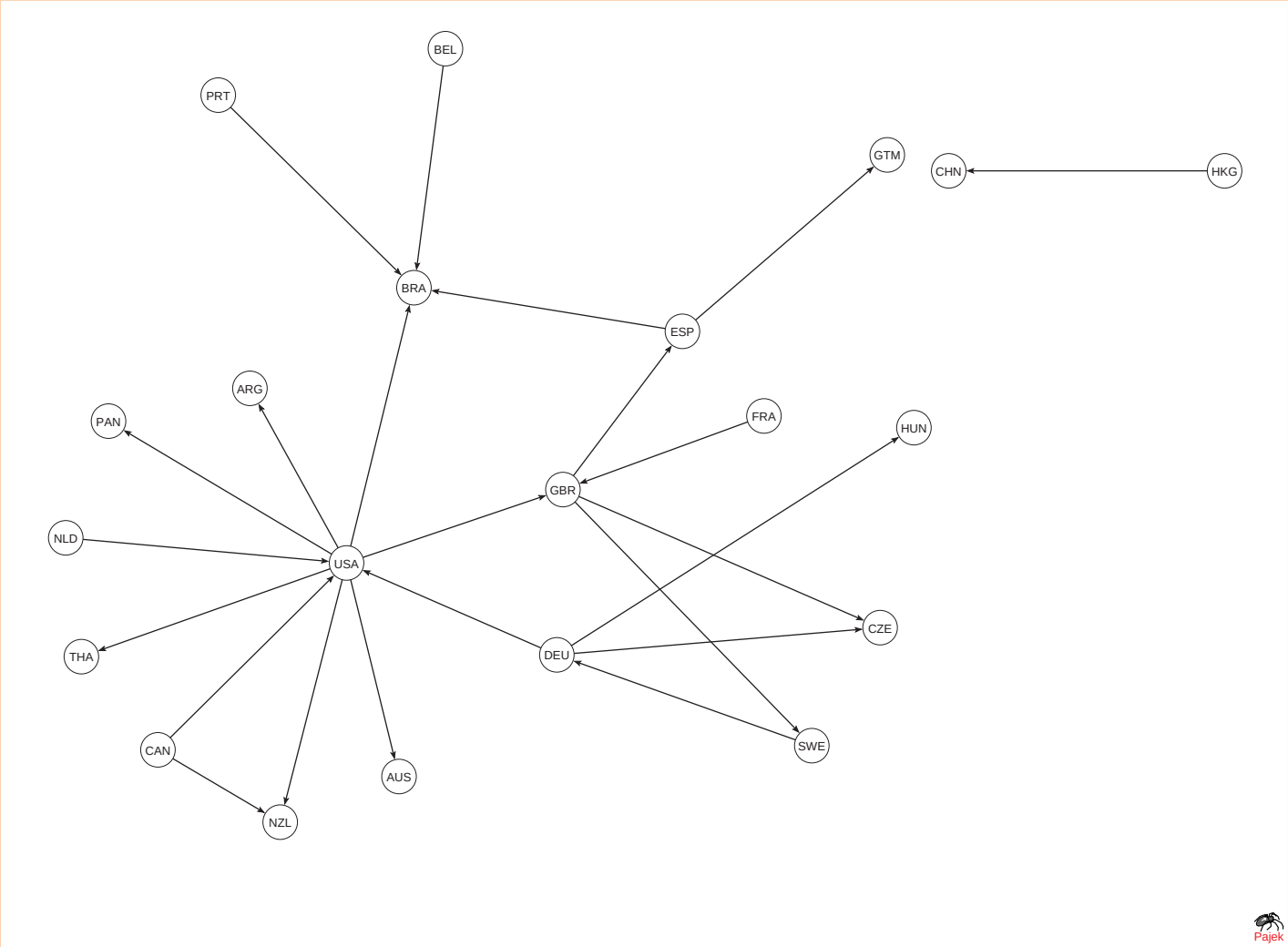
The nature of the failure: FDI data

- Investments are made by companies located in different countries.
- Each flow has a direction from the 'sending' country to the 'receiving' country
- Countries can be senders and/or receivers. Sometimes they are both in specific periods.
- This implies converting the seemingly 1-mode FDI flow data to 2-mode data. Treating these data as 1-mode data is worthless.
- Senders are one mode while receivers are in the other mode.
- The two modes are partitioned separately but at the same time.

