

# Blockmodeling 

## Vladimir Batagelj

University of Ljubljana

Figure 2. Participation of the Southern Women in Events

Networks Workshop
NICTA, Sydney, June 2005

## Outline

1 Matrix rearrangement view on blockmodeling ..... 1
3 Blockmodeling as a clustering problem ..... 3
12 Indirect Approach ..... 12
23 Direct Approach and Generalized Blockmodeling ..... 23
38 Pre-specified blockmodeling ..... 38
48 Blockmodeling in 2-mode networks ..... 48
51 Signed graphs ..... 51

## 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) or RCM - Reverse Cuthill-McKee algorithm.

## 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 $\mathcal{N}=(\mathcal{U}, R), R \subseteq \mathcal{U} \times \mathcal{U}$, 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 $\mathbf{C}=\left\{C_{1}, C_{2}, \ldots, C_{k}\right\}$ which is a partition of the set $\mathcal{U}$. 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

$$
R\left(C_{i}, C_{j}\right)=R \cap C_{i} \times C_{j}
$$

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 $R\left(C_{i}, C_{i}\right)$ 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:

$$
\begin{array}{llll}
\text { s1. } & \mathrm{X} R \mathrm{Y} \Leftrightarrow \mathrm{Y} R \mathrm{X} & \text { s3. } & \forall \mathrm{Z} \in \mathcal{U} \backslash\{\mathrm{X}, \mathrm{Y}\}:(\mathrm{X} R \mathrm{Z} \Leftrightarrow \mathrm{Y} R \mathrm{Z}) \\
\text { s2. } & \mathrm{X} R \mathrm{X} \Leftrightarrow \mathrm{Y} R \mathrm{Y} & \text { s4. } & \forall \mathrm{Z} \in \mathcal{U} \backslash\{\mathrm{X}, \mathrm{Y}\}:(\mathrm{Z} R \mathrm{X} \Leftrightarrow \mathrm{Z} R \mathrm{Y})
\end{array}
$$

## ...structural equivalence

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

| 0 | 0 | 0 | 0 | 0 |  |  |  | 1 | 0 | 0 | 0 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 0 | 0 | 0 | 0 | 0 |  |  | 0 | 1 | 0 | 0 |  |
| 0 | 0 | 0 | 0 | 0 |  |  | 0 | 0 | 1 | 0 |  |
| 0 | 0 | 0 | 0 | 0 |  |  | 0 | 0 | 0 | 1 |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
|  | 1 | 1 | 1 | 1 | 1 |  |  |  | 0 | 1 | 1 |
| 1 | 1 |  |  |  |  |  |  |  |  |  |  |
| 1 | 1 | 1 | 1 | 1 |  |  | 1 | 0 | 1 | 1 |  |
| 1 | 1 | 1 | 1 | 1 |  |  | 1 | 1 | 0 | 1 |  |
| 1 | 1 | 1 | 1 | 1 |  |  | 1 | 1 | 1 | 0 |  |

## 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 $\mathcal{U}$ is a regular equivalence on network $\mathcal{N}=(\mathcal{U}, R)$ if and only if for all $\mathrm{X}, \mathrm{Y}, \mathrm{Z} \in \mathcal{U}$, $\mathrm{X} \approx \mathrm{Y}$ implies both

$$
\begin{array}{ll}
\mathrm{R} 1 . & \mathrm{X} R \mathrm{Z} \Rightarrow \exists \mathrm{~W} \in \mathcal{U}:(\mathrm{Y} R \mathrm{~W} \wedge \mathrm{~W} \approx \mathrm{Z}) \\
\mathrm{R} 2 . & \mathrm{Z} R \mathrm{X} \Rightarrow \exists \mathrm{~W} \in \mathcal{U}:(\mathrm{W} R \mathrm{Y} \wedge \mathrm{~W} \approx \mathrm{Z})
\end{array}
$$

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

## ... regular equivalence

Theorem 1 (Batagelj, Doreian, Ferligoj, 1992) Let $\mathbf{C}=\left\{C_{i}\right\}$ be a partition corresponding to a regular equivalence $\approx$ on the network $\mathcal{N}=$ $(\mathcal{U}, R)$. Then each block $R\left(C_{u}, C_{v}\right)$ 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.

| 0 | 0 | 0 | 0 | 0 |  |  | 1 | 0 | 1 | 0 | 0 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| 0 | 0 | 0 | 0 | 0 |  | 0 | 0 | 1 | 0 | 1 |  |
| 0 | 0 | 0 | 0 | 0 |  | 0 | 1 | 0 | 0 | 0 |  |
| 0 | 0 | 0 | 0 | 0 |  |  | 1 | 0 | 1 | 1 | 0 |

## 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 $(\Phi, P)$ as follows:

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

$$
P\left(\mathbf{C}^{\star}\right)=\min _{\mathbf{C} \in \Phi} P(\mathbf{C})
$$

where $\Phi$ is the set of feasible clusterings and $P$ is a criterion function.
Since the set of units $\mathcal{U}$ is finite, the set of feasible clusterings is also finite. Therefore the set $\operatorname{Min}(\Phi, P)$ 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).


