

Blockmodeling

Vladimir Batagelj
University of Ljubljana, Slovenia

POLNET 2002 – Summer School on Political and Managerial Networks

University of Konstanz, July 29 – August 2, 2002

Outline

- matrix rearrangement view on blockmodeling
- blockmodeling and clustering
structural and regular equivalence
- generalized blockmodeling
- prespecified blockmodels
- 2-mode blockmodeling

Network

A *network* $\mathbf{N} = (\mathbf{U}, R)$ consists of set of *units* \mathbf{U} and of *relation* $R \subseteq \mathbf{U} \times \mathbf{U}$.

We write $(X, Y) \in R$ or $XR Y$ the fact that the unit X is related (by R) to Y .

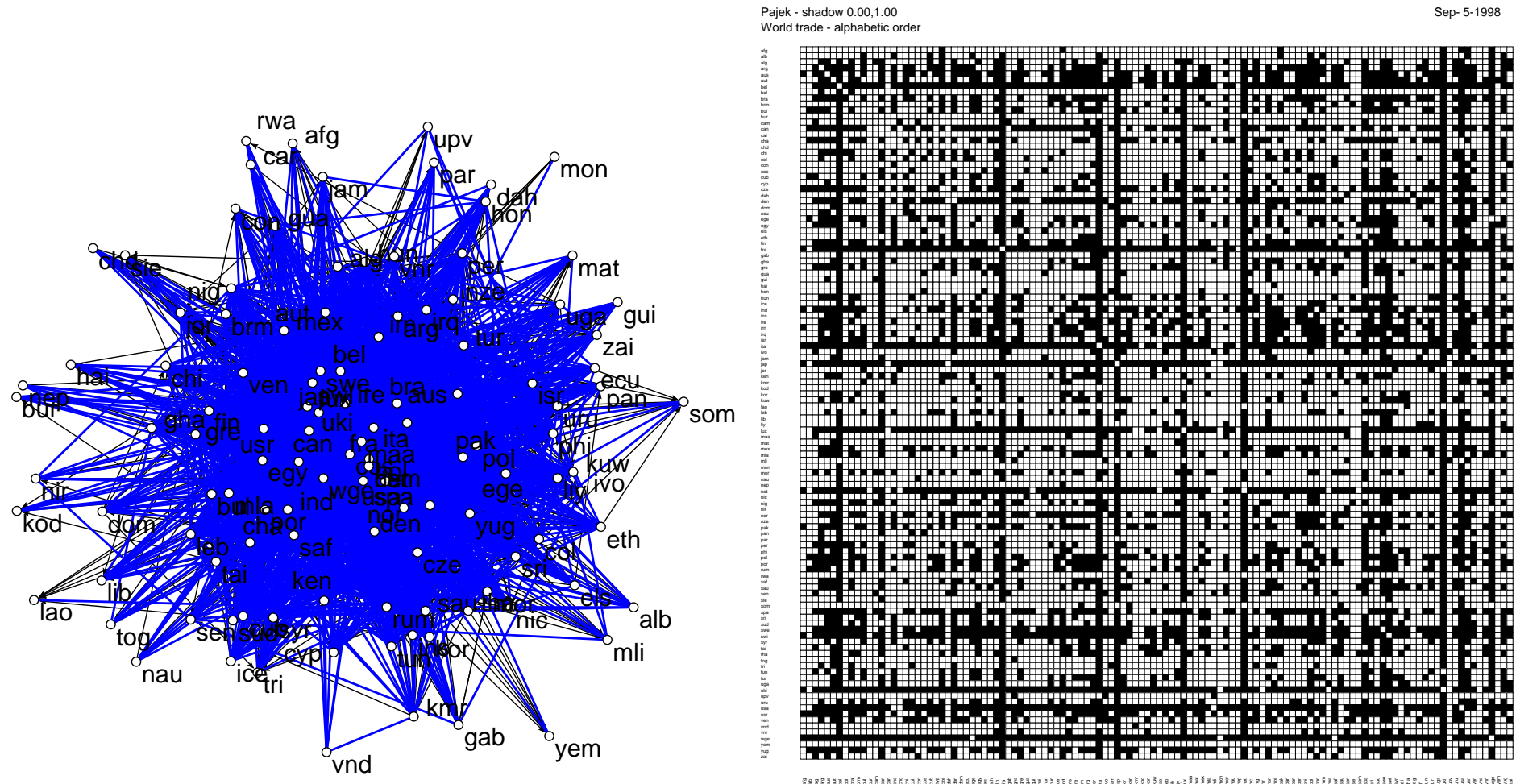
In some networks also a *weight* function $w : R \rightarrow \mathbb{R}$ is given.

Some networks can contain several relations on the same set of units.

In the following we shall deal with *simple* networks $\mathbf{N} = (\mathbf{U}, R)$.

