Generalized blockmodeling of large valued networks

Carl Nordlund

carl.nordlund@liu.se

www.carlnordlund.net

LINKÖPING UNIVERSITY

Institute for Analytical Sociology

Roadmap to presentation

- Short intro to blockmodeling and equivalences
- Direct vs. indirect approaches: pros and cons
  - Indirect: ✓Large ✓Valued ✘Structural equivalence only
  - Direct: ✓All equivalences ✓/✘Valued ✘Small networks only
- Direct generalized blockmodeling for valued networks
- Hybrid/sequential indirect-direct approach

Work in progress

Role-analysis?

"3^{rd} pillar" of Social Network Analysis

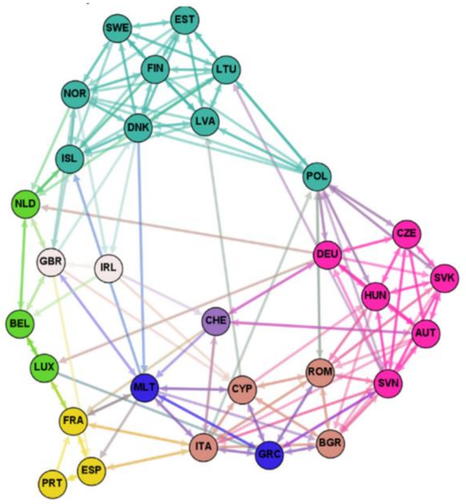
1^{st}: Node-centric properties and metrics (micro-level)
Measures and metrics that say something about individual actors, e.g. centrality measures

2^{nd}: Cohesive subgroups, triadic analysis etc (meso-level)
Measures and metrics relating to sets of actors (or more often: the sets themselves)

3^{rd}: Macro-level: measures and metrics that say something about the whole network
Density, average betweenness centrality (centralization), etc.
But more overarching and relevant: role analysis and blockmodeling

Indeed dealing with individual actors and subsets of actors, but the analysis and approach is about the network as a whole

Role-analysis



Identifying positions (aka clusters)

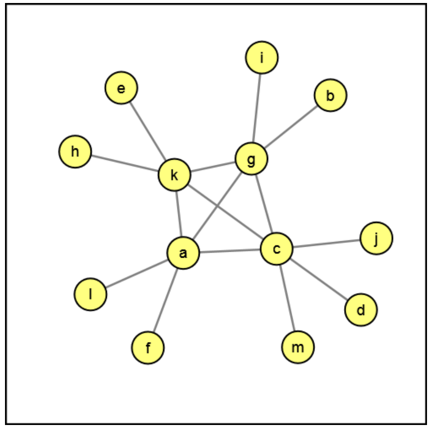
Partition the set of actors in a network into subsets (positions/clusters) based on a meaningful definition of equivalence

Mapping role-structure

Mapping overall relationships of actors within and between such positions

The quintessential technique for doing this: Blockmodeling!

Example: Galtung's feudal interaction structure



(Galtung 1971)

	a	b	c	d	e	f	g	h	l	j	k	l	m
a		0	1	0	0	1	1	0	0	0	1	1	0
b	0		0	0	0	0	1	0	0	0	0	0	0
c	1	0		1	0	0	1	0	0	1	1	0	1
d	0	0	1		0	0	0	0	0	0	0	0	0
e	0	0	0	0		0	0	0	0	0	1	0	0
f	1	0	0	0	0		0	0	0	0	0	0	0
g	1	1	1	0	0	0		0	1	0	1	0	0
h	0	0	0	0	0	0	0		0	0	1	0	0
l	0	0	0	0	0	0	1	0		0	0	0	0
j	0	0	1	0	0	0	0	0	0		0	0	0
k	1	0	1	0	1	0	1	1	0	0		0	0
l	1	0	0	0	0	0	0	0	0	0	0		0
m	0	0	1	0	0	0	0	0	0	0	0	0	

Sociomatrix representation

Intuitive understanding of this network

A set of 4 "central" and connected actors in the middle. Remaining actors hanging on to these, otherwise poorly connected.

