1,01

PSYCHOMETRIKA—VOL. 47, NO. 4.
DECEMBER, 1982

CLUSTERING WITH RELATIONAL CONSTRAINT

ANUŠKA FERLIGOJ AND VLADIMIR BATAGELJ

UNIVERSITY EDVARD KARDELJ

The paper deals with clustering problems where grouping is constrained by a symmetric and reflexive relation. For solving clustering problems with relational constraints two methods are adapted: the "standard" hierarchical clustering procedure based on the Lance and Williams formula, and local optimization procedure, CLUDIA. To illustrate these procedures, clusterings of the European countries are given based on the developmental indicators where the relation is determined by the geographical neighbourhoods of countries.

Key Word: optimization approach to clustering.

The clustering problem can be treated as an optimization problem over a set of clusterings. In some cases the set of (feasible) clusterings is determined by some additional conditions—constraints. In these cases we speak of clustering with constraints.

The paper deals with clustering methods where grouping is constrained by a symmetric and reflexive relation. For example, this is the case when the clusters of (geographical) regions also have to be internally connected. For solving this problem two methods are presented which are extensions of clustering methods for the usual clustering problem.

Clustering Problem

Let us start with the formal setting of the clustering problem (with constraints). First we introduce some basic notions [Batagelj, Note 1; Ferligoj & Batagelj, Note 2]:

E	—set of *units*
$C \subseteq E$	—*cluster*; $C \neq \varnothing$
$\mathscr{C} \subseteq \mathscr{P}(E)$	—set of clusters—*clustering*, where $\mathscr{P}(E)$ is a power set of E
$d: (C_1, C_2) \mapsto R^+ \cup \{0\}$	—*dissimilarity* between clusters
$P: \mathscr{C} \mapsto \mathscr{R}^+ \cup \{0\}$	—(clustering) *criterion function*

Usually the criterion function P takes the form

$$P(\mathscr{C}) = \sum_{C \in \mathscr{C}} p(C)$$

or

$$P(\mathscr{C}) = \max_{C \in \mathscr{C}} p(C)$$

where $p(C)$ is the contribution of the cluster $C \in \mathscr{C}$ to the value of the criterion function.

Extended version of the paper presented at the European meeting of the Psychometric Society, Groningen June, 19–21, 1980.

This work was supported in part by the Boris Kidrič Fund, Yugoslavia.

Requests for reprints should be sent to: Anuška Ferligoj, Faculty of Sociology, Political Sciences and Journalism, University Edvard Kardelj, Kardeljeva ploščad 5, 61000 Ljubljana, Yugoslavia.

0033-3123/82/1200-3032$00.75/0

To unify the description we introduce the operation & which stands for $+$, or for max, or for some other operation.

The criterion funcion P is *compatible* with dissimilarity d iff:

$$\text{(i)} \qquad P(\mathscr{C}) = \underset{C \in \mathscr{C}}{\&} \; p(C)$$

$$\text{(ii)} \qquad p(C) = \min_{\phi \subset D \subset C} (p(D) \;\&\; p(C - D) \;\&\; d(D, C - D)) \tag{1}$$

$$\text{(iii)} \qquad \forall X \in E\colon p(\{X\}) = 0$$

and $(R^+ \cup \{0\}, \&, 0, <)$ is an ordered abelian monoid.

Some examples (maximum; minimum and Ward's hierarchical clustering method; Sneath & Sokal, 1973; Everitt, 1974) of criterion functions compatible with dissimilarity d are presented in the Table 1.

With these notions we can express the *clustering problem* (*with constraints*) as follows: Find the clustering $\mathscr{C}^*$ for which

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

where Φ is the set of *feasible* clusterings, which is determined with the (additional) constraints, which are not expressed with criterion function P.

In the extreme case the set of feasible clusterings can also be empty—the clustering problem has no solution.

Some examples of sets of feasible clusterings are:

$\Phi_k = \{\mathscr{C} \mid \mathscr{C}$ is a partition of E into k sets$\}$
often called *complete clustering* into k clusters

or

$\Phi(R) = \{\mathscr{C} \mid \mathscr{C}$ is a partition of E and every
$C \in \mathscr{C}$ has to induce a connected subgraph
$(C, R \cap C \times C)$ in the graph $(E, R)\}$
called *clustering with relational constraint* R;
where $R \subseteq E \times E$ is a symmetric and reflexive relation

or

$\Phi[a, b] = \{\mathscr{C} \mid \mathscr{C}$ is a partition of E and for every
$C \in \mathscr{C} : a \leq |C| \leq b\}$

or combinations of them.

