Invariant Hierarchical Clustering Schemes

Ildar Batyrshin and Tamas Rudas

Summary. A general parametric scheme of hierarchical clustering procedures with invariance under monotone transformations of similarity values and invariance under numeration of objects is described. This scheme consists of two steps: correction of given similarity values between objects and transitive closure of obtained valued relation. Some theoretical properties of considered scheme are studied. Different parametric classes of clustering procedures from this scheme based on perceptions like "keep similarity classes," "break bridges between clusters," etc. are considered. Several examples are used to illustrate the application of proposed clustering procedures to analysis of similarity structures of data.

1 Introduction

At least two goals can be associated with cluster analysis of the set of objects based on information about similarity values between the objects: (1) decomposition of the set of objects into classes of similar objects and (2) analysis of similarity structure of this set. Unfortunately, many clustering algorithms seeking decomposition of given set of objects into given number of classes of similar objects do not bring out underlying structure but fit the data to some preconceived model [15, 26]. A user of cluster analysis packages can be very happy with good clusters obtained for his data by some standard clustering procedure but it is very possible that the obtained structure of clusters does not reflect intrinsic structure of data but imposed by specifications of clustering algorithm. One of the reasons of this disadvantage of many popular clustering procedures is their noninvariance under numeration of objects. The permutation of numeration of objects at the input of noninvariant clustering procedure often causes the change of results of clustering.

It means that the clustering obtained for given numeration of objects does not reflect the structure of the set of objects. A simple example of such noninvariance of classical algorithms is considered in Sect. 2. The requirement of invariance of clustering algorithms under numeration (permutation, ordering) of objects is considered in cluster analysis as most important requirement [1, 6, 8, 21], but, unfortunately, an overwhelming majority of popular clustering algorithms do not satisfy this property. This property is fulfilled for single linkage (also called nearest neighbor) algorithm discussed in many papers [15, 21–23]. This algorithm builds chains of clusters and for this reason reflects only a specific point of view on "cluster" which does not always acceptable. In this paper we consider parametric scheme of invariant clustering procedures which can vary the point of view on "cluster" and includes single linkage algorithm as a partial case.

Another important requirement on clustering algorithms is invariance under monotone transformation of similarity values between objects [18, 21, 22, 24]. This is a necessary requirement on clustering algorithm if similarity values are evaluated by experts in ordinal scale. This requirement is desirable also for insensitivity of the results of clustering to the choice of similarity or dissimilarity measure.

In this chapter we study a general scheme of hierarchical clustering procedures satisfying both invariance requirements considered above. This scheme initially proposed by Batyrshin [2–4] is based on the concept of a fuzzy equivalence relation introduced and studied in [28, 30]. Clustering procedure in this scheme consists in two steps: correction of given similarity values between objects and max–min transitive closure of obtained valued (fuzzy) relation. When a correction of similarity values does not used and only transitive closure of given similarity relation is applied then the clustering scheme gives clustering procedure proposed in [28] which is similar to single linkage algorithm [16]. Since transitive closure is invariant under numeration of objects and under monotone transformations of similarity values, the clustering procedure will satisfy both types of invariance if correction procedure satisfies them. Several schemes of such invariant parametric correction procedures are considered in this chapter.

To build a rational clustering procedure in considered scheme, it is necessary to propose suitable correction procedure. The chapter studies the properties of similarity relations and correction procedures related with the perceptions of "natural" cluster and "rational" clustering. Such relationships are formulated as propositions with main results given in Theorem 2 and Proposition 5. Theorem 2 gives reasons for construction of general

class of correction procedures as transformations decreasing similarity values in initially given similarity relation. Proposition 5 says that for some class of such transformations the resulting clustering procedure will satisfy the property "keep similarity classes." This result is used further for construction of clustering procedures "breaking" similarity classes considered as "bridges" between clusters.