## Indirect Approach



## Dissimilarities

The property $t: \mathcal{U} \rightarrow \mathbb{R}$ is structural property if, for every automorphism $\varphi$, of the relation $R$, and every unit $x \in \mathcal{U}$, it holds that $t(x)=t(\varphi(x))$. Some examples of a structural property include

$$
\begin{aligned}
& t(u)=\text { the degree of unit } u \\
& t(u)=\text { number of units at distance } d \text { from the unit } u ; \\
& t(u)=\text { number of triads of type } x \text { at the unit } u .
\end{aligned}
$$

Centrality measures are further examples of structural properties.
We can define the description of the unit $u$ as $[u]=\left[t_{1}(u), t_{2}(u), \ldots, t_{m}(u)\right]$. As a simple example, $t_{1}$ could be degree centrality, $t_{2}$ could be closeness centrality and $t_{3}$ could be betweenness centrality. The dissimilarity between units $u$ and $v$ could be defined as $d(u, v)=D([u],[v])$ where $D$ is some (standard) dissimilarity between real vectors. In the simple example, $D$ could be the Euclidean distance between the centrality profiles.

## Dissimilarities based on matrices

We consider the following list of dissimilarities between units $x_{i}$ and $x_{j}$ where the description of the unit consists of the row and column of the property matrix $\mathbf{Q}=\left[q_{i j}\right]$. We take as units the rows of the matrix

$$
\mathbf{X}=\left[\mathbf{Q} \mathbf{Q}^{T}\right]
$$

## ...Dissimilarities

Manhattan distance: $d_{m}\left(x_{i}, x_{j}\right)=\sum_{s=1}^{n}\left(\left|q_{i s}-q_{j s}\right|+\left|q_{s i}-q_{s j}\right|\right)$
Euclidean distance:

$$
d_{E}\left(x_{i}, x_{j}\right)=\sqrt{\sum_{s=1}^{n}\left(\left(q_{i s}-q_{j s}\right)^{2}+\left(q_{s i}-q_{s j}\right)^{2}\right)}
$$

Truncated Manhattan distance: $d_{s}\left(x_{i}, x_{j}\right)=\sum_{\substack{s=1 \\ s \neq i, j}}^{n}\left(\left|q_{i s}-q_{j s}\right|+\left|q_{s i}-q_{s j}\right|\right)$
Truncated Euclidean distance (Faust, 1988):

$$
d_{S}\left(x_{i}, x_{j}\right)=\sqrt{\sum_{\substack{s=1 \\ s \neq i, j}}^{n}\left(\left(q_{i s}-q_{j s}\right)^{2}+\left(q_{s i}-q_{s j}\right)^{2}\right)}
$$

## ...Dissimilarities

Corrected Manhattan-like dissimilarity $(p \geq 0)$ :

$$
d_{c}(p)\left(x_{i}, x_{j}\right)=d_{s}\left(x_{i}, x_{j}\right)+p \cdot\left(\left|q_{i i}-q_{j j}\right|+\left|q_{i j}-q_{j i}\right|\right)
$$

Corrected Euclidean-like dissimilarity (Burt and Minor, 1983):

$$
d_{e}(p)\left(x_{i}, x_{j}\right)=\sqrt{d_{S}\left(x_{i}, x_{j}\right)^{2}+p \cdot\left(\left(q_{i i}-q_{j j}\right)^{2}+\left(q_{i j}-q_{j i}\right)^{2}\right)}
$$

Corrected dissimilarity:

$$
d_{C}(p)\left(x_{i}, x_{j}\right)=\sqrt{d_{c}(p)\left(x_{i}, x_{j}\right)}
$$

The parameter, $p$, can take any positive value. Typically, $p=1$ or $p=2$, where these values count the number of times the corresponding diagonal pairs are counted.

