Generalized blockmodeling of 2-mode networks

Vladimir Batagelj, Anuška Ferligoj
University of Ljubljana, Slovenia

Patrick Doreian
University of Pittsburgh, USA

IFCS 2002

July 15-19, 2002, Cracow, Poland

Outline

- matrix rearrangement view on blockmodeling
- blockmodeling and clustering, 2-mode blockmodeling
- generalized blockmodeling
- prespecified blockmodels
- examples

2-mode networks

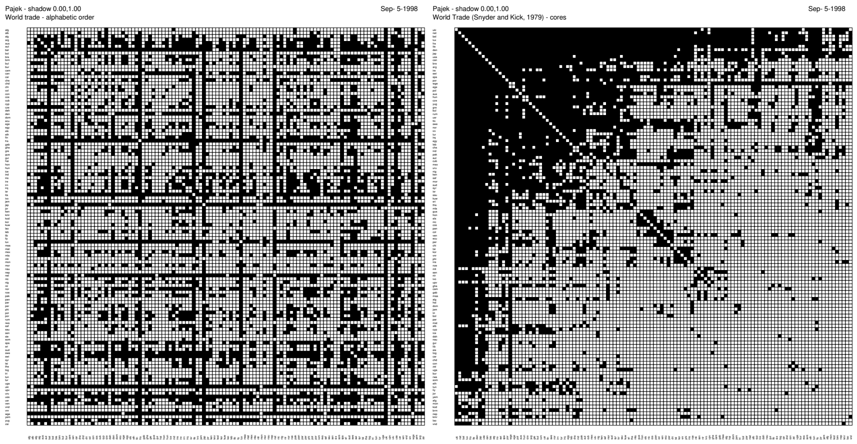
A *2-mode network* is a structure $(\mathbf{U}_1, \mathbf{U}_2, R, w)$, where $\mathbf{U}_1$ and $\mathbf{U}_2$ are disjoint sets of vertices (nodes, units), $R \subseteq \mathbf{U}_1 \times \mathbf{U}_2$ is a relation and $w : R \to \mathbb{R}$ is a weight. If no weight is defined we can assume a constant weight $w(u, v) = 1$ for all $(u, v) \in R$.

A 2-mode network can be viewed also as an ordinary (1-mode) network on the vertex set $\mathbf{U}_1 \cup \mathbf{U}_2$, divided into two sets $\mathbf{U}_1$ and $\mathbf{U}_2$, where the arcs can only go from $\mathbf{U}_1$ to $\mathbf{U}_2$ (a bipartite directed 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).

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 network 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:

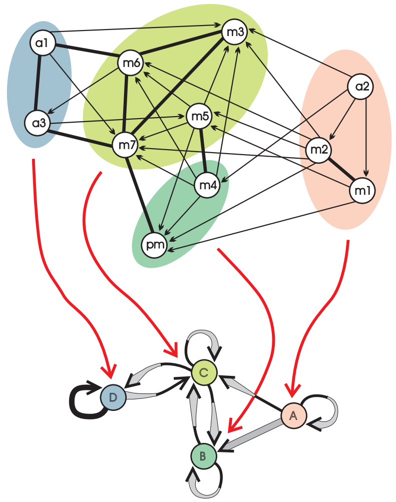
- ordering by (in/out)-degree;
- ordering by connected components;
- ordering by (in/out)-core numbers, connected components inside core levels, and degree;
- ordering according to a hierarchical clustering and some other property.

There exist also some special procedures to determine the ordering such us *seriation* and *clumping* (Murtagh 1985).

Another view is to look for *partitions* of rows and columns that determine the *blocks* inside the rearranged matrix.

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 (ordinary, 1-mode) blockmodeling is to identify, in a given network $\mathbf{N} = (\mathbf{U}, R, w)$, $R \subseteq \mathbf{U} \times \mathbf{U}$, *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* $\mathbf{C} = \{C_1, C_2, \ldots, C_k\}$ 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.

In blockmodeling of a 2-mode network $\mathbf{N} = (\mathbf{U}_1, \mathbf{U}_2, R, w)$ we are trying to identify two clusterings – $\mathbf{C}_r$ of $\mathbf{U}_1$ and $\mathbf{C}_c$ of $\mathbf{U}_2$ such that they induce selected blocks.

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 (Faust, 1988):

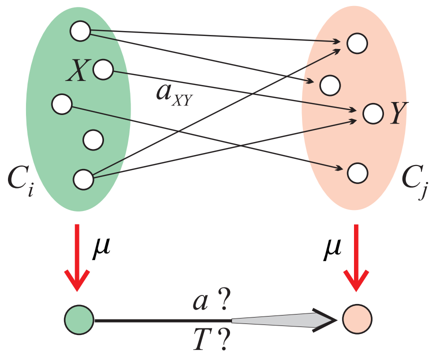
- the equivalent units have the same connection pattern to the **same** neighbors;
- the equivalent units have the same or similar connection pattern to (possibly) **different** neighbors.