TABLE 1
Examples of Criterion Functions Compatible with Dissimilarity

&	$d(C_1, C_2)$	$p(C)$
max	$\max_{X \in C_1, Y \in C_2} d(X, Y)$	$\max_{X, Y \in C} d(X, Y)$
+	$\min_{X \in C_1, Y \in C_2} d(X, Y)$	value of the minimal spanning tree over C with edge values $d(X, Y)$
+	$\frac{m_1 \cdot m_2}{m_1 + m_2} d_2^2(\bar{C}_1, \bar{C}_2)$	$\sum_{X \in C} d_2^2(X, \bar{C})$

where $m_i = |C_i|$, $\bar{C}_i = \sum_{X \in C_i} X/m_i$ and d_2 is the euclidean distance.

In some cases the clustering problem can be shown to be equivalent to an optimization problem, for which the efficient (polynomial in the size of the problem) exact algorithms are known. But it seems that most of the instances of the clustering problem are *NP*-complete: it is believed that there is no efficient exact algorithm for solving the problem [Garey & Johnson, 1979, pp. 281]. Therefore, usually approximative methods as: agglomerative (hierarchical), divisive, local optimization, reduction of the set of feasible clusterings to a small subset of "promising" clusterings, ... [Sneath & Sokal, 1973; Everitt, 1974; Hartigan, 1975; Späth, 1977; Lefkovitch, 1980] have to be used.

Let us try to explain in the proposed formalization the connection between the clustering problem and the agglomerative (hierarchical) methods for its solution.

The clustering $\mathscr{C}$ is a *tree clustering* iff

(i) $E \in \mathscr{C}$

(ii) $\forall C_1, C_2 \in \mathscr{C} : C_1 \subseteq C_2 \lor C_2 \subseteq C_1 \lor C_1 \cap C_2 = \varnothing$ (2)

and it is a *complete tree clustering* iff also

(iii) $\forall X \in E : \{X\} \in \mathscr{C}$

holds.

If the criterion function P is compatible with the dissimilarity d and the operation & distributes over min (this is obvious if & is + or max)

$$a \,\&\, \min_i b_i = \min_i (a \,\&\, b_i)$$

it can be shown [Batagelj, Note 1] that the equality

$$P(\mathscr{C}_k^*) = \min_{\mathscr{C} \in \Phi_k} P(\mathscr{C}) = \min_{C_1, C_2 \in \mathscr{C} \in \Phi_{k+1}} (P(\mathscr{C}) \,\&\, d(C_1, C_2)) \tag{3}$$

holds. From (3) we can see the following heuristic (approximation)

$$P(\mathscr{C}_k^*) \approx P(\mathscr{C}_{k+1}^*) \,\&\, \min_{C_1, C_2 \in \mathscr{C}_{k+1}^*} d(C_1, C_2) \tag{4}$$

which is the basis for usual (binary) hierarchical clustering method. This method generates, starting with $\mathscr{C}_n = \{\{X\} \mid X \in E\}$, a sequence of complete clusterings

$$\mathscr{C}_n, \quad \mathscr{C}_{n-1}, \quad \mathscr{C}_{n-2}, \ldots, \quad \mathscr{C}_2, \quad \mathscr{C}_1$$

which forms a complete tree clustering $\mathscr{C}_T = \bigcup_{i=1}^n \mathscr{C}_i$.

Clustering with Relational Constraint

As mentioned before for a given symmetric and reflexive relation $R \subseteq E \times E$ the set of feasible clusterings $\Phi_k(R)$ consists of complete clusterings into k clusters $\mathscr{C}$ for which the units of every cluster $C \in \mathscr{C}$ induce a connected subgraph $(C, R \cap C \times C)$ in the graph (E, R). Without loss of generality we can request that (E, R) has to be connected; if it is not, we can analyse each component separately.

For solving the corresponding *clustering problem with relational constraint* we adapted two methods:

—the "standard" hierarchical clustering procedure based on Lance and Williams [1967] formula:

$$d(C_p \cup C_q, C_s) = \alpha_1\, d(C_p, C_s) + \alpha_2\, d(C_q, C_s) + \beta\, d(C_p, C_q) + \gamma\, |d(C_p, C_s) - d(C_q, C_s)|. \tag{5}$$

The equality (3) holds also for $\Phi_k(R)$.