Matrix rearrangement view on blockmodeling

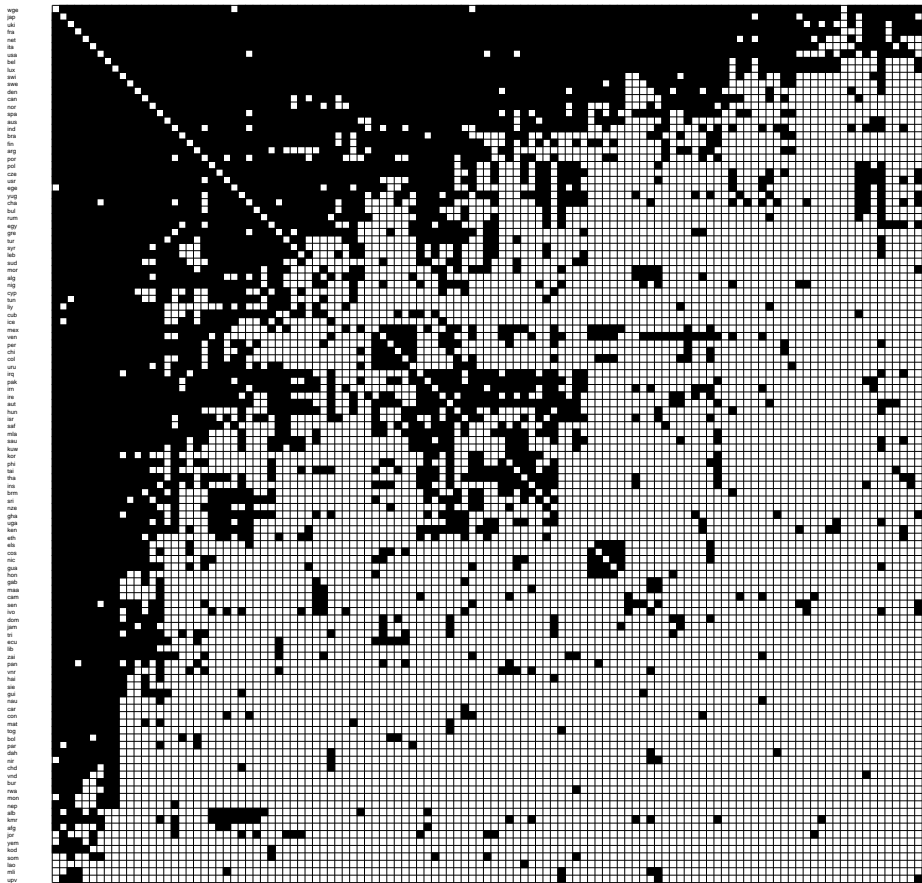
Snyder & Kick's World trade network / $n = 118, m = 514$



Graph drawing and Alphabetic order of countries

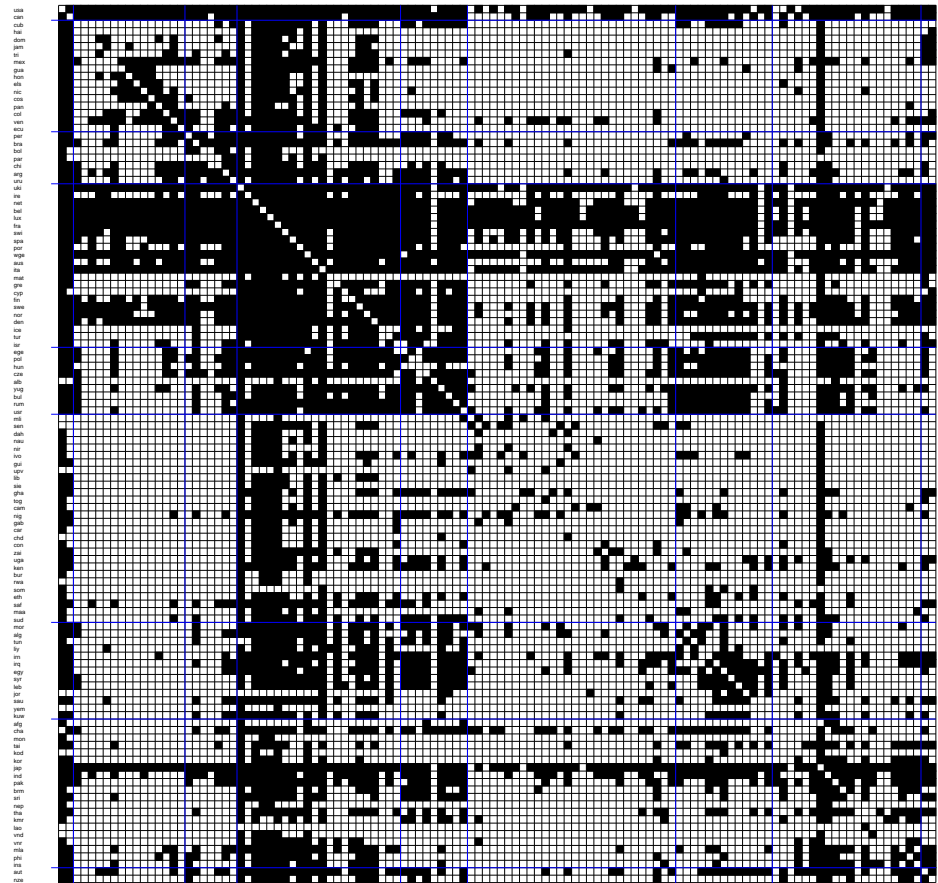
Snyder & Kick's World trade network / $n = 118, m = 514$

Pajek - shadow [0.00,1.00]



0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118

Pajek - shadow [0.00,1.00]



0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118

Ordering based on Ward clustering and on regional partition

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. For example seriation and clumping (Murtagh, 1985).

