SNA

Ferligoj

Introduction to Blockmodeling

Anuška Ferligoj

UL, Ljubljana, Slovenia
NRU HSE, Moscow, Russia

Outline

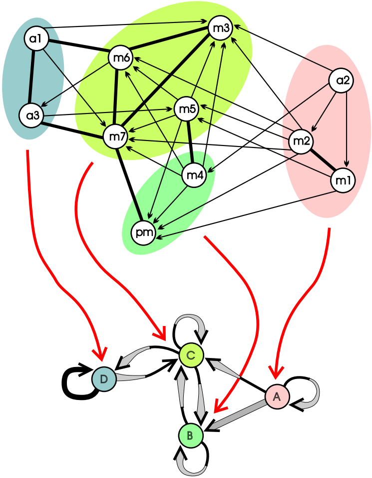
Introduction

SNA

Ferligoj

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 $\mathbf{N} = (\mathcal{U}, R)$, $R \subseteq \mathcal{U} \times \mathcal{U}$, *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* $\mathbf{C} = \{C_1, C_2, \dots, C_k\}$ which can be 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*

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

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.

The Everett Network

	a	b	c	d	e	f	g	h	i	j
a	0	1	1	1	0	0	0	0	0	0
b	1	0	1	0	1	0	0	0	0	0
c	1	1	0	1	0	0	0	0	0	0
d	1	0	1	0	1	0	0	0	0	0
e	0	1	0	1	0	1	0	0	0	0
f	0	0	0	0	1	0	1	0	1	0
g	0	0	0	0	0	1	0	1	0	1
h	0	0	0	0	0	0	1	0	1	1
i	0	0	0	0	0	1	0	1	0	1
j	0	0	0	0	0	0	1	1	1	0

	a	c	h	j	b	d	g	i	e	f
a	0	1	0	0	1	1	0	0	0	0
c	1	0	0	0	1	1	0	0	0	0
h	0	0	0	1	0	0	1	1	0	0
j	0	0	1	0	0	0	1	1	0	0
b	1	1	0	0	0	0	0	0	1	0
d	1	1	0	0	0	0	0	0	1	0
g	0	0	1	1	0	0	0	0	0	1
i	0	0	1	1	0	0	0	0	0	1
e	0	0	0	0	1	1	0	0	0	1
f	0	0	0	0	0	0	1	1	1	0

	A	B	C
A	1	1	0
B	1	0	1
C	0	1	1

Equivalences

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 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*.

In other words, x and y are structurally equivalent iff:

s1. $\mathrm{x}R\mathrm{y} \Leftrightarrow \mathrm{y}R\mathrm{x}$

s2. $\mathrm{x}R\mathrm{x} \Leftrightarrow \mathrm{y}R\mathrm{y}$

s3. $\forall \mathrm{z} \in \mathcal{U} \setminus \{\mathrm{x}, \mathrm{y}\} : (\mathrm{x}R\mathrm{z} \Leftrightarrow \mathrm{y}R\mathrm{z})$

s4. $\forall \mathrm{z} \in \mathcal{U} \setminus \{\mathrm{x}, \mathrm{y}\} : (\mathrm{z}R\mathrm{x} \Leftrightarrow \mathrm{z}R\mathrm{y})$

. . . 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 $\mathcal{U}$ is a *regular equivalence* on network $\mathbf{N} = (\mathcal{U}, R)$ if and only if for all $\mathrm{x}, \mathrm{y}, \mathrm{z}, \mathrm{w} \in \mathcal{U}$, $\mathrm{x} \approx \mathrm{y}$ implies both

R1. $\mathrm{x} R \mathrm{z} \Rightarrow \exists \mathrm{w} \in \mathcal{U} : (\mathrm{y} R \mathrm{w} \wedge \mathrm{w} \approx \mathrm{z})$

R2. $\mathrm{z} R \mathrm{x} \Rightarrow \exists \mathrm{w} \in \mathcal{U} : (\mathrm{w} R \mathrm{y} \wedge \mathrm{w} \approx \mathrm{z})$

. . . Regular Equivalence

The blocks for regular equivalence are null or 1-covered blocks (Batagelj, Doreian, Ferligoj, 1992):

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

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*.