—the local optimization procedure CLUDIA [Späth, 1977] for the cases

$$P(\mathscr{C}) = \sum_{C \in \mathscr{C}} p(C), \qquad p(C) = \frac{1}{2|C|} \sum_{X, Y \in C} d(X, Y)$$

and

$$P(\mathscr{C}) = \max_{C \in \mathscr{C}} p(C), \qquad p(C) = \max_{X, Y \in C} d(X, Y)$$

which allow quick updates of $p(C)$. The criterion function of the Ward's method is the special case of the first type of criterion function for $d(X, Y) = \| X - Y \|^2$.

In the descriptions of the procedures the (constraint) relation R is used in the form of the set $R(X)$ of neighbours of unit/cluster X

$$R(X) = \{Y \mid X\ R\ Y\}.$$

Procedure for Hierarchical Clustering with Relational Constraint

The procedure for hierarchical clustering with relational constraint is a straightforward adaption of ordinary hierarchical clustering procedure:

1. Each unit is a cluster:

$$C_i = \{X_i\}, \qquad X_i \in E, \qquad i = 1, 2, \ldots, n.$$

2. Repeat while there exist at least two neighbours:
2.1 Determine the nearest pair of neighbours (C_p, C_q)

$$d(C_p, C_q) = \min_{u, v} \{d(C_u, C_v) \mid C_u R\ C_v \wedge u \neq v\}$$

2.2. Fuse clusters C_p and C_q into a new cluster C_r:
2.2.1. Substitute clusters C_p and C_q by the cluster C_r;
2.2.2. Adjust the relation R:

$$R(C_r) = \{C_r\} \cup R(C_p) \cup R(C_q) - \{C_p, C_q\}$$

$$R(C_s) = \begin{cases} R(C_s) \cup \{C_r\} - \{C_p, C_q\} & C_s \in R(C_r) \\ R(C_s) & \text{otherwise} \end{cases}$$

2.2.3. Determine the dissimilarities between the cluster C_r and the other clusters according to the Lance and Williams formula (5).

The hierarchical solutions of clustering problems with relational constraints obtained by this procedure were often nonmonotonic. That means: let h be the (clustering) *level* of clusters (in the dendrogram) defined as follows:

$$\text{(i)} \qquad X \in E \Rightarrow h(\{X\}) = 0.$$

$$\text{(ii)} \qquad C_r = C_p \bigcup C_q \Rightarrow h(C_r) = d(C_p, C_q). \tag{6}$$

Then the clustering $\mathscr{C}$ is *monotonic* iff always:

$$h(C_r) \geq \max\ (h(C_p), h(C_q)).$$

We proved the following *theorem*:

The hierarchical clustering procedure based on the Lance and Williams formula (5) with coefficients $(\alpha_1, \alpha_2, \beta, \gamma)$ guarantees monotonic clusterings for each dissimilarity matrix D and for each relational constraint R ($R \neq E \times E$, (E, R) is a connected graph) iff at each step of the clustering procedure the following conditions hold:

$$\text{(i)} \qquad \alpha_1 + \alpha_2 \geq 0.$$

$$\text{(ii)} \qquad \gamma + \min(\alpha_1, \alpha_2) \geq 0.$$

$$\text{(iii)} \qquad \min(\alpha_1 + \alpha_2, \gamma + \min(\alpha_1, \alpha_2)) + \beta \geq 1. \qquad (7)$$

The proof of the theorem is given in the Appendix.

Among the common clustering strategies: minimum, maximum, centroid, median, group average, Ward's [Sneath & Sokal, 1973; Everitt, 1974], only maximum strategy fulfills the third condition (7iii). The Ward's strategy coefficients can be adapted (to fulfill also this condition) for example in the following ways:

$$\alpha_1 = \frac{m_p + m_s}{m}, \qquad \alpha_2 = \frac{m_q + m_s}{m}$$

where m_p, m_q and m_s are the numbers of units in clusters C_p, C_q and C_s; $m = m_p + m_q + m_s$; and

$$\beta = 1 - \min(\alpha_1, \alpha_2), \qquad \gamma = 0$$

or

$$\beta = -\frac{m_s}{m}, \qquad \gamma = \max(\alpha_1, \alpha_2).$$

The impact of these variations on the Ward's strategy was not studied till now.