Demo with Pajek

Read Network SaKT.net

Draw/Draw

Layout/Energy/Kamada-Kawai/Free

File/Network/Export Matrix to EPS/Only Black Borders/On

File/Network/Export Matrix to EPS/Original [test.eps]

Read/Partition SaKreg.clu

Draw/Draw-Partition

Partition/Make Permutation

File/Network/Export Matrix to EPS/Using permutation [test.eps, Yes]

Options/Blockmodel-Shrink/Degree density

Operations/Shrink Network/Partition [Yes, 0.5, 0]

Net/Transform/Arcs-->Edges/Bidirected Only/Max

Draw

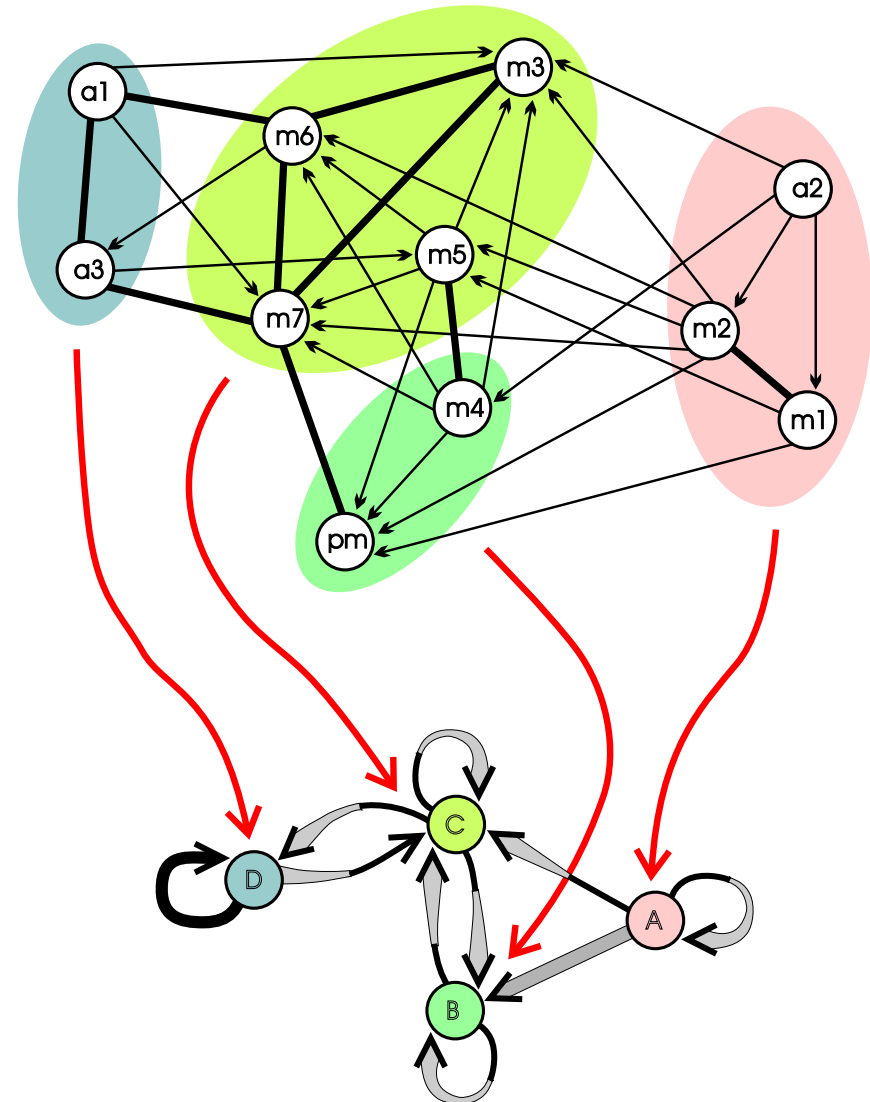
Net/Partition/Degree

Partition/Make Permutation

File/Network/Export Matrix to EPS/Using permutation [test.eps, Yes]

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 $\mathbf{N} = (\mathbf{U}, R)$, *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).

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 $X \equiv Y$, iff the permutation (transposition) $\pi = (X Y)$ is an automorphism of the relation R (Borgatti and Everett, 1992).

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

- | | | | |
|-----|-----------------------------|-----|--|
| s1. | $XR Y \Leftrightarrow YR X$ | s3. | $\forall Z \in U \setminus \{X, Y\} : (XR Z \Leftrightarrow YR Z)$ |
| s2. | $XR X \Leftrightarrow YR Y$ | s4. | $\forall Z \in U \setminus \{X, Y\} : (ZR X \Leftrightarrow ZR Y)$ |

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

Structural equivalence is too stringent for practical use.

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 $X, Y, Z \in \mathbf{U}$, $X \approx Y$ implies both

$$\text{R1. } XRZ \Rightarrow \exists W \in \mathbf{U} : (YRW \wedge W \approx Z)$$

$$\text{R2. } ZRX \Rightarrow \exists W \in \mathbf{U} : (WR Y \wedge W \approx Z)$$

