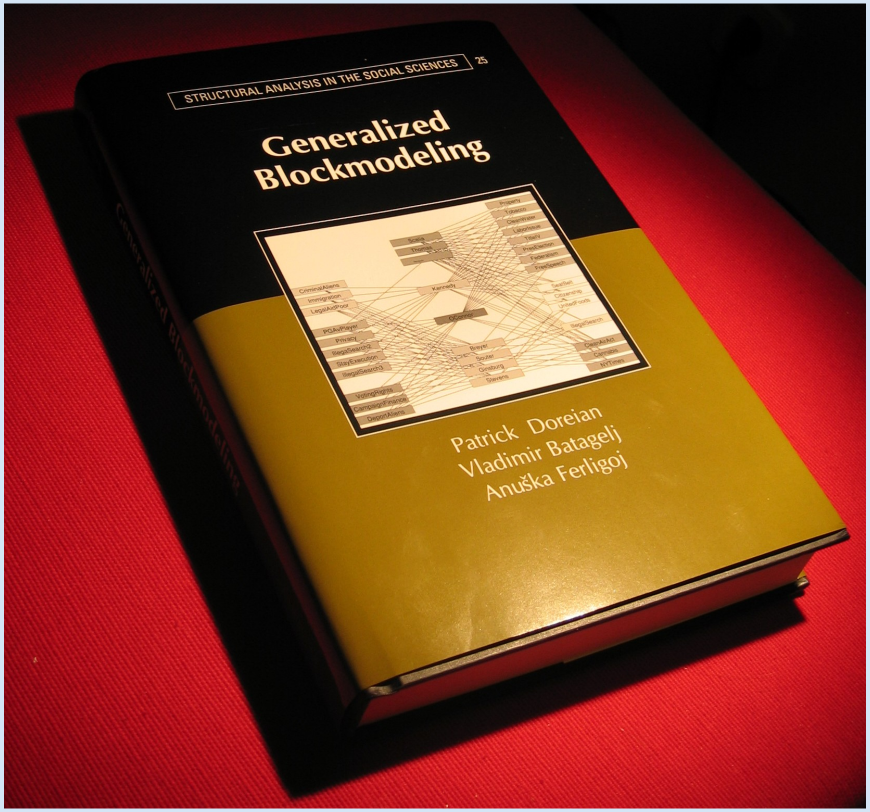
Advances in Generalized Blockmodeling
Pat Doreian

- V. Batagelj, P. Doreian, A. Ferligoj, A. Mrvar:
  Generalized Blockmodeling
- P. Doreian, Hyung Sam Park:
  Networks of Environmental Social Movement Organizations: The Turning Point Project
- A. Žiberna:
  Generalized Blockmodeling for Valued Networks
- N. Kejžar, V. Batagelj:
  Analysis of US Patents Network: Development of Patents over Time



Generalized Blockmodeling

Vladimir Batagelj
Patrick Doreian
Anuška Ferligoj
Andrej Mrvar

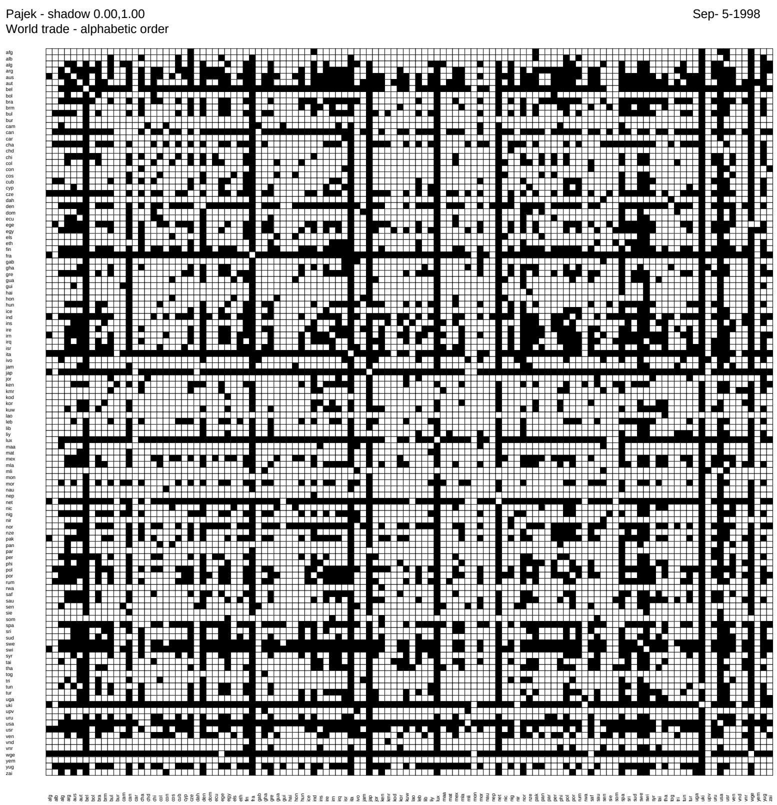
SUNBELT XXV

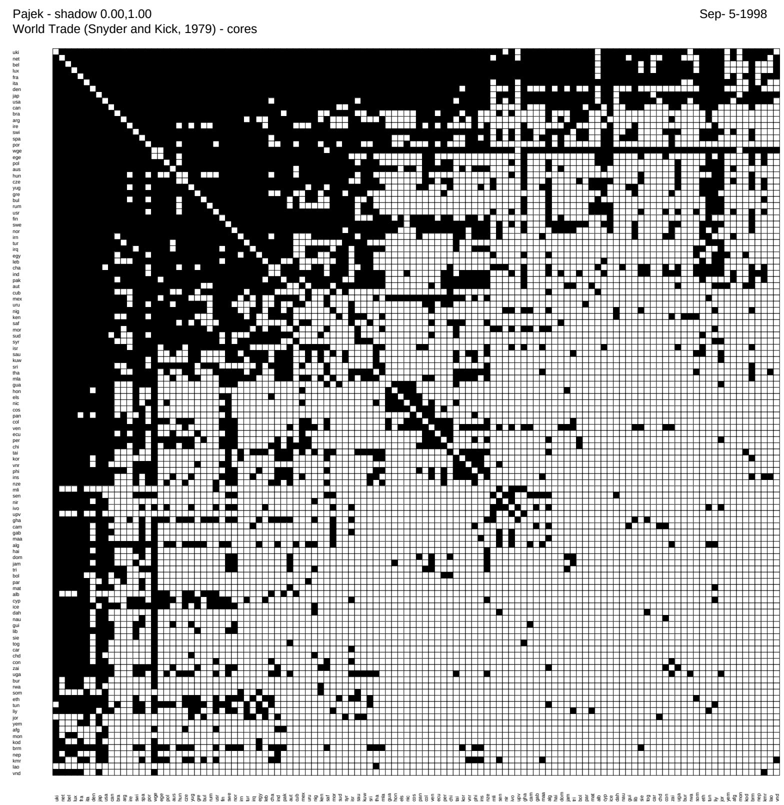
Redondo Beach, CA, February 18, 2005

Outline

Matrix rearrangement view on blockmodeling

Snyder & Kick's World trade network / $n = 118$, $m = 514$





Alphabetic order of countries (left) and rearrangement (right)

Ordering the matrix

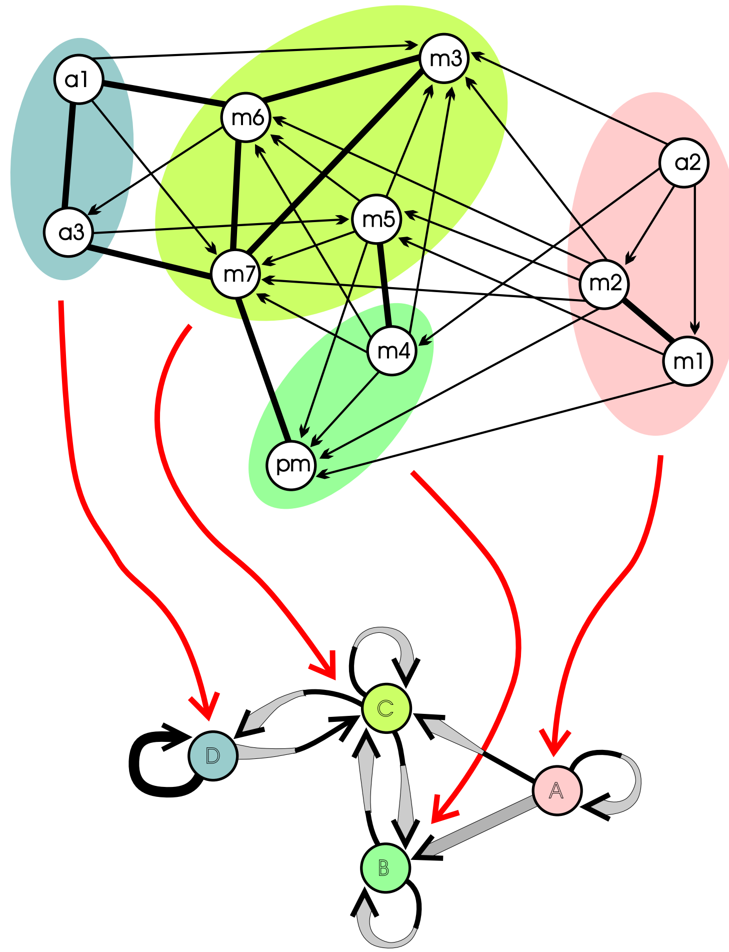
There are several ways how to rearrange a given matrix – determine an *ordering* or *permutation* of its rows and columns – to get some insight into its structure:

- ordering by degree;
- ordering by connected components;
- ordering by core number, connected components inside core levels, and degree;
- ordering according to a hierarchical clustering and some other property.