Local Optimization Procedure for Clustering with Relational Constraint

The main idea of a local optimization procedure is quite simple:

1. Determine (read or random generate) the initial clustering $\mathscr{C} \in \Phi_k(R)$.
2. While there exist $X \in C_p \in \mathscr{C}$ and $C_q \in \mathscr{C}$: $P(\mathscr{C}) > P(\mathscr{C}')$ where $\mathscr{C}' = (\mathscr{C} - \{C_p, C_q\}) \bigcup \{C_p - \{X\}, C_q \bigcup \{X\}\}$ and subgraphs induced by $C_p - \{X\}$ and $C_q \bigcup \{X\}$ in (E, R) are connected
   repeat:
   2.1. Substitute $\mathscr{C}$ by $\mathscr{C}'$.

In this procedure the (local optimization) *neighbourhood of clusterings* is based on the transformation which transfers the unit X from the cluster C_p to the cluster C_q. Another useful transformation is

$$\mathscr{C}' = (\mathscr{C} - \{C_p, C_q\}) \bigcup \{C_p \bigcup \{Y\} - \{X\}, C_q \bigcup \{X\} - \{Y\}\}$$

i.e., the clustering $\mathscr{C}'$ is obtained by interchanging units $X \in C_p$ and $Y \in C_q$ in the clustering $\mathscr{C}$. Because this transformation preserves the number of units in clusters it is especially suitable for solving clustering problems of the type $\Phi[a, b]$.

Two (sub)problems have still to be solved:

— testing for connectedness of the subgraph induced by a cluster;
— random generation of initial clustering.

The procedure for testing whether the cluster C is connected in (E, R) is the following:

1. $\Delta = C - \{X\}$, $S = \{X\}$ where $X \in C$ is arbitrarily chosen.

2. While $\Delta \neq \emptyset$ repeat:
2.1. If $\exists Y \in \Delta: R(Y) \bigcap S \neq \emptyset$ then
2.1.1. $S = S \bigcup R(Y)$
else
2.1.2. Return (not connected).
2.2. $\Delta = C - S$.
3. Return (connected).

For random generation of initial clustering $\mathscr{C} \in \Phi_k(R)$ we use the following modification of the procedure for hierarchical clustering with relational constraint:

1. $\mathscr{C} = \{\{X\} \mid X \in E\}$; the set of completed clusters is empty.
2. While $|\mathscr{C}| > k$ and not all clusters are completed repeat:
2.1. Select at random a noncompleted cluster C_p.
2.2. $\Delta = R(C_p) - \{C_p\}$.
2.3. If $\Delta \neq \emptyset$ then:
2.3.1. Select at random cluster $C_q \in \Delta$.
2.3.2. Fuse C_p and C_q into a new cluster $C_r = C_p \bigcup C_q$:
$\mathscr{C} = \mathscr{C} \bigcup \{C_r\} - \{C_p, C_q\}$.
2.3.3. Adjust the relation R:

$$R(C_r) = \{C_r\} \bigcup R(C_p) \bigcup R(C_q) - \{C_p, C_q\}$$

$$R(C_s) = \begin{cases} R(C_s) \bigcup \{C_r\} - \{C_p, C_q\} & C_s \in R(C_r) \\ R(C_s) & \text{otherwise.} \end{cases}$$

Else
2.3.4. Add C_p to the completed clusters.

Another way, suggested by one of the referees, to obtain random initial clustering is the following: Take k random seed units; add each of the remaining units in turn at random to one of the clusters they are connected to.

Note that each $\mathscr{C} \in \Phi_k(R)$ can be obtained by the described procedure; but not with the same probability. To approach this goal the random selections in 2.1. and 2.3.1. may be made to depend on some distributions. For example:

— on the number of units in clusters
— on the number of neighbours of clusters
— on the number of different neighbours of clusters.

All the described procedures are implemented in the collection of clustering programs CLUSE on CYBER 72 and DEC–10 [Batagelj, Note 3].

In the case of other types of constraints the local optimization technique with appropriately selected neighbourhood (transformations) of clusterings is a general method for solving such problems. In the program system CLUSE this approach is used to solve the clustering problems of the type $\Phi_k[a, b](R)$, i.e., the clustering problems with relation constraint R into k clusters where the number of units in each cluster has to be inside the interval $[a, b]$.

Some Other Approaches to the Clustering with Constraints

Recently Lefkovitch [1980] treated the clustering problem with constraints (conditional clustering) proposing a method for generating a limited number of subsets from which the optimal partitions and coverings can be obtained with exact methods. The constraints are considered while generating subsets.

