Symmetric-Acyclic Decompositions of Networks

Patrick Doreian
University of Pittsburgh

Vladimir Batagelj
University of Ljubljana

Anuška Ferligoj
University of Ljubljana

September 1, 1998

Abstract

This paper presents two new developments for partitioning networks. One is the symmetric-acyclic decomposition of a network and the other is a generalized blockmodeling approach to establish such decompositions of networks. Both are founded in the Davis and Leinhardt (1972) formulation of a ranked clusters model as a theoretical expectation concerning the structure of human groups and affect directed ties. The actual model can be applied in a variety of contexts and we illustrate the decomposition and generalized blockmodeling with the marriage network of noble families in Ragusa (Dubrovnik) for the 18th Century and early 19th Century.

1   Introduction

One major goal of social network analysis is to discern fundamental structure(s) of networks in a way that:

- allows us to know the structure and/or

- facilitates our understanding of network phenomena.

The most used tool (blockmodeling) does this through partitioning networks according to well specified criteria. Blockmodeling methods were established in social network analysis to partition units in a social network in terms of the structural information contained in their network ties. Hummon and Carley (1993)

1

showed the blockmodeling has been a major focus of network analysts. Methods were invented, for example, CONCOR (Breiger, Boorman, and Arabie, 1975) or adapted from extant clustering methods, for example STRUCTURE (Burt, 1976). This initial work focused on structural equivalence (Lorrain and White, 1971) where units are structurally equivalent if they are connected to the rest of the network in identical ways. Subsequent work, for example Sailer (1978) and White and Reitz (1983), generalized this concept to regular equivalence. Intuitively, two units are regularly equivalent if they are equivalently connected to equivalent others. They have the same types of neighbors. Further work dealt with automorphic equivalence (see, for example, Borgatti and Everett, 1992, and Pattison, 1988).

The partition of the units of a network into clusters generates a blockmodel where the (new) units are these clusters, called positions. (See Borgatti and Everett (1992) for a lucid discussion of this concept.) The set of ties between two positions forms a block. Those ties in a block (i.e. between all units in the two positions) are used to construct – and thereby summarize – the ties between positions. The *image* of a blockmodel is a matrix or pictorial representation of the ties between positions. Common to all of these efforts is the use of a definition of equivalence, a partitioning of networks according to this equivalence, and the use of a small number of established – within network analysis – clustering algorithms.

In this paper we will restrict attention to hierarchical blockmodels where each level (cluster of units) in the hierarchy has a specific structure: only symmetric ties are assumed. Our purpose here is to propose the symmetric-acyclic decomposition of networks as a new method which complements blockmodeling and vice versa. The complementary nature of these technical approaches means they can be used together to delineate the structure(s) of network.

2   Acyclic Structures

Distilling some essential ideas from Homans (1950), Davis and Leinhardt (1972) formulated two distinct structural features as general descriptions of small group structures for affective ties. One is the differentiation of these groups into cliques (in the sense of a maximal complete subgraph) while the other is the elaboration of ranks. The social processes generating these two structural features reinforce each other. As a result, cliques are found at distinct levels. A single clique may exist as the sole occupant of a level or multiple cliques can do so. In this case, there are no ties between members of the two cliques. Within cliques, every pair of members is linked by mutual ties. Concerning cliques at distinct levels, the hypothesized structure is that there will be asymmetric ties directed 'up' from lower ranked cliques to higher ranked cliques. No asymmetric ties are directed 'down' from higher ranked cliques to lower level cliques. Nor are there mutual
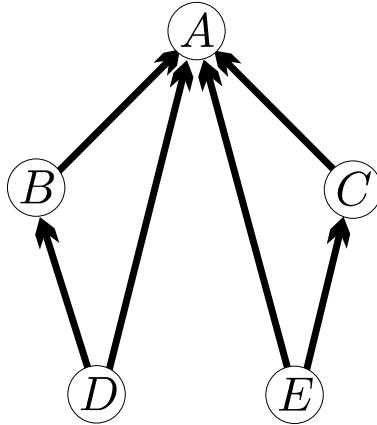
Figure 1: A Three Level Acyclic Model

ties between units at different levels. The network used here is for relation 'likes' and the liking ties between units go 'up'. If the relation is 'liked by' then the ties would be directed 'down'. The acyclic requirement is that these ties go in a single direction.

The 'ideal' model used by Davis and Leinhardt (1972) as an illustration of their ranked clusters model is shown in Figure 1 with five positions (as subgroups) and three levels. In the generalization of the ranked clusters model, we do *not* require subgroups to be maximal complete subgraphs. We use the term 'position' for the subgroups in such models. The highest ranking position is A with two positions, B and C, in the second level and two positions, D and E in the third level.

We present a new tool, called the symmetric-acyclic decomposition method, for verifying the presence (or not) of an acyclic structure characteristic of the ranked clusters model. The networks considered here have directed ties. The image of a 'pure' ranked clusters model is an acyclic directed graph. Any departures from such a structure are not allowed.

Table 1 contains an example that is used to illustrate the symmetric-acyclic decomposition described in the next section and blockmodeling approach. The sign '·' represents a 0. The italicized 1's represent ties that are inconsistent with a symmetric structure inside clusters and bold 1's represent ties that are inconsistent with the acyclic structure. Such ties fall into two categories. One concerns ties between blocks and the other concerns ties within (diagonal) blocks. The ties $a \rightarrow g$, and $i \rightarrow q$ belong to the first category and violate the acyclic requirement. Within the diagonal blocks, $j \rightarrow k$, $o \rightarrow n$, and $r \rightarrow p$ violation the symmetry requirement. See Figure 2.

