Optimizational Approach to Blockmodeling

Anuška Ferligoj[1], Patrick Doreian[2] and Vladimir Batagelj[3]

[1] Faculty of Social Sciences, University of Ljubljana, Ljubljana, Slovenia
[2] Department of Sociology, University of Pittsburgh, Pittsburgh, USA
[3] Vladimir Batagelj, Faculty of Mathematics and Physics, University of Ljubljana, Ljubljana, Slovenia

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, under some meaningful definition of equivalence. In the paper an optimizational approach to blockmodeling is discussed. Methods where a set of observed relations are fitted to a pre-specified blockmodel are also presented.

Keywords: Social network analysis, relations, local optimization, clustering.

1. Introduction

Social network analysis has attracted considerable interest from the social and behavioral science community in recent decades. Much of this interest can be attributed to the focus of social network analysis on *relationships* among social units, and on the patterns and implications of these relationships. Wasserman and Faust (1994) pointed out that the network perspective allows new leverage for answering standard social and behavioral science research questions by giving precise formal definitions to aspects of the political, economic, or social environment. From the view of social network analysis, the social environment can be expressed as patterns or regularities in relationships among interacting units.

The origins of social network analysis can be found in the social psychology of groups and at its subsequent development in sociological and social anthropological studies of factories and communities (e.g., Moreno 1934, Lewin 1936, Warner and Lunt 1941, Heider 1946, Bavelas 1948, Homans 1951, Cartwright and Harary 1956, Nadel 1957, Mitchell 1969). Scott (1991) gave a very nice discussion on the development of social network analysis from the theoretical ideas to growing technical complexity of the work carried out from the 1970s. In 1978 the International Network for Social Network Analysis (INSNA) was formed. This international group for the exchange of information and mutual intellectual support published a newsletter, Connections, and was involved in the foundation of the journal *Social Networks*. Within this institutionalized forum, network analysts have pursued a variety of topics with regard to the substance and development of network analytic tools.

There are several topics which are treated in social network analysis. Some of them are the following ones:

- connectedness and cohesive subgroups,
- boundaries of networks,
- centrality of units in a network,
- roles and positions in networks (blockmodeling).

Blockmodeling has been a main focus of network analysts (Hummon and Carley 1993) with position as a central concept (Borgatti and Everett 1992). Blockmodeling seeks to cluster units that have substantially similar patterns of relationships to others, and interpret the pattern of relationships among clusters.

2. Basic terms

Let $E = \{X_1, X_2, \ldots, X_n\}$ be a finite set of *units*. The units are related by binary *relations*

$$R_t \subseteq E \times E, \quad t = 1, \ldots, r$$

which determine a *network*

$$\mathcal{N} = (E, R_1, R_2, \ldots, R_r)$$

In the following we restrict our discussion to a single relation R described by a corresponding binary matrix $\mathbf{R} = [r_{ij}]_{n \times n}$ where

$$r_{ij} = \begin{cases} 1 & X_i R X_j \\ 0 & \text{otherwise} \end{cases}$$

In some applications r_{ij} can be a nonnegative real number expressing the strength of the relation R between units X_i and X_j.

One of the main procedural goals of network analysis is to identify, in a given network, *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*

$$\mathcal{C} = \{C_1, C_2, \ldots, C_k\}$$

which is a partition of the set E:

$$\bigcup_i C_i = E$$

$$i \neq j \Rightarrow C_i \cap C_j = \emptyset$$

Each partition determines an equivalence relation (and vice versa). Let us denote by $\sim$ the relation determined by partition $\mathcal{C}$.

A clustering $\mathcal{C}$ 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 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.

A *blockmodel* consists of structures obtained by identifying all units from the same cluster of the clustering $\mathcal{C}$. For an exact definition of a blockmodel we also have to be precise about which blocks produce an arc in the *reduced graph* and which do not. The reduced graph can also be presented by relational matrix, also called *image matrix*.

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, under some *meaningful* definition of equivalence.