A reviewer has pointed out that there are some similarities between our work and the work of some authors in the French literature. We are especially grateful to Dr. Christophe Perruchet who sent us copies of the related papers. Lebart [1978], Perruchet [Note 4], and Lechevallier [1980] are treating the clustering problem with contiguity constraint. For solving it Lebart and Perruchet proposed the algorithms based on a hierarchical clustering method, similar to our procedure for hierarchical clustering with relational constraint. This is the main similarity between these works and our paper.

Example

To illustrate the clustering with relational constraints we clustered the European countries on the basis of the developmental indicators, where the relation is determined with the geographical neighbourhood of countries. There are some difficulties to determine the neighbourhood between some littoral countries (for example: between Ireland and Spain). In our analysis we used the neighbourhood relation as it is presented in Table 2.

We considered only 27 European countries—because of lack of the data we excluded small countries: Andora, Liechtenstein, Vatican, San Marino, Monaco, and Malta.

TABLE 2

NEIGHBOURHOOD RELATION FOR EUROPEAN COUNTRIES

1	Albania	27	11	15						
2	Austria	27	12	5	10	24	15			
3	Belgium	8	16	10	17	26				
4	Bulgaria	27	11	21						
5	Czechoslovakia	2	12	25	19	9	10			
6	Denmark	18	23	9	10					
7	Finland	25	23	18						
8	France	26	22	15	24	10	16	3		
9	East Germany	10	5	19	23	6				
10	West Germany	17	3	16	24	2	5	9	6	8
11	Greece	15	1	27	4					
12	Hungary	2	27	21	25	5				
13	Iceland	18	26	14						
14	Ireland	26	13	22						
15	Italy	27	1	11	8	24	2			
16	Luxembourg	3	8	10						
17	Netherland	3	10	26						
18	Norway	6	26	23	7	25	13			
19	Poland	25	5	9	23					
20	Portugal	22								
21	Romania	25	12	27	4					
22	Spain	20	14	8						
23	Sweden	18	6	9	19	25	7			
24	Switzerland	8	15	2	10					
25	USSR	18	7	23	19	5	12	21		
26	United Kingdom	14	13	17	18	3	8			
27	Yugoslavia	2	15	1	11	4	21	12		

Among the available socio-economic and demographic variables we selected the following ones [The Hammond Almanac, 1980]:

— urban population per capita,
— density of the population,
— population in the largest city per capita,
— per capita income,
— industrial production per total production,
— birthrate,
— deathrate,
— life expectancy,
— number of inhabitants per physician,
— infant mortality,
— enrollment in higher education per capita,
— paved roads per area,
— number of motor vehicles per capita,
— railway mileage per area,
— number of radio-receivers per capita,
— number of television subscribers per capita,
— number of telephone subscribers per capita,
— number of newspaper copies per capita,
— number of inhabitants per hospital bed.

All variables were standardized. To measure the dissimilarities between countries we used the following coefficient:

$$d_{ij} = \frac{1 - r_{ij}}{2}$$

where r_{ij} is the Pearson correlation coefficient.

The Table 3 presents the results of local optimization procedures (criterion function of Ward's type) for ordinary and relational clustering into six groups for 20 random initial configurations and for initial configurations obtained by Ward's hierarchical clustering strategy. The values of the criterion function for the initial clusterings are denoted with P_0 and the values for the corresponding local minima with $P_{\min}$. The best obtained local minima of the criterion function P are indicated by an asterisk. The values of the criterion function of the (initial) clusterings obtained with hierarchical strategies are close to the obtained best local minima and therefore these clusterings provide a good starting point for further local optimization which corresponds to the experiences in other similar empirical analyses.

The obtained hiearchical and the best local optimization clusterings are presented in Table 4. Although the values of the criterion function of the clustering obtained with hierarchical strategy and of the corresponding best local minimum are very close, there are some differences between them. The comparison of the best local minima (clusterings) without and with relational constraint shows that they are quite similar. This means that the development in European countries is correlated with the geographical neighbourhood and tradition. For example, there are two identical groups: the "Scandinavian" group (Denmark, Finland, Iceland, Norway, and Sweden) and the "Austro-Hungarian" group (Austria, Czechoslovakia, East Germany, and Hungary). The influence of the relational constraint can be seen in the "Southern" group (Greece, Ireland, Portugal, Spain, and Yugoslavia) which in the relational case splits into two groups: "Balkan" and "Irish-Iberian" group because of lack of the geographical connection between them.