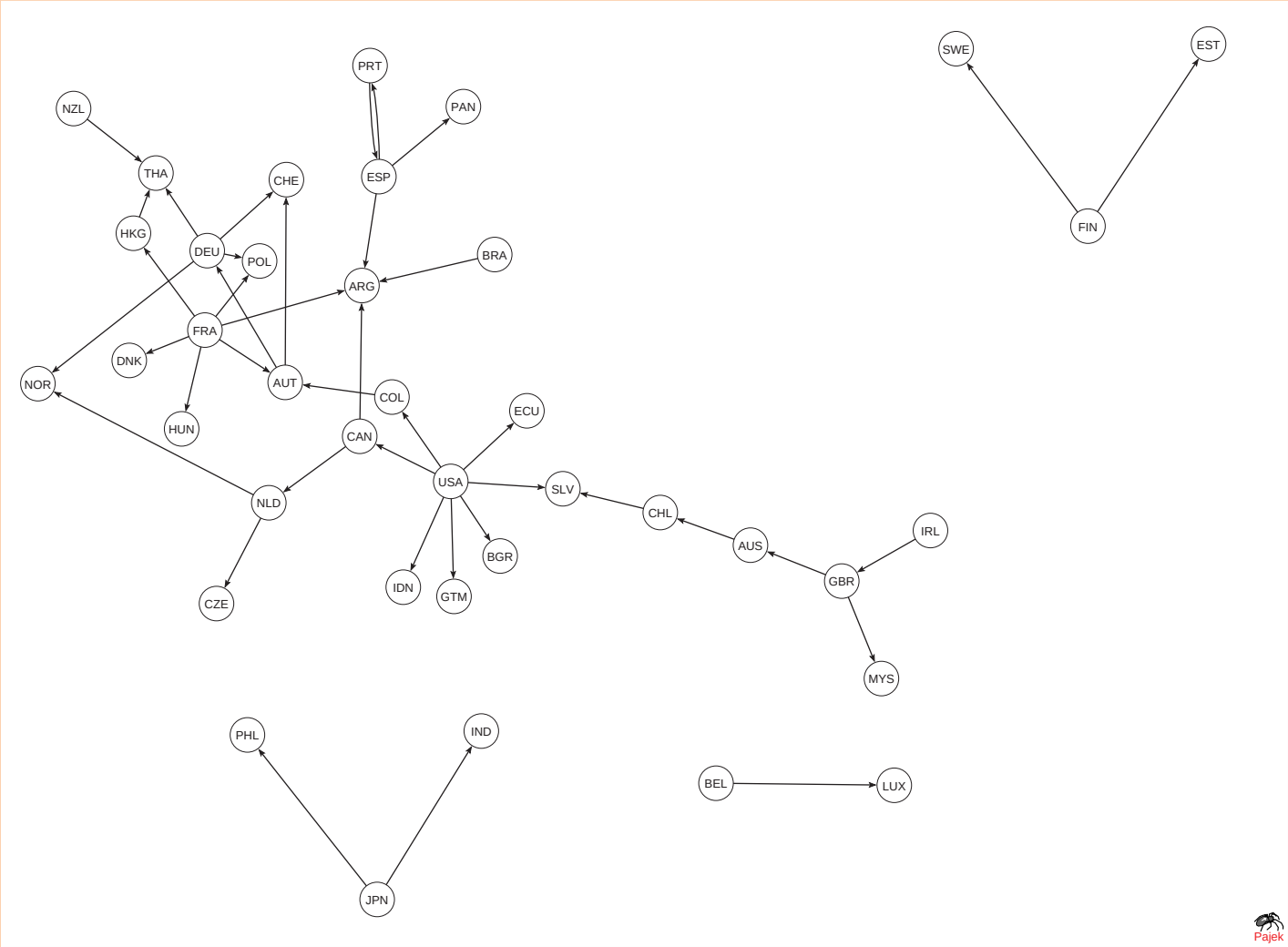
Change in the sizes of the CoW network

Time	Nations	P ties	N ties	All Ties
1946-1949	64	320	42	362
1953-1956	79	416	113	529
1966-1969	111	504	103	607
1976-1979	132	852	116	968
1990-1993	155	1160	128	1288
1996-1999	151	1100	147	1247