Some basic definitions and properties of valued relations are discussed in Sect. 3. In Sect. 4 we consider a general scheme of clustering procedures and its relation with a solution of a problem of approximation of valued similarity relations by valued equivalence relation. Section 5 discusses theoretical properties of the first version of this scheme [3] based on identity neighborhood functions and on perception "keep similarity classes." Section 6 discusses methods of extraction of valuable clusters from parametric dendrograms constructed by clustering scheme. Application of clustering procedures with identity neighborhood functions to clustering of Windham's data [13] is considered in Sect. 7. Section 8 considers clustering scheme with nonidentity neighborhood functions which is based on perception "break bridges between clusters." This scheme is illustrated on "butterfly" data. Example of clustering of time series from economics is considered in Sect. 9. In Sect. 10 we summarize results of the chapter and discuss possible extensions of considered clustering schemes.

2 Invariance and Noninvariance of Clustering Procedures

Let us consider a very simple example of a set of seven points symmetrically located on a circle (Fig. 1a). Initial information for clustering is given as a matrix of distances between the objects (see Appendix 1). For numeration of objects considered in Fig. 1b the average linkage clustering algorithm realized in Matlab 6.5 builds dendrogram shown in Fig. 1d. We can extract from this dendrogram a partition, for example, on two clusters. On the highest level of dendrogram we will obtain clusters $\{1,2,7\}$ and $\{3,4,5,6\}$ which corresponds to partition of the set of objects on subsets $\{a,b,g\}$ and $\{c,d,e,f\}$. The explorer, if he does not know the geometrical structure of data, could be very happy to obtain such clear partition of objects on two clusters. But if we change numeration of objects as shown in Fig. 1c then due to a symmetry of data the input matrix will not be changed and average linkage will give the same dendrogram as in Fig. 1d but now partition $\{1,2,7\}$ and $\{3,4,5,6\}$ will correspond to partition of
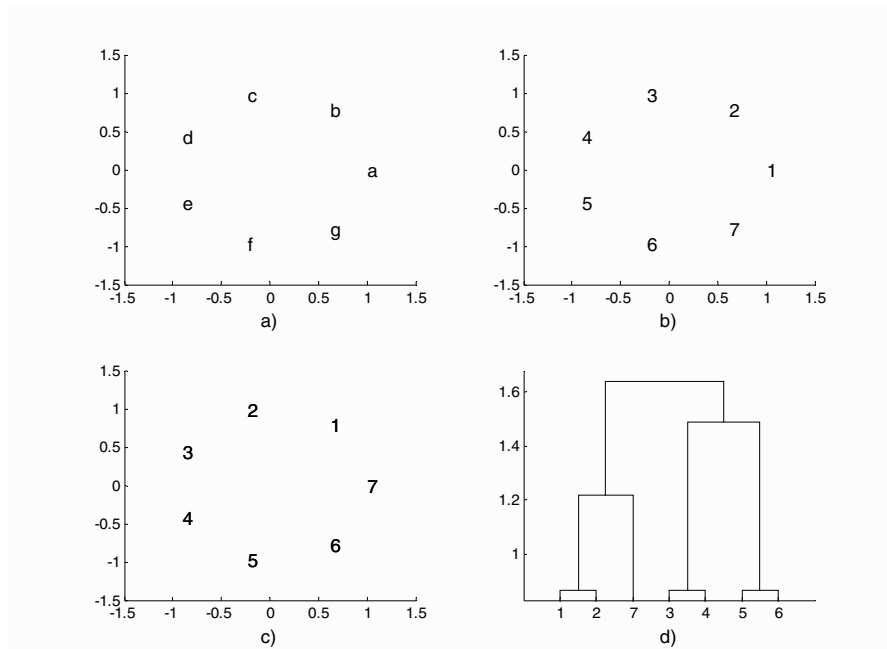
Fig. 1. Noninvariance of average linkage algorithm under initial numeration of objects: (**a**) seven objects symmetrically located on a circle; (**b–c**) two different numerations of objects; (**d**) dendrogram obtained for these numerations and resulting clustering of objects: *{{a,b},g},{{c,d},{e,f}}* and *{{b,c},a},{{d,e},{f,g}}*