Table 1: A Relation R

|	a	b	c	d	e	j	k	l	m	n	o	f	g	h	i	p	q	r	s
a	·	1	1	·	·	·	·	·	·	·	·	·	**1**	·	·	·	·	·	·
b	1	·	1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
c	1	1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
d	·	1	·	·	1	·	·	·	·	·	·	·	·	·	·	·	·	·	·
e	·	·	1	1	·	·	·	·	·	·	·	·	·	·	·	·	·	·	·
j	·	·	·	1	·	·	*1*	1	·	·	·	·	·	·	·	·	·	·	·
k	·	·	·	·	1	·	·	1	·	·	·	·	·	·	·	·	·	·	·
l	·	·	·	·	1	1	1	·	·	·	·	·	·	·	·	·	·	·	·
m	·	·	1	1	·	·	·	·	·	1	1	·	·	·	·	·	·	·	·
n	·	·	1	·	1	·	·	·	1	·	·	·	·	·	·	·	·	·	·
o	·	·	·	·	·	·	·	·	1	*1*	·	·	·	·	·	·	·	·	·
f	·	·	1	·	·	·	·	·	·	·	·	·	1	·	·	·	·	·	·
g	1	·	·	·	·	·	·	·	·	·	·	1	·	1	1	·	·	·	·
h	·	1	·	·	·	·	·	·	·	·	·	·	1	·	1	·	·	·	·
i	·	·	·	·	·	·	·	·	·	·	·	·	1	1	·	·	**1**	·	·
p	·	·	·	·	1	·	·	·	·	1	·	·	·	1	·	·	1	·	·
q	·	·	·	1	·	·	·	·	·	·	·	·	·	·	1	1	·	1	1
r	·	·	·	·	·	·	·	·	·	·	1	·	·	·	·	*1*	1	·	1
s	·	·	·	·	·	·	·	·	·	·	·	·	1	·	·	·	1	1	·
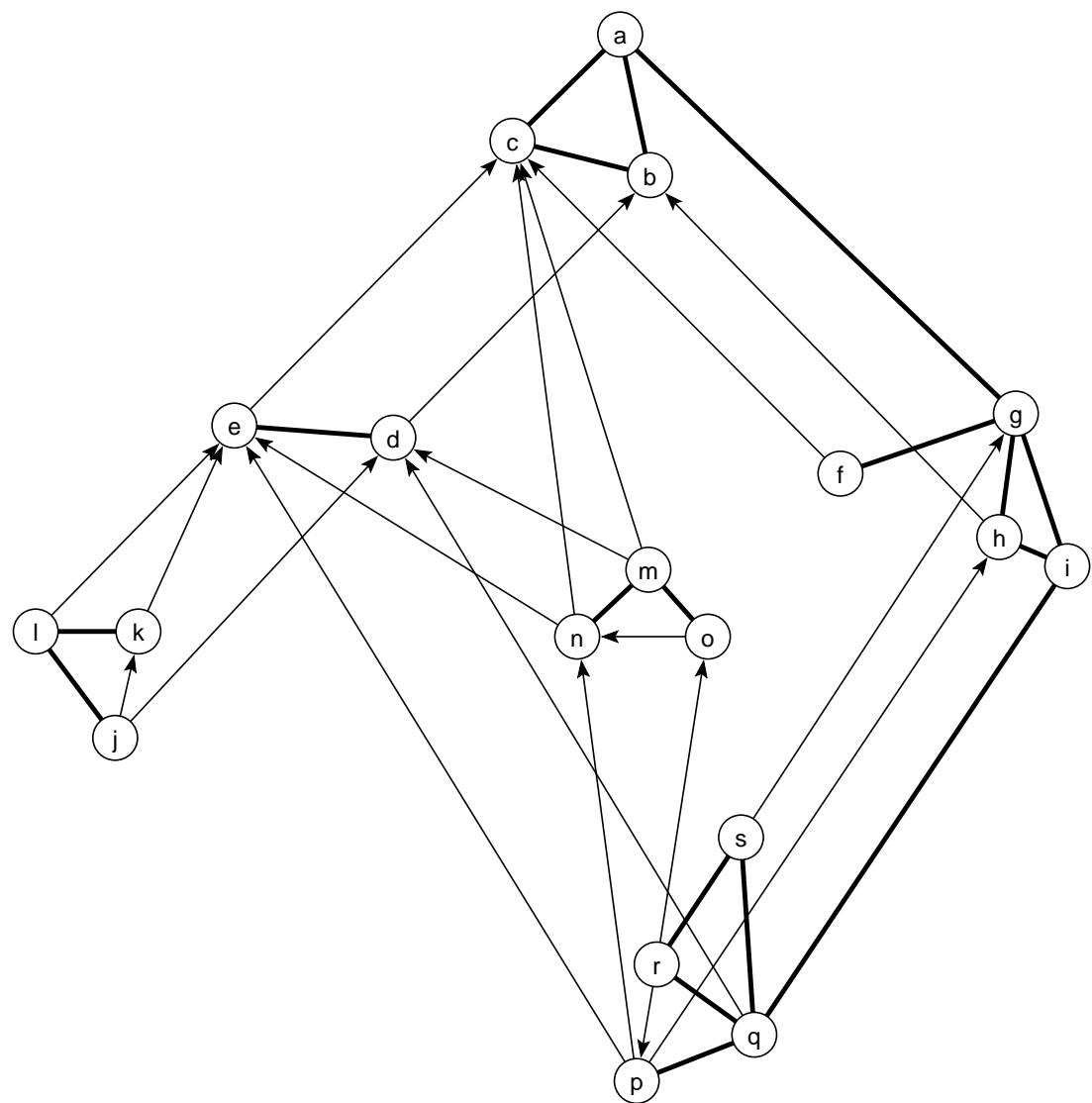
Figure 2: Graph of the Relation in Table 1

3 Establishing Symmetric-Acyclic Decompositions of Networks

Let $\mathcal{U} = \{x_1, x_2, \ldots, x_n\}$ be a finite set of *units*. The units are related by binary *relation*

$$R \subseteq \mathcal{U} \times \mathcal{U}$$

which determines a *network*

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

The network can be represented by a graph with units as vertices.

A *clustering* $\mathbf{C} = \{C_1, C_2, \ldots, C_k\}$, where $C_i \subseteq \mathcal{U}$ are *clusters*, partitions 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 the 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.

We say that a clustering $\mathbf{C}$ over the relation R is a *symmetric-acyclic clustering* iff

- all subgraphs induced by clusters from $\mathbf{C}$ contain only bidirectional arcs (*edges*); and

- each closed walk in $(\mathcal{U}, R)$ is entirely contained in a single cluster of $\mathbf{C}$.

3.1 Ideal Structures

A relation $R \subseteq \mathcal{U} \times \mathcal{U}$ has a *symmetric-acyclic decomposition* (S, Q) iff there exist relations $S, Q \subseteq \mathcal{U} \times \mathcal{U}$ such that

- (S, Q) is a partition of R, $S \cup Q = R$ and $S \cap Q = \emptyset$,

- S is symmetric, $S = S^{-1}$, and

- Q is acyclic in $(\mathcal{U}, R)$, $S^* \cap Q * R^* = \emptyset$,

where $I = \{(x, x) : x \in \mathcal{U}\}$ is the identity relation, R^* is the transitive and reflexive closure of relation R, and the operation $*$ denotes the product of relations $R_1 * R_2 = \{(x, y) : \exists z : (x R_1 z \wedge z R_2 y)\}$.

