Clustering Relational Data

Vladimir Batagelj[1] and Anuška Ferligoj[2]

[1] University of Ljubljana, Faculty of Mathematics and Physics, and Institute of Mathematics, Physics and Mechanics, Dept. of TCS, Jadranska 19, 1 000 Ljubljana, Slovenia
[2] University of Ljubljana, Faculty of Social Sciences, P.O. Box 47, 1 109 Ljubljana, Slovenia

Abstract. In the paper we show, based on the optimization approach to clustering, that clustering with relational constraint and blockmodeling (of social networks) problems are essentially special cases of the clustering relational data problem. An overview of the main results in this field, some open problems and directions for the future research are presented.

1 Introduction

There are at least two known approaches to clustering relational data: the relationally constrained clustering and blockmodeling. The first approach is based on attribute and relational data, and the second only on relational data.

For constrained clustering, grouping similar units into clusters, based on attribute data, has to satisfy some additional conditions. One of the 'classical' constrained clustering problems is the regionalization problem: clusters of similar (according to socio-economic development) geographical regions have to be determined such that the regions inside each cluster are also geographically connected. A number of approaches to this problem have been prposed. The majority of authors (e.g., Lebart, 1978; Lefkovitch, 1980; Ferligoj and Batagelj, 1982; Perruchet, 1983; Gordon, 1973, 1980, 1987; Legendre, 1987) solve this problem by adapting standard clustering procedures, especially agglomerative hierarchical algorithms, and local optimization clustering procedures, or by introducing penalties. The geographic contiguity is a special case of *relational constraint*. Ferligoj and Batagelj (1982, 1983) first treated this clustering problem for general symmetric relations and then for nonsymmetric relations. It is possible to work also with other, non-relational conditions. Murtagh (1985) provides a review of clustering with symmetric relational constraints. A more recent survey of constrained clustering was given by Gordon (1996).

Another, closely related problem, is the blockmodeling of (social) networks. A network consists of set of units and one or more binary relations on it. Blockmodeling seeks to cluster units that have substantially similar patterns of relationships with others, and interpret the pattern of relationships among clusters. 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). These first approaches to network decomposition concentrated on graph theory notions such as connectedness, cliques or other 'strong' substructures.

The main step towards today's blockmodeling was done by Lorrain and White in 1971 by introducing the notion of structural equivalence. Units are structurally equivalent if they are connected to the rest of the network in identical ways .

It soon become evident that the structural equivalence is to stringent for real networks. White and Reitz in 1983 proposed the regular equivalence as an attempt to better capture the network structure: intuitively, two units are regularly equivalent if they are equally connected to equivalent others.

Batagelj, Doreian and Ferligoj (1992) treated blockmodeling as a clustering problem with the criterion function measuring the departures of obtained blocks from ideal blocks for a selected type of 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 (1993, 1997) 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. In Batagelj (1997) and Batagelj, Ferligoj and Doreian (1998) they proposed a new type of relational constraint – pre-specified models.

2 Constrained clustering problem

Let $E = \{x_1, x_2, \ldots, x_n\}$ be a finite *set of units*. Its nonempty subset $C \subseteq E$ is called a *cluster*. A set of clusters $\mathbf{C} = \{C_i\}$ forms a *clustering*.

The clustering problem $(\mathbf{\Phi}, P, \min)$ can be expressed as follows (Ferligoj and Batagelj, 1982, 1983):
Determine the clustering $\mathbf{C}^* \in \mathbf{\Phi}$, for which

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

where $\mathbf{\Phi}$ is a *set of feasible clusterings* and $P : \mathbf{\Phi} \to \mathbb{R}_0^+$ is a *clustering criterion function*. We denote the *set of minimal solutions* by $\mathrm{Min}(\mathbf{\Phi}, P)$.

Let us introduce some notions which are needed in the following.