objects $\{b,c,a\}$ and $\{d,e,f,g\}$. Sequentially rotating a numeration of objects we can obtain five new partitions of objects on two clusters. It is clear that none of these partitions separately reflects symmetric structure of data. Only two trivial partitions $A=\{\{a\},\{b\},\{c\},\{d\},\{e\},\{f\},\{g\}\}$ and $B=\{\{a,b,c,d,e,f,g\}\}$ correspond to symmetry in data. These partitions are constructed by single linkage clustering algorithm which is invariant under numeration of objects. Most of known clustering algorithms also can decompose this set of objects onto 2, 3, 4, 5 or 6 clusters, these decompositions can be optimal, if some optimality criteria is used in clustering algorithm, but none of these decompositions will reflect similarity structure of data.

This simple example shows that explorer should be wary if he wants to use clustering algorithms for analysis of similarity structure of data. Most of the popular clustering procedures are noninvariant under numeration of objects. They can give "good" partition of data on clusters but this partition even if it optimizes some optimality criteria can be useless for analyzing a

similarity structure of data. The following sections consider parametric scheme of clustering procedures which are invariant under numeration of objects and showed good results on testing and experimental data.

3 Basic Definitions

Denote *X* a finite set of objects and *R* a set of non-negative real values. A function *S: X × X→R* satisfying *symmetry* condition

$$S(x,y) = S(y,x)$$

will be called *a proximity relation*. A proximity relation *D* is called *a dissimilarity relation* if *D(x,x)* = 0 for all *x* from *X*. In this case *D(x,y)* usually denotes dissimilarity or distance value between objects *x* and *y*. A proximity relation *S* is called *a similarity relation* if *S* satisfies *reflexivity* condition:

$$S(x,x) = I,$$

where $I = max_{y,z}(S(y,z))$ for all *x,y,z* from *X*. Similarity relation *S* and dissimilarity relation *D* can be obtained one from another, e.g., as follows: *D(x,y) = I - S(x,y)*. Note that in [30] similarity relation denotes reflexive and symmetric valued relation satisfying on *X (∨,∧) – transitivity* condition:

$$S(x,y) \geq min\{(S(x,z),S(z,y)\}.$$

Such relation will be called here *a valued equivalence relation.* The properties of valued equivalence relations were studied in [28, 30]. The property of (∨,∧) – transitivity is dual to the *ultrametric inequality*:

$$D(x,y) \leq max\{(D(x,z),D(z,y)\}.$$

If *S* is a valued equivalence relation then *D(x,y) = I - S(x,y)* is an ultrametric and vice versa. The properties of ultrametrics were studied in many works [1, 19, 20, 22, 17, 29].

For any value (level) *a* from *R* a valued relation *S* defines an ordinary relation $S_{[a]}$ and valued relation S_a as follows:

$$S_{[a]} = \{(x,y) \in X / S(x,y) \geq a\};$$
$$S_a(x,y) = 1, \text{ if } S(x,y) \geq a \text{ and } S_a(x,y) = 0, \text{ if } S(x,y) < a.$$

Valued relation S_a may be considered as a characteristic function of ordinary relation $S_{[a]}$. From $a < b$ it follows that $S_{[b]} \subseteq S_{[a]}$ and $S_b \subseteq S_a$. From reflexivity and symmetry of S it follows that for all $a \in R$ the ordinary relations $S_{[a]}$ also will be reflexive and symmetric. If S is $(\vee, \wedge)$ – transitive then all $S_{[a]}$ will be transitive relations. As result, valued equivalence relation defines a nested set of ordinary equivalence relations and hence a nested partition of X on equivalence classes.

The subset A of X will be called *a similarity class* of similarity relation S on X if

$$S(x,y) > S(x,z) \text{ for all } x,y \in A \text{ and all } z \notin A.$$

A similarity class A may be considered as a natural cluster in the set X. A value $s = min_{x,y \in A}\{S(x,y)\}$ will be called a strength of similarity class A.