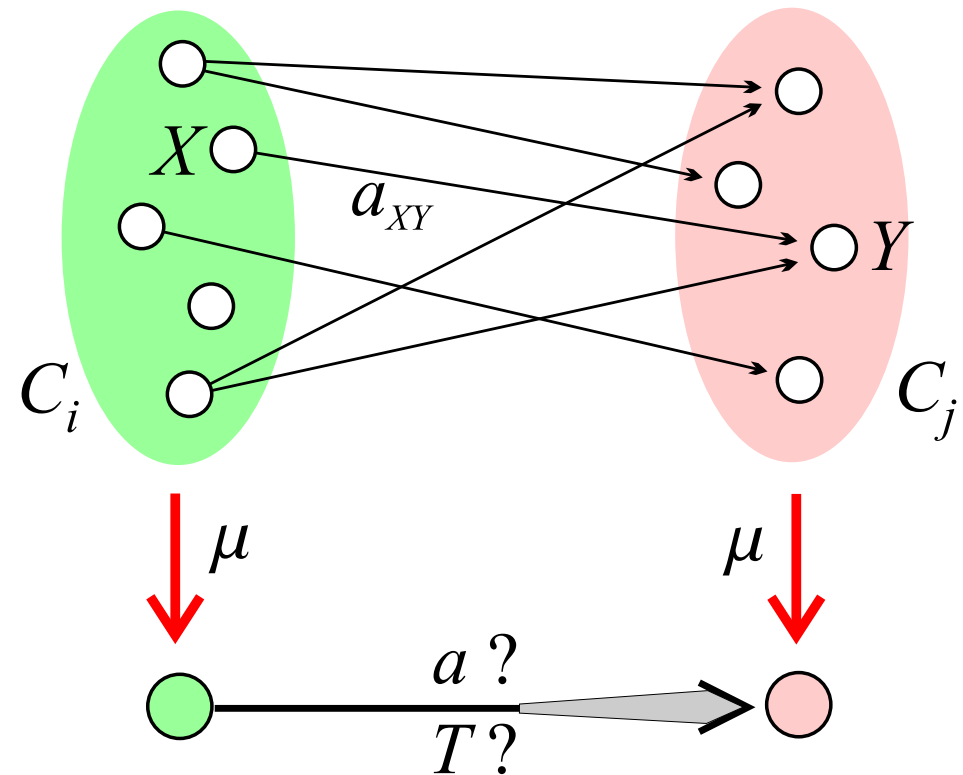
Another view of regular equivalence is based on colorings (Everett, Borgatti 1996).

Theorem 4.1 (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.

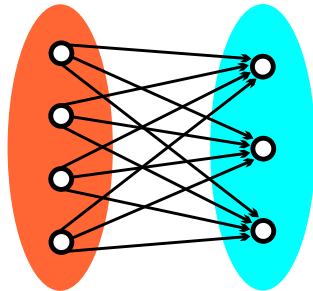
Generalized Blockmodeling

A *blockmodel* consists of structure 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*. For formal details see Batagelj (1997).

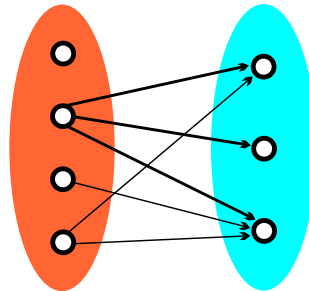


Block Types

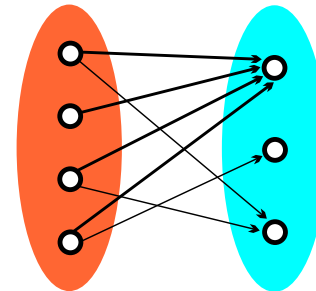
complete



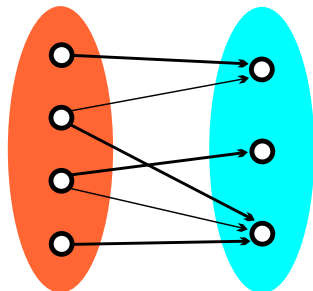
row-dominant



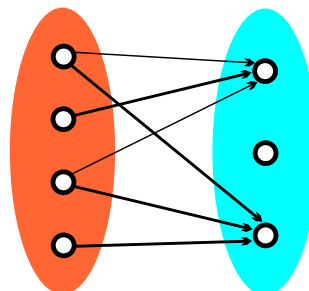
col-dominant



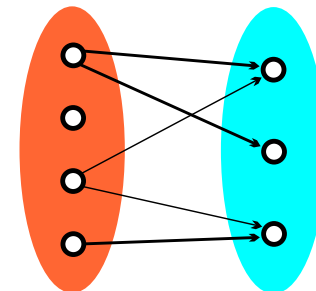
regular



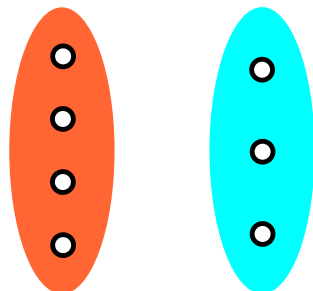
row-regular



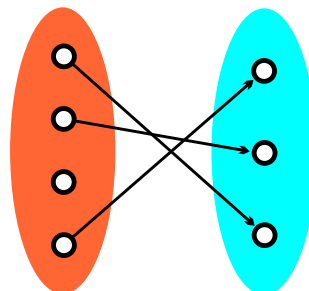
col-regular



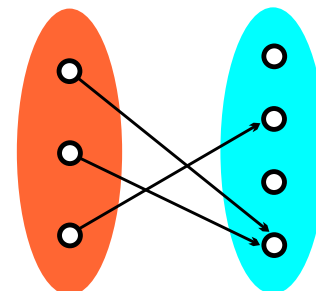
null





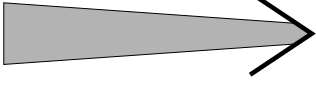







row-functional



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	

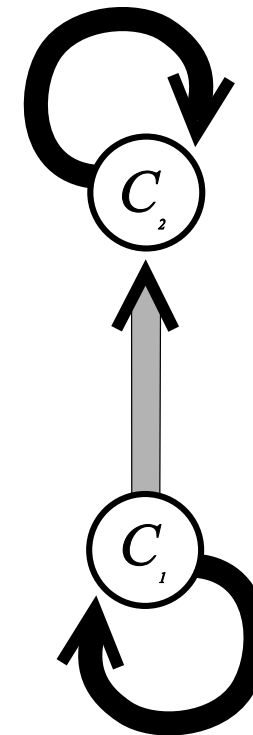
* except this may be diagonal

A block is *symmetric* iff $\forall X, Y \in C_i \times C_j : (XRY \Leftrightarrow YRX)$.