3. Equivalences

In general, and without surprise, different definitions of equivalence lead to distinct partitions. Regardless of the definition 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; the second by the notion of regular equivalence.

3.1. Structural equivalence

Units are structurally equivalent if they are connected to the rest of the network in identical ways (Lorrain and White 1971).

Formally: X and Y are structurally equivalent iff

s1. $XRY \Leftrightarrow YRX$
s2. $XRX \Leftrightarrow YRY$
s3. $\forall Z \in E \setminus \{X, Y\} : (XRZ \Leftrightarrow YRZ)$
s4. $\forall Z \in E \setminus \{X, Y\} : (ZRX \Leftrightarrow ZRY)$

or in the matrix form: $X_i \equiv X_j$ iff

s1'. $r_{ij} = r_{ji}$
s2'. $r_{ii} = r_{jj}$
s3'. $\forall k \neq i, j : r_{ik} = r_{jk}$
s4'. $\forall k \neq i, j : r_{ki} = r_{kj}$

0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	1

1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1

0	1	1	1
1	0	1	1
1	1	0	1
1	1	1	0

Fig. 1. Ideal blocks for structural equivalence

The matrix form of the definition of structural equivalence can also be extended to the case when r_{ij} are real numbers.

From the definition of structural equivalence it follows that there are four possible ideal blocks (Batagelj, Ferligoj, and Doreian 1992). They are presented in Figure 1.

3.2. Regular equivalence

Intuitively, two units are regularly equivalent if they are equally connected to equivalent others (White and Reitz 1983).

The equivalence relation $\approx$ on E is a *regular equivalence* on network $\mathcal{N} = (E, R)$ iff for all $X, Y, Z \in E, X \approx Y$ implies both

R1. $XRZ \Rightarrow \exists W \in E : (YRW \wedge W \approx Z)$
R2. $ZRX \Rightarrow \exists W \in E : (WRY \wedge W \approx Z)$

Batagelj, Doreian, and Ferligoj (1992) proved that regular equivalence produces two types of blocks: null blocks (which have all entries 0) and 1-covered blocks (which have in each row and in each column at least one 1). They are presented in Figure 2.

For a network there is, in general, a non-empty set of regular partitions which form a lattice (Borgatti and Everett 1989).

The following relation can be proven for these two equivalences

$$\equiv \subset \approx$$

Structural equivalence is a very stringent requirement in a network. Regular equivalence has weaker requirements and is more often found in a given network.

3.3. Generalized concepts of equivalence

An appropriate generalization of the equivalence idea is one where each block, of a particular partition, is free to form to a different equivalence idea. This led Batagelj (1993) and Doreian, Batagelj, and Ferligoj (1994) to the definition of several types of connection inside and between the clusters, or in another words, different types of ideal blocks. Some of them are presented in Table 1.

0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

1	0	1	0	0
0	0	1	0	1
0	1	0	0	0
1	0	1	1	0

Fig. 2. Ideal blocks for regular equivalence

Tab. 1. Generalized ideal blocks

0	0	0	0	0
0	0	0	0	0
0	0	0	0	0
0	0	0	0	0

Null

0	1	0	0	0
1	1	1	1	1
0	0	0	0	0
0	0	0	1	0

Row-Dominant

0	0	1	0	0
0	0	1	1	0
1	1	1	0	0
0	0	1	0	1

Col-Dominant

1	1	1	1	1
1	1	1	1	1
1	1	1	1	1
1	1	1	1	1

Complete

0	0	0	1	0
0	0	1	0	0
1	0	0	0	0
0	0	0	1	0

Row-Functional

1	0	0	0
0	1	0	0
0	0	1	0
0	0	0	0
0	0	0	1

Col-Functional

0	1	0	0	0
1	0	1	1	0
0	0	1	0	1
1	1	0	0	0

Regular

0	1	0	0	0
0	1	1	0	0
1	0	1	0	0
0	1	0	0	1

Row-Regular

0	1	0	1	0
1	0	1	0	0
1	1	0	1	1
0	0	0	0	0

Col-Regular

4. Establishing blockmodels

The problem of establishing a partition of a network in terms of a considered equivalence is a special case of **clustering problem** that can be formulated as an optimization problem: determine the clustering $\mathcal{C}^*$ for which

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

where $\mathcal{C}$ is a clustering of a given *set of units* E, Φ is the set of all possible clusterings and $P : \Phi \to \mathbb{R}$ the *criterion function*. The criterion function must reflect the considered equivalence.