Proposition 1. *A set of similarity classes of a valued equivalence relation S coincides with a set of equivalence classes of relations $S_{[a]}$, $a \in R$.*

The set $S(X)$ of all similarity relations defined on X is a partially ordered set with ordering relation $\subseteq$ given as follows:

$$S \subseteq T \quad \text{iff} \quad S(x,y) \leq T(x,y) \text{ for all } x,y \text{ from } X.$$

We will write $S \subset T$ if and only if $S \subseteq T$ and $S \neq T$. $S(X)$ is a distributive lattice [14] with operations $\cap$ and $\cup$ defined on $S(X)$ as follows:

$$(S \cap T)(x,y) = min(S(x,y), T(x,y)),$$
$$(S \cup T)(x,y) = max(S(x,y), T(x,y)).$$

Note that the intersection of valued equivalence relations will give valued equivalence relation but for union operation the similar property generally does not hold.

$(\vee, \wedge)$-*composition* $S \circ T$ of valued relations S and T on X is defined as follows:

$$(S \circ T)(x,y) = \vee_{z \in X}(S(x,z) \wedge T(z,y)).$$

$(\vee, \wedge)$ – transitivity of S can be written as $S \supseteq S \circ S$. $(\vee, \wedge)$-*transitive closure* $\hat{S}$ of S is defined as follows: $\hat{S} = \bigcup\limits_{k=1}^{\infty} S^k$, where $S^k = S^{k-1} \circ S$, for all $k > 1$, and $S^1 = S$. From reflexivity of S and from $/X/=n$ it follows $S \subseteq S^2 \subseteq S^3 \subseteq ... \subseteq S^{n-1} = S^n = ...$ and hence $\hat{S} = S^{n-1}$. A transitive closure $\hat{S}$ of S will be denoted also as $TC(S)$.

For transitive closure of similarity relation S the following properties are fulfilled:

(1) $\hat{S}$ is a valued equivalence relation, i.e., $\hat{S}$ is transitive;

(2) S is transitive if and only if $S = \hat{S}$;

(3)   if $S \subseteq T$ then $\hat{S} \subseteq \hat{T}$;

(4) $S \subseteq \hat{S}$ and $\hat{S}$ is a least transitive-valued relation containing S, i.e., if $S \subseteq T$ and T is transitive then $\hat{S} \subseteq T$.

4 General Scheme of Hierarchical Clustering Procedures

A hierarchical clustering procedure can be considered as a transformation of a given similarity relation S into a valued equivalence relation E which defines a nested partition of X on equivalence classes. In terms of ultra-metric a clustering procedure can be considered as a transformation of a dissimilarity relation into an ultrametric [17, 20, 22]. In terms of valued equivalence relations, there exists a natural relationship between the concepts of equivalence relation, partition and clustering. This approach was used in [28], where the transitive closure of the given similarity relation was used as such a transformation. The method proposed in [28] is equivalent to the single linkage clustering algorithm [16]. We will use here a more general approach, where transitive closure is applied to a corrected similarity relation. We will consider the following *general scheme of clustering procedures* [3, 4]:

$$E = Q(S) = TC(F(S)) = \hat{F(S)}, \tag{1}$$

where F is some "correction" of given similarity relation S and TC is a procedure of transitive closure of valued similarity relations. The

procedure of transitive closure is studied in the theory of fuzzy relations, in graph theory and in cluster analysis and may be realized by single linkage clustering method [16] or by special algorithms [25, 27]. This procedure possesses both types of invariance discussed above. When a correction procedure *F* also exhibits both types of invariance, then the clustering procedure *Q* will also satisfy both invariance properties. A clustering procedure consisting of these two procedures *F* and *TC* will be called *a relational clustering procedure*.

In [3, 4] it was required also that a reasonable correction procedure *F* should satisfy the following constraint:

$$F(S) \subseteq S, \tag{2}$$

where $\subseteq$ is a partial ordering of valued relations. This constraint follows from the following formal considerations.