There exists also some special procedures to determine the ordering such as seriation and clumping (Murtagh).

Blockmodeling as a clustering problem

The goal of *blockmodeling* is to reduce a large, potentially incoherent network to a smaller comprehensible structure that can be interpreted more readily. Blockmodeling, as an empirical procedure, is based on the idea that units in a network can be grouped according to the extent to which they are equivalent, according to some *meaningful* definition of equivalence.



Cluster, clustering, blocks

One of the main procedural goals of blockmodeling is to identify, in a given network $\mathbf{N} = (\mathbf{U}, R)$, $R \subseteq \mathbf{U} \times \mathbf{U}$, *clusters* (classes) of units that share structural characteristics defined in terms of R. The units within a cluster have the same or similar connection patterns to other units. They form a *clustering* $\mathbf{C} = \{C_1, C_2, \dots, C_k\}$ which is a *partition* of the set $\mathbf{U}$. Each partition determines an equivalence relation (and vice versa). Let us denote by $\sim$ the relation determined by partition $\mathbf{C}$.

A clustering $\mathbf{C}$ partitions also the relation R into *blocks*

$$R(C_i, C_j) = R \cap C_i \times C_j$$

Each such block consists of units belonging to clusters C_i and C_j and all arcs leading from cluster C_i to cluster C_j. If $i = j$, a block $R(C_i, C_i)$ is called a *diagonal* block.

Structural and regular equivalence

Regardless of the definition of equivalence used, there are two basic approaches to the equivalence of units in a given network (compare Faust, 1988):

- the equivalent units have the same connection pattern to the **same** neighbors;
- the equivalent units have the same or similar connection pattern to (possibly) **different** neighbors.

The first type of equivalence is formalized by the notion of structural equivalence and the second by the notion of regular equivalence with the latter a generalization of the former.

Structural equivalence

Units are equivalent if they are connected to the rest of the network in *identical* ways (Lorrain and White, 1971). Such units are said to be *structurally equivalent*.

The units X and Y are *structurally equivalent*, we write $\mathrm{X} \equiv \mathrm{Y}$, iff the permutation (transposition) $\pi = (\mathrm{X}\,\mathrm{Y})$ is an automorphism of the relation R (Borgatti and Everett, 1992).

In other words, X and Y are structurally equivalent iff:

s1.	$\mathrm{X}R\mathrm{Y} \Leftrightarrow \mathrm{Y}R\mathrm{X}$	s3.	$\forall \mathrm{Z} \in \mathbf{U} \setminus \{\mathrm{X}, \mathrm{Y}\} : (\mathrm{X}R\mathrm{Z} \Leftrightarrow \mathrm{Y}R\mathrm{Z})$
s2.	$\mathrm{X}R\mathrm{X} \Leftrightarrow \mathrm{Y}R\mathrm{Y}$	s4.	$\forall \mathrm{Z} \in \mathbf{U} \setminus \{\mathrm{X}, \mathrm{Y}\} : (\mathrm{Z}R\mathrm{X} \Leftrightarrow \mathrm{Z}R\mathrm{Y})$

...structural equivalence

The blocks for structural equivalence are null or complete with variations on diagonal in diagonal blocks.

$$\begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix} \qquad \begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}$$

$$\begin{matrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{matrix} \qquad \begin{matrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{matrix}$$

Regular equivalence

Integral to all attempts to generalize structural equivalence is the idea that units are equivalent if they link in equivalent ways to other units that are also equivalent.

White and Reitz (1983): The equivalence relation $\approx$ on $\mathbf{U}$ is a *regular equivalence* on network $\mathbf{N} = (\mathbf{U}, R)$ if and only if for all $\mathrm{X}, \mathrm{Y}, \mathrm{Z} \in \mathbf{U}$, $\mathrm{X} \approx \mathrm{Y}$ implies both

R1. $\mathrm{X}R\mathrm{Z} \Rightarrow \exists \mathrm{W} \in \mathbf{U} : (\mathrm{Y}R\mathrm{W} \wedge \mathrm{W} \approx \mathrm{Z})$

R2. $\mathrm{Z}R\mathrm{X} \Rightarrow \exists \mathrm{W} \in \mathbf{U} : (\mathrm{W}R\mathrm{Y} \wedge \mathrm{W} \approx \mathrm{Z})$

Another view of regular equivalence is based on colorings (Everett, Borgatti 1996).

. . . regular equivalence

Theorem 1 (Batagelj, Doreian, Ferligoj, 1992) *Let* $\mathbf{C} = \{C_i\}$ *be a partition corresponding to a regular equivalence* $\approx$ *on the network* $\mathbf{N} = (\mathbf{U}, R)$. *Then each block* $R(C_u, C_v)$ *is either null or it has the property that there is at least one 1 in each of its rows and in each of its columns. Conversely, if for a given clustering* $\mathbf{C}$, *each block has this property then the corresponding equivalence relation is a regular equivalence.*

The blocks for regular equivalence are null or 1-covered blocks.

$$\begin{matrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix} \qquad \begin{matrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \end{matrix}$$

Establishing Blockmodels

The problem of establishing a partition of units in a network in terms of a selected type of equivalence is a special case of *clustering problem* that can be formulated as an optimization problem (Φ, P) as follows:

Determine the clustering $\mathbf{C}^\star \in \Phi$ for which

$$P(\mathbf{C}^\star) = \min_{\mathbf{C} \in \Phi} P(\mathbf{C})$$

where Φ is the set of *feasible clusterings* and P is a *criterion function*.

Since the set of units $\mathbf{U}$ is finite, the set of feasible clusterings is also finite. Therefore the set $\mathrm{Min}(\Phi, P)$ of all solutions of the problem (optimal clusterings) is not empty.

Criterion function

Criterion functions can be constructed

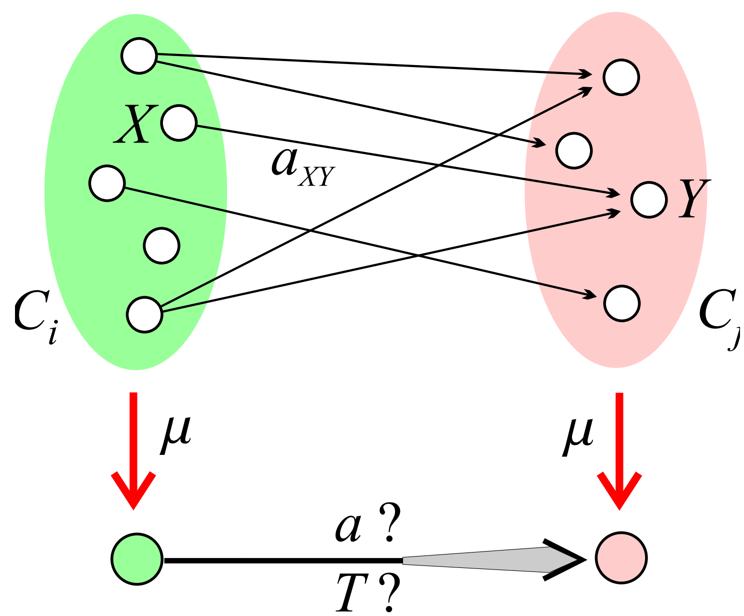
- *indirectly* as a function of a compatible (dis)similarity measure between pairs of units, or
- *directly* as a function measuring the fit of a clustering to an ideal one with perfect relations within each cluster and between clusters according to the considered types of connections (equivalence).