Theorem 1 *If a relation $R \subseteq \mathcal{U} \times \mathcal{U}$ has a symmetric-acyclic decomposition (S, Q) then*

$$S = R \cap R^{-1} \qquad and \qquad Q = R \setminus S$$
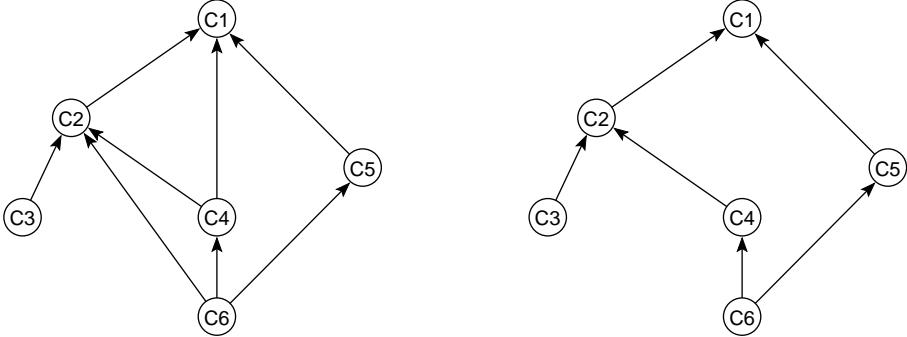
Figure 3: A Four Level Acyclic Graph for Relation in Table 1 and its Hasse Graph

The reflexive and transitive closure S^* of relation S is an equivalence relation, equal to the strong connectivity relation in $(\mathcal{U}, R)$. Let $\mathcal{U}'$ denotes the factor set $\mathcal{U}/S^*$. $\mathcal{U}'$ is a symmetric-acyclic clustering. The relation $\sqsubset$ defined on $\mathcal{U}'$ by

$$X \sqsubset Y \equiv \exists x \in X \exists y \in Y : xQy$$

is acyclic.

If the relation shown in Table 1 had 0s instead of the italicized 1s and the bold 1s it would have a symmetric-acyclic decomposition. As such, it is an ideal model and we confine attention to it in this section. The ideal model for this relation is drawn on the left in Figure 3 with six positions: $C_1 = \{a, b, c\}$, $C_2 = \{d, e\}$, $C_3 = \{j, k, l\}$, $C_4 = \{m, n, o\}$, $C_5 = \{f, g, h, i\}$, and $C_6 = \{p, q, r, s\}$.

The *cover* relation $\sqsubset\!\!\cdot$, analogous to the Hasse relation of an order, can be defined also for an acyclic relation $\sqsubset$ by

$$\sqsubset\!\!\cdot \; = \sqsubset \setminus \sqsubset * \overline{\sqsubset}$$

where $\overline{\sqsubset}$ denotes the transitive closure of relation $\sqsubset$.

We obtain $\sqsubset\!\!\cdot$ from $\sqsubset$ by deleting each arc (x, y) with the property that there exists a path of length at least 2 from x to y.

Using the Hasse graph $\mathbf{H} = (\mathcal{U}', \sqsubset\!\!\cdot)$ we can describe the set Σ of all acyclic clusterings with symmetric clusters over R :

- the minimum number of clusters is equal to $1 + d$, where d is the length of the longest path in $\mathbf{H}$;

- the maximum number of clusters is equal to $\operatorname{card} \mathcal{U}'$;

- card $\Sigma = 1$ iff **H** is a path;

- each k-clustering (partition into k clusters) is determined by a k-set (a set with k elements) of "compatible independent" vertices of **H**.

Two vertices $u, v \in \mathcal{U}'$ are *independent* iff they are not connected by a path in **H**. A pair of sets $X, Y \subset \mathcal{U}'$ is *compatible* iff

$$\neg\, \exists x_1, x_2 \in X\, \exists y_1, y_2 \in Y : (x_1 \sqsubset y_1 \wedge y_2 \sqsubset x_2)$$

All pairs of sets in a k-set have to be compatible. The clusters are unions of the sets corresponding to vertices from the sets of a k-set. We note that, as a special case, an asymmetric tie down a ranked clusters model leads to an incompatibility.

For the idealized relation shown in Table 1, the graph of the relation $\sqsubset$ is presented on the left side of Figure 3. To obtain the Hasse graph **H** (see the right side of Figure 3) two arcs (C_4, C_1) (given the path $C_4 C_2 C_1$) and (C_6, C_2) (given the path $C_6 C_4 C_2$) have to be deleted. The longest path in **H** has length 3 and therefore the number of levels is 4. From **H** we see that the following symmetric-acyclic clusterings (where only compatible clusters can be grouped together) exist:

$$
\begin{aligned}
\mathbf{C}_1^4 &= \{C_1, C_2 \cup C_5, C_3 \cup C_4, C_6\} \\
\mathbf{C}_2^4 &= \{C_1, C_2, C_3 \cup C_4 \cup C_5, C_6\} \\
\mathbf{C}_3^4 &= \{C_1, C_2 \cup C_5, C_4, C_3 \cup C_6\} \\
\mathbf{C}_4^4 &= \{C_1, C_2, C_4 \cup C_5, C_3 \cup C_6\} \\
\mathbf{C}_1^5 &= \{C_1, C_2 \cup C_5, C_3, C_4, C_6\} \\
\mathbf{C}_2^5 &= \{C_1, C_2, C_3, C_4 \cup C_5, C_6\} \\
\mathbf{C}_3^5 &= \{C_1, C_2, C_3 \cup C_5, C_4, C_6\} \\
\mathbf{C}_4^5 &= \{C_1, C_2, C_3 \cup C_6, C_4, C_5\} \\
\mathbf{C}_5^5 &= \{C_1, C_2, C_3 \cup C_4, C_5, C_6\} \\
\mathbf{C}_1^6 &= \{C_1, C_2, C_3, C_4, C_5, C_6\}
\end{aligned}
$$

We can introduce, in Σ, an operation

$$\mathbf{C}_1 \sqcap \mathbf{C}_2 \equiv \{C_1 \cap C_2 : C_1 \in \mathbf{C}_1, C_2 \in \mathbf{C}_2\} \setminus \{\emptyset\}$$

For example

$$\mathbf{C}_1^4 \sqcap \mathbf{C}_3^4 = \mathbf{C}_1^5$$

It is easy to verify that $(\Sigma, \sqcap)$ is a semilattice (associative, commutative, idempotent, with an absorption element).