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).

4.1. Indirect approach

Figure 3 presents the process of establishing a blockmodel by the indirect approach (Ferligoj, Batagelj, and Doreian 1994). In this process the most important requirement is that the selected dissimilarity measure is *compatible* with the considered equivalence (Batagelj, Ferligoj, and Doreian 1992). 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

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

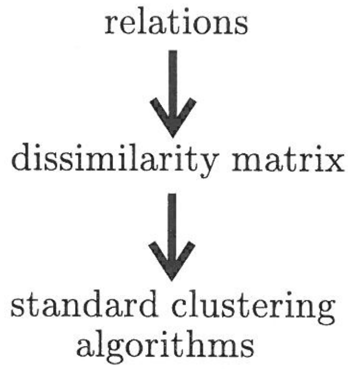

Fig. 3. Indirect approach

is compatible with structural equivalence.

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.

4.2. Direct approach

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

Criterion function $P(\mathcal{C})$ has to be **sensitive** to considered equivalence:

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

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 $\mathcal{C} = \{C_1, C_2, \ldots, C_k\}$, 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 $\mathcal{C}$ can be expressed as

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

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. The function d has to be compatible with the selected type of equivalence.

In the case of direct clustering approach where an appropriate criterion function is constructed we can use one of the local optimization procedures for solving the given blocking problem (Batagelj, Doreian and Ferligoj 1992). E.g., relocation algorithm:

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 $\mathcal{C}'$.

Usually, 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*).

5. Fitting to a pre-specified blockmodel

In the previous sections inductive approaches in establishing blockmodels for a set of social relations defined over a set of units were discussed. Some forms of equivalence are 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. Batagelj, Ferligoj, and Doreian (1996) presented the methods where a set of observed relations are fitted to a pre-specified blockmodel. In this case, given a network, set of types of ideal blocks, and a reduced model, a solution can be determined which minimizes the criterion function. It is also possible to fit the network to a partial model and analyze the residual afterward. Different constraints on the model can also be introduced. For example: units X and Y are of the same type; or, types of units X and Y are not connected.

6. Example

The analyzed network consists of social support exchange relation among thirteen students of the Social Science Informatics second year class (1992/1993) at the Faculty of Social Sciences, University of Ljubljana (Ferligoj and Hlebec, 1995). Data were collected through CAPI (Computer Assisted Personal Interviewing) supported by the program INTERV (de Pijper and Saris, 1986). Interviews were conducted in May 1993. 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.)

The measured support relation is presented by the matrix in Table 2.

We assumed three clusters and three types of blocks: null, complete and regular. Using the direct approach we obtained one optimal solution with 6 errors (the value of the criterion function). The obtained clustering into three clusters is the following:

$$\mathcal{C} = \{\{a, f, g, j, l, m\}, \{b, c, e, h, k\}, \{d, i\}\}$$

In Table 3 the matrix partitioned into nine blocks is presented. It is constructed by permuting the rows and columns of the relational matrix in such a way that first, the units of the first cluster $A = \{a, f, g, j, l, m\}$ are listed, then the units of the second cluster $B = \{b, c, e, h, k\}$ and finally the units of the last cluster $C = \{d, i\}$. In the table the six errors can be seen: there is a tie between units b and j, but it should not be in the null block and there is no tie in the first row of the regular block from A to B (5 errors).