  Criterion function $P(\mathbf{C})$ has to be *sensitive* to considered equivalence:

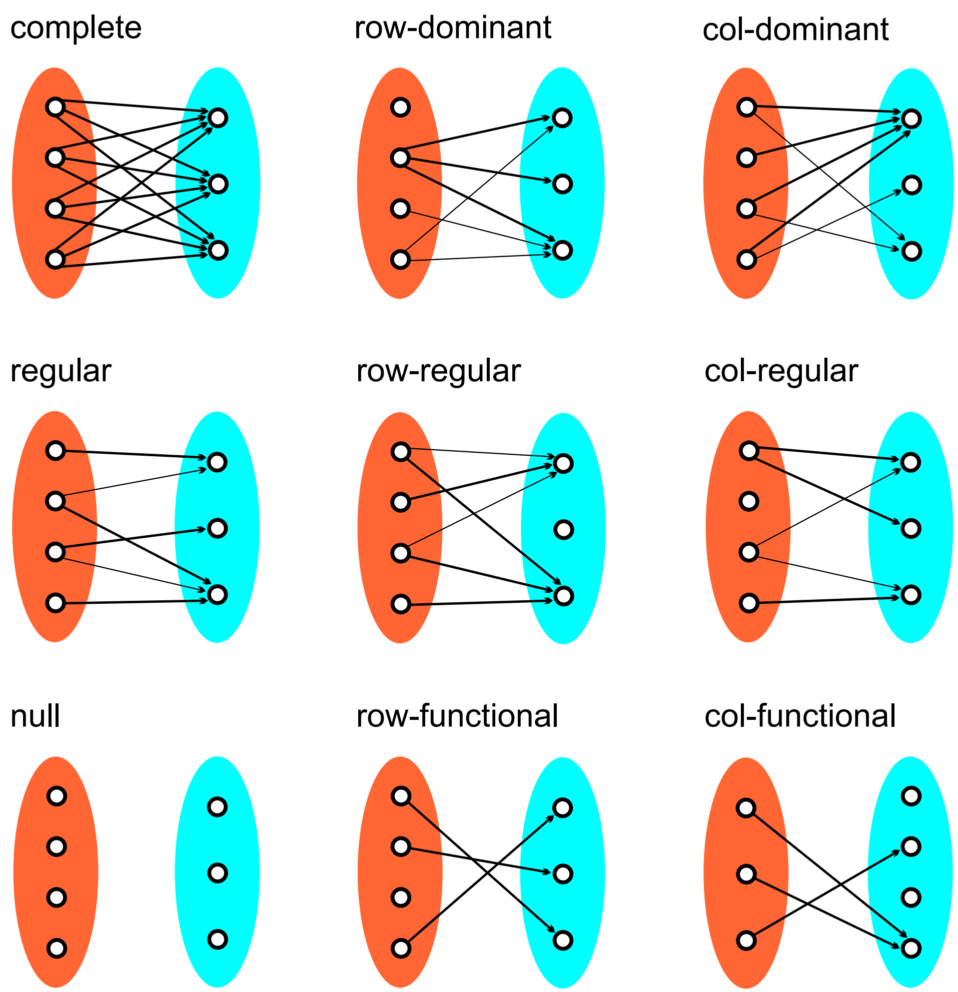
$$P(\mathbf{C}) = 0 \Leftrightarrow \mathbf{C} \text{ defines considered equivalence.}$$

Generalized Blockmodeling

A *blockmodel* consists of structures obtained by identifying all units from the same cluster of the clustering **C**. For an exact definition of a blockmodel we have to be precise also about which blocks produce an arc in the *reduced graph* and which do not, and of what *type*. Some types of connections are presented in the figure on the next slide. The reduced graph can be represented by relational matrix, called also *image matrix*.



Block Types


complete
row-dominant
col-dominant
regular
row-regular
col-regular
null
row-functional
col-functional

Generalized equivalence / Block Types

$$\begin{array}{c|ccccc|} & & & Y & & \\ \hline & 1 & 1 & 1 & 1 & 1 \\ X & 1 & 1 & 1 & 1 & 1 \\ & 1 & 1 & 1 & 1 & 1 \\ & 1 & 1 & 1 & 1 & 1 \\ \hline \end{array}$$

complete

$$\begin{array}{c|ccccc|} & & & Y & & \\ \hline & 0 & 1 & 0 & 0 & 0 \\ X & 1 & 1 & 1 & 1 & 1 \\ & 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 1 & 0 \\ \hline \end{array}$$

row-dominant

$$\begin{array}{c|ccccc|} & & & Y & & \\ \hline & 0 & 0 & 1 & 0 & 0 \\ X & 0 & 0 & 1 & 1 & 0 \\ & 1 & 1 & 1 & 0 & 0 \\ & 0 & 0 & 1 & 0 & 1 \\ \hline \end{array}$$

col-dominant

$$\begin{array}{c|ccccc|} & & & Y & & \\ \hline & 0 & 1 & 0 & 0 & 0 \\ X & 1 & 0 & 1 & 1 & 0 \\ & 0 & 0 & 1 & 0 & 1 \\ & 1 & 1 & 0 & 0 & 0 \\ \hline \end{array}$$

regular

$$\begin{array}{c|ccccc|} & & & Y & & \\ \hline & 0 & 1 & 0 & 0 & 0 \\ X & 0 & 1 & 1 & 0 & 0 \\ & 1 & 0 & 1 & 0 & 0 \\ & 0 & 1 & 0 & 0 & 1 \\ \hline \end{array}$$

row-regular

$$\begin{array}{c|ccccc|} & & & Y & & \\ \hline & 0 & 1 & 0 & 1 & 0 \\ X & 1 & 0 & 1 & 0 & 0 \\ & 1 & 1 & 0 & 1 & 1 \\ & 0 & 0 & 0 & 0 & 0 \\ \hline \end{array}$$

col-regular

$$\begin{array}{c|ccccc|} & & & Y & & \\ \hline & 0 & 0 & 0 & 0 & 0 \\ X & 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 & 0 \\ \hline \end{array}$$

null

$$\begin{array}{c|ccccc|} & & & Y & & \\ \hline & 0 & 0 & 0 & 1 & 0 \\ X & 0 & 0 & 1 & 0 & 0 \\ & 1 & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 1 & 0 \\ \hline \end{array}$$

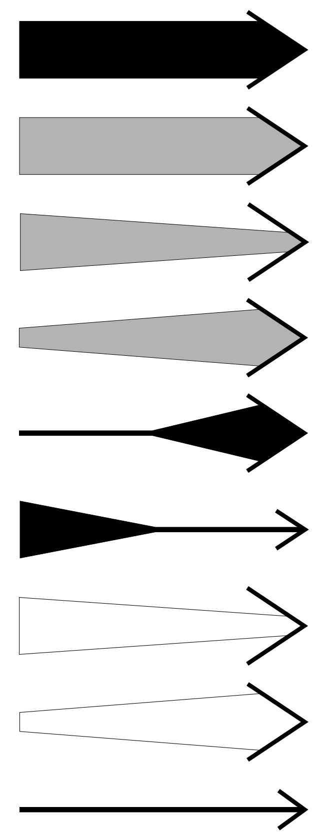
row-functional

$$\begin{array}{c|cccc|} & & Y & & \\ \hline & 1 & 0 & 0 & 0 \\ & 0 & 1 & 0 & 0 \\ X & 0 & 0 & 1 & 0 \\ & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 1 \\ \hline \end{array}$$

col-functional

Characterizations of Types of Blocks

null	nul	all 0 *
complete	com	all 1 *
regular	reg	1-covered rows and columns
row-regular	rre	each row is 1-covered
col-regular	cre	each column is 1 -covered
row-dominant	rdo	$\exists$ all 1 row *
col-dominant	cdo	$\exists$ all 1 column *
row-functional	rfn	$\exists!$ one 1 in each row
col-functional	cfn	$\exists!$ one 1 in each column
non-null	one	$\exists$ at least one 1



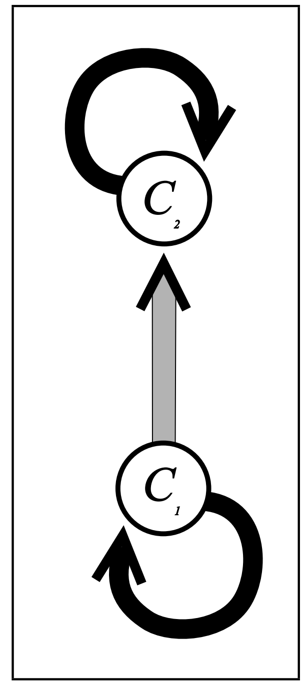
* except this may be diagonal

A block is *symmetric* iff $\forall \mathrm{X}, \mathrm{Y} \in C_i \times C_j : (\mathrm{X}R\mathrm{Y} \Leftrightarrow \mathrm{Y}R\mathrm{X})$.

Block Types and Matrices

$$
\begin{array}{cccc|cccc}
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 \\
1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 0
\end{array}
$$

	C_1	C_2
C_1	complete	regular
C_2	null	complete



Criterion function

One of the possible ways of constructing a criterion function that directly reflects the considered equivalence is to measure the fit of a clustering to an ideal one with perfect relations within each cluster and between clusters according to the considered equivalence.

Given a clustering $\mathbf{C} = \{C_1, C_2, \ldots, C_k\}$, let $\mathcal{B}(C_u, C_v)$ denote the set of all ideal blocks corresponding to block $R(C_u, C_v)$. Then the global error of clustering $\mathbf{C}$ can be expressed as

$$P(\mathbf{C}) = \sum_{C_u, C_v \in \mathbf{C}} \min_{B \in \mathcal{B}(C_u, C_v)} d(R(C_u, C_v), B)$$

where the term $d(R(C_u, C_v), B)$ measures the difference (error) between the block $R(C_u, C_v)$ and the ideal block B. d is constructed on the basis of characterizations of types of blocks. The function d has to be compatible with the selected type of equivalence.