So let's hypothesize that these two represent two different positions:

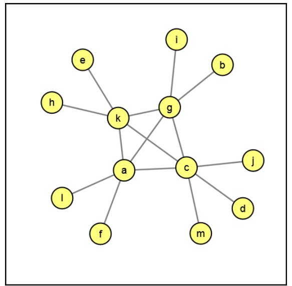
P1={a,c,g,k}
P2={b,d,e,f,h,i,j,l,m}

We sort the original matrix according to these positions

		P1				P2								
		a	g	k	c	l	f	l	b	e	h	m	j	d
P1	a		1	1	1	1	1	0	0	0	0	0	0	0
	g	1		1	1	0	0	1	1	0	0	0	0	0
	k	1	1		1	0	0	0	0	1	1	0	0	0
	c	1	1	1		0	0	0	0	0	0	1	1	1
P2	l	1	0	0	0		0	0	0	0	0	0	0	0
	f	1	0	0	0	0		0	0	0	0	0	0	0
	l	0	1	0	0	0	0		0	0	0	0	0	0
	b	0	1	0	0	0	0	0		0	0	0	0	0
	e	0	0	1	0	0	0	0	0		0	0	0	0
	h	0	0	1	0	0	0	0	0	0		0	0	0
	m	0	0	0	1	0	0	0	0	0	0		0	0
	j	0	0	0	1	0	0	0	0	0	0	0		0
	d	0	0	0	1	0	0	0	0	0	0	0	0	

Galtung’s feudal interaction structure

A hypothetical blockmodel of the Galtung network



		P1				P2								
		a	g	k	c	l	f	l	b	e	h	m	j	d
P1	a		1	1	1	1	1	0	0	0	0	0	0	0
	g	1		1	1	0	0	1	1	0	0	0	0	0
	k	1	1		1	0	0	0	0	1	1	0	0	0
	c	1	1	1		0	0	0	0	0	0	1	1	1
P2	l	1	0	0	0		0	0	0	0	0	0	0	0
	f	1	0	0	0	0		0	0	0	0	0	0	0
	l	0	1	0	0	0	0		0	0	0	0	0	0
	b	0	1	0	0	0	0	0		0	0	0	0	0
	e	0	0	1	0	0	0	0	0		0	0	0	0
	h	0	0	1	0	0	0	0	0	0		0	0	0
	m	0	0	0	1	0	0	0	0	0	0		0	0
	j	0	0	0	1	0	0	0	0	0	0	0		0
	d	0	0	0	1	0	0	0	0	0	0	0	0	

I.e. a specific partition of the actors (with the sociomatrix ordered according this this partition)

Two (2) positions, four (2x2) blocks

Having sorted the original sociomatrix according to our 2 positions (P1 and P2), a total of four (2x2) blocks can be delineated, capturing the ties within and between actors of the positions.

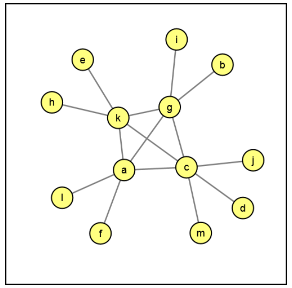
Blocks: P1-P1, P1-P2, P2-P1, P2-P2

In this state, we observe certain patterns in these blocks!

P1-P1: completely filled with ties (except for the diagonal)
P2-P2: totally missing ties
P1-P2, P2-p1: not empty, not filled, something in-between (with ties on each row and each column of the blocks)

Galtung’s feudal interaction structure

A hypothetical blockmodel of the Galtung network

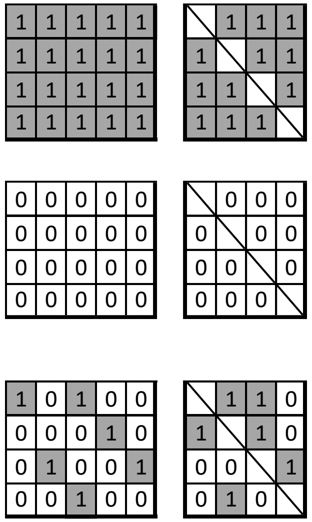


		P1				P2								
		a	g	k	c	l	f	l	b	e	h	m	j	d
P1	a		1	1	1	1	1	0	0	0	0	0	0	0
	g	1		1	1	0	0	1	1	0	0	0	0	0
	k	1	1		1	0	0	0	0	1	1	0	0	0
	c	1	1	1		0	0	0	0	0	0	1	1	1
P2	l	1	0	0	0		0	0	0	0	0	0	0	0
	f	1	0	0	0	0		0	0	0	0	0	0	0
	l	0	1	0	0	0	0		0	0	0	0	0	0
	b	0	1	0	0	0	0	0		0	0	0	0	0
	e	0	0	1	0	0	0	0	0		0	0	0	0
	h	0	0	1	0	0	0	0	0	0		0	0	0
	m	0	0	0	1	0	0	0	0	0	0		0	0
	j	0	0	0	1	0	0	0	0	0	0	0		0
	d	0	0	0	1	0	0	0	0	0	0	0	0	

Interpreting the blocks

Turns out that the blocks we identify in this hypothetical blockmodel represent three ideal blocks in blockmodeling:

Structural equivalence blockmodeling



The complete block (com)

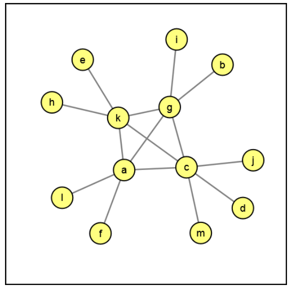
The null block (nul)