The obtained blockmodel is presented by the model matrix and by the three-node graph in

Tab. 2. Relational matrix

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

Tab. 3. Partition into blocks

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

	A	B	C
A	nul	reg	reg
B	nul	nul	reg
C	nul	nul	com

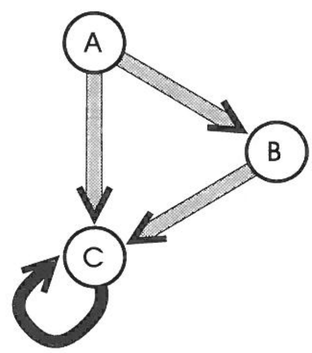

Fig. 4. Blockmodel

Figure 4. The elements of the model matrix are nul (null block), reg (regular block) and com (complete block). In Figure 4 the first cluster {a, f, g, j, l, m} is represented by the vertex A, the second cluster {b, c, e, h, k} by the vertex B and the third cluster {d, i} by the vertex C. There is no tie between vertices if there is a null block, otherwise there is a tie. This simple graph shows a very comprehensible structure of the network. Each student from the cluster *A* is borrowing studying material from at least one student of the cluster *B* and cluster *C* but not from the students of the same cluster *A*. Students from the cluster *B* are borrowing studying material only from students of the cluster *C* and students of the cluster *C* are borrowing material only from the other students of the same cluster.

7. Conclusion

To establish a partition of networks in terms of a considered equivalence, appropriate clustering algorithms can be successfully applied. We can choose between direct and indirect approaches, and second, for indirect approaches, there is a choice with regard to a dissimilarity measure. Concerning the choice of a similarity measure, it is necessary to use a measure that is compatible with the selected form of equivalence.

The direct approach uses a special goodness-of-fit measure as a criterion function to obtain partitions based on a selected equivalence. We used a local optimization procedure to minimize the appropriately constructed criterion function.

There are still many open problems in establishing blockmodel, e.g., how to consider multiple networks, find efficient algorithms for large networks, and develop approaches for establishing blockmodels for valued networks.

References

[1] Batagelj, V. (1993): "Notes on block modelling". In: *Abstracts and Short Versions of Papers*, 3rd European Conference on Social Network Analysis, Munchen: DJI, 1-9. Extended version to appear in *Social Networks*.

[2] Batagelj, V. (1996): "Ragusan families marriage networks." In: A. Ferligoj and A. Kramberger (Eds.): *Developments in Data Analysis*. Ljubljana: FDV, 217-228.

[3] Batagelj, V., P. Doreian, and A. Ferligoj (1992): "An optimizational approach to regular equivalence". *Social Networks* **14**: 121-135.

[4] Batagelj, V., A. Ferligoj, and P. Doreian (1992): "Direct and indirect methods for structural equivalence". *Social Networks* **14**: 63-90.

[5] Batagelj, V., A. Ferligoj, and P. Doreian (1996): "Fitting pre-specified blockmodels". In: *Abstracts*, 5th Conference of International Federation of Classification Societies – *Data Science, Classification and Related Methods*, Kobe. To appear in the conference proceedings.

[6] Bavelas, A. (1948): "A mathematical model for group structure". *Applied Anthropology* **7**.

[7] Borgatti, S.P. and M.G. Everett (1989): "The class of all regular equivalences: Algebraic structure and computation". *Social Networks* **11**: 65-88.

[8] Cartwright, D. and F. Harary (1956): "Structural balance: a generalisation of Heider's theory". *Psychological Review* **63**.

[9] de Pijper, N.M. and Saris W.E. (1986): "The formulation of intervews using the program INTERV". Amsterdam: Sociometric Research Foundation.

[10] Doreian, P., V. Batagelj, and A. Ferligoj (1994): "Partitioning networks on generalized concepts of equivalence". *Journal of Mathematical Sociology*, **19**: 1-27.

[11] Everett, M.G. and S.P. Borgatti (1988): "Calculating role similarities: An algorithm that helps determine the orbits of a graph ". *Social Networks* **10**: 71-91.