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

RELATION R

DESCRIPTIONS OF UNITS Q — original relation, path matrix, triads

DISSIMILARITY MATRIX D

STANDARD CLUSTERING ALGORITHMS — hierarchical algorithms, relocation algorithm, k-means algorithm, etc.

Dissimilarities

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

$$x_i \sim x_j \Leftrightarrow d(x_i, x_j) = 0$$

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

$$d(x_i, x_j) = \sqrt{(r_{ii} - r_{jj})^2 + (r_{ij} - r_{ji})^2 + \sum_{\substack{s=1 \\ s \neq i,j}}^{n} ((r_{is} - r_{js})^2 + (r_{si} - r_{sj})^2)}$$

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.

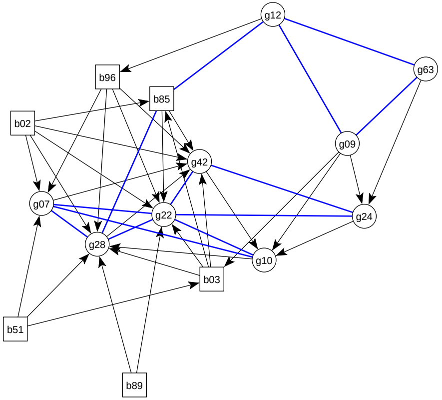
Example: Support network among informatics students

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:

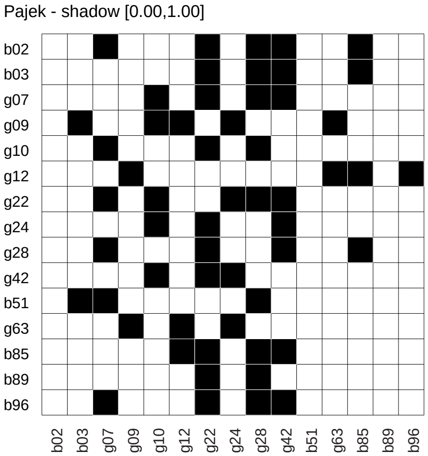
> *Introduction: You have done several exams since you are in the second class now. Students usually borrow studying material from their colleagues.*
> *Enumerate (list) the names of your colleagues that you have most often borrowed studying material from. (The number of listed persons is not limited.)*

Class network - graph

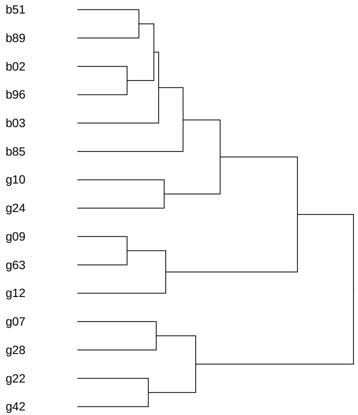


Vertices represent students in the class: circles – girls, squares – boys.
Reciprocated arcs are represented by edges.

Class network – matrix



Indirect Approach



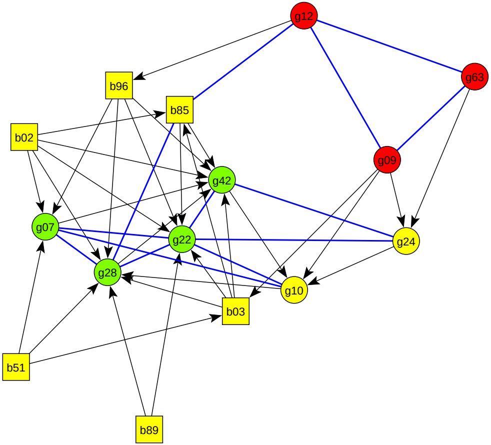
Using *Corrected Euclidean-like dissimilarity* and *Ward clustering method* we obtain the following dendrogram.

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**.

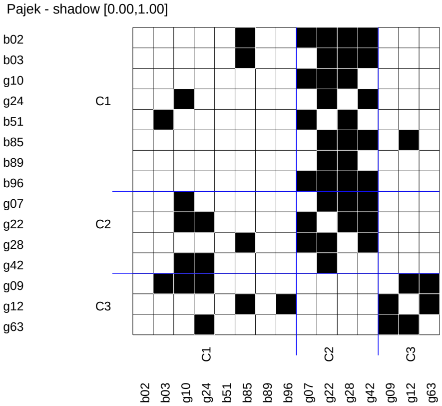
$$\mathbf{C} = \{\{b51, b89, b02, b96, b03, b85, g10, g24\}, \{g09, g63, g12\}, \{g07, g28, g22, g42\}\}$$

Partition into three clusters (Indirect approach)



On the picture, vertices in the same cluster are of the same color.