TABLE 3

LOCAL OPTIMIZATION

Dissimilarity: Pearson correlation coefficient
Criterion function: Ward´s
N. of groups = 6

initial configuration	ORDINARY CLUSTERING P_o	P_{min}	CLUSTERING WITH RELATIONAL CONSTRAINT P_o	P_{min}
1	5.461	2.801 *	5.209	3.604
2	5.413	3.005	5.113	3.099
3	5.900	2.801 *	4.777	3.023 *
4	5.447	2.875	4.898	3.313
5	5.708	2.879	4.919	3.807
6	5.334	2.988	4.901	3.543
7	5.956	2.991	5.022	3.099
8	5.814	2.806	5.047	3.138
9	4.737	2.801 *	4.815	3.940
10	5.255	2.806	5.212	3.654
11	5.550	2.961	5.279	3.090
12	5.522	3.026	5.245	3.311
13	5.122	2.843	4.493	3.698
14	5.744	2.875	5.327	3.875
15	5.034	2.875	4.758	3.280
16	5.571	2.951	5.254	3.099
17	5.042	2.986	4.797	3.058
18	5.901	2.801 *	5.054	3.031
19	5.024	2.801 *	5.135	3.023 *
20	5.614	2.801 *	5.465	3.775
Ward´s h. c.	2.897	2.801 *	3.031	3.031

Conclusion

In the paper we have developed an optimization approach to clustering (with constraints). In this framework we have treated the problem of clustering with relational constraints and for solving it we have proposed two methods which are extensions of the existing techniques. The relational constraint in the example is the geographical neighbourhood but other types of problems can be formalized as clustering problems with relational constraint.

Appendix: *The Proof of the Monotonicity Theorem*

Let us first prove that the conditions (7) are sufficient: if the method fulfills the conditions (7) then in each step of the procedure the monotonicity condition holds

$$d_{k(ij)} \geq d_{ij}, \tag{8}$$

where duv = d(Cu, Cv).

TABLE 4

CLUSTERINGS OF EUROPEAN COUNTRIES

Dissimilarity: Pearson correlation coefficient
Criterion function: Ward´s
N. of groups = 6

ORDINARY CLUSTERINGS		RELATIONAL CLUSTERINGS	
hierarchical	local optimization	hierarchical	local optimization
BEL	BEL	BEL	BEL
W.G.	W.G.	E.G.	FRA
ITA	ITA	W.G.	W.G.
NET	NET		ITA
SWI	SWI		LUX
		FRA	NET
		ITA	SWI
LUX	FRA	LUX	U.K.
U.K.	LUX	NET	
	U.K.	SWI	
		U.K.	AUS
AUS			CZE
BUL	AUS		E.G.
CZE	CZE	AUS	HUN
E.G.	E.G.	CZE	
HUN	HUN	HUN	
			DEN
			FIN
DEN	DEN	DEN	ICE
FIN	FIN	FIN	NOR
FRA	ICE	ICE	SWE
ICE	NOR	NOR	
NOR	SWE	SWE	
SWE			BUL
			POL
	ALB	ALB	ROM
POL	BUL	BUL	USS
ROM	POL	GRE	
USS	ROM	POL	
	USS	ROM	ALB
		USS	GRE
ALB		YUG	YUG
GRE	GRE		
IRE	IRE		
POR	POR	IRE	IRE
SPA	SPA	POR	POR
YUG	YUG	SPA	SPA
2.897	2.801	3.031	3.023

Let C_p be the farthest of the clusters C_i and C_j from the cluster C_k, and let C_q be the nearest of them, i.e.,

$$d_{kp} \geq d_{kq} \quad \text{and} \quad \{i, j\} = \{p, q\}. \tag{9}$$

Then

$$|d_{ki} - d_{kj}| = d_{kp} - d_{kq}. \tag{10}$$

Let us also define

$$\alpha'_1 = \begin{cases} \alpha_1 & p = i \\ \alpha_2 & p = j \end{cases} \quad \text{and} \quad \alpha'_2 = \begin{cases} \alpha_1 & q = i \\ \alpha_2 & q = j \end{cases}. \tag{11}$$

Then we can rewrite the formula (5):

$$\begin{aligned} d_{k(pq)} = d_{k(ij)} &= \alpha'_1 d_{kp} + \alpha'_2 d_{kq} + \gamma(d_{kp} - d_{kq}) + \beta d_{pq} \\ &= (\alpha'_1 + \gamma) d_{kp} + (\alpha'_2 - \gamma) d_{kq} + \beta d_{pq}. \end{aligned} \tag{12}$$