Block Types and Matrices

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
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

	C_1	C_2
C_1	complete	regular
C_2	null	complete



Optimization approach to blockmodeling

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 $\mathbf{C}^* \in \Phi$ for which

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

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

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 $\mathbf{C} = \{C_1, C_2, \dots, C_k\}$, 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.

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

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

where r_{XY} is the observed tie and b_{XY} 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 $P(\mathbf{C}) = 0 \Leftrightarrow$ the block model determined by \mathbf{C} is exact. For all presented block types sensitive criterion functions can be constructed (Batagelj, 1997).

The obtained optimization problem can be solved by local optimization. Once an optimal partitioning \mathbf{C} and types of connection 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): Given a network \mathbf{N} , a model \mathcal{M} , and partition \mathbf{C} , 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 \mathbf{C} 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 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:

*	*
*	0

center -
periphery

*	0
*	*

hierarchy

*	0
0	*

clustering

0	*
*	0

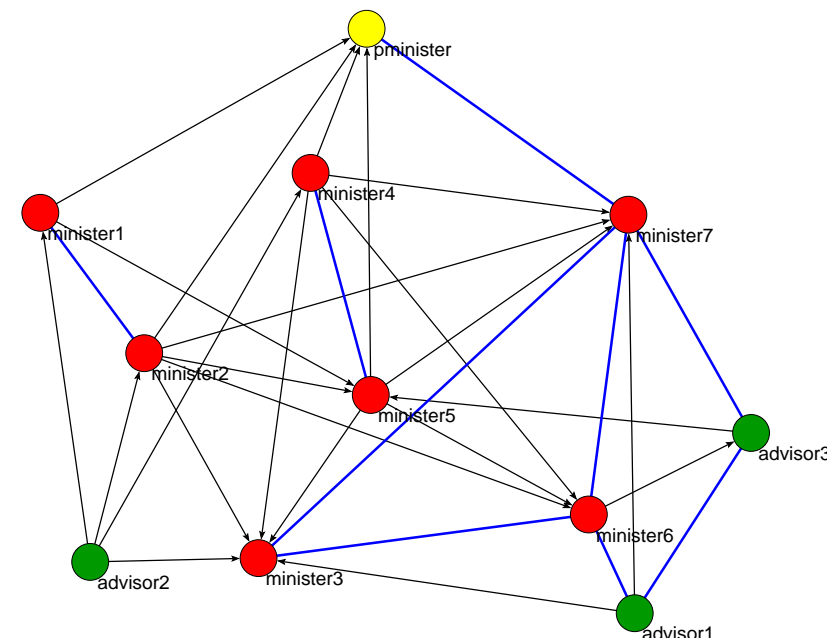
bipartition

The Student Government at the University of Ljubljana in 1992

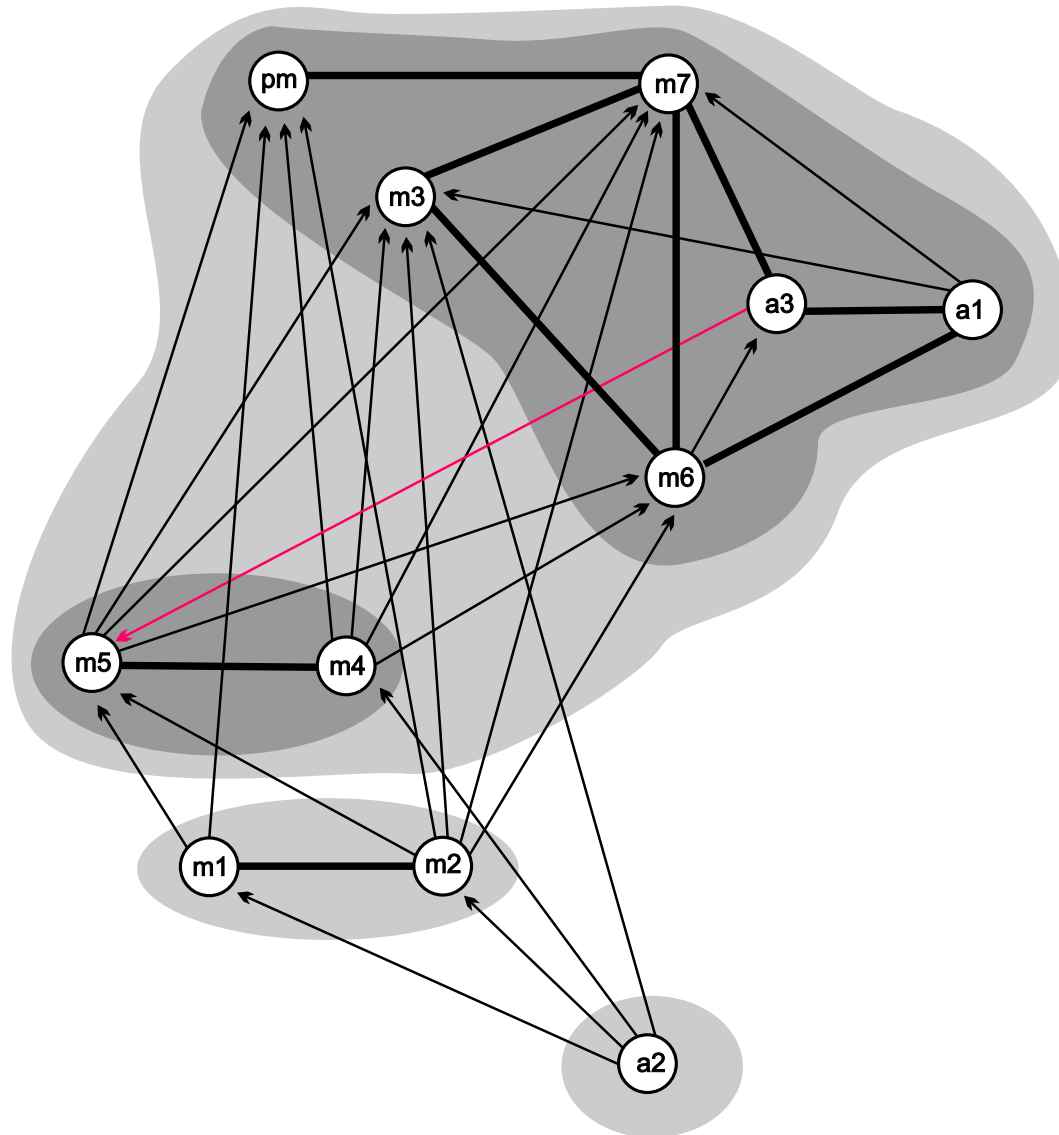
The relation is determined by the following question (Hlebec, 1993):