... criterion function

For example, for structural equivalence, the term $d(R(C_u, C_v), B)$ can be expressed, for non-diagonal blocks, as

$$d(R(C_u, C_v), B) = \sum_{\mathrm{X} \in C_u, \mathrm{Y} \in C_v} |r_{\mathrm{X\,Y}} - b_{\mathrm{X\,Y}}|.$$

where $r_{\mathrm{X\,Y}}$ is the observed tie and $b_{\mathrm{X\,Y}}$ is the corresponding value in an ideal block. This criterion function counts the number of 1s in erstwhile null blocks and the number of 0s in otherwise complete blocks. These two types of inconsistencies can be weighted differently.

Determining the block error, we also determine the type of the best fitting ideal block (the types are ordered).

The criterion function $P(\mathbf{C})$ is *sensitive* iff $P(\mathbf{C}) = 0 \Leftrightarrow \mu$ (determined by $\mathbf{C}$) is an exact blockmodeling. For all presented block types sensitive criterion functions can be constructed (Batagelj, 1997).

Solving the blockmodeling problem

The obtained optimization problem can be solved by local optimization.

Once a partitioning μ and types of connection π are determined, we can also compute the values of connections by using averaging rules.

Benefits from Optimization Approach

- *ordinary / inductive blockmodeling*: Given a network $\mathbf{N}$ and set of types of connection $\mathcal{T}$, determine the model $\mathcal{M}$;
- *evaluation of the quality of a model, comparing different models, analyzing the evolution of a network* (Sampson data, Doreian and Mrvar 1996; states / continents): Given a network $\mathbf{N}$, a model $\mathcal{M}$, and blockmodeling μ, compute the corresponding criterion function;
- *model fitting / deductive blockmodeling*: Given a network $\mathbf{N}$, set of types $\mathcal{T}$, and a family of models, determine μ which minimizes the criterion function (Batagelj, Ferligoj, Doreian, 1998).
- we can fit the network to a partial model and analyze the residual afterward;
- we can also introduce different constraints on the model, for example: units X and Y are of the same type; or, types of units X and Y are not connected; …

Pre-specified blockmodeling

In the previous slides the inductive approaches for establishing blockmodels for a set of social relations defined over a set of units were discussed. Some form of equivalence is specified and clusterings are sought that are consistent with a specified equivalence.

Another view of blockmodeling is deductive in the sense of starting with a blockmodel that is specified in terms of substance prior to an analysis.

In this case given a network, set of types of ideal blocks, and a reduced model, a solution (a clustering) can be determined which minimizes the criterion function.

Pre-Specified Blockmodels

The pre-specified blockmodeling starts with a blockmodel specified, in terms of substance, ***prior to an analysis***. Given a network, a set of ideal blocks is selected, a family of reduced models is formulated, and partitions are established by minimizing the criterion function.

The basic types of models are:

$$\begin{array}{|c|c|}\hline * & * \\ \hline * & 0 \\ \hline\end{array}$$

center - periphery

$$\begin{array}{|c|c|}\hline * & 0 \\ \hline * & * \\ \hline\end{array}$$

hierarchy

$$\begin{array}{|c|c|}\hline * & 0 \\ \hline 0 & * \\ \hline\end{array}$$

clustering

$$\begin{array}{|c|c|}\hline 0 & * \\ \hline * & 0 \\ \hline\end{array}$$

bipartition

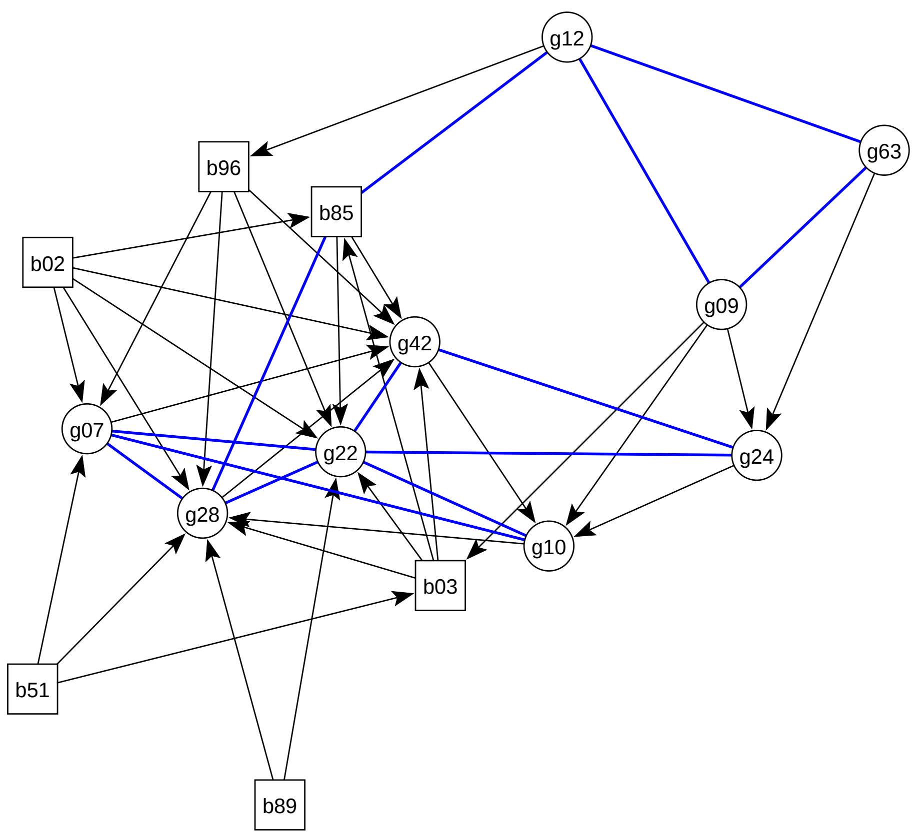
Prespecified blockmodeling example
Support network among informatics students

The analyzed network consists of social support exchange relation among fifteen students of the Social Science Informatics fourth year class (2002/2003) at the Faculty of Social Sciences, University of Ljubljana. Interviews were conducted in October 2002.

Support relation among students was identified by the following question:

> Introduction: You have done several exams since you are in the second class now. Students usually borrow studying material from their colleagues.
>
> Enumerate (list) the names of your colleagues that you have most often borrowed studying material from. (The number of listed persons is not limited.)

Class network



class.net

Vertices represent students in the class; circles – girls, squares – boys. Opposite pairs of arcs are replaced by edges.

Prespecified blockmodeling example

We expect that center-periphery model exists in the network: some students having good studying material, some not.

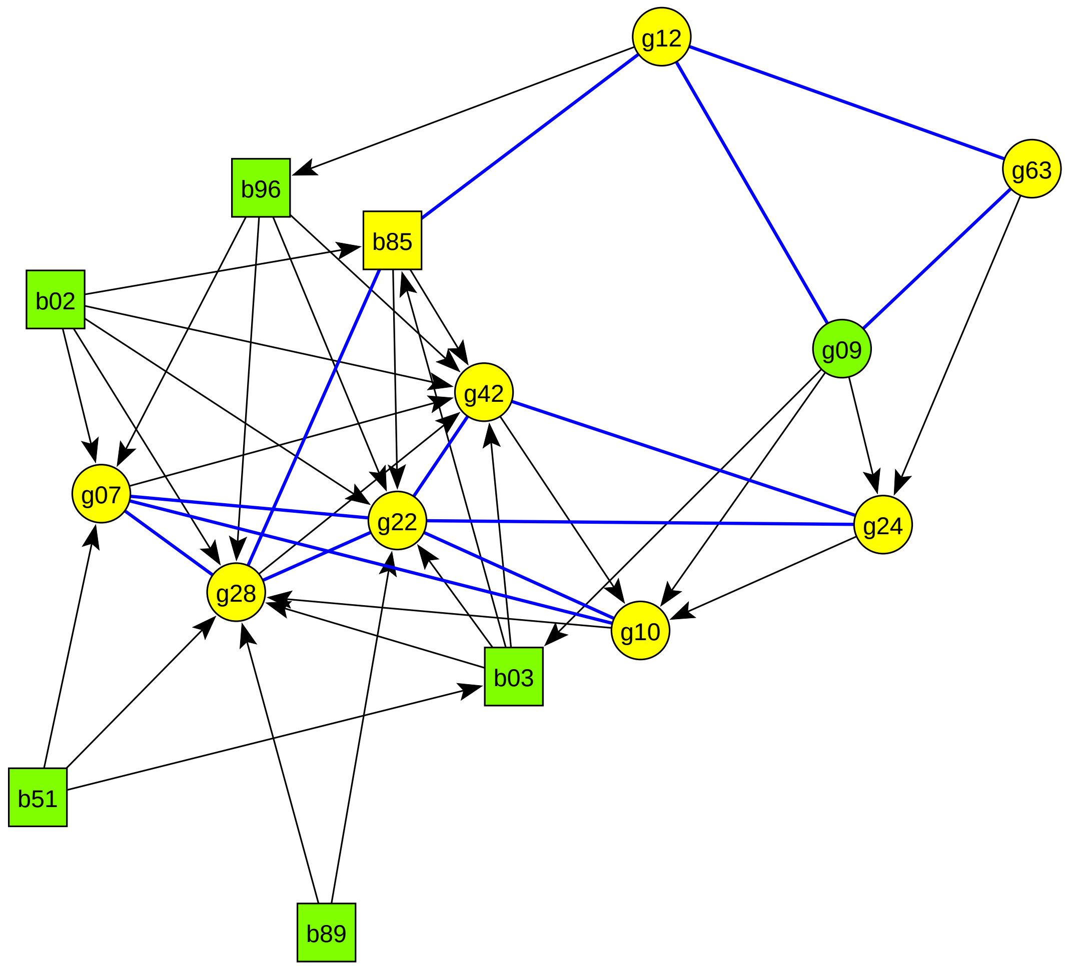
Prespecified blockmodel: (com/complete, reg/regular, -/null block)