The semilattice of clusterings for our example is presented in Figure 4.
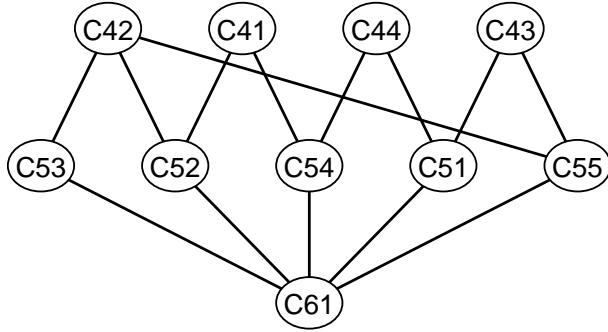
Figure 4: Semilattice

3.2 Relations Without a Symmetric-Acyclic Decomposition

There are several possible approaches to networks that do not have an exact ranked clusters structure (and so are not ideal).

The symmetric-acyclic decomposition (S, Q) described above can be improved by setting

$$S = R \cap R^{-1} \qquad \text{and} \qquad Q = R \setminus \overline{S}$$

where $\overline{S}$ replaces S in the definition of Q. In the ideal case this still produces the asymmetric-acyclic decomposition; but in the nonideal case it removes all 'asymmetric errors' from diagonal blocks.

3.2.1 Deleting Arcs

This provides useful information even though R has no symmetric-acyclic decomposition. The set

$$\Delta = \overline{S} \cap (R \setminus S)$$

gives us asymmetric arcs inside the otherwise symmetric classes of $\mathcal{U}$. For Table 1, Δ identifies exactly the three asymmetric ties, the italicized 0s, within the diagonal blocks $\Delta = \{(j, k), (o, n), (r, p)\}$. If the only violations of the symmetric-acyclic assumption concerned symmetry, the analysis would be complete. In the example, Q is not acyclic and we have to delete some arcs to make it acyclic. Here, the italicized 1s above the diagonal would be located and deleted leaving an exact model (with the violations noted). In general, however, the identification of the smallest number of violations of acyclicity is not an easy problem.

9

3.2.2 Iteration of the Symmetric-Acyclic Decomposition

Also, we can iterate the above decomposition procedure until we obtain a graph without edges.

Let $\mathcal{U}_0 := \mathcal{U}$, $R_0 := R$ and set $i := 0$. Then repeat the following steps:

- determine the symmetric part of R_i (the edges)

$$S_i := (R_i \cap R_i^{-1}) \setminus I_{\mathcal{U}_i}$$

- if no edge exists, $S_i = \emptyset$, stop iterating,

- shrink symmetric components, producing a new reduced graph $(\mathcal{U}_{i+1}, R_{i+1})$ determined by:

$$\mathcal{U}_{i+1} := \mathcal{U}_i / S_i^*$$

$$R_{i+1} := \{(X, Y) : \exists x \in X \exists y \in Y : xRy\}$$

- increase the counter of steps $i := i + 1$.

The obtained clusterings are nested (and form a hierarchy) and all clusters are strongly connected. If the final clustering is not acyclic we make an additional step in which we factorize the graph according to strong connectivity. The final graph is the condensation of original network (Harary, Norman, and Cartwright, 1965).

This procedure is implemented in Pajek – a program for large network analysis (Batagelj and Mrvar, 1998).

3.3 Generalized Blockmodeling Approach

As a third approach generalized blockmodeling can be used. Blockmodeling, as initially proposed, seeks to cluster together units having substantially similar patterns of relationships with the rest of the units of the network.

A *blockmodel* consists of structures 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 about which blocks produce an arc in the *reduced graph* and which do not. Here, the reduced graph is the graph version of the blockmodel and can be presented also by a relational matrix, called also an *image matrix*.

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. In general, and without surprise, different definitions of equivalence lead to distinct partitions. Two definitions of equivalence were extensively treated in the last three decades: structural

Table 2: Characterizations of Types of Blocks

null	nul	all 0 (except may be diagonal)
complete	com	all 1 (except may be diagonal)
row-regular	rre	each row is 1-covered
col-regular	cre	each column is 1 -covered
row-dominant	rdo	$\exists$ all 1 row (except may be diagonal)
col-dominant	cdo	$\exists$ all 1 column (except may be diagonal)
regular	reg	1-covered rows and 1-covered columns
non-null	one	$\exists$ at least one 1

and regular equivalence. An appropriate generalization of the equivalence idea is one where each block, of a particular partition, is free to conform to a different equivalence idea. This led Batagelj (1997) and Doreian, Batagelj, and Ferligoj (1994) to the definition of several types of connection inside and between the clusters as different types of blocks. Some of them are presented in Table 2. From the definition of structural equivalence it follows that there are two basic blocks: null and complete (Batagelj, Ferligoj, and Doreian, 1992). Batagelj, Doreian, and Ferligoj (1992) proved that regular equivalence produces two types of blocks: null and regular (see Table 2).

Another block type, introduced here, is necessary for the symmetric-acyclic decomposition of networks. This is the *symmetric block*. A block is symmetric if

$$\forall x, y \in C_i \times C_j : (xRy \Leftrightarrow yRx)$$

Note that for nondiagonal blocks this condition involves a pair of blocks $R(C_i, C_j)$ and $R(C_j, C_i)$.

3.3.1 Formalization of Blockmodeling

The point of departure is, as before, a network with a set of units, $\mathcal{U}$, and a relation $R \subseteq \mathcal{U} \times \mathcal{U}$. Let $\mathcal{Z}$ be a set of positions or images of clusters of units. Let $\mu : \mathcal{U} \to \mathcal{Z}$ denote a mapping which maps each unit to its position. The cluster of units $C(t)$ with the same position $t \in \mathcal{Z}$ is

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

Therefore

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

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

A *(generalized) blockmodel* is an ordered quadruple $\mathbf{M} = (\mathcal{Z}, K, \mathcal{T}, \pi)$ where:

11

- $\mathcal{Z}$ is a set of positions;

- $K \subseteq \mathcal{Z} \times \mathcal{Z}$ is a set of *connections* between positions;

- $\mathcal{T}$ is a set of predicates used to describe the types of connections between clusters in a network; we assume that $\mathrm{nul} \in \mathcal{T}$;

- A mapping $\pi : K \to \mathcal{T} \setminus \{\mathrm{nul}\}$ assigns predicates to connections.