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

		m	p	m	m	m	m	m	m	a	a	a
		1	2	3	4	5	6	7	8	9	10	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

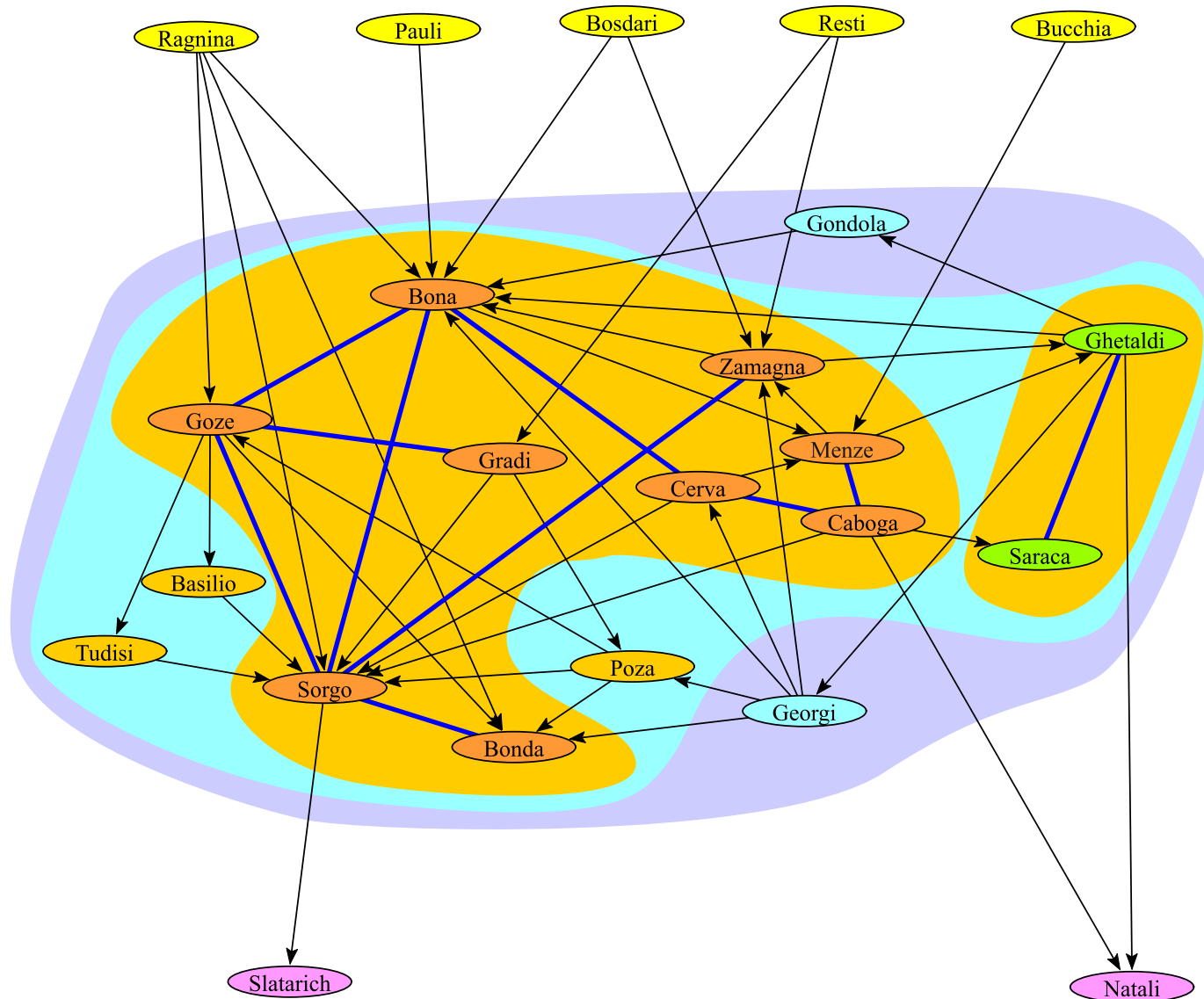


The obtained clustering in 4 clusters is almost exact. The only error is produced by the arc $(a3, m5)$.