The first type of equivalence is formalized by the notion of *structural* equivalence (Lorrain and White, 1971) and the second by the notion of *regular* equivalence (White and Reitz, 1983) with the latter a generalization of the former.

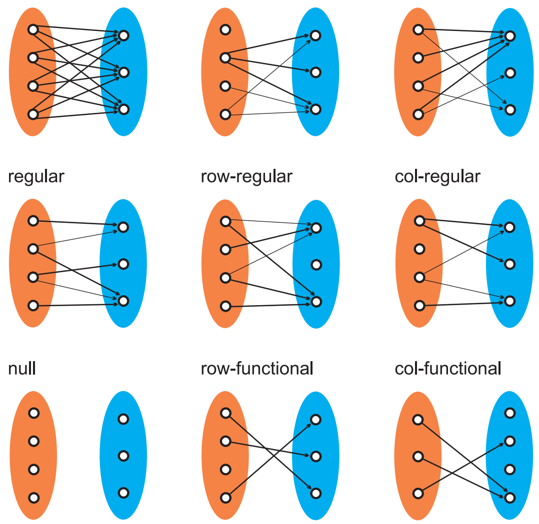
Generalized Blockmodeling

A *blockmodel* consists of structures obtained by identifying (shrinking into a model vertex) all units from the same cluster. 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 2-mode blockmodeling $\mathbf{C}_i \subset \mathbf{U}_1$ and $\mathbf{C}_j \subset \mathbf{U}_2$.

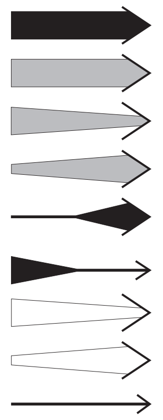


Block Types



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	∃ all 1 row *
col-dominant	cdo	∃ all 1 column *
row-functional	rfn	∃! one 1 in each row
col-functional	cfn	∃! one 1 in each column
non-null	one	∃ at least one 1



* except this may be diagonal

Equivalences

Let $\sim$ be an equivalence relation over $\mathbf{U}$ and $[X] = \{Y \in \mathbf{U} : X \sim Y\}$. We say that $\sim$ is *compatible* with the set of types $\mathcal{T}$ over a network $\mathbf{N}$ iff

$$\forall X, Y \in \mathbf{U} \exists T \in \mathcal{T} : T([X], [Y]).$$

It is easy to verify that the notion of compatibility for $\mathcal{T} = \{\text{nul}, \text{reg}\}$ reduces to the usual definition of regular equivalence. Similarly, compatibility for $\mathcal{T} = \{\text{nul}, \text{com}\}$ reduces to structural equivalence.

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

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

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

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

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 solution $(\mathbf{C}_r, \mathbf{C}_c)$, 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

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

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

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

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 $P(\mathbf{C}_r, \mathbf{C}_c)$ is *sensitive* iff $P(\mathbf{C}_r, \mathbf{C}_c) = 0 \Leftrightarrow$ the blockmodel determined by $(\mathbf{C}_r, \mathbf{C}_c))$ 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 a partitions $(\mathbf{C}_r, \mathbf{C}_c)$ and types of blocks 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 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 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.

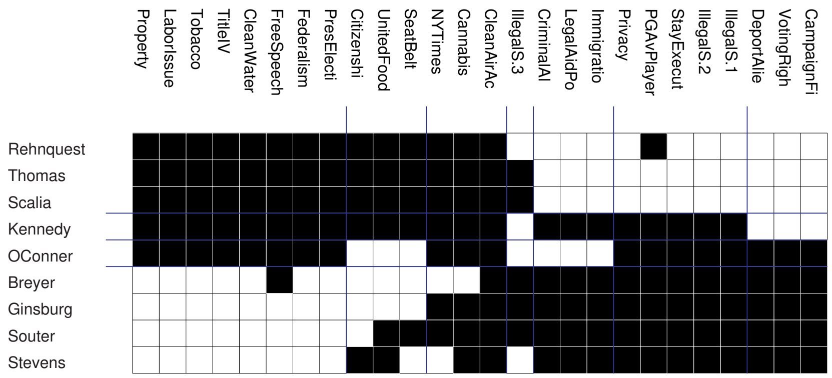
Supreme Court Voting for Twenty-Six Important Decisions

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).

The generalized blockmodeling of ordinary (1-mode) networks is 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/`

For blockmodeling of 2-mode networks a special program `TwoModel` is available

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

The current version of these slides is available at:

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