	1	2
1	[com reg]	-
2	[com reg]	-

Using local optimization we get the partition:

$$\begin{aligned} \mathbf{C} &= \{\{b02, b03, b51, b85, b89, b96, g09\}, \\ &\quad \{g07, g10, g12, g22, g24, g28, g42, g63\}\} \end{aligned}$$

2 clusters solution



Model

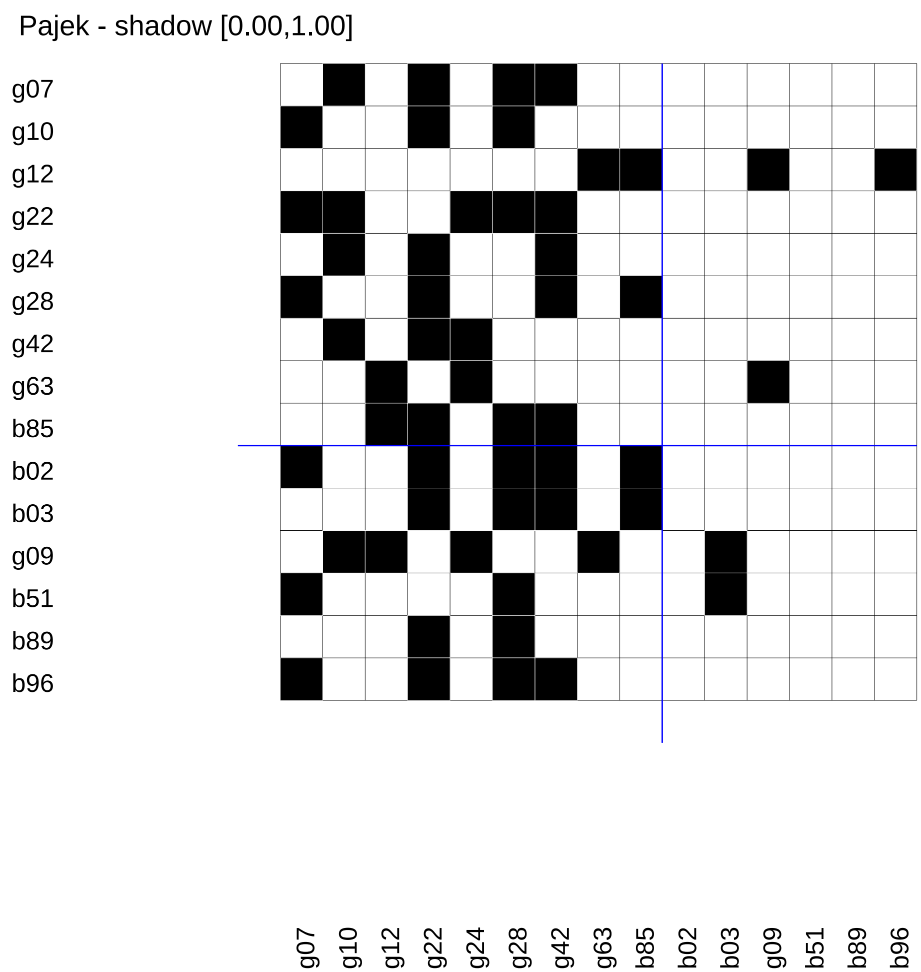


Image and Error Matrices:

	1	2
1	reg	-
2	reg	-

	1	2
1	0	3
2	0	2

Total error = 5

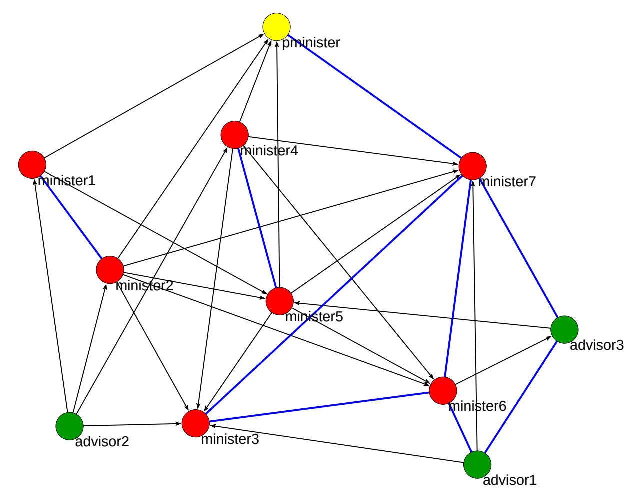
center-periphery

The Student Government at the University of Ljubljana in 1992

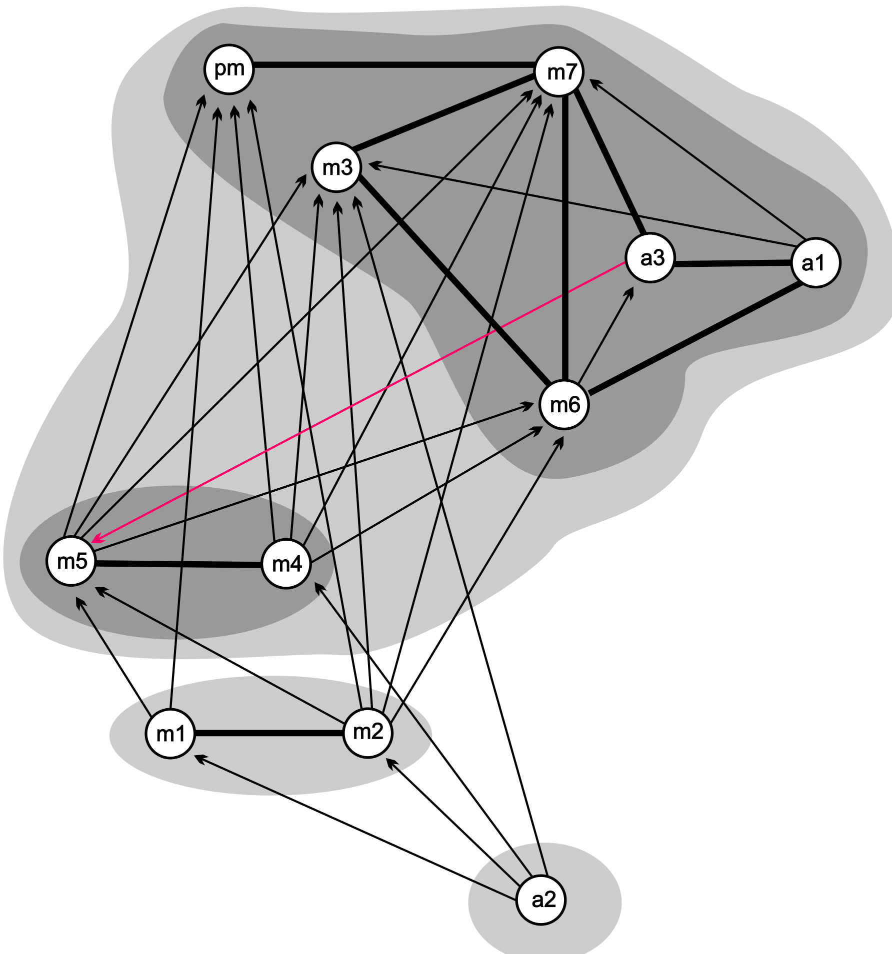
The relation is determined by the following question (Hlebec, 1993):

> Of the members and advisors of the Student Government, whom do you most often talk with about the matters of the Student Government?

		m 1	p 2	m 3	m 4	m 5	m 6	m 7	m 8	a 9	a 10	a 11
minister 1	1	·	1	1	·	·	1	·	·	·	·	·
p.minister	2	·	·	·	·	·	·	·	1	·	·	·
minister 2	3	1	1	·	1	·	1	1	1	·	·	·
minister 3	4	·	·	·	·	·	·	1	1	·	·	·
minister 4	5	·	1	·	1	·	1	1	1	·	·	·
minister 5	6	·	1	·	1	1	·	1	1	·	·	·
minister 6	7	·	·	·	1	·	·	·	1	1	·	1
minister 7	8	·	1	·	1	·	·	1	·	·	·	1
adviser 1	9	·	·	·	1	·	·	1	1	·	·	1
adviser 2	10	1	·	1	1	1	·	·	·	·	·	·
adviser 3	11	·	·	·	·	·	1	·	1	1	·	·



A Symmetric Acyclic Blockmodel of Student Government



The obtained clustering in 4 clusters is almost exact. The only error is produced by the arc $(a3, m5)$.