Ragusan Noble Families Marriage Network, 18th and 19th Century

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	
Basilio	1	1	.	.	
Bona	2	1	.	.	.	2	.	2	1	.	.	
Bonda	3	2	.	.	
Bosdari	4	.	1	1	
Bucchia	5	1	
Caboga	6	1	1	1	1	.	1	.	.	
Cerva	7	.	1	.	.	.	1	1	1	.	.	
Georgi	8	.	1	2	.	.	.	1	4	1	
Ghetaldi	9	.	1	1	.	1	.	.	1	1	
Gondola	10	.	1	
Goze	11	1	2	1	2	2	2	1	.	
Gradi	12	1	1	3	.	.	
Menze	13	1	.	.	1	1	
Natali	14	
Pauli	15	.	1	
Poza	16	.	.	2	1	1	1	.	.	
Ragnina	17	.	1	1	1	1	.	.	
Resti	18	1	1	
Saraca	19	1	
Slatarich	20	
Sorgo	21	.	2	1	1	1	1	.	1
Tudisi	22	1	.	.
Zamagna	23	.	1	2	1	.	.

A Symmetric-Acyclic Decomposition of the Ragusan Families Network



Demo with Pajek

Read Network Tina.net

Net/Transform/Arcs-->Edges/Bidirected Only/Max

Draw/Draw

Layout/Energy/Kamada-Kawai/Free

Operations/Blockmodeling/Restricted Options [On]

Operations/Blockmodeling/Random Start

[4, Ranks.MDL], [Repetitions, 100], [Clusters, 4], [RUN]

extend the dialog box to see the model

Draw/Draw-Partition

Blockmodeling in 2-mode networks

A *2-mode network* is a network (\mathbf{U}, R, w) in which the vertices are partitioned into two sets \mathbf{U}_1 and \mathbf{U}_2 such that every link connects a vertex from \mathbf{U}_1 to a vertex from \mathbf{U}_2 (bipartite graph).

Some examples of 2-mode networks are:

(authors, papers; cites the paper), (authors, papers; is cited in the paper),
(authors, papers; is the (co)author of the paper), (people, events; was present at),
(people, institutions; is member of), (customers, products/services; consumption),
(articles, shopping lists; is on the list), (delegates, proposals; voting YES),
(country, country; exports to).

For 2-mode blockmodeling the solutions are pairs of clusterings (partitions) $(\mathbf{C}_r, \mathbf{C}_c)$, row-partition \mathbf{C}_r of \mathbf{U}_1 and column-partition \mathbf{C}_c of \mathbf{U}_2 .

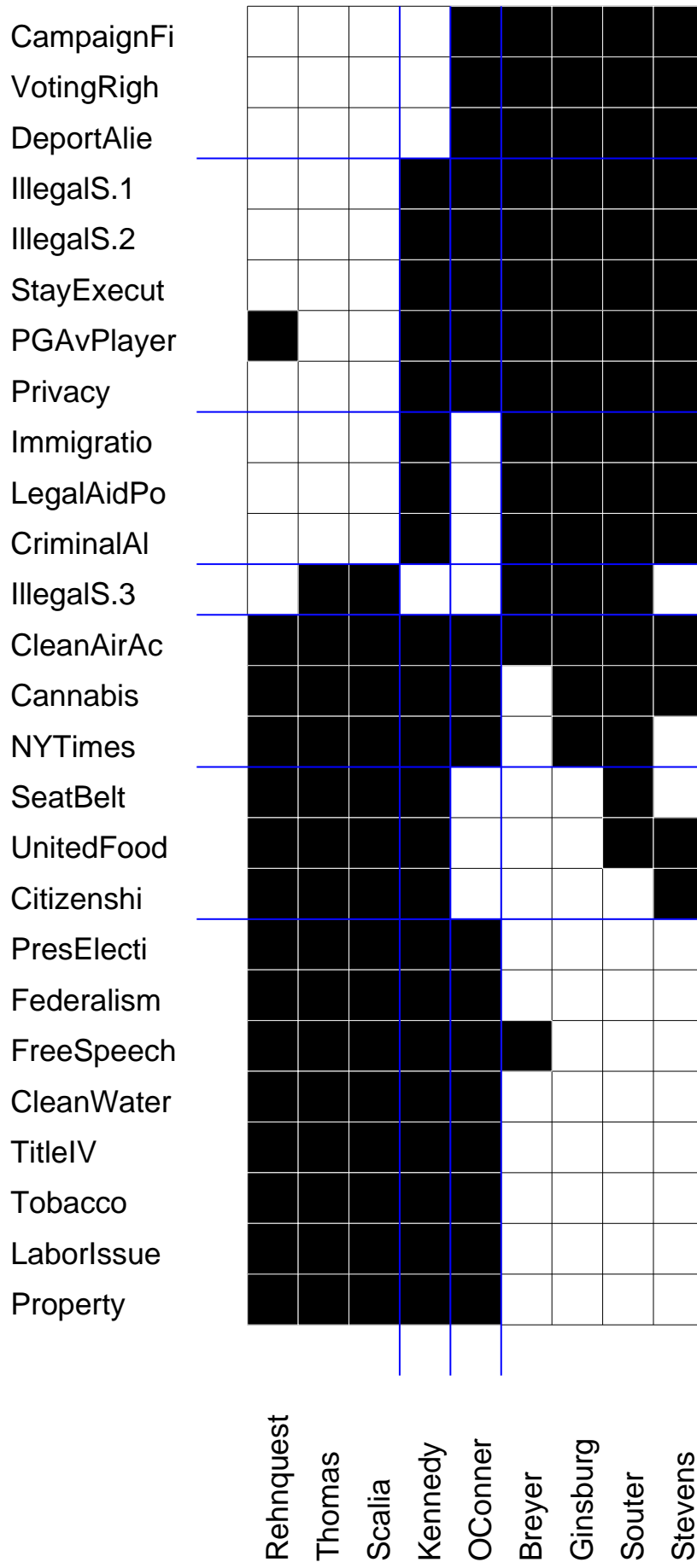
Supreme Court Voting for Twenty-Six Important Decisions