The regular block (reg)

Regular equivalence blockmodeling

Galtung's feudal interaction structure

A hypothetical blockmodel of the Galtung network



		P1				P2								
		a	g	k	c	l	f	l	b	e	h	m	j	d
P1	a		1	1	1	1	1	0	0	0	0	0	0	0
	g	1		1	1	0	0	1	1	0	0	0	0	0
	k	1	1		1	0	0	0	0	1	1	0	0	0
	c	1	1	1		0	0	0	0	0	0	1	1	1
P2	l	1	0	0	0		0	0	0	0	0	0	0	0
	f	1	0	0	0	0		0	0	0	0	0	0	0
	l	0	1	0	0	0	0		0	0	0	0	0	0
	b	0	1	0	0	0	0	0		0	0	0	0	0
	e	0	0	1	0	0	0	0	0		0	0	0	0
	h	0	0	1	0	0	0	0	0	0		0	0	0
	m	0	0	0	1	0	0	0	0	0	0		0	0
	j	0	0	0	1	0	0	0	0	0	0	0		0
	d	0	0	0	1	0	0	0	0	0	0	0	0	

Collapsing the blockmodel

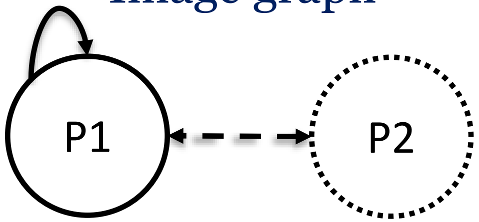
Having modeled each empirical block in terms of an ideal block, we can now "collapse" the blockmodel into a block image, or an image graph.

Here, nodes represent positions (subsets of equivalent actors)

Block image

	P1	P2
P1	com	reg
P2	reg	nul

Image graph



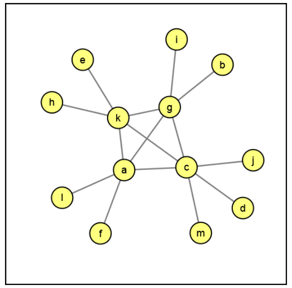
From this: roles

Role of P1 is to be connected other in P1, and having sporadic ties to and from P2

Role of P2 is to have sporadic ties with P1, but lack ties to other in P2

Galtung’s feudal interaction structure

But why do we do blockmodeling?



		P1				P2								
		a	g	k	c	l	f	l	b	e	h	m	j	d
P1	a		1	1	1	1	1	0	0	0	0	0	0	0
	g	1		1	1	0	0	1	1	0	0	0	0	0
	k	1	1		1	0	0	0	0	1	1	0	0	0
	c	1	1	1		0	0	0	0	0	0	1	1	1
P2	l	1	0	0	0		0	0	0	0	0	0	0	0
	f	1	0	0	0	0		0	0	0	0	0	0	0
	l	0	1	0	0	0	0		0	0	0	0	0	0
	b	0	1	0	0	0	0	0		0	0	0	0	0
	e	0	0	1	0	0	0	0	0		0	0	0	0
	h	0	0	1	0	0	0	0	0	0		0	0	0
	m	0	0	0	1	0	0	0	0	0	0		0	0
	j	0	0	0	1	0	0	0	0	0	0	0		0
	d	0	0	0	1	0	0	0	0	0	0	0	0	

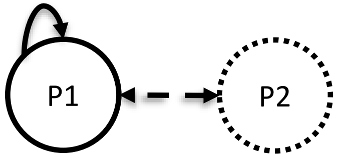
Comparisons with structural types

The Galtung network turned out to be a core-periphery structure

…but there are a multitude of other possibilities: hierarchies, transitive structures, cohesive subgroups.

Community structures: a special case (one possible outcome) of blockmodeling!

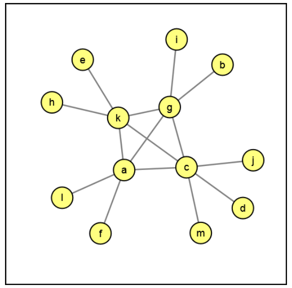
	P1	P2
P1	com	reg
P2	reg	nul



	P1	P2	P3
P1	com	nul	nul
P2	nul	com	nul
P3	nul	nul	com

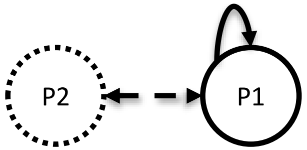
Galtung’s feudal interaction structure

But why do we do blockmodeling?