The clustering $\mathbf{C}$ is a *complete* clustering if it is a partition of the set of units E. We shall denote by $\mathbf{\Pi}(E)$ the set of all complete clusterings of E. Two among them $\mathbf{O} \equiv \{\{X\} : X \in E\}$ and $\mathbf{I} \equiv \{E\}$ deserve to be denoted by special symbols. The set of feasible clusterings $\mathbf{\Phi}$ can be decomposed into "strata" (layers)

$$\mathbf{\Phi}_k = \{\mathbf{C} \in \mathbf{\Phi} : \mathrm{card}(\mathbf{C}) = k\}$$

Usually the criterion function $P(\mathbf{C})$ has the form:

P1s. $\quad P(\mathbf{C}) = \sum_{C \in \mathbf{C}} p(C) \qquad$, or

P1m. $\quad P(\mathbf{C}) = \max_{C \in \mathbf{C}} p(C)$

where

P2. $\quad p(C) \geq 0 \qquad$, and

P3. $\quad \forall X \in E : p(\{X\}) = 0$

Such a criterion function is called *simple*. The function $p(C)$ is a *cluster error* function and expresses the error (tension) produced by fusing units from C into a cluster. Using some algebra (ordered monoids) conditions **P1**s and **P1**m can be generalized to a single condition.

Let $a \colon E \to A$ be a function that assigns a value to each unit (descriptions of units - can be vector of values of attributes or some other structure). For a selected dissimilarity d defined on A we determine a dissimilarity matrix $\mathbf{D}_a = [d_{XY}]_{E \times E}$ where $d_{xy} = d(a(x), a(y))$. Using it we can define several cluster error functions. For example

$$p(C) = \frac{1}{\mathrm{card}(C)} \sum_{x,y \in C : x < y} d_{xy}$$

or

$$p(C) = \max_{x,y \in C : x < y} d_{xy}$$

For almost all criterion functions used in applications, it holds also:

P4s. $\quad p(C_1 \cup C_2) \geq p(C_1) + p(C_2) \qquad$, or

P4m. $\quad p(C_1 \cup C_2) \geq \max(p(C_1), p(C_2))$.

For a simple criterion function satisfying also the appropriate condition **4**, it holds for $k < n$: $\forall \mathbf{C} \in \mathbf{\Pi}_k \exists \mathbf{C}' \in \mathbf{\Pi}_{k+1} : P(\mathbf{C}') \leq P(\mathbf{C})$. Since $P(\mathbf{C}) \geq 0$ and $P(\mathbf{O}) = 0$, it holds that $\mathbf{O} \in \mathrm{Min}(\mathbf{\Pi}, P)$. To avoid this trivial problem we usually introduce the obvious constraint – we restrict the problem to $\mathbf{\Pi}_k$, where k is a given number of clusters.

2.1 Clustering with relational constraints

Suppose that the units are described by attribute data $a \colon E \to A$ and related by a binary *relation* $R \subseteq E \times E$ that determine the *relational data*

$$\mathcal{N} = (E, R, a)$$

The relation R is often described by the corresponding binary matrix $\mathbf{R} = [r_{xy}]_{E \times E}$ where

$$r_{xy} = \begin{cases} 1 \; x \, R \, y \\ 0 \text{ otherwise} \end{cases}$$

Generally, the set of feasible clusterings for this type of constraint can be defined as:

$$\mathbf{\Phi}(R) = \begin{array}{l}\{\mathbf{C} \in \mathbf{\Pi} : \text{each cluster } C \in \mathbf{C} \text{ is a subgraph } (C, R \cap C \times C) \\ \text{in the graph } (E, R) \text{ with the required type of connectedness}\}\end{array}$$

We can define different types of sets of feasible clusterings for the same relation R if it is nonsymmetric (Ferligoj and Batagelj, 1983). Some examples of clusterings with (nonsymmetric) relational constraint $\Phi^i(R)$ are

type of clusterings	type of connectedness
$\Phi^1(R)$	weakly connected units
$\Phi^2(R)$	weakly connected units that contain at most one center
$\Phi^3(R)$	strongly connected units
$\Phi^4(R)$	clique
$\Phi^5(R)$	the existence of a trail containing all the units of the cluster

A center of a cluster C in the clustering type $\Phi^2(R)$ is the set of units $L \subseteq C$ iff the subgraph induced by L is strongly connected and $R(L) \cap (C \setminus L) = 0$ where $R(L) = \{y : \exists x \in L : xRy\}$.