In the case where nonmonotonicity results from the fusion of the cluster C_k with the cluster $C_{(ij)}$ there must be at least one of the clusters C_i and C_j in the relation with the cluster C_k. Because of the fusion of the clusters C_i and C_j, at least one of the inequalities $d_{ki} \geq d_{ij}$ or $d_{kj} \geq d_{ij}$ holds. The last two inequalities can be combined in the inequality $\max(d_{ki}, d_{kj}) \geq d_{ij}$ or

$$d_{kp} \geq d_{pq}. \tag{13}$$

From $a \geq \min(a, b)$ and the condition (7iii) it follows:

$$\gamma + \min(\alpha_1, \alpha_2) + \beta \geq 1 \tag{14}$$

and

$$\alpha_1 + \alpha_2 + \beta \geq 1. \tag{15}$$

To prove that from the conditions (7) the monotonicity (8) follows, we shall consider two cases, which appear in the analysis of the formula (12):

1. $\alpha'_2 - \gamma \geq 0;$

From the supposition $\alpha'_2 - \gamma \geq 0$ it follows $(\alpha'_2 - \gamma) d_{kq} \geq 0$. Considering it in (12) we get

$$d_{k(pq)} \geq (\alpha'_1 + \gamma) d_{kp} + \beta d_{pq},$$

and, further, from (13) and the relation $\alpha'_1 \geq \min(\alpha_1, \alpha_2)$, we have

$$d_{k(pq)} \geq (\alpha'_1 + \gamma + \beta) d_{pq} \geq (\gamma + \min(\alpha_1, \alpha_2) + \beta) d_{pq}$$

which gives, combined with the condition (14), the monotonicity condition:

$$d_{k(pq)} \geq d_{pq}.$$

2. $\alpha'_2 - \gamma < 0;$

From $\alpha'_2 - \gamma < 0$ and (9) it follows

$$(\alpha'_2 - \gamma) d_{kq} \geq (\alpha'_2 - \gamma) d_{kp}.$$

Considering the last inequality and $\alpha'_1 + \alpha'_2 = \alpha_1 + \alpha_2$ in (12) we get

$$d_{k(pq)} \geq (\alpha_1 + \alpha_2) d_{kp} + \beta d_{pq}$$

and further from the inequalities (7i) and (13)

$$d_{k(pq)} \geq (\alpha_1 + \alpha_2 + \beta)d_{pq}$$

which gives combined with the condition (15) the monotonicity condition.

This completes the first part of the proof.

In the second part of the proof we have to prove that the conditions (7) are also necessary: if the method is monotonic then the conditions (7) hold. We shall follow the logically equivalent way: if the method does not fulfill the conditions (7) then the method does not guarantee monotonic clusterings. In this case it is sufficient to find at least one dissimilarity matrix D and relation R, for which the method is not monotonic.

We shall use the same notations as in the first part of the proof and let us suppose $(k, q) \notin R$. Without loss of generality of the proof we can suppose also

$$\alpha_1' \leq \alpha_2' \tag{16}$$

or

$$\alpha_1' = \min(\alpha_1, \alpha_2).$$

Then the nonmonotonicity condition $d_{k(pq)} < d_{pq}$ can be written as follows:

$$(\alpha_1' + \gamma)d_{kp} + (\alpha_2' - \gamma)d_{kq} < (1 - \beta)d_{pq}. \tag{17}$$

To prove the second part of the theorem we have to show that the inequality (17) has at least one solution (d_{kp}, d_{kq}), $d_{kp} \geq d_{kq}$, $d_{kp} \geq d_{pq}$ as soon as the conditions (7) do not hold. There are three cases to be considered:

1. $$\alpha_1 + \alpha_2 < 0.$$

In this case we put $d_{kp} = d_{kq}$ in the inequality (17) and we get

$$(\alpha_1 + \alpha_2)\, d_{kp} < (1 - \beta)\, d_{pq}$$

from where we obtain the solutions

$$d_{kp} = d_{kq} > d_{pq} \max\left(1, \frac{1 - \beta}{\alpha_1 + \alpha_2}\right).$$

2. $$\alpha_1 + \alpha_2 \geq 0; \qquad \gamma + \min(\alpha_1, \alpha_2) < 0.$$

Putting $d_{kq} = d_{pq}$ in the inequality (17) we get

$$(\alpha_1' + \gamma)d_{kp} < (1 - \beta + \gamma - \alpha_2')d_{pq}.$$

The inequality (17) has at least the following solutions (d_{kp}, d_{kq}):

$$d_{kq} = d_{pq}$$

and

$$d_{kp} > d_{pq} \max\left(1, \frac{1 - \beta + \gamma - \alpha_2'}{\alpha_1' + \gamma}\right).$$