		P1				P2								
		a	g	k	c	l	f	l	b	e	h	m	j	d
P1	a		1	1	1	1	1	0	0	0	0	0	0	0
	g	1		1	1	0	0	1	1	0	0	0	0	0
	k	1	1		1	0	0	0	0	1	1	0	0	0
	c	1	1	1		0	0	0	0	0	0	1	1	1
P2	l	1	0	0	0		0	0	0	0	0	0	0	0
	f	1	0	0	0	0		0	0	0	0	0	0	0
	l	0	1	0	0	0	0		0	0	0	0	0	0
	b	0	1	0	0	0	0	0		0	0	0	0	0
	e	0	0	1	0	0	0	0	0		0	0	0	0
	h	0	0	1	0	0	0	0	0	0		0	0	0
	m	0	0	0	1	0	0	0	0	0	0		0	0
	j	0	0	0	1	0	0	0	0	0	0	0		0
	d	0	0	0	1	0	0	0	0	0	0	0	0	

	P1	P2
P1	com	reg
P2	reg	nul



Reduction of complexity

A blockmodel reduces a complex network into its fundamental structure. We tease out the “functional anatomy”.

In the Galtung network, we find a blockimage that corresponds to a classical core-periphery structure.

For a network twice (or 100 times) the size, the same block image could appear, with different relative sizes of P1 and P2. Still the same underlying functional anatomy.

Allows for comparing complex networks: finding similarities in what could appear as dissimilar

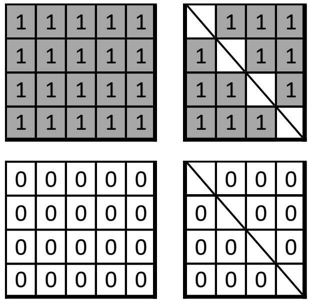
Types of equivalences

What is meant by equivalence in a network? Different notions and usages exist.

Structural equivalence

Two actors are structurally equivalent if they have identical sets of ties to the other actors.

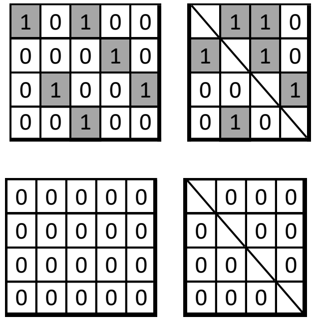
Under this notion of equivalence, blockmodels would ideally consist of complete and null blocks



Regular equivalence

Two actors are regular equivalent if they have ties to actors that in turn are also regular equivalent (!?)

Under this notion of equivalence, blockmodels would ideally consist of null blocks and regular blocks (at least one tie in each row, and at least one tie in each column)

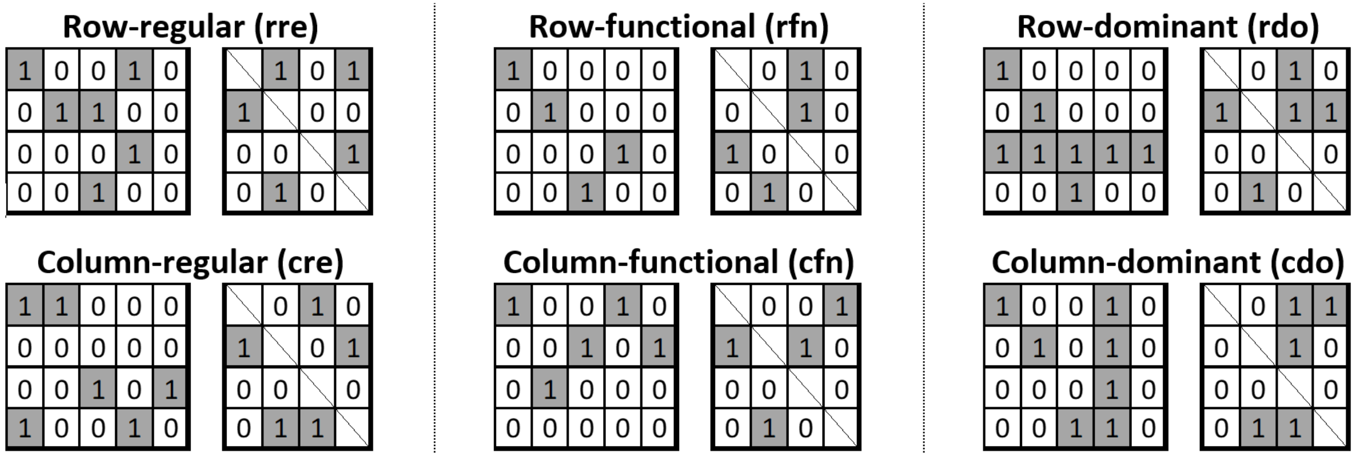


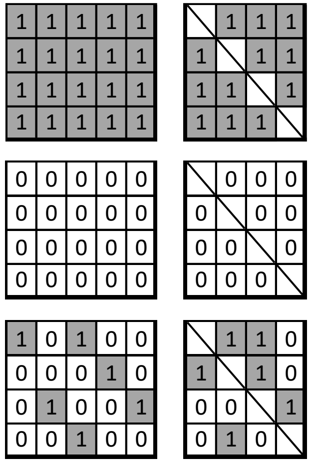
Types of equivalences

What is meant by equivalence in a network? Different notions and usages exist.