It is desirable to use a correction procedure *F* such that the distance between the initial similarity relation *S* and a final equivalence relation *E* will be small. The small transformation of initial similarity relation produced by clustering algorithm gives reasons to suppose that clusters corresponding to final valued equivalence relation reflect intrinsic structure of data. Of course, for some reasonable clustering algorithm that extracts clusters with specific form this distance may be sufficiently large. Nevertheless, the small distance between initial and final valued relations may be considered as a desirable property for any clustering algorithm. Formally this requirement can be formulated as follows:

$$\text{Find } E^* \in E(X): d(S,E^*) = min_{E \in E(X)} d(S,E), \tag{3}$$

where *S* is a given similarity relation on *X*, *E(X)* is a set of all possible valued equivalence relations defined on *X* and *d* is some distance measure defined on the set *S(X)* of all similarity relations on *X*. The problem (3) is studied in more general form in [3, 5] as a problem of approximation in a partially ordered set with closure operation.

A function *d: S(X) × S(X)→R* is called *a positive distance function* on *S(X)* if it satisfies on *S(X)* the following properties:

A1. $d(S,S) = 0$.
A2. $d(S,T) = d(S \cap T, S \cup T)$.
A3a. If $P \subseteq S \subset T$ then $d(P,S) < d(P,T)$,
A3b. If $P \subset S \subseteq T$ then $d(S,T) < d(P,T)$.

It is easy to see that d satisfies also the properties:

$$d(S,T) = d(T,S), \text{ and}$$
$$d(S,T) > 0 \text{ if and only if } S \neq T.$$

A function d will be called *an isotonic distance function* on $S(X)$ if it satisfies the properties *A1, A2* and the property

A3*. If $P \subseteq S \subseteq T$ then $max(d(P,S),d(S,T)) \leq d(P,T)$.

As an example of a positive distance function on $S(X)$ we can use any metric d defined as: $d(S,T) = v(S \cup T) - v(S \cap T)$, where v is a positive valuation on $S(X)$ [14], i.e., v is a real-valued function $v:S(X) \rightarrow R$ satisfying the properties:

$$v(S \cup T) + v(S \cap T) = v(S) + v(T),$$
$$\text{if } S \subset T \text{ then } v(S) < v(T).$$

For example, the function $v(S) = \sum_x \sum_y S(x,y)$ will be a positive valuation on the set of all similarity relations and hence the metric $d(S,T) = \sum_x \sum_y |S(x,y) - T(x,y)|$ defined by this valuation will be a positive distance function on $S(X)$. Most of the known metrics are positive distance functions but the metric $d(S,T) = max_{x,y} |S(x,y) - T(x,y)|$ will be only isotonic one.

Theorem 2. *If d is a positive distance function on $S(X)$ then the solution of (3) has representation*

$$E^* = TC(S_c), \tag{4}$$

where S_c is some element of $S(X)$ such that

$$S_c \subseteq S. \tag{5}$$

Theorem 2 gives reasons for the constraint (2) on the correction procedure F in general scheme of clustering procedures (1).

Several parametric correction procedures F satisfying condition (2) were proposed in [3, 4, 7] such that the resulting clustering procedures showed good results on many real and testing data [4, 7, 9].

From $F(S) \subseteq S$ it follows that correction procedure should decrease some similarity values $S(x,y)$. To be invariant under numeration of objects a correction procedure should be applied to all pairs of objects (x,y) simultaneously and independently on their numeration. To be invariant

under monotone transformations of similarity values a correction proce-dure should take into account only mutual linear ordering of similarity values $S(x,y)$. Of course, when the last condition of invariance does not required from clustering algorithm then correction procedure can use some quantitative measures depending on similarity values $S(x,y)$.

Below is a description of parameterized correction procedure given in more general form in [11] than it was proposed initially in [3].