## ...Dissimilarities

It is easy to verify that all expressions from the list define a dissimilarity (i.e. that $d(x, y) \geq 0 ; d(x, x)=0$; and $d(x, y)=d(y, x)$ ). Each of the dissimilarities from the list can be assessed to see whether or not it is also a distance: $d(x, y)=0 \Rightarrow x=y$ and $d(x, y)+d(y, z) \geq d(x, z)$.

The dissimilarity measure $d$ is compatible with a considered equivalence $\sim$ if for each pair of units holds

$$
X_{i} \sim X_{j} \Leftrightarrow d\left(X_{i}, X_{j}\right)=0
$$

Not all dissimilarity measures typically used are compatible with structural equivalence. For example, the corrected Euclidean-like dissimilarity is compatible with structural equivalence.

The indirect clustering approach does not seem suitable for establishing clusterings in terms of regular equivalence since there is no evident way how to construct a compatible (dis)similarity measure.

## 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.

## Indirect approach



Using Corrected Euclidean-like dissimilarity and Ward clustering method we obtain the following dendrogram.
From it we can determine the number of clusters: 'Natural' clusterings correspond to clear 'jumps' in the dendrogram.

If we select 3 clusters we get the partition C.

$$
\begin{aligned}
\mathbf{C}= & \{\{b 51, b 89, b 02, b 96, b 03, b 85, g 10, g 24\}, \\
& \{g 09, g 63, g 12\},\{g 07, g 28, g 22, g 42\}\}
\end{aligned}
$$



## Matrix

Pajek - shadow [0.00,1.00]
b02
b03
g10
g24
b51
b85
b89
b96
g07
g22
g28
g42
g09
g12
g63


The partition can be used also to reorder rows and columns of the matrix representing the network. Clusters are divided using blue vertical and horizontal lines.

## Direct Approach and Generalized Blockmodeling

The second possibility for solving the blockmodeling problem is to construct an appropriate criterion function directly and then use a local optimization algorithm to obtain a 'good' clustering solution.

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 $\mathbf{C}$. 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.


## Generalized equivalence / Block Types




|  | $Y$ |  | $Y$ |
| :---: | :---: | :---: | :---: |
| $X$ | $0 \begin{array}{lllll}0 & 0 & 0 & 0 & 0\end{array}$ | $X$ | $\begin{array}{lllll}0 & 0 & 0 & 1 & 0\end{array}$ |
|  | $\begin{array}{lllll}0 & 0 & 0 & 0 & 0\end{array}$ |  | $\begin{array}{llllll}0 & 0 & 1 & 0 & 0\end{array}$ |
|  | $\begin{array}{llllll}0 & 0 & 0 & 0 & 0\end{array}$ |  | $\begin{array}{lllll}1 & 0 & 0 & 0 & 0\end{array}$ |
|  | $\begin{array}{llllll}0 & 0 & 0 & 0 & 0\end{array}$ |  | $\begin{array}{lllll}0 & 0 & 0 & 1 & 0\end{array}$ |
|  | null |  | ow-functional |


|  |  | $Y$ |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
|  | 1 | 0 | 0 | 0 |  |
| $X$ | 0 | 1 | 0 | 0 |  |
|  | 0 | 0 | 1 | 0 |  |
|  | 0 | 0 | 0 | 0 |  |
|  | 0 | 0 | 0 | 1 |  |
| col-functional |  |  |  |  |  |

## Characterizations of Types of Blocks

| null | nul | all $0^{*}$ |
| :--- | :--- | :--- | :--- |
| complete | com | all $1^{*}$ |
| regular | 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 |  |

* except this may be diagonal

A block is symmetric iff $\forall \mathrm{X}, \mathrm{Y} \in C_{i} \times C_{j}:(\mathrm{X} R \mathrm{Y} \Leftrightarrow \mathrm{Y} R \mathrm{X})$.


## Formalization of blockmodeling

Let $V$ be a set of positions or images of clusters of units. Let $\mu: \mathcal{U} \rightarrow V$ denote a mapping which maps each unit to its position. The cluster of units $C(t)$ with the same position $t \in V$ is

$$
C(t)=\mu^{-1}(t)=\{\mathrm{X} \in \mathcal{U}: \mu(\mathrm{X})=t\}
$$

Therefore

$$
\mathbf{C}(\mu)=\{C(t): t \in V\}
$$

is a partition (clustering) of the set of units $\mathcal{U}$.