Generalized equivalence

In generalized blockmodeling, ideal blocks consists of null, complete and regular blocks – and a host of other block types:





Finding positions

There are two broad approaches for finding equivalent partitions in a network: direct and indirect methods.

Actually three ways: hypothesis testing as well

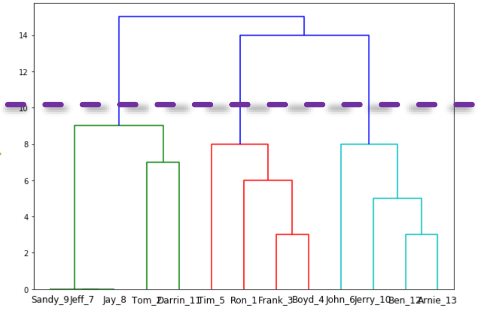
Indirect blockmodeling

First, calculate/determine a suitable indirect measure of equivalence. How these are calculated depend on the notion of equivalence we are interested in.

Apply hierarchical clustering to this matrix, Choose suitable number of positions, Create blockmodel

	Ron_1	Frank_3	Boyd_4	Tim_5	Tom_2	Darrin_11	John_6	Ben_12	Arnie_13	Jeff_7	Jay_8	Sandy_9	Jerry_10
Ron_1		1	1	1	0	0	0	0	0	0	0	0	0
Frank_3	1		1	0	0	1	0	0	0	0	0	0	0
Boyd_4	1	1		0	1	0	0	0	0	0	0	0	0
Tim_5	1	1	1		0	0	0	0	0	0	0	0	0
Tom_2	1	1	0	0		1	0	0	0	0	0	0	0
Darrin_11	1	0	0	0	1		0	0	0	0	0	0	1
John_6	0	0	0	1	0	0		1	1	0	0	0	0
Ben_12	1	0	0	0	0	0	1		0	0	0	0	1
Arnie_13	0	0	0	1	0	0	1	0		0	0	0	0
Jeff_7	0	0	0	0	1	0	0	0	0		1	1	0
Jay_8	0	0	0	0	1	0	0	0	0	1		1	0
Sandy_9	0	0	0	0	1	0	0	0	0	1	1		0
Jerry_10	1	0	0	0	0	0	1	0	0	0	0	0	

dist	Ron_1	Frank_3	Boyd_4	Tim_5	Tom_2	Darrin_11	John_6	Ben_12	Arnie_13	Jeff_7	Jay_8	Sandy_9	Jerry_10
Ron_1	0	5	6	8	12	11	10	14	11	15	15	15	10
Frank_3	5	0	3	7	9	8	13	9	10	12	12	12	9
Boyd_4	6	3	0	6	10	5	12	8	9	9	9	9	8
Tim_5	8	7	6	0	10	9	10	6	7	11	11	11	8
Tom_2	12	9	10	10	0	7	14	10	11	9	9	9	8
Darrin_11	11	8	5	9	7	0	11	5	8	8	8	8	7
John_6	10	13	12	10	14	11	0	6	3	11	11	11	8
Ben_12	14	9	8	6	10	5	6	0	3	9	9	9	4
Arnie_13	11	10	9	7	11	8	3	3	0	8	8	8	5
Jeff_7	15	12	9	11	9	8	11	9	8	0	0	0	9
Jay_8	15	12	9	11	9	8	11	9	8	0	0	0	9
Sandy_9	15	12	9	11	9	8	11	9	8	0	0	0	9
Jerry_10	10	9	8	8	8	7	8	4	5	9	9	9	0