Suppose $f_1, f_2, f_3 : R \to R$ are monotone functions. A correction procedure depends on the following sets and functions:

$$V_y(x) = \{z \in X \backslash \{x,y\} \mid S(x,z) \geq f_1(S(x,y))\},$$
$$V_x(y) = \{z \in X \backslash \{x,y\} \mid S(y,z) \geq f_1(S(x,y))\}.$$

The sets $V_y(x)$ and $V_x(y)$ denote the sets of objects "similar" to x and to y, respectively, when the value $f_1(S(x,y))$ serves as a criterion of this similarity. The set

$$V(x,y) = \{z \in X \backslash \{x,y\} \mid max\{S(x,z),S(y,z)\} \geq f_2(S(x,y))\},$$

contains the objects from X which are "similar" at least to one of the objects x and y. When $f_1 \equiv f_2$ we have $V(x,y)=V_y(x) \cup V_x(y)$. This set will be considered as the set of "neighbors" of x and y. The objects in $V(x,y)$ will be taken into account when decision about correction of the value $S(x,y)$ will be made. The set

$$W(x,y) = \{z \in V(x,y) \mid min\{S(x,z),S(y,z)\} \geq f_3(S(x,y))\},$$

denotes the set of "strong" or "common" neighbors, i.e., objects which are "similar" to both objects x and y. The objects from $W(x,y)$ will "support" the value $S(x,y)$. When $f_1 \equiv f_3$ we have $W(x,y)= V_y(x) \cap V_x(y)$. The functions f_1, f_2, f_3 used in clustering procedure will be called *neighborhood functions*.

The decision about correction of the value $S(x,y)$ will depend on the relative part of objects "supporting" the similarity value $S(x,y)$. One can consider the following methods to calculate for each pair of objects x and y this relative part denoted as h_i:

$$h_1 = \frac{|W(x,y)|}{\min(|V_y(x)|,|V_x(y)|)} , \quad h_2 = \frac{|W(x,y)|}{\max(|V_y(x)|,|V_x(y)|)} , \quad h_3 = \frac{|W(x,y)|}{|V(x,y)|} ,$$

$$h_4 = \frac{|X \backslash V| + |W| - 2}{|X| - 2} , \text{ etc.}$$

where, by definition, $h_i = 1$ if denominator of h_i is equal to 0.

The correction procedure $F(S)$ in the clustering procedure Q may be defined as follows:

$$F(S(x,y)) = \begin{cases} S(x,y) & if \ h_i \geq p \\ F_j(x,y) & otherwise \end{cases},$$

where $p \in [0,1]$, j are parameters and $F_j(x,y)$ is a corrected value such that $F_j(x,y) \leq S(x,y)$. We will suppose that $F_j(x,y)$ depends on the values $S(x,z)$, $S(y,z)$ for all objects z belonging to the sets of neighbors $V_y(x)$, $V_x(y)$ and $V(x,y)$. We require also that $F_j(x,y) \geq min_{z \in V} \{S(x,z), S(y,z)\}$, where $V = V_y(x) \cup \mathcal{N}_x(y) \cup \mathcal{N}(x,y)$. The specific definition of $F_j(x,y)$ will be discussed later. When $p = 0$, from $h_j \geq 0$ it follows that $F(S(x,y)) = S(x,y)$, i.e., for all x,y from X the values $S(x,y)$ will be uncorrected, and $Q(S) = TC(F(S)) = TC(S)$, i.e., clustering procedure will coincide with single linkage method and method considered in [16, 28].

Instead of relative part of supporting neighbors h_i it is possible to consider the number of supporting neighbors which can be calculated as follows: $g_1 = |W(x,y)|$ or $g_2 = |W(x,y)| + |X \backslash V|$ -2. In this case the correction procedure can be defined as follows:

$$F(S(x,y)) = \begin{cases} S(x,y) & if \ g_i \geq t \\ F_j(x,y) & otherwise \end{cases},$$

with parameter $t \in \{0,1,\ldots, n-2\}$, $n = |X|$.

To be invariant under numeration of objects a correction procedure should contain the same parameters for all pairs of objects or these parameters should be independent from the numeration of the objects.