Matrix



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

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:

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

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.

Example

Empirical blocks | Ideal blocks

	a	b	c	d	e	f	g		a	b	c	d	e	f	g
a	0	1	1	0	1	0	0	a	0	1	1	0	0	0	0
b	1	0	1	0	0	0	0	b	1	0	1	0	0	0	0
c	1	1	0	0	0	0	0	c	1	1	0	0	0	0	0
d	1	1	1	0	0	0	0	d	1	1	1	0	0	0	0
e	1	1	1	0	0	0	0	e	1	1	1	0	0	0	0
f	1	1	1	0	1	0	1	f	1	1	1	0	0	0	0
g	0	1	1	0	0	0	0	g	1	1	1	0	0	0	0

Number of inconsistencies for each block

	A	B
A	0	1
B	1	2

$P = 4.$

Local Optimization

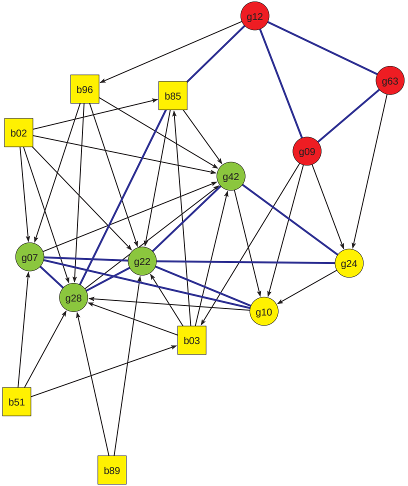
For solving the blockmodeling problem the relocation algorithm can be used:

Determine the initial clustering $\mathcal{C}$;
repeat:
  if in the neighborhood of the current clustering $\mathcal{C}$
  there exists a clustering $\mathcal{C}'$ such that $P(\mathcal{C}') < P(\mathcal{C})$
  then move to clustering $\mathcal{C}'$.

The neighborhood in this local optimization procedure is determined by the following two transformations:

- *moving* a unit x_k from cluster C_p to cluster C_q (*transition*);
- *interchanging* units x_u and x_v from different clusters C_p and C_q (*transposition*).

Partition into three clusters: Direct solution

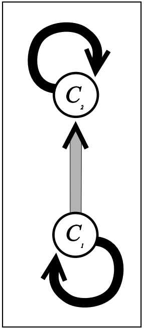


This is the same partition and has the number of inconsistencies.

Generalized Blockmodeling

$$\left[\begin{array}{cccc|cccc}
1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 0 & 1 & 0 & 1 \\
1 & 1 & 1 & 1 & 0 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\
\hline
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 0
\end{array}\right]$$

	C_1	C_2
C_1	complete	regular
C_2	null	complete



Generalized Equivalence / Block Types

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

complete

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

row-dominant

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

col-dominant

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

regular

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

row-regular

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

col-regular

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

null

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

row-functional

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

col-functional

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.

Types of 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 blockmodels are:

*	*	*
*	0	0
*	0	0

core - periphery

*	0	0
*	*	0
?	*	*

hierarchy

*	0	0
0	*	0
0	0	*

clustering

Example 1: Class network

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

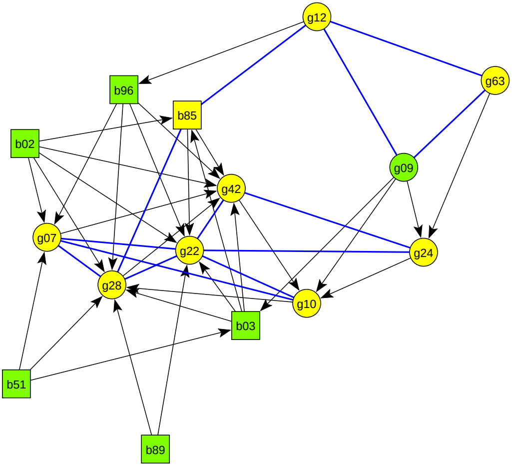
Pre-specified blockmodel: (com/complete, reg/regular, -/null block)