## Blockmodel

A blockmodel is an ordered sextuple $\mathcal{M}=(V, K, \mathcal{T}, Q, \pi, \alpha)$ where:

- $V$ is a set of types of units (images or representatives of classes);
- $K \subseteq V \times V$ is a set of connections;
- $\mathcal{T}$ is a set of predicates used to describe the types of connections between different classes (clusters, groups, types of units) in a network. We assume that nul $\in \mathcal{T}$. A mapping $\pi: K \rightarrow \mathcal{T} \backslash\{$ nul $\}$ assigns predicates to connections;
- $Q$ is a set of averaging rules. A mapping $\alpha: K \rightarrow Q$ determines rules for computing values of connections.

A (surjective) mapping $\mu: \mathcal{U} \rightarrow V$ determines a blockmodel $\mathcal{M}$ of network $\mathcal{N}$ iff it satisfies the conditions: $\forall(t, w) \in K: \pi(t, w)(C(t), C(w))$ and $\forall(t, w) \in V \times V \backslash K: \operatorname{nul}(C(t), C(w))$.

## Equivalences

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

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

It is easy to verify that the notion of compatibility for $\mathcal{T}=\{$ nul, reg $\}$ reduces to the usual definition of regular equivalence (White and Reitz 1983). Similarly, compatibility for $\mathcal{T}=\{$ nul, com $\}$ reduces to structural equivalence (Lorrain and White 1971).

For a compatible equivalence $\sim$ the mapping $\mu: \mathrm{X} \mapsto[\mathrm{X}]$ determines a blockmodel with $V=\mathcal{U} / \sim$.

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.

## 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 $\mathbf{C}=\left\{C_{1}, C_{2}, \ldots, C_{k}\right\}$, let $\mathcal{B}\left(C_{u}, C_{v}\right)$ denote the set of all ideal blocks corresponding to block $R\left(C_{u}, C_{v}\right)$. Then the global error of clustering $\mathbf{C}$ can be expressed as

$$
P(\mathbf{C})=\sum_{C_{u}, C_{v} \in \mathbf{C}} \min _{B \in \mathcal{B}\left(C_{u}, C_{v}\right)} d\left(R\left(C_{u}, C_{v}\right), B\right)
$$

where the term $d\left(R\left(C_{u}, C_{v}\right), B\right)$ measures the difference (error) between the block $R\left(C_{u}, C_{v}\right)$ 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 $d\left(R\left(C_{u}, C_{v}\right), B\right)$ can be expressed, for non-diagonal blocks, as

$$
d\left(R\left(C_{u}, C_{v}\right), B\right)=\sum_{\mathrm{X} \in C_{u}, \mathrm{Y} \in C_{v}}\left|r_{\mathrm{XY}}-b_{\mathrm{XY}}\right|
$$

where $r_{\mathrm{XY}}$ is the observed tie and $b_{\mathrm{XY}}$ is the corresponding value in an ideal block. This criterion function counts the number of 1 s in erstwhile null blocks and the number of 0 s 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).

## Deviations Measures for Types of Blocks

We can efficiently test whether a block $R(X, Y)$ is of the type $T$ by making use of the characterizations of block types. On this basis we can construct the corresponding deviation measures. The quantities used in the expressions for deviations have the following meaning:

Throughout the number of elements in a block is $n_{r} n_{c}$.

## ... Deviations Measures for Types of Blocks

| Connection | $\delta(X, Y ; T)$ |  |
| :---: | :---: | :---: |
| null | $\left\{\begin{array}{l} s_{t} \\ s_{t}+d-s_{d} \end{array}\right.$ | nondiagonal diagonal |
| complete | $\left\{\begin{array}{l} n_{r} n_{c}-s_{t} \\ n_{r} n_{c}-s_{t}+d+s_{d}-n_{r} \end{array}\right.$ | nondiagonal diagonal |
| row-dominant | $\left\{\begin{array}{l} \left(n_{c}-m_{r}-1\right) n_{r} \\ \left(n_{c}-m_{r}\right) n_{r} \end{array}\right.$ | diagonal, $s_{d}=0$ otherwise |
| col-dominant | $\left\{\begin{array}{l} \left(n_{r}-m_{c}-1\right) n_{c} \\ \left(n_{r}-m_{c}\right) n_{c} \end{array}\right.$ | $\text { diagonal }, s_{d}=0$ otherwise |
| row-regular | $\left(n_{r}-p_{r}\right) n_{c}$ |  |
| col-regular | $\left(n_{c}-p_{c}\right) n_{r}$ |  |
| regular | $\left(n_{c}-p_{c}\right) n_{r}+\left(n_{r}-p_{r}\right) p_{c}$ |  |
| row-functional | $s_{t}-p_{r}+\left(n_{r}-p_{r}\right) n_{c}$ |  |
| col-functional | $s_{t}-p_{c}+\left(n_{c}-p_{c}\right) n_{r}$ |  |
| density $\gamma$ | $\max \left(0, \gamma n_{r} n_{c}-s_{t}\right)$ |  |

For the null, complete, row-dominant and column-dominant blocks it is necessary to distinguish diagonal blocks and non-diagonal blocks.