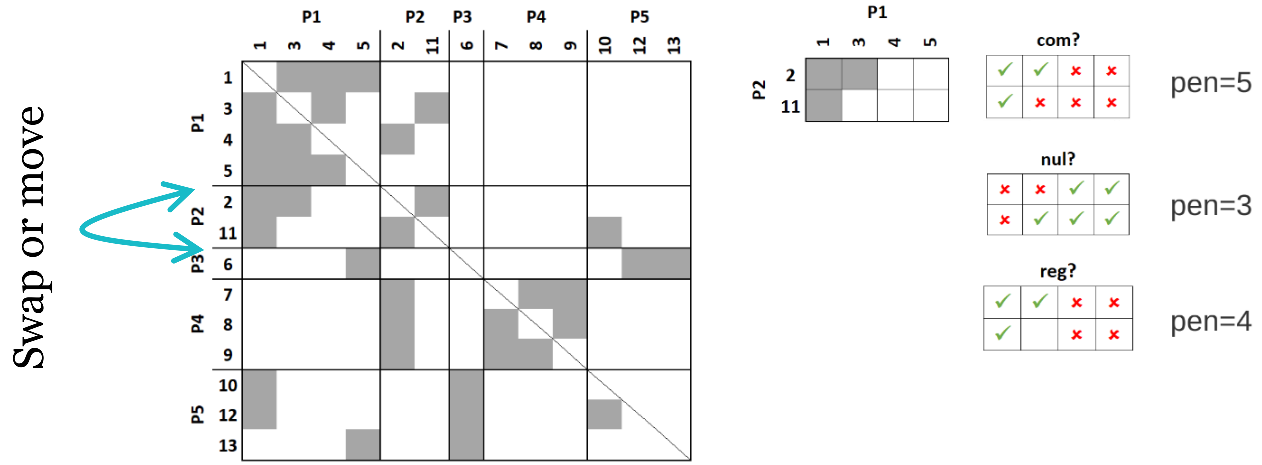
Advantage:
Works for large networks
Works for valued networks

Disadvantage:
Restricted to structural equivalence
Regular equivalence: sort of (REGE)
Generalized blockmodeling: nope!
Lots of choices in the analysis

Finding positions

Direct blockmodeling (semi-supervised)

Given a random starting partition, given number of positions, and a set of ideal blocks, the direct approach uses genetic algorithms (local optimization) to move/swap actors between positions. A criteria function is used (e.g. number of inconsistencies), testing among the various types of ideal blocks.



Advantage:
Deemed better
Allows for capturing complicated ideal blocks (generalized blockmodeling)

Disadvantage:
Computationally heavy
Potentially finding local minima
More suited for binary networks

Finding positions

Direct blockmodeling for valued networks

Dilemma with direct blockmodeling: ideal binary blocks vs. valued data: how to count nbr of inconsistencies?

	1	3	4	5
2	17	2	0	4
11	3	817	41	0

1	1	1	1
1	1	1	1

?

0	0	0	0
0	0	0	0

Existing workarounds

- Dichotomization of valued networks
    - Data reduction; choosing threshold etc.
- Introducing novel (ideal) types
    - Valued blockmodeling (Ziberna 2007)
    - Novel ways to identify regular blocks (Nordlund 2007)
    - Yet both of these still apply thresholds of some kind

Finding positions

Direct blockmodeling for valued networks

Proposed solution for direct generalized blockmodeling of valued networks (Nordlund 2020)

Replace the goodness-of-fit function!

In conventional direct blockmodeling: counting the number of inconsistencies between ideal and empirical blocks (Hamming distance), minimize nbr of inconsistencies

Inspired by the classical core-periphery metric and heuristic of Borgatti & Everett (1999) and extension of this in Nordlund (2018):

Goodness-of-fit measure using weighted correlation coefficient

The B-E core-periphery approach

	Core	Peri
Core	com	(dnc)
Peri	(dnc)	nul

		C			P		
		a	b	c	d	e	f
C	a		1	1	0	1	1
	b	0		1	0	0	0
	c	1	1		0	0	1
P	d	0	1	1		0	0
	e	1	0	0	0		0
	f	0	0	1	0	1	

Block	C->C (Complete)						P->P (Null)					
Cell	a->b	a->c	b->a	b->c	c->a	c->b	d->e	d->f	e->d	e->f	f->d	f->e
Observed (X)	1	1	0	1	1	1	0	0	0	0	0	1
Ideal (Y)	1	1	1	1	1	1	0	0	0	0	0	0

Nbr of inconsistencies: 2 corr(X,Y): 0.67

- Y: ideal (binary), X: observed (binary, valued)
- Standardized measure of fit
- Point-biserial correlation: slight bias towards 50% dense blockmodels

Finding positions

Direct generalized blockmodeling for valued networks (Nordlund 2020)

Borgatti-Everett c-p approach: direct SE blockmodeling with fixed blockimage

	Core	Peri
Core	com	(dnc)
Peri	(dnc)	nul

Generalize this: create a list of value-triplets (L) representing observed (X) and ideal (Y) values, and a weight (W) attached to each of these (x,y) pairs

For a given partition and hypothetical block image, populate this list with value-triplets as defined by the empirical and ideal blocks

$f_{COMPLETE}(A, P_i, P_j)$:	$f_{NULL}(A, P_i, P_j)$:

$L_{i,j}$={ }
foreach r in P_i:
  foreach c in P_j:
    if r<>c:
      $L_{i,j} \leftarrow (a_{r,c}; 1; 1)$
return $L_{i,j}$
	$L_{i,j}$={ }
	foreach r in P_i:
	  foreach c in P_j:
	    if r<>c:
	      $L_{i,j} \leftarrow (a_{r,c}; 0; 1)$
	return $L_{i,j}$