2.2 Solving constrained clustering problems

With few exceptions the clustering problem is too hard to be exactly solved efficiently (Garey and Johnson, 1979; Shamos, 1976; Brucker, 1978). Therefore, approximative/heuristic methods have to be used. Among these, agglomerative (hierarchical) and local optimization (relocation) methods are the most popular.

Usually, for clustering problems, the neighborhood relation for local optimization is determined by the following two transformations: *moving* a unit X from cluster C_p to cluster C_q (*transition*); and *interchanging* units X and Y from different clusters C_p and C_q (*transposition*).

If the constraints are not too stringent, the relocation method can be applied directly on $\mathbf{\Phi}$; otherwise, we can transform (penalty function method) the problem to an equivalent nonconstrained problem $(\mathbf{\Pi}_k, Q, \min)$ with $Q(\mathbf{C}) = P(\mathbf{C}) + \alpha K(\mathbf{C})$ where $\alpha > 0$ is a large constant and

$$K(\mathbf{C}) = \begin{cases} 0 & \Phi(\mathbf{C}) \\ > 0 & \text{otherwise} \end{cases}$$

There exist several improvements of the basic relocation algorithm: simulated annealing, tabu search, ... (Aarts and Lenstra, 1997).

In a *multicriteria clustering problem* $(\mathbf{\Phi}, P_1, P_2, \ldots, P_k, \min)$ we have several criterion functions $P_t, t = 1, \ldots, k$ over the same set of feasible clusterings $\mathbf{\Phi}$, and our aim is to determine the clustering $\mathbf{C} \in \mathbf{\Phi}$ in such a way that $P_t(\mathbf{C}) \to \min$, $t = 1, \ldots, k$. For these problems we usually try to identify clusterings from the set of *Pareto efficient* clusterings: a clustering is Pareto efficient if it cannot be improved on any criterion without sacrificing on some other criterion. A multicriteria clustering problem can be approached in different ways (Ferligoj and Batagelj, 1992). The direct approach is to apply repeatedly local optimization combined with a sieve, filtering out Pareto clusterings. It can be solved also by using constrained clustering algorithms where a selected criterion is considered as the clustering criterion and all other criteria determine the (optimizational) constraints. And conversely: a constrained clustering problem can be transformed to a multiciteria clustering problem by expressing the deviations from constraints by penalty functions.

For some problems the *dynamic programming* approach can be used (Batagelj, Korenjak and Klavžar, 1994; Lebbe and Vignes, 1996). Some other optimizational approaches for solving constrained clustering problems can be found in Klauer (1994) and Hansen, Jaumard, and Sanlaville (1994).

Existing methods for solving constrained clustering problems are at least of order $O(n^2)$, n is the number of units. Therefore they can be used only for datasets of moderate size (up to some thousands of units).

The relationally constrained clustering methods are implemented in program package CLUSE: HICLUR – hierarchical clustering, and CLUDER – local optimization (Batagelj 1981).

3 Blockmodeling

A clustering $\mathbf{C}$ 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 arcs leading from cluster C_i to cluster C_j (see upper part of Figure 1). If $i = j$, a block $R(C_i, C_i)$ is called a *diagonal* block.

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.