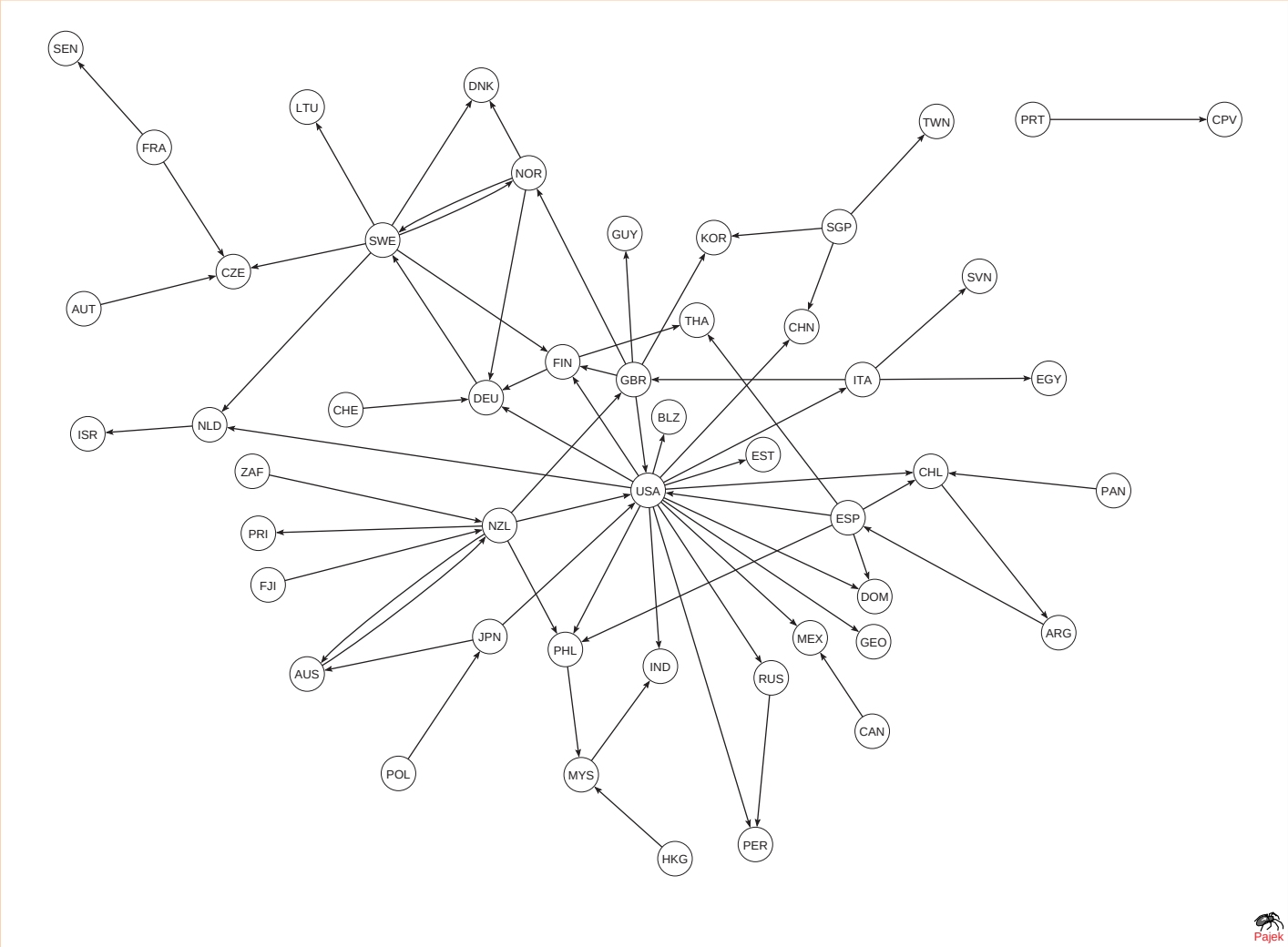
FDI network 1: ties remaining 1999



FDI network 2: ties vanishing 1999



FDI network 3: new ties 1999



Pre-specification as a potential solution

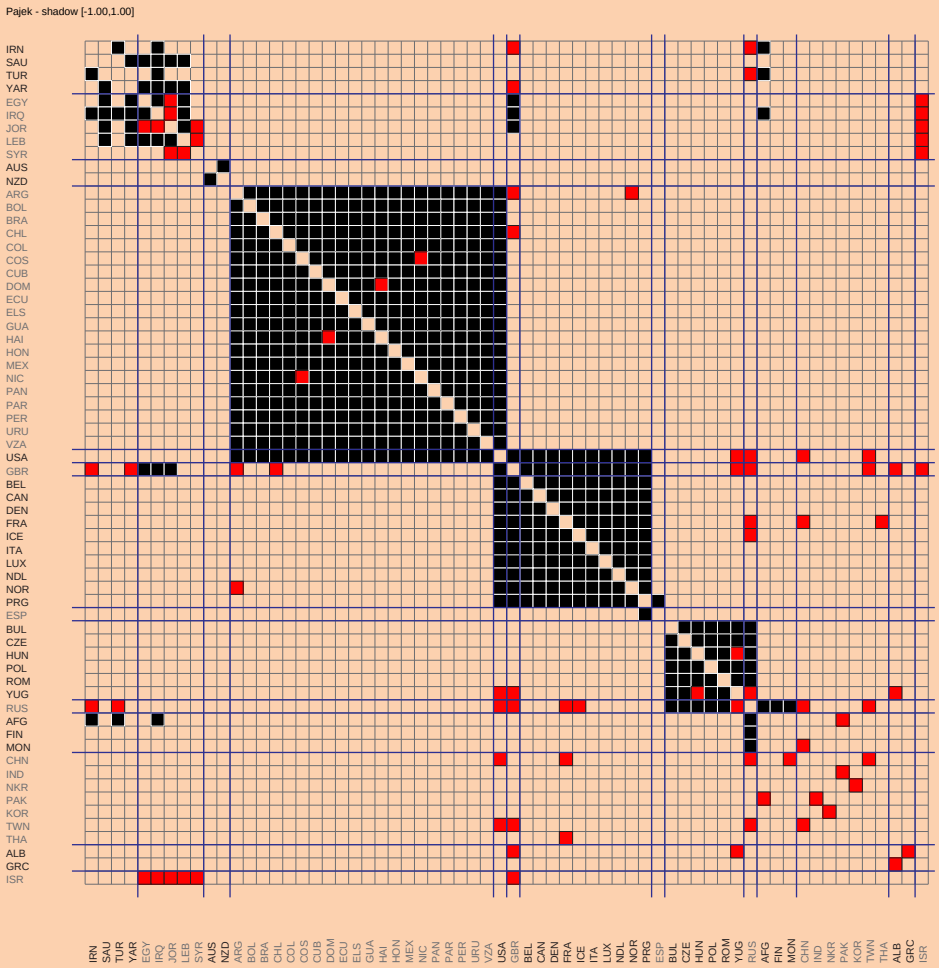
- Specify (or allow) singletons as postions.
  - For the CoW data: USA; USSR; Great Britain (after WWII); Israel and Yugoslavia (early - Tito v. Stalin) then late (following its breakup after 1991).
  - For the FDI data: having the USA and Germany in both modes as receivers and senders. Also allowing the Netherlands and Spain as senders plus Argentina and Chile as receivers.
- Specify position memberships;
- Specify the block types and for the CoW data and the complete blockmodel. For the FDI data use multiple block types. But . . .
- The first potentially huge downside is that this activity is very labor intensive.