A (surjective) mapping $\mu : \mathcal{U} \to \mathcal{Z}$ determines a blockmodel, **M**, of a network **N** iff it satisfies the conditions:

$$\forall (t, w) \in K : \pi(t, w)(C(t), C(w))$$

and

$$\forall (t, w) \in \mathcal{Z} \times \mathcal{Z} \setminus K : \mathrm{nul}(C(t), C(w))$$

Let $\approx$ be an equivalence relation over $\mathcal{U}$. It partitions the set of units $\mathcal{U}$ into clusters

$$[x] = \{y \in \mathcal{U} : x \approx y\}$$

We say that $\approx$ is *compatible* with $\mathcal{T}$ or a $\mathcal{T}$-equivalence over a network **N** iff

$$\forall x, y \in \mathcal{U}, \ \exists T \in \mathcal{T} : T([x], [y])$$

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

For a compatible equivalence $\approx$ the mapping $\mu \colon x \mapsto [x]$ determines a block-model with $\mathcal{Z} = \mathcal{U}/\approx$.

3.3.2 Optimization

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: determine the clustering $\mathbf{C}^*$ for which

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

where $\mathbf{C}$ is a clustering of a given *set of units* $\mathcal{U}$, Φ is the set of all feasible clusterings and $P : \Phi \to \mathbb{R}$ the *criterion function*.

Criterion function $P(\mathbf{C})$ has to be **sensitive** to selected type of equivalence:

$$P(\mathbf{C}) = 0 \Leftrightarrow \mathbf{C} \text{ determines an equivalence of the selected type.}$$

12

3.3.3 Criterion Functions

One of the possible ways of constructing a criterion function that directly reflects the selected type of 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 selected type of 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

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

where the term $d\big(R(C_u, C_v), B\big)$ measures the difference (error) between the block $R(C_u, C_v)$ and the ideal block B. The function d has to be compatible with the selected type of equivalence.

Given a set of types of connection $\mathcal{T}$ and a block $R(C_u, C_v)$, $C_u, C_v \subseteq \mathcal{U}$, we can determine the strongest (according to the ordering of the set $\mathcal{T}$) type T which is satisfied the most by $R(C_u, C_v)$. In this case we set

$$\pi\big(\mu(C_u), \mu(C_v)\big) = T$$

The obtained optimization problem can be solved by local optimization, using, for example a relocation algorithm (Batagelj, Doreian and Ferligoj, 1992).

3.3.4 Pre-Specified Blockmodels

The pre-specified blockmodeling starts with a blockmodel specified, in terms of substance, *prior to an analysis*. Batagelj, Ferligoj, and Doreian (1998) presented methods where a set of observed relations are fitted to a pre-specified blockmodel. Given a network, a set of ideal blocks is selected, a reduced model is formulated, and partitions are established by minimizing the criterion function. The pre-specified blockmodeling is supported by the program MODEL 2 (Batagelj, 1996b).

In the example of the next section, the pre-specified blockmodel is acyclic with symmetric diagonal blocks. For example for a clustering into 4 clusters symmetric-acyclic type of models can be specified in the following way:

sym	nul	nul	nul
nul,one	sym	nul	nul
nul,one	nul,one	sym	nul
nul,one	nul,one	nul,one	sym

In concrete analyses, sym on the diagonal, and its subtypes nul and com, usually are added.

In our discussion thus far, all errors are alike in the sense that they contribute in the same fashion to the total number of errors. Given the ranked clusters model, it is possible to argue that some errors are more consequential than others. Following this logic, asymmetric ties down from a higher to a lower level are the most serious. The computed criterion function can include having a higher penalty for these ties (e.g., 100). Asymmetric ties within clusters seem the next most important type of error. The modification to the criterion function can have a milder penalty for these ties (e.g., 10). Finally, errors in the blocks below the main diagonal can be specified as minor (e.g., 1). Therefore, the penalty matrix for the example above can be the following one

10	100	100	100
1	10	100	100
1	1	10	100
1	1	1	10

4 Applications

4.1 Constructed Example

If the inconsistencies with the ranked clusters model in the relation in Table 1 are not present, all methods delineate the true structure correctly. Also, if only the violations of the symmetry condition were present, all methods are successful. The (optimized) blockmodel correctly reports the three violations of the symmetry condition as errors. In contrast, the presence of the violations of the acyclic condition are consequential.

With all of the inconsistencies shown in Table 1 in the data, the optimization approach locates the clusters correctly and reports all of the 5 inconsistencies as errors. The model specification and penalties were determined as described in the previous section. However, the symmetric-asymmetric decomposition goes not fare as well. The reasons for this are evident in Figure 2 representing the network from Table 1. The six clusters are separated in the figure and the two line types are distinguished. The thick lines represent symmetric ties while the thin lines represent asymmetric ties (with the direction shown). In the decomposition, the algorithm establishes in the first step $\{j, k, l\}$, $\{d, e\}$ and $\{m, n, o\}$ correctly as subsets internally connected through mutual ties and not connected in that way to any other units. The remaining three clusters are identified also as being linked internally through mutual ties. The presence of the mutual tie between a and g connects the cluster $\{a, b, c\}$ to cluster $\{f, g, h, i\}$ and the mutual tie between

14

i and q connects the second of these clusters to $\{p, q, r, s\}$. Therefore, all three clusters are connected via symmetric ties. When the symmetric components are shrunk, there are only four units for the next iteration. As three of them have symmetric ties the algorithm ends after second iteration with only two levels. The first consists of cluster $\{j, k, l\}$ and the second one of all other units.

One methodological interpretation is that the optimization approach is robust in the presence of violations of the acyclic condition while the symmetric-acyclic decomposition is not. The latter procedure assumes perfect acyclic data. In contrast, the blockmodeling optimization concedes the presence of some errors and reports partitions with errors rather than be overly affected by the errors.

4.2   Ragusan Families Marriage Network

Krivošić (1990) collected interesting data about the population of Ragusa (Dubrovnik), a republic for most of its history. He also constructed two matrices describing the marriage networks of the Ragusan noble families in the 16th century and in the 18th century (and the beginning of 19th century). The matrix of marriage relation for the second period is presented in Table 3. While these data are similar to the well known Padgett marriage network for the elite Florentine families in the 15th century, they are directed and valued. Men marrying women and women marrying men are distinguished. The relation in Table 3 is 'men marrying women'.