Signed graphs

A *signed graph* is an ordered pair (G, σ) where

- $G = (V, R)$ is a directed graph (without loops) with set of vertices V and set of arcs $R \subseteq V \times V$;
- $\sigma : R \to \{p, n\}$ is a *sign* function. The arcs with the sign p are *positive* and the arcs with the sign n are *negative*. We denote the set of all positive arcs by R^+ and the set of all negative arcs by R^-.

The case when the graph is undirected can be reduced to the case of directed graph by replacing each edge e by a pair of opposite arcs both signed with the sign of the edge e.

Balanced and clusterable signed graphs

The signed graphs were introduced in Harary, 1953 and later studied by several authors. Following Roberts (1976, p. 75–77) a signed graph (G, σ) is:

- *balanced* iff the set of vertices V can be partitioned into two subsets so that every positive arc joins vertices of the same subset and every negative arc joins vertices of different subsets.
- *clusterable* iff the set of V can be partitioned into subsets, called *clusters*, so that every positive arc joins vertices of the same subset and every negative arc joins vertices of different subsets.

. . . Properties

The (semi)walk on the signed graph is *positive* iff it contains an even number of negative arcs; otherwise it is *negative*.

The balanced and clusterable signed graphs are characterised by the following theorems:

THEOREM 1. A signed graph (G, σ) is balanced iff every closed semiwalk is positive.

THEOREM 2. A signed graph (G, σ) is clusterable iff G contains no closed semiwalk with exactly one negative arc.

Clusterability and blockmodeling

To the sign graph clusterability problemu corespond three types of blocks:

- *null* all elements in a block are 0;
- *positive* all elements in a block are positive or 0;
- *negative* all elements in a block are negative or 0;

If a graph is clusterable the blocks determined by the partition are: positive or null on the diagonal; and negative or null outside the diagonal.

The clusterability of partition $\mathbf{C} = \{C_1, C_2, \ldots, C_k\}$ can be therefore measured as follows ($0 \leq \alpha \leq 1$):

$$P_\alpha(\mathbf{C}) = \alpha \sum_{C \in \mathbf{C}} \sum_{u,v \in C} \max(0, -w_{uv}) + (1-\alpha) \sum_{\substack{C, C' \in \mathbf{C} \\ C \neq C'}} \sum_{u \in C, v \in C'} \max(0, w_{uv})$$

The blockmodeling problem can be solved by local optimization.

Slovenian political parties 1994 (S. Kropivnik)

		1	2	3	4	5	6	7	8	9	10
SKD	1	0	-215	114	-89	-77	94	-170	176	117	-210
ZLSD	2	-215	0	-217	134	77	-150	57	-253	-230	49
SDSS	3	114	-217	0	-203	-80	138	-109	177	180	-174
LDS	4	-89	134	-203	0	157	-142	173	-241	-254	23
ZSESS	5	-77	77	-80	157	0	-188	170	-120	-160	-9
ZS	6	94	-150	138	-142	-188	0	-97	140	116	-106
DS	7	-170	57	-109	173	170	-97	0	-184	-191	-6
SLS	8	176	-253	177	-241	-120	140	-184	0	235	-132
SPS-SNS	9	117	-230	180	-254	-160	116	-191	235	0	-164
SNS	10	-210	49	-174	23	-9	-106	-6	-132	-164	0

SKD – Slovene Christian Democrats; ZLSD – Associated List of Social Democrats; SDSS – Social Democratic Party of Slovenia; LDS – Liberal Democratic Party; ZSESS and ZS – two Green Parties, separated after 1992 elections; DS – Democratic Party; SLS – Slovene People's Party; SNS – Slovene National Party; SPS SNS – a group of deputies, former members of SNS, separated after 1992 elections

Network Stranke94.

Slovenian political parties 1994 / reordered

		1	3	6	8	9	2	4	5	7	10
SKD	1	0	114	94	176	117	-215	-89	-77	-170	-210
SDSS	3	114	0	138	177	180	-217	-203	-80	-109	-174
ZS	6	94	138	0	140	116	-150	-142	-188	-97	-106
SLS	8	176	177	140	0	235	-253	-241	-120	-184	-132
SPS-SNS	9	117	180	116	235	0	-230	-254	-160	-191	-164
ZLSD	2	-215	-217	-150	-253	-230	0	134	77	57	49
LDS	4	-89	-203	-142	-241	-254	134	0	157	173	23
ZSESS	5	-77	-80	-188	-120	-160	77	157	0	170	-9
DS	7	-170	-109	-97	-184	-191	57	173	170	0	-6
SNS	10	-210	-174	-106	-132	-164	49	23	-9	-6	0

S. Kropivnik, A. Mrvar: An Analysis of the Slovene Parliamentary Parties Network. in Developments in data analysis, MZ 12, FDV, Ljubljana, 1996, p. 209-216.

Blockmodeling in 2-mode networks

We already presented some ways of rearranging 2-mode network matrices at the beginning of this lecture.

It is also possible to formulate this goal as a generalized blockmodeling problem where the solutions consist of two partitions — row-partition and column-partition.

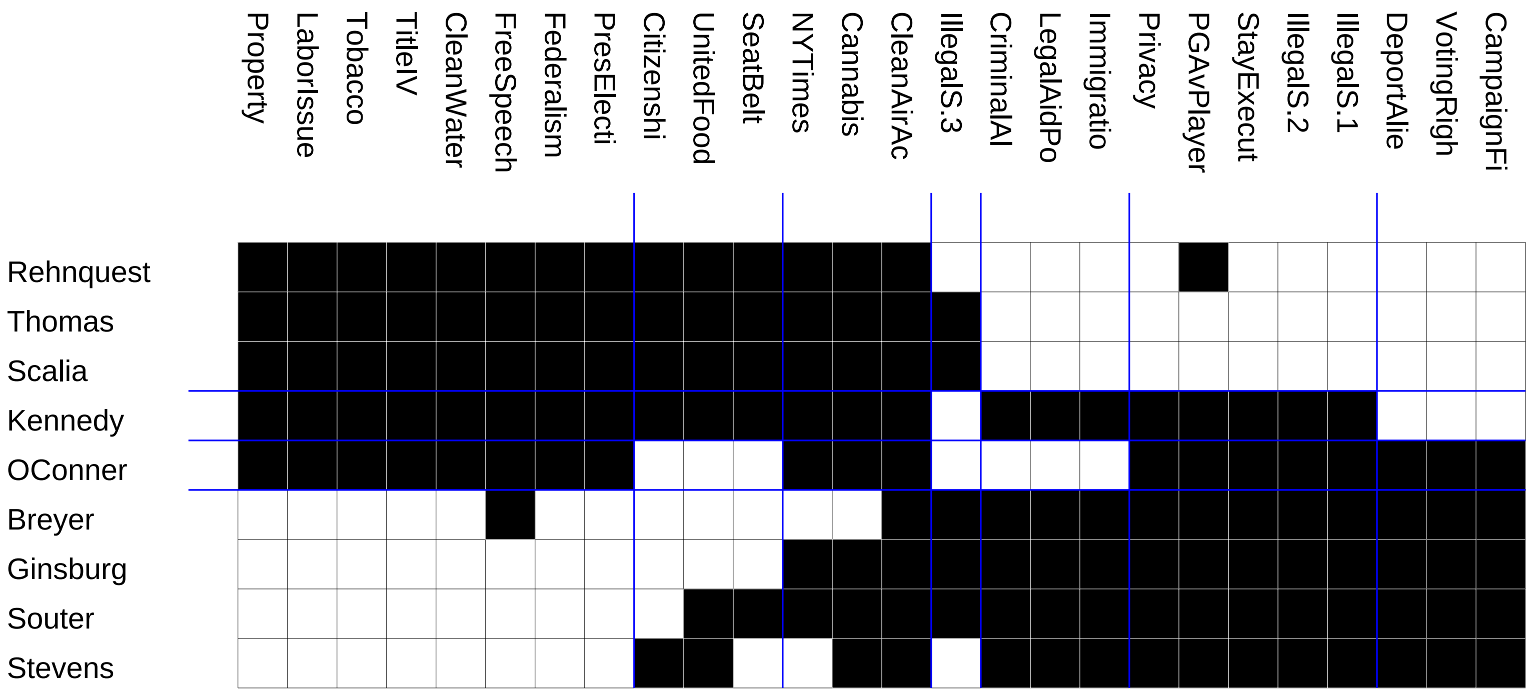
Supreme Court Voting for Twenty-Six Important Decisions