Compete and Null blocks

Similar to BE c-p approach: one data point per cell (each weighted by 1)

Finding positions

Direct generalized blockmodeling for valued networks (Nordlund 2020)

$f_{REGULAR}(A, P_i, P_j)$:

```
Li,j={ }
w=(|Pi|·(|Pj| - [i=j])) / (|Pi| + |Pj|)
foreach r in Pi:
  Li,j ←(maxc ar,c; 1; w)
foreach c in Pj:
  Li,j ←(maxr ar,c; 1; w)
return Li,j
```

$f_{ROW\text{-}REGULAR}(A, P_i, P_j)$:

```
Li,j={ }
w=|Pj| - [i=j]
foreach r in Pi:
  Li,j ←(maxc ar,c; 1; w)
return Li,j
```

$f_{COLUMN\text{-}REGULAR}(A, P_i, P_j)$:

```
Li,j={ }
w=|Pi| - [i=j]
foreach c in Pj:
  Li,j ←(maxr ar,c; 1; w)
return Li,j
```

```
  if max_r = 0:
    Li,j ←(0; 1; w)
  else:
    index = -1
    foreach c in Pj:
      if ar,c = max_r:
        index_c = c
    foreach c in Pj:
      if c = index_c:
        Li,j ←(ar,c; 1; 1)
      else:
        Li,j ←(ar,c; 0; 1)
return Li,j
```

```
  if max_c = 0:
    Li,j ←(0; 1; w)
  else:
    index = -1
    foreach r in Pi:
      if ar,c = max_c:
        index_r = r
    foreach r in Pi:
      if r = index_r:
        Li,j ←(ar,c; 1; 1)
      else:
        Li,j ←(ar,c; 0; 1)
return Li,j
```

Regular blocks

Get max value in each row and column, respectively; correlate with unity

Weight each data point by block size! (w_i != 1)

Row/col-regular, functional etc

Find characteristic patterns, weigh data points so this segment represents block size (w_i != 1)

Finding positions

Direct generalized blockmodeling for valued networks (Nordlund 2020)

Get List L of value triplets: calculate the weighted correlation coefficient (WCC) formula; use as goodness-of-fit score

$$r_w = \frac{\sum w_i (x_i - \bar{x})(y_i - \bar{y})}{\sum w_i \sqrt{\sum w_i (x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

Advantages

- Direct blockmodeling of binary and valued networks (no special 'valued' approach)
- Using standard set of ideal (binary) blocks in all forms of equivalences (SE, RE, generalized)
- No dichotomization of valued networks (but could be transformed)
- Slight bias towards solutions whose block densities lean towards 0.5 (where non-null blocks are deemed as 'filled', i.e. like complete blocks); distinguish between solutions that in Hamming approach would be equally fitting
- Comparable goodness-of-fit measures, across sizes, valued/binary, types of equivalences etc