Ragusa was settled in 7th century, as reported by Constantine Porphyrogenite, by fugitives from Epidaurum after its destruction. Ragusa was for a time under Byzantine protection, becoming a free commune as early as 12th century. Ragusa quickly grew into a free city-state. They prospered unhindered thanks primarily to their clever diplomacy and great skill in balancing the great powers, formally recognizing and paying tribute alternately to one then another. Napoleon, having destroyed the Venetian Republic in 1797, put an end to the Republic of Ragusa in 1806, which subsequently came under Austrian control until the fall of the Austro-Hungarian monarchy in 1918.

The Ragusan nobility evolved in the 12th century through the 14th century and was finally established by statute in 1332. After 1332 no new family was accepted until the large earthquake in 1667.

In Ragusa all political power was in the hands of male nobles older than 18 years. They were members of the Great Council (*Consilium majus*) which had the legislative function. Every year, 11 members of the Small Council (*Consilium minus*) were elected. Together with a duke – who was elected for a period of one month – it had both executive and representative functions. The main power was in the hands of the Senat (*Consilium rogatorum*) which had 45 members elected for one year.

This organization prevented any single family, unlike the Medici in Florence,

from prevailing. Nevertheless the historians agree that the Sorgo family was all the time among the most influential. For example:

- in the 17th century, 50 % of dukes and senators were from the following five families – Bona, Gondola, Goze, Menze, and Sorgo;

- in the 18th century, 56 % of senators were from five families – Sorgo, Goze, Zamagna, Caboga, and Georgi;

- in the last 8 years of Republic, 50 % of dukes were from five families: Sorgo, Goze, Gradis, Bona, and Ragnina.

A major problem facing the Ragusan noble families was also that by decreases of their numbers and the lack of noble families in the neighboring areas (which were under Turkish control), they became more and more closely related (1566 – *"quasi tutti siamo congionti in terzo et in quarto grado di consanguinita et affinita"*) – the marriages between relatives of only 3rd and 4th removes were frequent.

4.2.1 Earlier Analyses

Batagelj (1996a) found that the most influential families, according to the indices of centrality, in the second period (18th and 19th century) were Sorgo, Bona, and also Zamagna, Cerva and Menze. He obtained two basic clusters for this period by using the generalized blockmodeling approach:

Cluster 1: Basilio, Bona, Bonda, Caboga, Cerva, Georgi, Ghetaldi, Gondola, Goze, Gradi, Menze, Poza, Saraca, Sorgo, Tudisi, and Zamagna.

Cluster 2: Bosdari, Bucchia, Natali, Pauli, Ragnina, Resti, and Slatarich.

The second cluster contains all new families, accepted after the earthquake.

The structure obtained by Batagelj is an example of a center-periphery model. Most marriages were among the families of the first cluster, there was no marriage among families of the second cluster, and there were only few marriages between the two clusters.

4.2.2 Network Decomposition

In this section only the Ragusan families marriage network of the second period is analyzed. The relational matrix from Table 3 was binarized.

By the acyclic decomposition with respect to strong connectivity, three levels were obtained (see Figure 5). In the first and the third level only single families are present. The middle level consists of a cluster C with several families. The men from the families of the first level choose wives from the families of the second

Table 3: 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	.	.	.	.	.	.	.	.	.	.	2	.	.	.	.	.	.	.	.	1	.	.	.
Bonda	3	.	.	.	.	.	.	.	.	.	.	2	2	.	.	.	.	.	.	.	2	.	.	.
Bosdari	4	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
Bucchia	5	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.
Caboga	6	.	.	.	.	.	.	1	.	.	.	.	1	1	1	.	.	.	1	.	1	.	.	.
Cerva	7	.	1	.	.	.	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	.	.	.	.	.	.	.	.	2	2	1	.
Gradi	12	.	.	.	.	.	.	.	.	1	.	1	.	.	.	.	1	.	.	.	3	.	.	.
Menze	13	.	.	.	.	.	1	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
Natali	14	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Pauli	15	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Poza	16	.	.	2	.	.	.	.	.	1	.	1	.	.	.	.	1	.	.	.	1	.	.	.
Ragnina	17	.	1	1	.	.	.	.	.	1	.	1	.	.	.	.	.	.	.	.	1	.	.	.
Resti	18	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	1
Saraca	19	.	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
Slatarich	20	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
Sorgo	21	.	2	1	.	.	.	.	.	1	.	1	.	.	.	.	.	.	.	1	.	1	1	.
Tudisi	22	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
Zamagna	23	.	1	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.
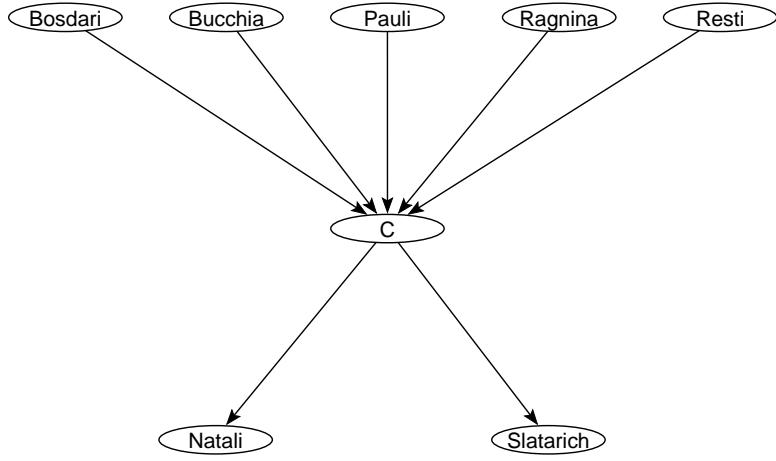
Figure 5: Acyclic Model of Ragusan Families

level and the men from the second level choose wives from the families of the third level. It is worth noting that five families are *sources* in the network providing husbands to the large 'middle' cluster of families, among which husbands (and wives) circulate. These families are *transmitters* and there are two families that are *sinks*.