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 (see Figure 1) 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 Figure 2. A block is *symmetric* if

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

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

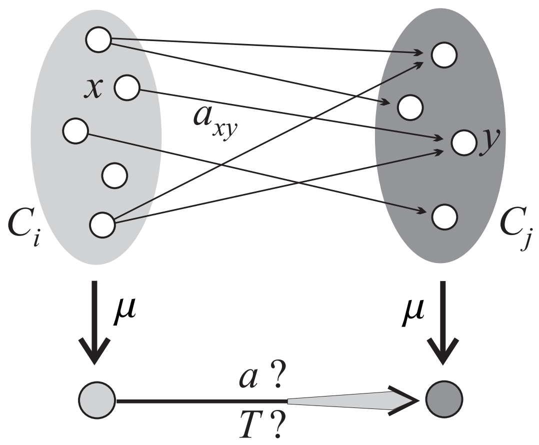

Fig. 1. Blockmodeling scheme.

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

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

$\mathbf{C}(\mu) = \{C(t) : t \in U\}$ is a partition (clustering) of the set of units E.

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

- U is a set of *positions* (types of units);
- $K \subseteq U \times U$ is a set of *connections*;
- $\mathcal{T}$ is a set of predicates used to describe the types of connections between different 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;
- Q is a set of *averaging rules*. A mapping $\alpha : K \to Q$ determines rules for computing values of connections.

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

$$\forall (t, w) \in K : \pi(t, w)(C(t), C(w)) \qquad \text{, and}$$

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

The reduced graph can be presented by a matrix $\mathbf{M}$, called also *image* or *model* matrix.

$$m_{ij} = \text{type of block } R(C_i, C_j)$$

Let $\approx$ be an equivalence relation over E and $[x] = \{y \in E : x \approx y\}$. We say that $\approx$ is *compatible* with $\mathcal{T}$ over a network $\mathcal{N}$ iff

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

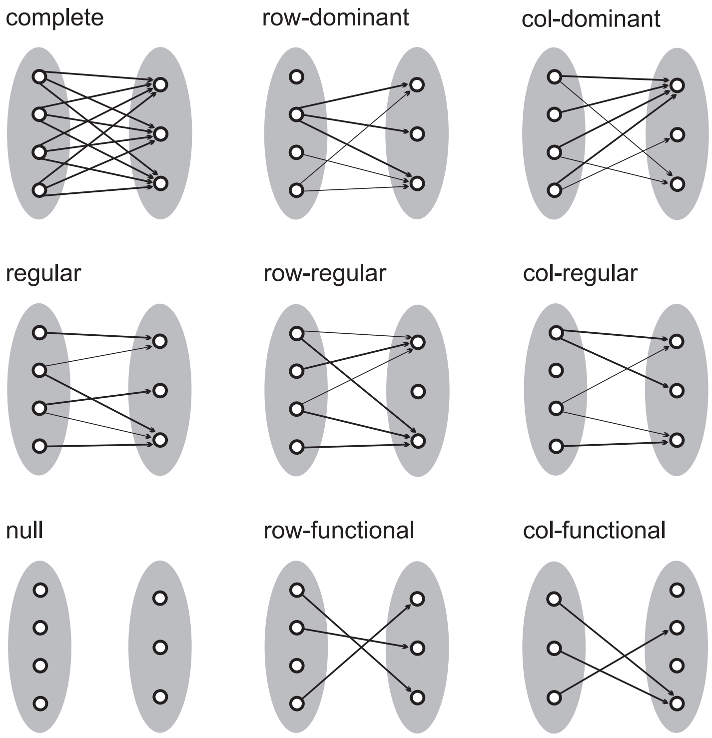

Fig. 2. Types of connection between two sets; the left set is the ego-set.

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

For a compatible equivalence $\approx$ the mapping $\mu\colon x \mapsto [x]$ determines a blockmodel with $U \equiv E/\approx$.

3.1 Optimization approach to blockmodeling

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.