Issue	Label	Br	Gi	So	St	OC	Ke	Re	Sc	Th
Presidential Election	PE	-	-	-	-	+	+	+	+	+
Criminal Law Cases										
Illegal Search 1	CL1	+	+	+	+	+	+	-	-	-
Illegal Search 2	CL2	+	+	+	+	+	+	-	-	-
Illegal Search 3	CL3	+	+	+	-	-	-	-	+	+
Seat Belts	CL4	-	-	+	-	-	+	+	+	+
Stay of Execution	CL5	+	+	+	+	+	+	-	-	-
Federal Authority Cases										
Federalism	FA1	-	-	-	-	+	+	+	+	+
Clean Air Action	FA2	+	+	+	+	+	+	+	+	+
Clean Water	FA3	-	-	-	-	+	+	+	+	+
Cannabis for Health	FA4	0	+	+	+	+	+	+	+	+
United Foods	FA5	-	-	+	+	-	+	+	+	+
NY Times Copyrights	FA6	-	+	+	-	+	+	+	+	+
Civil Rights Cases										
Voting Rights	CR1	+	+	+	+	+	-	-	-	-
Title VI Disabilities	CR2	-	-	-	-	+	+	+	+	+
PGA v. Handicapped Player	CR3	+	+	+	+	+	+	+	-	-
Immigration Law Cases										
Immigration Jurisdiction	Im1	+	+	+	+	-	+	-	-	-
Deporting Criminal Aliens	Im2	+	+	+	+	+	-	-	-	-
Detaining Criminal Aliens	Im3	+	+	+	+	-	+	-	-	-
Citizenship	Im4	-	-	-	+	-	+	+	+	+
Speech and Press Cases										
Legal Aid for Poor	SP1	+	+	+	+	-	+	-	-	-
Privacy	SP2	+	+	+	+	+	+	-	-	-
Free Speech	SP3	+	-	-	-	+	+	+	+	+
Campaign Finance	SP4	+	+	+	+	+	-	-	-	-
Tobacco Ads	SP5	-	-	-	-	+	+	+	+	+
Labor and Property Rights Cases										
Labor Rights	LPR1	-	-	-	-	+	+	+	+	+
Property Rights	LPR2	-	-	-	-	+	+	+	+	+

The Supreme Court Justices and their 'votes' on a set of 26 "important decisions" made during the 2000-2001 term, Doreian and Fujimoto (2002).

The Justices (in the order in which they joined the Supreme Court) are: Rehnquist (1972), Stevens (1975), O'Conner (1981), Scalia (1982), Kennedy (1988), Souter (1990), Ginsburg (1993) and Breyer (1994).

…Supreme Court Voting / a (4,7) partition



upper – conservative / lower – liberal

Open problems

- GBM of valued networks
- GBM of multirelational networks
- GBM of temporal networks
- GBM of large networks

GBM of valued networks

Can the clustering with relational constraint and blockmodeling problem be generalized to a common problem?

Batagelj V., Ferligoj A.: Clustering relational data. Data Analysis (ed.: W. Gaul, O. Opitz, M. Schader), Springer, Berlin 2000, 3-15.

General problem of clustering relational data

The relationally constrained clustering problem with simple criterion function considers only the diagonal blocks that should be of one of the types $\Phi^i(R)$. It also takes into account the dissimilarity matrix on units (derived from attribute data).

The blockmodeling problem deals only with relational data. The proposed optimization approach essentially expresses the constraints with a penalty function.

Both problems can be expressed as special cases of a clustering problem with a general criterion function of the form

- **G1s**. $P(\mathbf{C}) = \sum_{(C_1,C_2)\in\mathbf{C}\times\mathbf{C}} q(C_1, C_2)$, or
- **G1m**. $P(\mathbf{C}) = \max_{(C_1,C_2)\in\mathbf{C}\times\mathbf{C}} q(C_1, C_2)$

where q is a block error satisfying

- **G2**. $q(C_1, C_2) \geq 0$

. . . clustering relational data

The set of feasible clusterings $\Phi_k(R)$ for this problem is determined by the relation R and additional requirements, such as:

- the blocks should be of selected types
- the model graph should be of specified form (prespecified)
- selected units should / should not be in the same cluster
- selected unit should / should not be in the selected cluster

Approaches to the problem

There are different types of relational data (valued networks). In the following we shall assume $\mathbf{N} = (V, R, a, b)$ where $a : V \to A$ assigns a value to each unit/vertex and $b : R \to B$ assigns a value to each arc (link) of R. A and B are sets of values.

The function b determines a matrix $\mathbf{B} = [b_{ij}]_{n \times n}, b_{ij} \in B \cup \{0\}$ and $b_{ij} = 0$ if units i and j are not connected by an arc.

We can approach the problem of clustering relational data by *indirect* (transformation to standard data analysis problems) or *direct* (formulating the problem as an optimization problem and solving it) approach.

Indirect approach

A scenario for the indirect approach is to transform attribute data a into dissimilarity matrix D_a and network data b into dissimilarity matrix D_b and build criterion functions P_a and P_b based on them (they can be defined also directly from a and b). Then we apply the multicriteria relationally constrained clustering methods on these functions.

We can also first combine D_a and D_b into a joint matrix D_{ab} and apply relationally constrained clustering methods on it.

In a special case, when D_b is defined as some corrected dissimilarity (see Batagelj, Ferligoj, Doreian, 1992) between descriptions $\mathbf{b}(x) = [\mathbf{B}(x), \mathbf{B}^T(x)]$, the relational data are built into D_b and we can apply on the combined matrix D_{ab} all standard methods for analysis of dissimilarity matrices.

Direct approach

Again there are different possibilities:

1. Structural approach: used in program MODEL (Batagelj, 1996): Important is the structure (relation). Determine the best clustering $\mathbf{C}$ and the corresponding model. On the basis of a, b and the obtained model compute values of model connections.

2. Multicriteria approach: construct two criterion functions: one based on values, the second based on structure. Solve the obtained multicriteria problem (Ferligoj, Batagelj, 1992).

3. Implicit approach: the types of connections are built into the criterion function combined with values.

Only the last approach needs some further explanations.

Implicit approach

Let $\approx$ be an equivalence over set of units V, and $\mathcal{T}$ given types. We construct on blocks deviation functions $\delta(C_1, C_2; T), T \in \mathcal{T}$ such that $\approx$ is compatible with $\mathcal{T}$ over the network $\mathbf{N}$ iff

$$\forall u, v \in V \exists \delta(., .; T), T \in \mathcal{T} : \delta([u], [v]; T) = 0$$

Applying also an adequate normalization of δs we can construct a criterion function

$$P(\mathbf{C}) = \sum_{C_1, C_2 \in \mathbf{C}} \min_{T \in \mathcal{T}} \delta(C_1, C_2; T)$$

Evidently, $P(\mathbf{C}) = 0 \Leftrightarrow \mathbf{C}$ is compatible with $\mathcal{T}$ — all blocks of $\mathbf{C}$ are compatible with $\mathcal{T}$.

Some examples

Assume that a and b are transformed into a matrix $\mathbf{A} = [a_{uv}]_{V \times V}$, $a_{uv} \geq 0$. Then

$$\begin{aligned}
\delta(X, Y; \mathrm{nul}) &= \frac{\sum_{x \in X, y \in Y} a_{xy}}{|X| \cdot |Y| \cdot \max\{a_{xy} : x \in X, y \in Y\}} \\
\delta(X, Y; \mathrm{rdo}) &= 1 - \max_{x \in X} \frac{\sum_{y \in Y} a_{xy}}{|Y| \cdot \max\{a_{xy} : y \in Y\}} \\
\delta(X, Y; \mathrm{cre}) &= 1 - \frac{\sum_{y \in Y} \max_{x \in X} a_{xy}}{|Y| \cdot \max\{a_{xy} : x \in X, y \in Y\}}
\end{aligned}$$

If $\max$ in the denominator equals to 0 then also the fraction has value 0.

Some ideas how to approach the GBM of valued networks will be presented by Aleš Žiberna.

GBM of multirelational networks

Multiple networks are networks with more than one relation defined on the same set of vertices.

Assume that we have relations $R_1, R_2, \ldots, R_s$ and corresponding criterion functions $P_1, P_2, \ldots, P_s$ compatible respectively with $\mathcal{T}_1, \mathcal{T}_2, \ldots, \mathcal{T}_s$. Then we get the following multicriteria optimization problem: determine $\mathbf{C}^\star \in \Phi$ that 'minimizes'

$$(\Phi, P_1, P_2, \ldots, P_s)$$

This problem can be approched by (Ferligoj, Batagelj, 1992)

- multicriteria optimization. The solutions are Pareto points.
- transformation to single criterion optimization
  - by combining criterion functions: $\sum \alpha_i P_i$ or $\max \alpha_i P_i$, where $\alpha_i \geq 0$ and $\sum \alpha_i = 1$;
  - by combining relations.