The internal structure of the second cluster (the middle level in Figure 5) can be revealed by the iteration of the symmetric-acyclic decomposition method. The symmetric components of the middle cluster C are:

- $C_1 = \{$ Bona, Bonda, Caboga, Cerva, Goze, Gradi, Menze, Sorgo, Zamagna $\}$

- $C_2 = \{$ Ghetaldi, Saraca $\}$

- $C_3 = \{$ Basilio $\}$

- $C_4 = \{$ Georgi $\}$

- $C_5 = \{$ Gondola $\}$

- $C_6 = \{$ Poza $\}$

- $C_7 = \{$ Tudisi $\}$
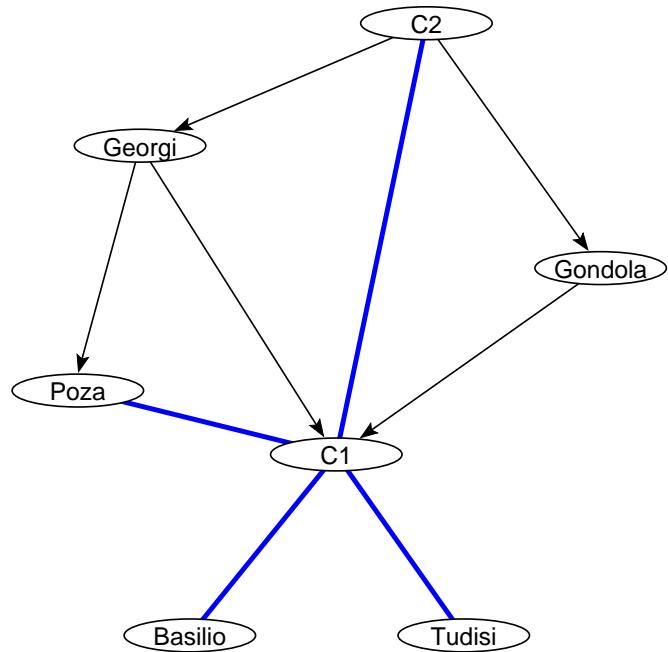
Figure 6: The Graph of Seven First Order Symmetric Components of the Middle Cluster
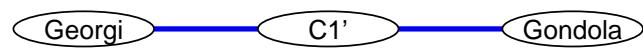
Figure 7: The Graph of Three Second Order Symmetric Components of the Middle Cluster
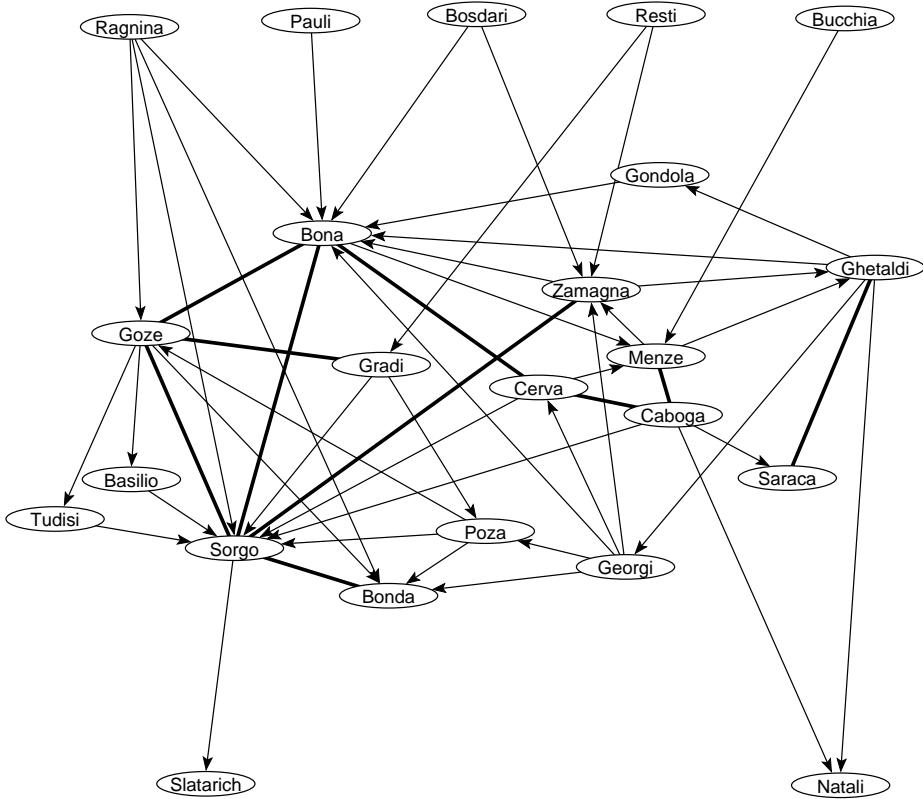
Figure 8: Acyclic-Symmetric Decomposition of Ragusan Families Network

The graph of the seven (first order) symmetric components (clusters) of the cluster C is presented in Figure 6. This reduced graph has three (second order) symmetric components:

- $C'_1 = \{C_1, C_2, C_3, C_6, C_7\}$

- $C'_2 = \{ \text{ Georgi } \}$

- $C'_3 = \{ \text{ Gondola } \}$

Their reduced graph (see Figure 7) is a single symmetric component. All three steps of the hierarchical decomposition can be seen in Figure 8.

4.2.3 Blockmodeling Approach

Based on three levels network decomposition two pre-specified models can be assumed. The first one is

nul,reg,sym	nul	nul
cre,rre	nul,reg,sym	nul
cre,rre	cre,rre	nul,reg,sym

The second pre-specified model is similar to the first one. The only difference is that the type 'reg' is excluded from the diagonal elements (this model has more precisely defined blocks in the lower triangle than in the general case from Section 3.3.4):

nul,sym	nul	nul
cre,rre	nul,sym	nul
cre,rre	cre,rre	nul,sym

Only the second model is symmetric-acyclic. The penalties in each cell were assumed in accordance with Section 3.3.4. For both models the same clustering of families was obtained (compare Figure 9):

Cluster 1: Natali, Slatarich.

Cluster 2: Basilio, Bona, Bonda, Caboga, Cerva, Georgi, Ghetaldi, Gondola, Goze, Gradi, Menze, Poza, Saraca, Sorgo, Tudisi, Zamagna.

Cluster 3: Bosdari, Bucchia, Pauli, Ragnina, Resti.

The permuted original relational matrix for this clustering is given in Table 4. The obtained models are

nul	nul	nul
cre	reg	nul
nul	rre	nul

and

nul	nul	nul
cre	sym	nul
nul	rre	nul

The first one has no errors and in the second one there are 28 errors in the middle diagonal block. These errors are produced by non-symmetric ties.

For the clusterings with larger number of clusters (4, 5 or 6) the solutions for the first model have no errors. All obtained clusterings have the same second cluster. There are some splits, as we can expect from the theory of decomposition, of the first and/or the third cluster.