[12] Faust, K. (1988): "Comparison of methods for positional analysis: Structural and general equivalences". *Social Networks*, **10**: 313-341.

[13] Ferligoj, A., V. Batagelj, and P. Doreian (1993): "On connecting network analysis and cluster analysis". In: G.H. Fischer and D. Laming (Eds.): *Contributions to Mathematical Psychology, Psychometrics, and Methodology*, New York: Springer Verlag, 323-338.

[14] Ferligoj, A. and Hlebec, V. (1995): "Evaluating network data quality: An application of MTMM approach". In: W.E. Saris and A. Munnich (Eds.): *The Multitrait-Multimethod Approach to Evaluate Measurement Instruments*. Budapest: Eotvos University Press, 155-172.

[15] Heider, F. (1946): "Attitudes and cognitive orientation". *Journal of Psychology* **21**.

[16] Homans, G.C. (1951): "The human Group". London: Routledge and Kegan Paul.

[17] Lewin, K. (1936): "Principles of Topological Psychology". New York: McGraw-Hill.

[18] Lorrain, F. and H.C. White (1971): "Structural equivalence of individuals in social networks". *Journal of Mathematical Sociology*, **1**: 49-80.

[19] Mitchell, J.C. (1969): "The concept and use of social networks". In: J.C. Mitchell (Ed.): *Social Networks in Urban Situations*, Manchester: Manchester University Press.

[20] Moreno, J. (1934): "Who Shall Survive?". New York: Beacon Press.

[21] Nadel, S.F. (1957): "The Theory of Social Structure". London: Cohen and West.

[22] Scott, J. (1991): "Social Network Analysis: A Handbook". London: Sage.

[23] Warner, W.L. and P.S. Lunt (1941): "The Social Life of a Modern Community". New Haven, Conn.: Yale University Press.

[24] Wasserman, S. and K. Faust (1994): "Social Network Analysis". Melbourne: Cambridge University Press.

[25] White, D.R. and K.P. Reitz (1983): "Graph and semigroup homomorphisms on networks of relations". *Social Networks* **5**: 193-234.

Received: October, 1996
Accepted: December, 1996

Contact address:
Anuška Ferligoj
Faculty of Social Sciences
University of Ljubljana
P.O. Box 47, 1109 Ljubljana
Slovenia
e-mail: anuska.ferligoj@uni-lj.si

Patrick Doreian
Department of Sociology
University of Pittsburgh
Pittsburgh, PA 15260
USA
e-mail: PITPAT@vsm.cis.pitt.edu

Vladimir Batagelj
Department of Mathematics
Faculty of Mathematics and Physics
University of Ljubljana
Jadranska 19, 1000 Ljubljana
Slovenia
e-mail: vladimir.batagelj@uni-lj.si
WWW: http://vlado.fmf.uni-lj.si/

ANUŠKA FERLIGOJ is Professor of Statistics at the University of Ljubljana and head of the Center of Methodology and Informatics at the Institute of Social Scence in Ljubljana. She has approximately 100 publications including books edited or coedited, in cluster analysis, research methodology and network analysis, with applications in social sciences, economy, epidemiology, and education. She has been planary speaker, invited speaker, member of the programme committee or organizer of several international conferences.

PATRICK DOREIAN is Professor of Sociology at the University of Pittsburgh. He is editor of the Journal of Mathematical Sociology. His research and teaching interests are in social network analysis (primarily blockmodeling, the evolution of social networks, network and spatial autocorrelation), inter-organizational networks, mathematical sociology (e.g., the theory of solidarity), macro social change, and statistical methods (primarily applied regression and structural equation modeling).

VLADIMIR BATAGELJ is Associate Professor of Discrete and Computational Mathematics at the University of Ljubljana. His main research interests are in mathematics and computer science, combinatorics with emphasis on graph theory, algorithms on graphs and networks, combinatorial optimization, algorithms and data structures, cluster analysis and applications of information technology in education.