Pre-specification for the CoW data: one time point

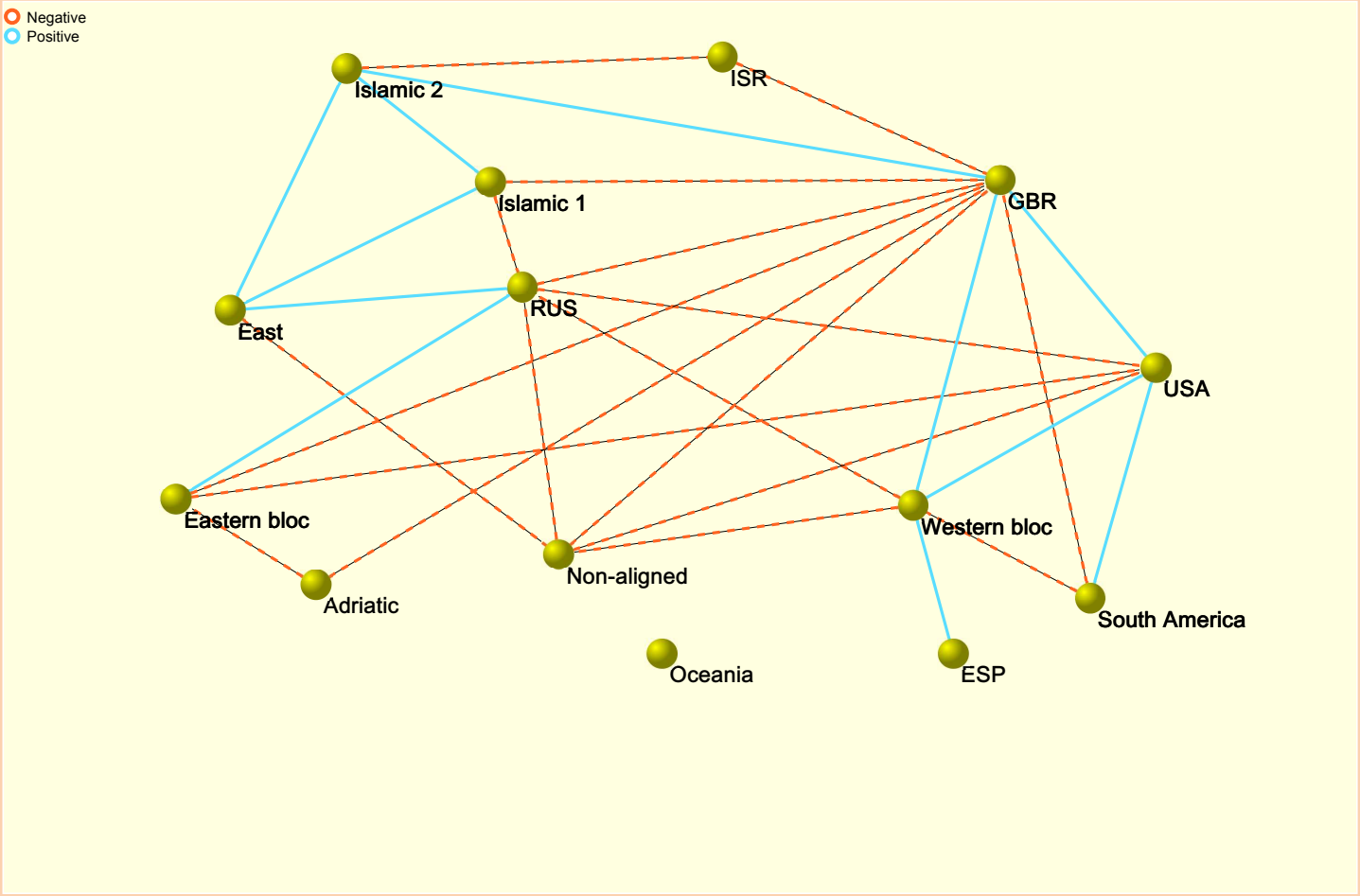
P	P	0	0	0	N	0	0	0	N	P	0	0	0
P	P	0	0	0	P	0	0	0	0	P	0	0	N
0	0	P	0	0	0	0	0	0	0	0	0	0	0
0	0	0	P	P	N	N	0	0	0	0	0	0	0
0	0	0	P	0	P	P	0	N	N	0	N	0	0
N	P	0	N	P	0	P	0	N	N	0	N	N	N
0	0	0	N	P	P	P	P	0	N	0	N	0	0
0	0	0	0	0	0	P	0	0	0	0	0	0	0
0	0	0	0	N	N	0	0	P	P	0	0	N	0
N	0	0	0	N	N	N	0	P	0	P	N	0	0
P	P	0	0	0	0	0	0	0	P	0	N	0	0
0	0	0	0	N	N	N	0	0	N	N	N	0	0
0	0	0	0	0	N	0	0	N	0	0	0	N	0
0	N	0	0	0	N	0	0	0	0	0	0	0	0

Note the presence of P (positive), N (negative) and null blocks (0) - both on and off the main diagonal.