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 set of types of connection $\mathcal{T}$ we can introduce the set of *ideal blocks* for a given type $T \in \mathcal{T}$ by

$$\mathcal{B}(C_i, C_j; T) = \{B \subseteq C_i \times C_j : T(B)\}$$

Table 1. Characterizations of types of blocks.

null	nul	all 0*
complete	com	all 1*
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*
regular	reg	1-covered rows and 1-covered columns
non-null	one	$\exists$ at least one 1

* except may be diagonal

Using Table 1 we can efficiently test whether the block $R(C_i, C_j)$ is of the type T; and define the *deviation* $\delta(C_i, C_j; T)$ of a block $R(C_i, C_j)$ from the nearest ideal block. For example

$$\delta(C_i, C_j; \mathrm{reg}) = |C_i| \cdot (|C_j| - c_j) + |C_j| \cdot (|C_i| - r_i)$$

where c_j is the number of non-zero columns, and r_i is the number of non-zero rows in the block $R(C_i, C_j)$.

We combine block-deviations into a *total error* – blockmodeling *criterion function*

$$P(\mathbf{C}(\mu); \mathcal{T}) = \sum_{(t,w) \in U \times U} \min_{T \in \mathcal{T}} \delta(C(t), C(w); T).$$

For the proposed types from Table 1 it is possible to construct a criterion function P such that

$$P(\mathbf{C}(\mu)) = 0 \Leftrightarrow \mu \text{ is an exact blockmodeling}$$

The obtained optimization problem can be solved by local optimization. Once a partitioning μ and types of connection π are determined, we can also compute the values of connections by using *averaging rules*. For further details see (Batagelj 1997).

The pre-specified blockmodeling starts with a class of blockmodels $\mathcal{M}$ specified (constraints!), in terms of substance, *prior to an analysis*. Given a network, a set of ideal blocks is selected, a reduced model is formulated, and partitions are established by minimizing the criterion function (Batagelj, Ferligoj, Doreian, 1998). The pre-specified blockmodeling is supported by the program MODEL 2 (Batagelj, 1996).

A special class of blockmodeling problems are symmetric-acyclic decompositions (Doreian, Batagelj, Ferligoj, 1998) for which also an algorithm for large networks was developed.

4 General problem of clustering relational data

Can the clustering with relational constraint and blockmodeling problem be generalized to a common problem?

The relationally constrained clustering problem with simple criterion function considers only the diagonal blocks that should be of one of the types $\Phi^i(R)$. It also takes into account the dissimilarity matrix on units (derived from attribute data).

The blockmodeling problem deals only with relational data. The proposed optimization approach essentially expresses the constraints with a penalty function.

Both problems can be expressed as special cases of a clustering problem with a *general criterion function* of the form

G1s. $$P(\mathbf{C}) = \sum_{(C_1,C_2)\in\mathbf{C}\times\mathbf{C}} q(C_1, C_2), \qquad \text{or}$$

G1m. $$P(\mathbf{C}) = \max_{(C_1,C_2)\in\mathbf{C}\times\mathbf{C}} q(C_1, C_2)$$

and

G2. $q(C_1, C_2) \geq 0$

The *set of feasible clusterings* $\Phi_k(R)$ for this problem is determined by the relation R and additional requirements, such as:

- the blocks should be of selected types
- the model graph should be of specified form (prespecified)
- selected units should / should not be in the same cluster
- selected unit should / should not be in the selected cluster

4.1 Approaches to the problem

There are different types of relational data (valued networks). In the following we shall assume

$$\mathcal{N} = (E, R, a, b)$$

where $a\colon E \to A$ assigns a value to each unit and $b\colon R \to B$ assigns a value to each arc (link) of R. A and B are sets of values.

The function b determines a matrix $\mathbf{B} = [b_{ij}]_{n\times n}$, $b_{ij} \in B \cup \{0\}$ and $b_{ij} = 0$ if units i and j are not connected by an arc.

There are two main approaches to solve the problem of clustering relational data:

- *Indirect approach*: transformation to standard data analysis problems;
- *Direct approach*: formulating the problem as an optimization problem and solving it.

Indirect approach. A 'scenario' for the indirect approach is to transform attribute data a into dissimilarity matrix D_a and network data b into dissimilarity matrix D_b and build criterion functions P_a and P_b based on them (they can be defined also directly from a and b). Then we apply the multicriteria relationally constrained clustering methods on these functions. We can also first combine D_a and D_b into a joint matrix D_{ab} and apply relationally constrained clustering methods on it.