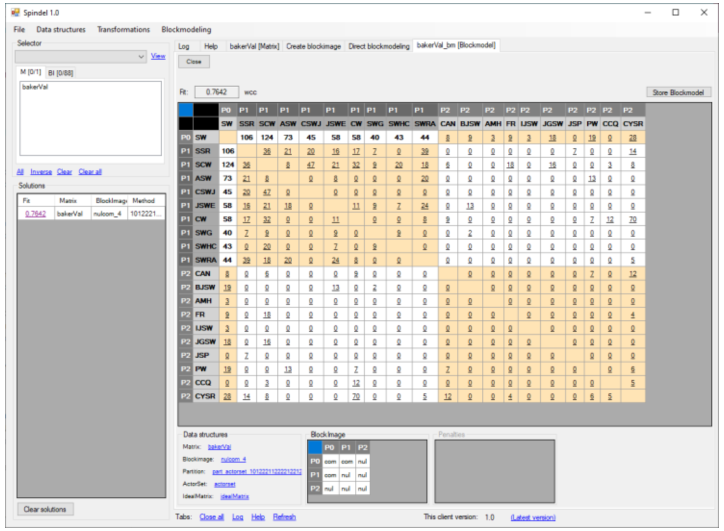
Disadvantages

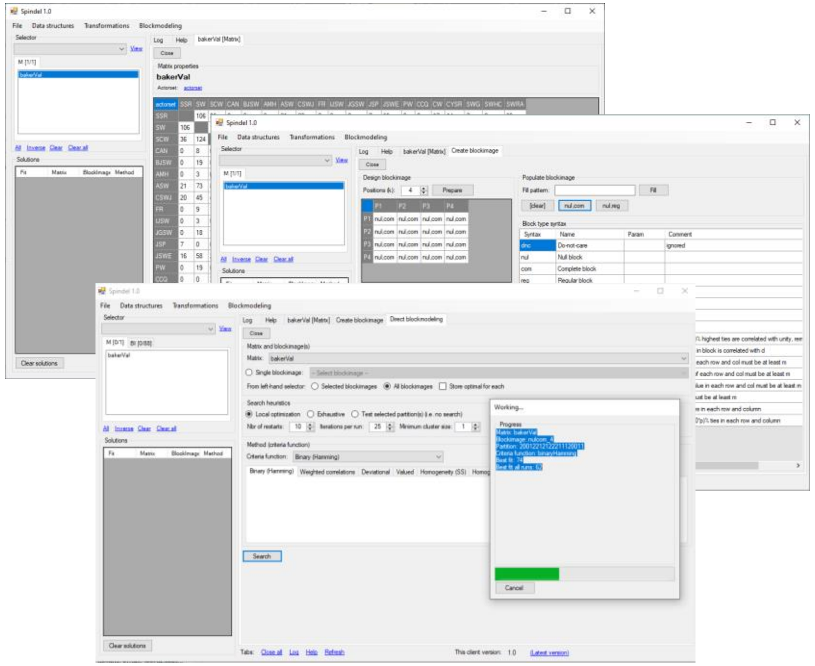
- In conventional (Hamming-based) inconsistency-counting: fit one block at a time
    - In WCC approach: need to calculate correlation for whole blockmodel (as involving mean values)
- This implies: far more computationally intensive; test multiple hypothetical block images at once

Spindel (‘spider’ in Swedish)

Software client for direct generalized valued and binary blockmodeling

- Equivalences: structural, regular, generalized, Ziberna-blocks, power-relational blocks
- Goodness-of-fit: Hamming, WCC, Ziberna-valued, homogeneity (SS, AD)
- Generate (non-isomorphic) blockimages (need to fit each individually)





https://demesta.com/spindel/

Weighted correlation coefficient (WCC) approach

- Might solve some existing dilemmas: no dichotomization, standardized GoF, standard set of ideal blocks, all equivalence types
- Still: **restricted to smaller networks**
  - …and perhaps more restricted than the conventional direct (Hamming-based) approach due to need to test all potential (non-isomorphic) block images!

How to expand this to larger networks?

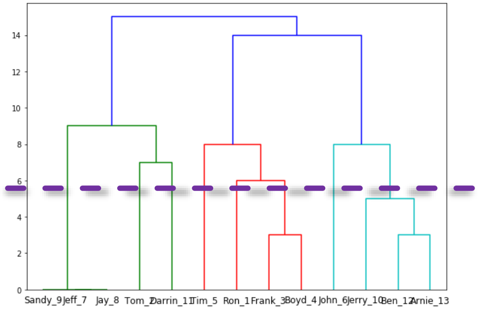
Sequential indirect-direct approach

Sequential indirect-direct approach

Work in progress

1st step: indirect structural equivalence

- Based on a suitable indirect structural equivalence metric, identify positions in a large network based on structural equivalence
- Necessary to be fairly ‘strict’: preferably find a partition where the blocks are close to the ideal (complete and null) blocks
- Partition the network on the basis of chosen partition; create the resulting image graph
- Document/store the goodness-of-fit of *each individual block* in this indirectly derived blockmodel



2nd step: direct blockmodeling using a modification of the WCC approach

- View the image graph from the 1st (indirect) step as a network!
  - Directional, self-ties, valued (block-specific goodness-of-fit)
- Use this collapsed network as input to a modified version of the weighted correlation-coefficient-based direct approach (Nordlund 2020)
  - Particularly: utilize the block-specific goodness-of-fit in the fitting function
  - Adjust (weigh) the significance of fitting 2nd stage blocks according to ‘sharpness’ of 1st-stage block densities

Sequential indirect-direct approach

Work in progress

Will this work?

- First: not all large networks can be reduced to ‘sharp’ structural equivalence blockmodels
    - Could be many networks where this approach won’t work
    - The image graph might still end up being too large for direct blockmodeling
- Might need to modify the Pearson approach as indirect measure of structural equivalence

Way forward

- Modify the direct WCC-based blockmodeling approach
    - Take self-ties into account (communities observed in 1st step)
    - Integrate block-wise fitting scores from the indirect 1st step in the direct BM step
- (Potential modification of indirect formula)
- Interpretation after 2nd step
    - Partition from direct approach: 2nd level partition
    - Expand to original network size
    - Additional tests for goodness of fit? Minor tweaks? Singleton positions?

Thank you!

Linköping University

Institute for Analytical Sociology

analytical.sociology@liu.se

www.liu.se/ias

www.liu.se

LINKÖPING UNIVERSITY