When the structure inside the cluster in one of the levels of the symmetric-acyclic decomposition is not clearly symmetric some other types of the structure can be tested by applying the blockmodeling approach to this cluster separately. In this case an appropriate model should be pre-specified. As the iterative decomposition is based on the symmetric components of the cluster the blockmodeling

21

Table 4: Permuted Ragusan Families Marriage Network, 18th and 19th Century

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

22

approach is more appropriate to search for also other types of structures (e.g., center-periphery structure).

Using pre-specified blockmodeling on the subgraph induced by the cluster C the following results were obtained:

- **symmetric clusters model** ({nul} on out-diagonal and {com,sym} on the diagonal with penalties 1) gives for 7 clusters the same clustering as was obtained at the first step of the iterative procedure;

- **center-periphery model** ({nul,one} with penalty 1 on the out-diaginal, {com} with penalty 10 on the diagonal, {nul} with penalty 100 as the first diagonal element) gives the following 4 clusters of C:

    Periphery: Basilio, Bonda, Cerva, Gondola, Gradi, Saraca, Tudisi, Zamagna.
    Center 1: Caboga, Menze.
    Center 2: Bona, Georgi, Ghetaldi.
    Center 3: Goze, Poza, Sorgo.

    which are considered also in the ordering of the second cluster in Table 4.

5   Discussion

The above analyses can all be located within the domain of blockmodeling in so far as they were designed to partition social networks. However, several conclusions can be drawn from these analyses and are worth stressing. On the blockmodeling side, we emphasize that the analyses were done with the generalized version of blockmodeling and pre-specified models were used. Both represent major departures from conventional blockmodeling. On the symmetric-acyclic side, this is a new method that forces analysts to look inside the blocks of a blockmodel.

With the blockmodeling specification of the ranked clusters model, and networks that do not fit exactly, attention has to be paid to asymmetric ties within diagonal blocks and the ties that violate the acyclic specification. The blockmodeling approach enables us to weight each type of violation. It is also appropriate when the network data do not fit the ranked clusters model perfectly. On the other side the symmetric-acyclic approach is very efficient for very large networks, but is less robust to the violations of the acyclic assumption of the models.

Finally, the symmetric-acyclic decomposition method has its origins firmly in substance. It was a consideration of the ranked clusters model that lead us to formulate both the new block type and the decomposition itself. This seems very appropriate as the initial formulation of blockmodeling was also grounded in theoretical concerns.

All computations were done by programs Pajek and MODEL 2 which are freely available for noncommercial use at the address:
`http://vlado.fmf.uni-lj.si/pub/networks`.

References

[1] BATAGELJ, V. (1996a), "Ragusan Families Marriage Networks," in *Developments in Data Analysis*, Eds. A. Ferligoj and A. Kramberger, Ljubljana: FDV, p.p. 217–228.

[2] BATAGELJ, V. (1996b), "MODEL 2 – Program for Generalized Pre-Specified Blockmodeling", manual, Ljubljana.

[3] BATAGELJ, V. (1997), "Notes on Blockmodeling," *Social Networks*, *19*, 143–155.

[4] BATAGELJ, V., DOREIAN, P., and FERLIGOJ, A. (1992), "An Optimizational Approach to Regular Equivalence," *Social Networks*, *14*, 121–135.

[5] BATAGELJ, V., FERLIGOJ, A., and DOREIAN, P. (1992), "Direct and Indirect Methods for Structural Equivalence," *Social Networks*, *14*, 63–90.

[6] BATAGELJ, V., FERLIGOJ, A., and DOREIAN, P. (1998), "Fitting Pre-Specified Blockmodels," in *Data Science, Classification, and Related Methods*, Eds., C. Hayashi, N. Ohsumi, K. Yajima, Y. Tanaka, H. H. Bock, and Y. Baba", Tokyo: Springer Verlag, p.p. 199-206.

[7] BATAGELJ, V., and MRVAR, A. (1998), "Pajek – Program for Large Network Analysis," *Connections* (to appear).

[8] BORGATTI, S.P., and EVERETT, M.G. (1992), "Notions of Positions in Social Network Analysis," in *Sociological Methodology*, Ed., P.V. Marsden, San Francisco: Jossey-bass, pp. 1–35.

[9] BREIGER, R.L., BOORMAN, S.A., and ARABIE, P. (1975), "An Algorithm for Clustering Relational Data with Applications to Social Network Analysis and Comparison to Multidimensional Scaling," *Journal of Mathematical Psychology*, *12*, 328–383.

[10] BURT, R.S. (1976), "Positions in Networks," *Social Forces*, *55*, 93–122.

[11] DAVIS, J.A., and LEINHARDT, S. (1972), "The Structure of Positive Interpersonal Relations in Small Groups," in *Sociological Theories in Progress*, *Volume 2*, Ed., J. Berger, Boston: Houghton Mifflin, pp. 218–251.

[12] DOREIAN, P., BATAGELJ, V., and FERLIGOJ, A. (1994), "Partitioning Networks on Generalized Concepts of Equivalence," *Journal of Mathematical Sociology*, *19*, 1–27.

[13] HARARY, F, NORMAN, R.Z., and CARTWRIGHT, D. (1965), *Structural Models: An Introduction to the Theory of Directed Graphs*, New York: John Wiley & Sons.

[14] HOMANS, G.C. (1950), *The Human Group*, London: Routledge and Kegan Paul.

[15] HUMMON, N.P., and CARLEY, K. (1993), "Social Networks as Normal Science," *Social Networks*, *15*, 71–106.

[16] KRIVOŠIĆ, S. (1990), *Stanovništvo Dubrovnika i demografske promjene u prošlosti*, Dubrovnik: Zavod za povjesne znanosti JAZU u Dubrovniku.

[17] LORRAIN, F., and WHITE, H.C. (1971), "Structural Equivalence of Individuals in Social Networks," *Journal of Mathematical Sociology*, *1*, 49–80.

[18] PATTISON, P. (1988), "Network Models; Some Comments on Papers in this Special Issue," *Social Networks*, *10*, 383–411.

[19] SAILER, L.D. (1978), "Structural Equivalence: Meaning and Definition, Computation and Application," *Social Networks*, *1*, 73–90.

[20] WHITE, D.R., and REITZ, K.P. (1983), "Graph and Semigroup Homomorphisms on Networks of Relations," *Social Networks*, *5*, 193–234.