In a special case, when D_b is defined as some 'corrected' dissimilarity (see Batagelj, Ferligoj, Doreian, 1992) between descriptions $\mathbf{b}(x) = [\mathbf{B}(x), \mathbf{B}^T(x)]$, the relational data are built into D_b and we can apply on the combined matrix D_{ab} all standard methods for analysis of dissimilarity matrices.

Direct approach. Again there are different possibilities:

1. *Structural approach*: used in program `MODEL` (Batagelj, 1996): Important is the structure (relation). Determine the best clustering $\mathbf{C}$ and the corresponding model. On the basis of a, b and the obtained model compute values of model connections.
2. *Multicriteria approach*: construct two criterion functions: one based on values, the second based on structure. Solve the obtained multicriteria problem (Ferligoj, Batagelj, 1992).
3. *Implicit approach*: the types of connections are built into the criterion function combined with values.

Only the last approach needs some further explanations.

Implicit approach. Let $\approx$ be an equivalence over set of units E, and $\mathcal{T}$ given types. We construct on blocks deviation functions $\delta(C_1, C_2; T), T \in \mathcal{T}$ such that $\approx$ is compatible with $\mathcal{T}$ over the network $\mathcal{N}$ iff

$$\forall x, y \in E \ \exists \delta(.,.;T), \ T \in \mathcal{T} : \delta([x],[y];T) = 0$$

Applying also an adequate normalization of δs we can construct a criterion function

$$P(\mathbf{C}) = \sum_{X,Y \in \mathbf{C}} \min_{T \in \mathcal{T}} \delta(X, Y; T)$$

Evidently, $P(\mathbf{C}) = 0 \quad \Leftrightarrow \quad \mathbf{C}$ is compatible with $\mathcal{T}$ – all blocks of $\mathbf{C}$ are compatible with $\mathcal{T}$.

Some examples. Assume that a and b are transformed into a matrix $\mathbf{A} = [a_{xy}]_{E \times E}$, $a_{xy} \geq 0$. Then

$$\delta(X, Y; \mathrm{nul}) = \frac{\sum_{x \in X, y \in Y} a_{xy}}{|X| \cdot |Y| \cdot \max\{a_{xy} \neq 0 : x \in X, y \in Y\}}$$

$$\delta(X, Y; \mathrm{rdo}) = 1 - \max_{x \in X} \frac{\sum_{y \in Y} a_{xy}}{|Y| \cdot \max\{a_{xy} \neq 0 : y \in Y\}}$$

$$\delta(X, Y; \mathrm{cre}) = 1 - \frac{\sum_{y \in Y} \max_{x \in X} a_{xy}}{|Y| \cdot \max\{a_{xy} \neq 0 : x \in X, y \in Y\}}$$

If max in the denominator equals 0 also the fraction has value 0.

5 Conclusion

In the paper we presented an overview of clustering with relational constraints and blockmodeling problems. We showed that both problems are special cases of the general clustering of relational data problem. For this problem we proposed some possible approches to solve it. The software support along these lines is still to be developed. Of special interest are methods for clustering large relational datasets.

The proposed *general clustering of relational data problem* provides a framework in which several new types of clustering problems can be defined, studied and solved. Perhaps, also other multivariate methods can be extended to analyse such data.

Some related papers are available at
`http://vlado.fmf.uni-lj.si/pub/cluster/`

References

AARTS, E. and LENSTRA, J.K. (Eds.)(1997). *Local Search in Combinatorial Optimization*. Wiley, Chichester.

BATAGELJ, V. (1981), CLUSE – Program package for cluster analysis, manual, Department of Mathematics, University of Ljubljana.

BATAGELJ, V. (1984). Agglomerative methods in clustering with constraints. *Preprint Series Dept. Math. Univ. Ljubljana*, 22(1984)102, 5-19.