GBM of temporal networks

If the clustering $\mathbf{C}$ is the same for all time points we can treat the GBM of temporal networks problem as a GBM of multirelational networks for relations determined by the time slices.

In general, however, also the clustering $\mathbf{C}$ can change through time – the solution is a sequence of clusterings $(\mathbf{C}_1, \mathbf{C}_2, \ldots, \mathbf{C}_s)$, $\mathbf{C}_i \in \Phi_i$ (or $\mathbf{C}_i \in \Phi_i(\mathbf{C}_{i-1})$).

The Sampson Data

The affect data from Sampson (1968) provide a useful source for examining structural balance through time. (See Doreian and Mrvar, 1996, for a more complete description and use of these data.) There are three subgroups: One group was the ‘Young Turks’ made up of Gregory, John Bosco, Mark, Winfrid, Hugh, Boniface and Albert.

A second subgroup was labeled the ‘Loyal Opposition’ and comprised Peter, Bonaventure, Berthold, Victor, Ambrose, Romuald, Louis and Amand.

The remaining three individuals - Basil, Simplicius and Elias - were labelled as the ‘Outcasts’.

We consider the partitions reported by Doreian and Mrvar (1996).

Sampson reported data for these actors for 3 periods in time. Each actor was asked to name the three other actors he liked the most and the three he disliked the most. The ties $\{3, 2, 1\}$ are the labels for the positive ties while $\{-3, -2, -1\}$ denote the negative ties.

k	T_2	T_3	T_4
1	48.5	48.0	47.0
2	21.5	16.0	12.5
3	17.5	11.0	10.5
4	19.0	13.5	12.5
5	20.5	16.0	15.0

Table reports the values of the criterion function for partitions having 1 to 5 clusters. Partitions consistent with structural balance (into two plus-sets) show a consistent drop through time of the criterion function. The same is true for generalized of balance ($k > 2$). The lowest values of the criterion function are those for a partition into 3 plus-sets. table shows strong support for the fundamental structural balance hypothesis. This group evolved through time towards a balanced form.

Partition of Sampson Monks at T_2, T_3, and T_4

T_2

		1	2	7	12	14	15	16	3	13	17	18	4	5	6	8	9	10	11
John Bosco	1	·	·	-1	·	1	·	·	2	·	·	·	·	3	-2	·	·	-3	·
Gregory	2	3	·	2	·	1	·	·	·	-3	-2	·	·	·	·	·	·	-1	·
Mark	7	·	2	·	·	·	·	3	·	·	·	·	-3	-1	-2	1	·	·	·
Winfrid	12	3	2	·	·	1	·	·	-1	·	·	·	-3	·	·	·	·	·	-2
Hugh	14	3	·	·	2	·	2	·	·	-3	-1	·	·	·	·	-2	·	·	1
Boniface	15	3	2	·	·	1	·	·	-2	-3	-1	-1	·	·	·	·	·	·	·
Albert	16	1	2	3	·	·	·	·	·	-1	-3	-2	·	·	·	·	·	·	·
Basil	3	2	3	·	·	·	·	·	·	·	1	·	-1	·	·	-3	-2	·	·
Amand	13	·	-3	1	-1	·	·	·	·	·	·	3	·	2	-2	·	·	·	·
Elias	17	·	·	·	·	·	·	·	3	2	·	1	-3	-2	·	·	·	·	-1
Simplicius	18	2	3	1	·	·	·	-1	·	·	·	·	-3	·	-2	·	·	·	·
Peter	4	·	·	-3	·	·	·	·	-2	·	·	-1	·	3	1	·	·	2	·
Bonaventure	5	·	·	·	·	·	·	·	·	1	·	·	3	·	·	·	·	·	2
Berthold	6	1	·	-3	-2	·	·	·	·	·	·	·	3	·	·	-1	2	·	·
Victor	8	3	2	·	·	-2	·	·	-3	·	-1	·	·	·	·	·	1	·	·
Ambrose	9	·	·	·	·	·	·	1	-3	·	-2	-1	·	2	·	3	·	·	·
Romuald	10	·	·	·	·	2	·	·	·	·	·	·	3	·	·	1	·	·	·
Louis	11	·	·	·	·	2	·	·	-1	-3	-2	·	·	3	·	1	·	·	·

T_3

		1	2	7	12	14	15	16	3	13	17	18	4	5	6	8	9	10	11
John Bosco	1	·	2	·	·	·	·	·	·	·	·	·	-2	1	·	3	·	-3	-1
Gregory	2	3	·	1	2	·	1	·	·	-1	·	·	-3	·	·	-2	·	·	·
Mark	7	1	2	·	·	·	·	3	·	·	·	·	-3	·	-2	-1	·	·	·
Winfrid	12	3	1	·	·	2	·	·	-3	·	-1	·	·	·	-2	·	·	·	·
Hugh	14	3	1	·	1	·	2	·	·	·	-1	·	-3	·	·	·	·	-2	·
Boniface	15	2	3	·	·	1	·	·	-2	-3	-1	·	·	1	·	·	·	-1	·
Albert	16	·	2	3	1	·	·	·	-1	·	-2	·	-3	·	·	·	·	·	·
Basil	3	3	-1	·	·	·	·	·	·	·	1	2	·	·	-2	-3	·	·	·
Amand	13	·	-3	1	-1	·	·	·	·	·	·	3	·	2	-2	·	·	·	·
Elias	17	·	1	·	·	·	·	·	2	·	·	3	-2	·	-1	·	·	·	-3
Simplicius	18	·	1	·	·	·	·	·	·	2	3	·	-3	·	-2	·	·	·	-1
Peter	4	-2	-3	·	·	·	·	·	·	·	·	·	·	3	2	·	·	·	1
Bonaventure	5	2	·	·	·	·	·	·	·	·	·	·	3	·	·	·	·	·	1
Berthold	6	1	·	·	-2	·	·	·	-3	·	·	·	3	·	·	-1	2	·	·
Victor	8	-2	-3	·	·	·	·	·	-1	·	·	·	3	·	1	·	·	2	·
Ambrose	9	·	·	·	2	·	·	·	-3	·	-2	-1	·	1	·	3	·	·	·
Romuald	10	·	·	·	·	·	·	·	·	2	·	·	3	1	·	·	·	·	·
Louis	11	-1	-3	·	·	·	·	1	-2	·	·	·	2	3	·	·	·	·	·

T_4																			
		1	2	7	12	14	15	16	3	13	17	18	4	5	6	8	9	10	11
John Bosco	1	·	-2	-3	1	2	·	·	3	·	·	·	·	·	·	·	·	-1	·
Gregory	2	3	·	1	2	·	·	·	·	-1	·	·	-3	·	·	-2	·	·	·
Mark	7	·	3	·	1	·	·	2	·	·	·	·	-3	·	-2	-1	·	·	·
Winfrid	12	3	2	1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
Hugh	14	3	·	·	1	·	2	·	·	·	-1	·	-3	·	·	-2	·	·	·
Boniface	15	·	3	1	2	·	·	·	-2	-3	·	·	-1	·	·	·	·	·	·
Albert	16	·	3	2	·	·	1	·	-1	·	-2	·	-3	·	·	·	·	·	·
Basil	3	3	-2	·	·	·	·	-1	·	2	1	2	-3	·	-2	·	·	·	·
Amand	13	·	-3	1	-1	·	·	·	·	·	·	3	·	2	-2	·	·	·	·
Elias	17	·	1	·	·	·	·	·	2	·	·	3	-1	·	-3	-2	·	·	·
Simplicius	18	·	1	·	·	·	·	·	2	·	3	·	-1	·	·	-3	·	-2	·
Peter	4	-2	-3	·	·	-1	·	·	·	·	·	·	·	3	1	·	·	·	2
Bonaventure	5	·	·	·	·	·	·	·	·	·	·	·	3	·	·	·	1	·	2
Berthold	6	·	-1	-2	·	·	·	·	-3	·	-2	·	3	1	·	·	2	·	·
Victor	8	·	-3	·	·	-1	·	·	-2	·	·	·	3	·	2	·	1	·	·
Ambrose	9	·	·	·	2	·	·	·	-3	·	-2	-1	·	1	·	3	·	·	·
Romuald	10	·	·	·	·	·	·	·	·	2	·	·	3	1	·	·	1	·	·
Louis	11	-1	-3	·	·	1	·	·	-2	·	·	·	·	2	·	3	·	·	·

GBM of large networks

Large networks are usually sparse $m << n^2$.

In the case of given partition $\mathbf{C}$ (for example partition of contries to continents, patents to categories, $\ldots$) it is easy to determine the corresponding GBM that minimizes the criterion function.

A special class of GBM problems are symmetric-acyclic decompositions (Doreian, Batagelj, Ferligoj, 2000) for which also an algorithm for large networks was developed.

. . . GBM of large networks

For some GBM problems on large networks the clustering methods can be used:

- if we have to compute the dissimilarities between attribute data for vertices we may consider methods that require only dissimilarities between the linked vertices;
- if we base the clustering on the descriptions of vertices using selected structural properties we can consider some variant of leader's method.