	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

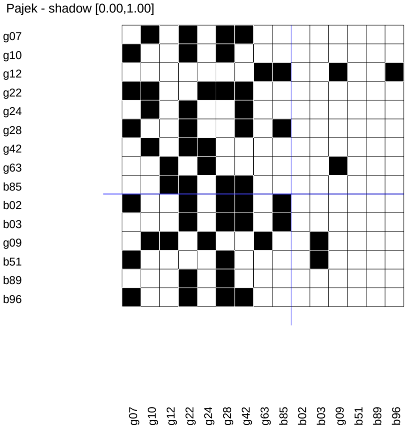


Image and error matrices:

	1	2
1	reg	-
2	reg	-

	1	2
1	0	3
2	0	2

Total error $= 5$
core-periphery

Example 2: Co-authorship networks of Slovenian researchers

Blockmodeling of Vauled Networks

	1	7	13	2	5	6	10	11	12	3	4	8	9
1									3		15	1	8
7										6	14	14	6
13					8						1	8	3
2		5					10	1	3	2	3	5	10
5			3			5		5			16	7	16
6			1		4			7	3	1		7	3
10				16		1		1	2	2	16	16	
11					2	2			2	2	8	5	14
12	2	2		2	2	2	2	11		8	2	2	6
3											19	3	1
4	2				1			1		6		1	19
8											5		6
9											19	1	

Žiberna (2007) proposed several approaches to generalized blockmodeling of valued networks, where values of the ties are assumed to be measured on at least interval scale. One of them is homogeneity blockmodeling. The basic idea of homogeneity blockmodeling is that the inconsistency of an empirical block with its ideal block can be measured by within block variability of appropriate values.

Blockmodeling of Three-Way Networks

The indirect approach to *structural equivalence blockmodeling in 3-mode networks* was proposed by Batagelj et al. (2007).

A *3-mode network* **N** over the basic sets X, Y and Z is determined by a ternary relation $R \subseteq X \times Y \times Z$. The notion of structural equivalence depends on which of the sets X, Y and Z are (considered) the same.

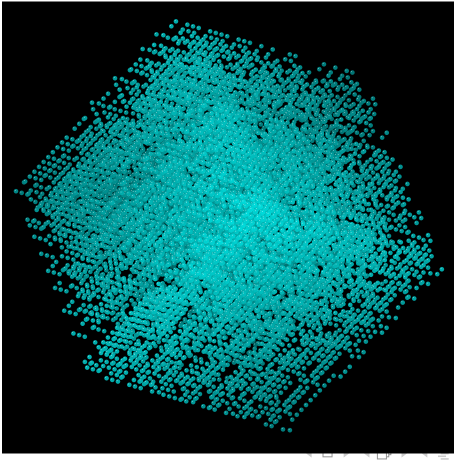
There are three basic cases:

- all three sets are different
- two sets are the same
- all three sets are the same

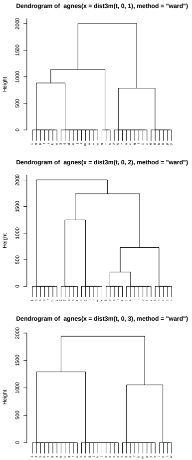
Example: Artificial Dataset

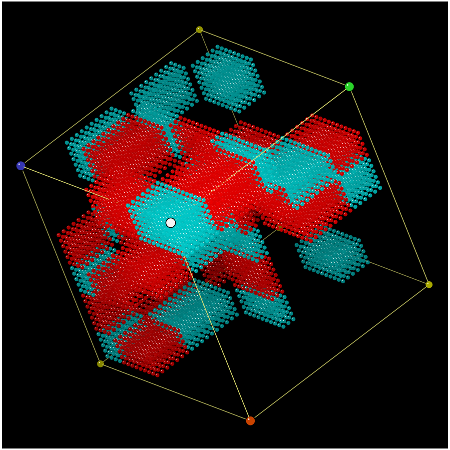
Randomly generated ideal structure

`rndTest(c(5,6,4),c(35,35,35))`:



Example: Solutions





Open Problems in Generalized Blockmodeling

- Boundary problems
- Measurement errors
- Assesing fits of blockmodels
- Blockmodeling large networks
- Number of positions
- Dynamic blockmodels

See more in Ferligoj, Doreian, Batagelj (2011).