5 Clustering Procedures with Identity Neighborhood Functions

Clustering procedures with identity functions f_1–f_3 were considered in the first clustering scheme [3, 4], where the relative part h_3 and correction procedure $F_1(x,y) = min_{z \in V} min\{S(x,z), S(y,z)\}$ were considered. Generally it can be considered some aggregation of values $S(x,z)$, $S(y,z)$, $(z \in V)$, that less than $S(x,y)$. As such aggregation function it can be used correction

procedure $F_j(x,y)$ equal to the *mean* or *max* [7]. These procedures were introduced heuristically and yielded good results on different experimental and testing data [4, 7, 9]. The ordinal versions of correction procedures F_j were considered in [11].

The correction procedure formalizes the following idea. We can say that two objects *x* and *y* will be considered as *identical in S*

if $S(x,y) = I$ and $S(x,z) = S(y,z)$ for all objects *z* from $X\backslash\{x,y\}$.

More generally, two objects *x* and *y* will be called *indistinguishable on the level $a \in R$*

if $S(x,y) \geq a$ and for any $z \in X$ it is fulfilled $S(x,z) \geq a$ if $S(y,z) \geq a$.

It is clear that two objects indistinguishable on some level *a* will be identical in similarity relation S_a. It is clear also that all objects are indistinguishable on the minimal possible level 0. Two objects *x* and *y* will be called *indistinguishable* in *S* if they are indistinguishable on the level $a = S(x,y)$.

Proposition 3. *A similarity relation S defined on X is a valued equivalence relation if and only if all objects of X are indistinguishable in S.*

From the properties of transitive closure procedure it follows that *TC* transforms any similarity relation *S* into valued equivalence relation *E* such that $S \subseteq E$ and *E* is the minimal valued equivalence relation including *S*. Hence transitive closure procedure produces minimal increase of values $S(x,y)$ when transforms *S* into valued equivalence relation *E*. From Proposition 3 we can conclude that this procedure transforms nonindistinguishable pairs of objects into indistinguishable. Hence we can suppose that the total value of transformation of *S* into *E* produced by *TC* depends on the number of nonindistinguishable pairs of elements in *S* and on the "degree of indistinguishability" of these elements, if we can measure it. Hence, the correction procedure *F* decreasing similarity values $S(x,y)$ should produce such minimal corrections of these values which will increase the number of indistinguishable pairs of objects or increase the "degree of indistinguishability" of pairs of objects. In this case the transformation $TC(F(S))$ produced by transitive closure will be small.

For construction of suitable correction procedure it is desirable to decide: for what pairs of objects *(x,y)* the similarity values $S(x,y)$ should be

corrected and how these values should be decreased. For these purposes the following evaluation of indistinguishability may be used. We will say that two objects *x* and *y* are *indistinguishable with respect to object z*

$$\text{if from } S(x,z) \geq S(x,y) \text{ it follows } S(y,z) \geq S(x,y).$$

In this case we will say that object *z* "supports" similarity value *S(x,y)*. The more the objects in *X* support similarity value *S(x,y)*, the more the degree of indistinguishability of *x* and *y* will be. Our goal is to change the value of *S(x,y)* such that the number of objects supporting similarity between *x* and *y* and hence the degree of indistinguishability of these objects will increase. We can say, that if the objects *x* and *y* are indistinguishable only with respect to small part of objects and hence they show different behavior on large part of objects then the similarity value *S(x,y)* does not confirmed or supported by objects of the set *X* and, as result, the similarity value *S(x,y)* can be corrected (decreased).

This idea of correcting procedure illustrated by Fig. 2 where the nodes of graph denote the objects of a set *X* and the presence of an edge between two nodes denotes that these objects are "similar" with degree *I*. Reflexive edges in nodes are omitted. For simplicity we consider ordinary relation when all weights of edges equal to 0 or *I*.