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

Journal Citation Matrix with the (5,4)-Partition

	Id	a	c	f	d	m	e	g	h	i	j	k	l	n	o	b	p	q	r
CW	a	187	32	10	58	0	11	0	0	0	7	0	0	0	0	6	0	0	7
CYSR	b	70	8	14	28	0	0	0	0	5	12	0	0	0	5	26	0	0	6
SCW	c	17	149	36	124	8	21	8	6	18	6	8	6	0	0	0	6	0	0
SW	d	52	58	53	356	15	33	15	43	8	0	0	9	0	0	0	0	0	19
JSWE	e	0	18	16	58	9	104	0	7	16	0	0	0	0	0	0	0	0	0
SSR	f	17	30	105	106	7	9	0	0	25	0	0	0	0	0	0	0	0	0
SWG	g	0	9	7	40	0	9	41	9	0	0	0	0	0	0	0	0	0	0
SWHC	h	0	20	0	26	0	0	0	86	0	0	0	0	0	0	0	0	0	0
SWRA	i	8	8	39	44	0	24	0	0	40	0	0	0	0	0	0	0	0	0
CAN	j	9	6	0	8	0	0	0	0	0	109	0	0	0	0	0	0	0	0
CSWJ	k	0	47	20	45	0	0	0	0	0	0	40	0	0	0	0	0	0	0
FR	l	0	18	0	9	0	0	0	0	0	0	0	205	0	0	0	0	0	0
ASW	m	0	0	21	73	70	18	0	0	7	0	0	0	0	0	0	0	0	13
BJSW	n	0	0	0	19	0	13	0	0	0	0	0	0	95	0	0	0	0	0
CCQ	o	12	0	0	0	0	0	0	0	0	0	0	0	0	92	0	0	0	0
JGSW	p	0	16	0	18	0	0	0	0	0	0	0	0	0	0	0	9	0	0
JSP	q	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	35	0
PW	r	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	9

com	reg	cre	rre
reg	reg	nul	nul
reg	nul	rfn	nul
rre	rre	nul	rfn
cre	nul	nul	rfn

Final Remarks

The current, local optimization based, programs for generalized blockmodeling can deal only with networks with at most some hundreds of units. What to do with larger networks is an open question. For some specialized problems also procedures for (very) large networks can be developed (Doreian, Batagelj, Ferligoj, 1998; Batagelj, Zaveršnik, 2002). Another interesting problem is the development of *blockmodeling of valued networks* or more general *relational data analysis* (Batagelj, Ferligoj, 2000).

Most of described procedures are implemented in Pajek – program for analysis and visualization of large networks. It is freely available, for noncommercial use, at:

`http://vlado.fmf.uni-lj.si/pub/networks/pajek/`

Several papers and presentations on blockmodeling and network analysis are available at:

`http://vlado.fmf.uni-lj.si/pub/networks/doc/`

The author's e-mail address is: **`vladimir.batagelj@uni-lj.si`**

References

- [1] Batagelj, V., Ferligoj, A., and Doreian, P. (1992), Direct and Indirect Methods for Structural Equivalence. *Social Networks* **14**, 63-90.
- [2] Batagelj, V., Doreian, P., and Ferligoj, A. (1992), An Optimizational Approach to Regular Equivalence. *Social Networks* **14**, 121-135.
- [3] Batagelj, V. (1997), Notes on blockmodeling. *Social Networks* **19**, 143-155.
- [4] De Nooy, W., Mrvar A., Batagelj, V. (2002), Exploratory Social Network Analysis with Pajek. Manuscript of a textbook.
- [5] Doreian, P., Batagelj, V., Ferligoj, A. (2000), Symmetric-acyclic decompositions of networks. *J. classif.*, **17**(1), 3-28.
- [6] Doreian, P., Batagelj, V., Ferligoj, A. (2002), Generalized blockmodeling. Manuscript of a book.
- [7] Everett, M.E and Borgatti, S.P. (1996), Exact Colorations of Graphs and Digraphs. *Social Networks*, **18**, 319-331.
- [8] Faust, K. (1988), Comparison of Methods for Positional Analysis: Structural and General Equivalences, *Social Networks*, **10**, 313-341.
- [9] Lorrain, F. and H.C. White (1971): Structural equivalence of individuals in social networks. *Journal of Mathematical Sociology*, **1**:49–80.
- [10] Murtagh, F. (1985), Multidimensional Clustering Algorithms, *Compstat lectures*, **4**, Vienna: Physica-Verlag.
- [11] White, D.R. and K.P. Reitz (1983), Graph and semigroup homomorphisms on networks of relations. *Social Networks*, **5**:193–234.