3. $$\alpha_1 + \alpha_2 \geq 0; \qquad \gamma + \min(\alpha_1, \alpha_2) \geq 0;$$

$$\min(\alpha_1 + \alpha_2, \gamma + \min(\alpha_1, \alpha_2)) + \beta < 1.$$

Considering which coefficient dominates in min in the last inequality, we split the analysis in two cases:

3a. $$\alpha_1 + \alpha_2 \geq \gamma + \min(\alpha_1, \alpha_2);$$

In this case the condition *not* (7iii) is the following

$$\gamma + \alpha_1' + \beta < 1. \tag{18}$$

Putting $d_{kq} = 0$ in the inequality (17) we get:

$$(\alpha_1' + \gamma)d_{kp} < (1 - \beta)d_{pq}.$$

Supposing $\alpha_1' + \gamma = 0$ it follows from (18) $1 - \beta > 0$ and therefore each $d_{kp} \geq d_{pq}$ solves the inequality; otherwise from inequality (18) it follows $1 < (1 - \beta)/(\alpha_1' + \gamma)$ and the inequality is solved by all d_{kp} which satisfy the following condition:

$$d_{pq} \leq d_{kp} < \frac{1 - \beta}{\alpha_1' + \gamma} d_{pq}.$$

3b. $$\alpha_1 + \alpha_2 < \gamma + \min(\alpha_1, \alpha_2);$$

In this case the condition *not* (7iii) takes the form

$$\alpha_1 + \alpha_2 + \beta < 1. \tag{19}$$

Let us set $d_{kp} = d_{kq}$. Considering it in (17) we get

$$(\alpha_1 + \alpha_2)d_{kp} < (1 - \beta)d_{pq}.$$

The case $\alpha_1 + \alpha_2 = 0$ can be treated as in the case 3a; otherwise from (19) follows $1 < (1 - \beta)/(\alpha_1 + \alpha_2)$. Therefore the inequality is solved by all d_{kp} which fulfill the following condition:

$$d_{pq} \leq d_{kp} < \frac{1 - \beta}{\alpha_1 + \alpha_2} d_{pq}.$$

We exhausted all possible cases and in each case we found the solutions of the nonmonotonicity inequality (17). Therefore also the second part of the theorem holds. This completes the proof.

REFERENCE NOTES

1. Batagelj, V. Clustering—basic notions. *Seminar for numerical mathematics and computer science*, 156. Ljubljana: DMFA SRS, 1979 (in Slovene).
2. Ferligoj, A., & Batagelj, V. *Clustering methods in the social sciences*. (report). Ljubljana: RIFSPN, 1980 (in Solvene).
3. Batagelj, V. CLUSE. (manual). Ljubljana, 1980.
4. Perruchet, C. *Classification sous contrainte de contiguité continue* (Application aux sciences de la terre). Thesis. Paris: 1979 (in French).

REFERENCES

Batagelj, V. Note on ultrametric hierarchical clustering algorithms. *Psychometrika*, 1981, *46*, 351–352.

Everitt, B. *Cluster analysis*. London: Heinemann Educational Books, 1974.

Garey, M. R., & Johnson, D. S. *Computer and intractability*. San Francisco: Freeman, 1979.

The Hammond almanac 1980. Maplewood, New Jersey: Hammond Almanac, Inc., 1980.

Hartigan, J. A. *Cluster algorithms*. New York: Wiley, 1975.

Lance, G. N., & Williams, W. T. A general theory of classificatory sorting strategies, 1. Hierarchical systems. *The Computer Journal*, 1967, *9*, 373–380.

Lebart, L. Programme d'Agrégation avec Contraintes (C.A.H. Contiguité). *Les Cahiers de l'Analyse des Données*, 1978, *3*, 275–287 (in French).

Lechevallier, Y. Classification sous contraintes. In Diday, E. & others (Eds.). *Optimisation en Classification Automatique*. Paris: INRIA, 1980, 677–696 (in French).

Lefkovitch, L. P. Conditional clustering. *Biometrics*, 1980, *36*, 43–58.
Sneath, P. H. A., & Sokal, R. R. *Numerical taxonomy*. San Francisco: Freeman, 1973.
Späth, H. *Cluster analyse algorithmen.* München: R. Oldenbourg Verlag, 1977 (in German)

Manuscript received 10/8/80
First revision received 9/9/81
Final version received 4/20/82