

Model 2

Vladimir Batagelj
University of Ljubljana, FMP, Department of Mathematics
e-mail: vladimir.batagelj@uni-lj.si

January 27, 1996

a: March 17, 1996 / b: May 16, 1998 / c: May 24, 1998 /
d: August 7, 1998 / e: August 18, 1998 / g: April 05, 2000
printed: May 20, 2000

MODEL 2 is a program from package STRAN – STRucture ANalysis. It is freely available for noncommercial use at

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

MODEL 2 seeks for the best clustering of a given network satisfying a given model.

MODEL 2 introduced MDL files to describe the class of possible models. If the MDL file is missing a { nul, com, reg } model is assumed.

Version a added a don't care type; version b a penalty to each MDL file entry; version c extended the maximum size of networks to 100 units; and version d introduced two new types: non-null and symmetric.

1 MDL files

The structure of a MDL file is evident from the following example

```
*MODEL Tina
9
0 3 100 0 1 2 3 4
*CONSTRAINTS
1 100 2 1
4 100 1 3
*EOM
```

The first character in each line should be a star * or a blank.

The last character in a line should not be a blank.

A number in the second line is the maximal number of allowed clusters. The following lines have the structure

i j penalty t₁ t₂...t_k

When $i, j > 0$ the line prescribes that the block (i, j) can be of types $t_1, t_2 \dots t_k$. The types are coded as follows

0	-	-	null	7	rfn	-	row-function
1	com	-	complete	8	cfm	-	col-function
2	rdo	-	row-dominant	9	den	-	density
3	cdo	-	col-dominant	10	dnc	-	do not care
4	reg	-	regular	11	one	-	non-null
5	rre	-	row-regular	12	sym	-	symmetric
6	cre	-	col-regular				

Lines with $i = 0$ defines the types of parts of model matrix:

- $j = 0$: diagonal;
- $j = 1$: upper triangle;
- $j = 2$: lower triangle;
- $j = 3$: complete matrix.

Constraints have the form

$$k \text{ penalty } i \ j$$

with the following meaning

- $k = 1$: $i \in C_j$ – unit i belongs to cluster C_j ;
- $k = 2$: $i \notin C_j$ – unit i does not belong to cluster C_j ;
- $k = 3$: $C(i) = C(j)$ – units i and j belong to the same cluster;
- $k = 4$: $C(i) \neq C(j)$ – units i and j belong to different
- $k = 5$: $i \leq |C(j)|$ – cluster C_j has at least i units;
- $k = 6$: $i \geq |C(j)|$ – cluster C_j has at most i units.

The violations of constraints contribute to criterion function with a term

$$+ \text{ \# of violations } \times \text{ penalty}$$

The values of penalties have to be in the range 0 to 1000.

In the case of several lines describing a block the last prescription prevails.

In applications the first *no-of-clusters* rows and columns of model are considered.

2 Examples of MDL files

2.1 Regular blocks

```
*MODEL Regular
10
0 3 1 0 1 4
*EOM
```

2.2 Diagonal blocks (clustering)

```
*MODEL Diagonal
10
0 3 100 0
0 0 1 0 1 4
*EOM
```

2.3 Acyclic model (up)

```
*MODEL Hierarchy
9
0 1 1 0 5 6
0 0 10 0 1 4 12
0 2 100 0
*EOM
```

2.4 Acyclic model with symmetric clusters (down)

```
*MODEL SymHiera
9
0 0 10 0 1 12
0 1 100 0
0 2 1 0 11
*EOM
```

2.5 Center-Periphery

```
*MODEL Center-Periphery
2
0 3 1 0 11
2 2 10 0
1 1 100 0 1 4
*EOM
```

2.6 Regular path

```
*MODEL Regular Path
9
0 0 10 0 1 4
1 2 10 0 1 4
2 3 10 0 1 4
3 4 10 0 1 4
4 5 10 0 1 4
5 6 10 0 1 4
6 7 10 0 1 4
7 8 10 0 1 4
8 9 10 0 1 4
*EOM
```

2.7 Regular chain

```
*MODEL Regular Chain
9
0 0 10 0 1 4
1 2 10 0 1 4
2 3 10 0 1 4
3 4 10 0 1 4
4 5 10 0 1 4
5 6 10 0 1 4
6 7 10 0 1 4
7 8 10 0 1 4
8 9 10 0 1 4
2 1 10 0 1 4
3 2 10 0 1 4
4 3 10 0 1 4
5 4 10 0 1 4
6 5 10 0 1 4
7 6 10 0 1 4
8 7 10 0 1 4
9 8 10 0 1 4
*EOM
```