## Solving the blockmodeling problem

The obtained optimization problem can be solved by local optimization.
Once a partitioning $\mu$ and types of connection $\pi$ are determined, we can also compute the values of connections by using averaging rules.

## Benefits from Optimization Approach

- ordinary / inductive blockmodeling: Given a network $\mathcal{N}$ 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 $\mathcal{N}$, a model $\mathcal{M}$, and blockmodeling $\mu$, compute the corresponding criterion function;
- model fitting / deductive blockmodeling: Given a network $\mathcal{N}$, set of types $\mathcal{T}$, and a family of models, determine $\mu$ 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:

| $*$ | $*$ |
| :---: | :---: |
| $*$ | 0 |

center periphery

hierarchy

clustering bipartition

## 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:

$$
\begin{aligned}
\mathbf{C}= & \{\{b 02, b 03, b 51, b 85, b 89, b 96, g 09\}, \\
& \{g 07, g 10, g 12, g 22, g 24, g 28, g 42, g 63\}\}
\end{aligned}
$$




## The Student Government at the University of Ljubljana in

 1992The 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 | . | . | . | . | $\cdot$ |
| 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 $\operatorname{arc}(a 3, m 5)$.

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

|  |  | 1 | 2 | 3 | 4 | 5 | 6 | 7 |  |  | 9 | 10 |  | 12 |  |  |  |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 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 |  |  |  | . |
| 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 |
| 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

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



## Signed graphs

A signed graph is an ordered pair $(\mathcal{G}, \sigma)$ where

- $\mathcal{G}=(\mathcal{V}, R)$ is a directed graph (without loops) with set of vertices $\mathcal{V}$ and set of $\operatorname{arcs} R \subseteq \mathcal{V} \times \mathcal{V}$;
- $\sigma: R \rightarrow\{p, n\}$ 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 $(\mathcal{G}, \sigma)$ is:

- balanced iff the set of vertices $\mathcal{V}$ 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 $\mathcal{V}$ 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 $(\mathcal{G}, \sigma)$ is balanced iff every closed semiwalk is positive.

THEOREM 2. A signed graph $(\mathcal{G}, \sigma)$ is clusterable iff $\mathcal{G}$ contains no closed semiwalk with exactly one negative arc.

## Chartrand's example - graph



In the figure the graph from Chartarand (1985, p. 181) and its value matrix are given. The positive edges are drawn with solid lines, and the negative edges with dotted lines.

## Chartrand's example - closures

|  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |  | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1 | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | 1 | p | $n$ | $n$ | $n$ | p | p | $n$ | $n$ | p |
| 2 | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | 2 | $n$ | p | p | p | $n$ | $n$ | $n$ | $n$ | $n$ |
| 3 | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | 3 | $n$ | p | p | p | $n$ | $n$ | $n$ | $n$ | $n$ |
| 4 | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | 4 | $n$ | p | p | p | $n$ | $n$ | $n$ | $n$ | $n$ |
| 5 | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | 5 | p | $n$ | $n$ | $n$ | p | p | $n$ | $n$ | p |
| 6 | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | 6 | p | $n$ | $n$ | $n$ | p | p | $n$ | $n$ | p |
| 7 | $a$ | $a$ | $a$ | $a$ | $a$ | a | $a$ | $a$ | $a$ | 7 | $n$ | $n$ | $n$ | $n$ | $n$ | $n$ | p | p | $n$ |
| 8 | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | 8 | $n$ | $n$ | $n$ | $n$ | $n$ | $n$ | p | p | $n$ |
| 9 | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | $a$ | 9 | p | $n$ | $n$ | $n$ | p | p | $n$ | $n$ | p |

On the left side of the table the corresponding balance-closure is given the graph is not balanced. From the cluster-closure on the right side of the table we can see that the graph is clusterable and it has the clusters

$$
\mathcal{V}_{1}=\{1,5,6,9\}, \quad \mathcal{V}_{2}=\{2,3,4\}, \quad \mathcal{V}_{3}=\{7,8\}
$$

## 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 $\mathbf{C}=\left\{C_{1}, C_{2}, \ldots, C_{k}\right\}$ can be therefore measured as follows ( $0 \leq \alpha \leq 1$ ):

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

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.