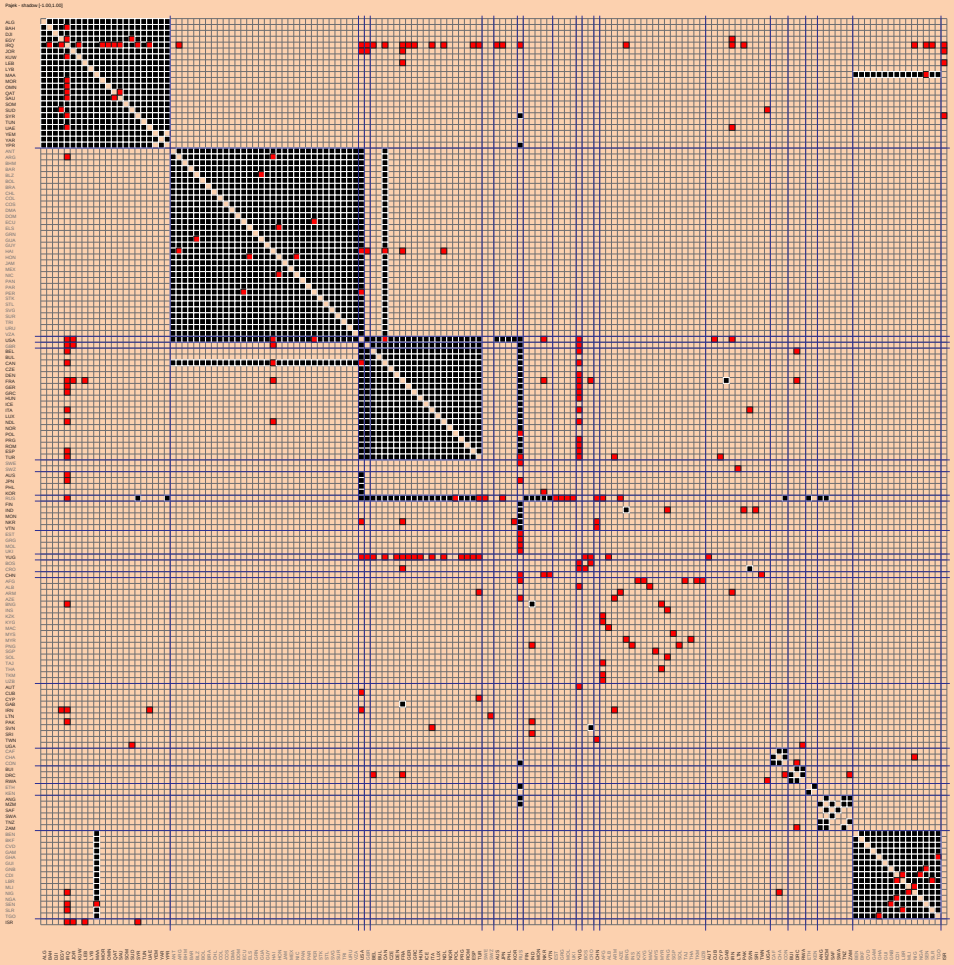
Results: CoW data blockmodel array 1946



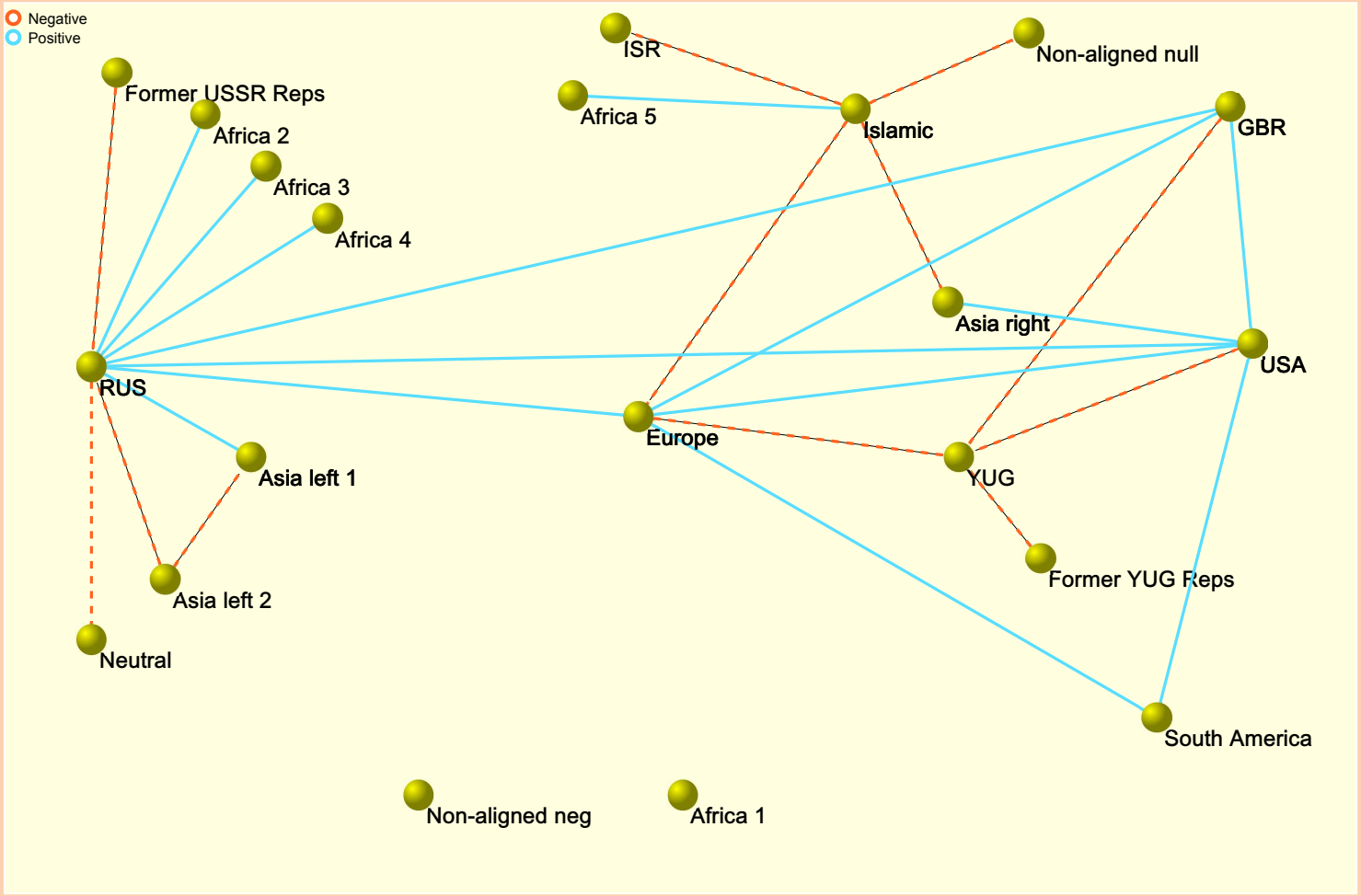
Results: CoW image 1946



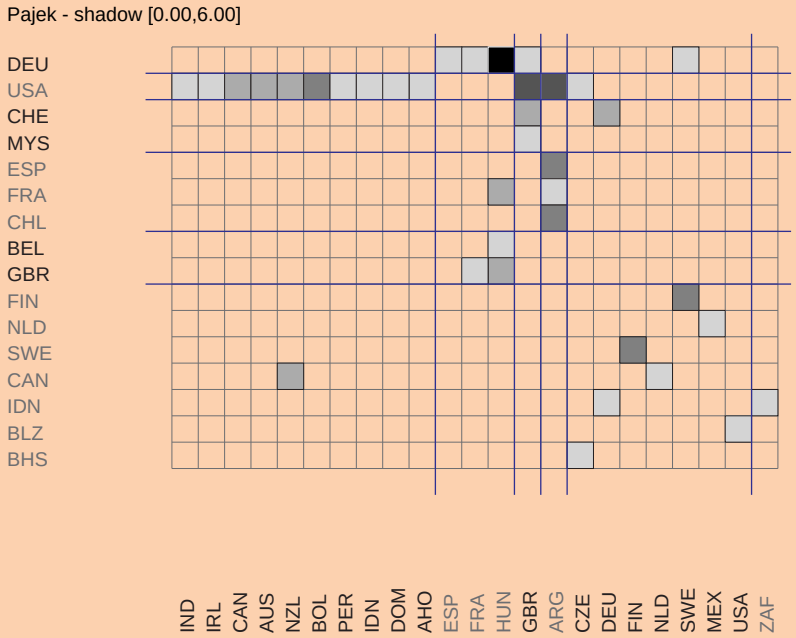
Results: CoW blockmodel array 1990



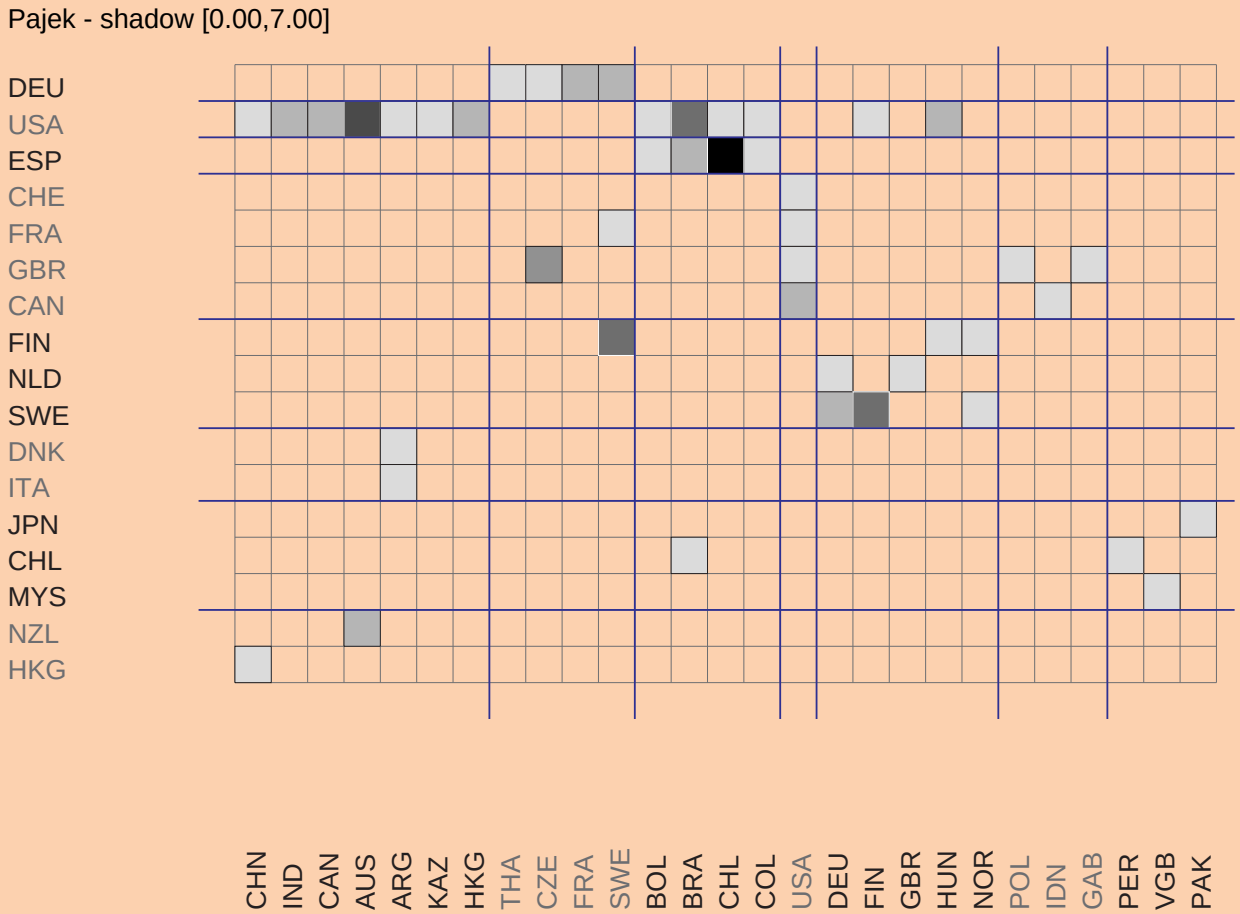
Results: CoW image 1990



Results: FDI blockmodel array 1995



Results:FDI data blockmodel array 1997



Provisional conclusions

- There are data arrays posing problems for discerning the structure of networks. (Maybe the majority of networks have this feature?)
- Unless we are *truly ignorant* about what we study, the inductive approach has *little or no* value.
- Utilizing the knowledge we have is far more preferable.
- This implies pre-specifying blockmodels. But . . .
  - It can be a very labor intensive process.
  - There is a risk of 'over-specifying' blockmodels.
- This is a risk I am more than willing to make - but with one exception:
- If the blockmodel is specified *solely* on substantive grounds, the data *cannot* be examined ahead of time.