BATAGELJ, V. (1996), MODEL 2 – Program for Generalized Pre-Specified Blockmodeling, manual, Department of Mathematics, University of Ljubljana.

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

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

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

BATAGELJ, V., FERLIGOJ, A. and DOREIAN, P. (1998). Fitting Pre-specified Blockmodels. C. Hayashi, and (Eds.): *Data Science, Classification, and Related Methods / IFCS'96*, Springer, Tokyo, 199–206.

BATAGELJ, V., KORENJAK-ČERNE, S. and KLAVŽAR, S. (1994). Dynamic Programming and Convex Clustering. *Algorithmica* **11**, 93–103.

BORGATTI, S.P. and EVERETT, M.G. (1989): The class of all regular equivalences: Algebraic structure and computation. *Social Networks*, 11:65–88.

BRUCKER, P. (1978). On the complexity of clustering problems, in: R.Henn, B. Korte, W. Oettli (Eds.): *Optimization and Operations Research*. Lecture Notes in Economics and Mathematical Systems, Vol. 157, Springer-Verlag, Berlin.

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

DOREIAN, P., BATAGELJ, V. and FERLIGOJ, A. (1998): Symmetric-Acyclic Decompositions of Networks. To appear in *Journal of Classification*.

DOREIAN, P. and MRVAR, A. (1996) A Partitioning Approach to Structural Balance. *Social Networks* 18:149–168.

FAUST, K. (1988): Comparison of methods for positional analysis: Structural and general equivalences. *Social Networks*, 10:313–341.

FERLIGOJ, A. and BATAGELJ, V. (1982). Clustering with relational constraint. *Psychometrika*, **47**, 4, 413–426.

FERLIGOJ, A. and BATAGELJ, V. (1983). Some types of clustering with relational constraints. *Psychometrika*, **48**, 4, 541-552.

FERLIGOJ, A. and BATAGELJ, V. (1992). Direct Multicriteria Clustering Algorithms. *Journal of Classification* **9**, 1, 43–61.

FERLIGOJ, A., BATAGELJ, V. and DOREIAN, P. (1994): On Connecting Network Analysis and Cluster Analysis. In *Contributions to Mathematical Psychology, Psychometrics, and Methodology* (G.H. Fischer, D. Laming Eds.), New York: Springer.

GAREY, M.R. and JOHNSON, D.S. (1979). *Computer and intractability*. Freeman, San Francisco.

GORDON, A.D. (1996). A survey of constrained classification. *Computational Statistics and Data Analysis* **21**, 17–29.

HANSEN, P., JAUMARD, B. and SANLAVILLE, E. (1994). A Review of Mathematical Programming Approaches, in: P. Bertrand, B. Burtschy (Eds.): *New Approaches in Classification and Data Analysis / IFCS'93*, Springer-Verlag, Berlin, 228–240.

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

KLAUER, K.C. (1994). Representing proximities by network models, in: P. Bertrand, B. Burtschy (Eds.): *New Approaches in Classification and Data Analysis / IFCS'93*, Springer-Verlag, Berlin, 493–501.

LEBBE, J. and VIGNES, R. (1996). Optimal Hierarchical Clustering with Order Constraint, in: E. Diday, Y. Lechevallier, O. Opitz (Eds.): *Ordinal and Symbolic Data Analysis / OSDA'95*, Springer-Verlag, Berlin, 265–276.

LORRAIN, F. and WHITE, H.C. (1971): Structural equivalence of individuals in social networks. *Journal of Mathematical Sociology*, 1:49–80.

MURTAGH, F. (1985). A survey of algorithms for contiguity-constrained clustering and related problems. *The Computer Journal* **28**, 82–88.

SHAMOS, M.I. (1976). Geometry and statistics: Problems at the interface, in: J. Traub (Ed.): *Algorithms and Complexity (New directions and recent results)*. Academic Press, New York, 251–288.

WHITE, D.R. and REITZ, K.P. (1983): Graph and semigroup homomorphisms on networks of relations. *Social Networks*, 5:193–234.