Objects *x, z* and *v, w* are identical in relation presented in Fig. 2 by graph. We have $V_u(y) = \{x,z\}$, $V_y(u) = \{v,w\}$, $W(y,u) = \varnothing$, i.e., the similarity between objects *y* and *u* does not supported by neighboring objects hence the respective edge can be deleted. The graph of valued equivalence relation nearest to *S* can be obtained by deleting the edge *(y,u)* and then by transitive closure of resulting graph, i.e., by adding the edge *(u,t)*. The resulting equivalence relation presented by graph in Fig. 2b will contain two equivalence classes *{x,y,z}* and *{u,v,w,t}*. Since the correction procedure considered in Sect. 4 depends on parameter *p* (or *t*), for some values of this parameter the correction procedure can delete in initial similarity relation also edges *(u,v)* and *(u,w)* (Fig. 2c) or even delete all edges except the edges *(x,z)* and *(y,w)* between identical objects (Fig. 2d). If correction will be not applied then all objects will be joined in one cluster. Analysis of all possible similarity structures generated by clustering procedure for different parameter values will give the following nontrivial clusters: *{x,y,z}*, *{u,v,w,t}*, *{v,w,t}*, *{v,w}*, *{x,z}*. All of these clusters describe similarity structures existing in data but different arguments in favor of these clusters can be used.
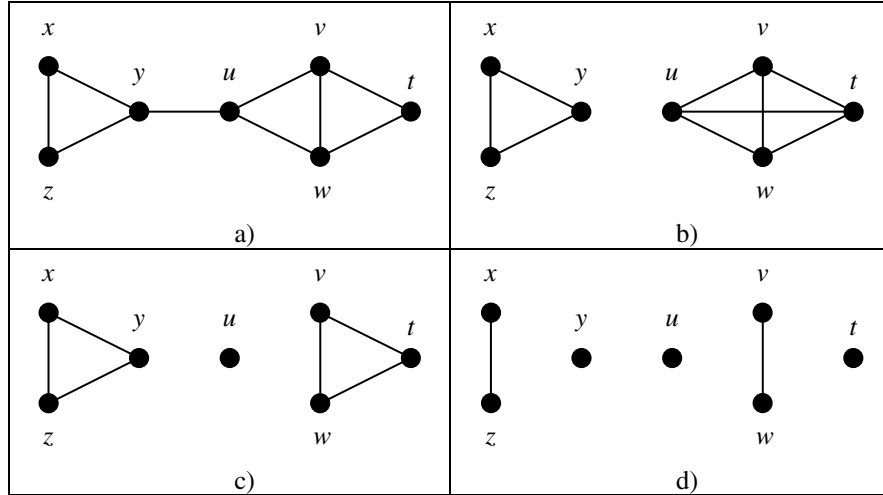
Fig. 2. (**a**) Graph of initial similarity relation *S*; (**b–d**) Graphs of possible equivalence relations *E* obtained by clustering procedure from *S*. A selection of edges for correction depends on the value of parameter *p* and hence on requirement on "similarity" of objects joined in one cluster

For valued relations and graphs the situation is more complicated because the correction procedure instead of deleting edges will decrease the weights of edges. Some methods of analysis of similarity structures generated by parametric scheme are discussed in Sect. 6.

One of the desirable properties of a clustering procedure is "to keep equivalence relations" and "to keep similarity classes." If such clusters exist in the initially given similarity relation *S* then these clusters should also exist in the clustering obtained by the clustering procedure. It can be proved that a clustering procedure *Q* with identity neighborhood functions f_1–f_3 "keeps equivalence relations" and "keeps similarity classes."

Proposition 4. *For clustering procedures Q with identity functions f_1–f_3 it is fulfilled Q(S) = S if and only if S is a valued equivalence function.*

Proposition 5. *Clustering procedures from proposed scheme "keeps similarity classes" if the neighborhood functions f_1 and f_2 used in this procedures are identity functions.*

Let $L_V(x,y)$ denote the list of all values *S(x,z), S(y,z), (z∈V),* which are less than *S(x,y),* ordered in descending order. Denote the number of elements in $L_V(x,y)$ as $m = |L_V(x,y)|$ and the elements of $L_V(x,y)$ as l_k (*k*=1, *m*). If *m* > 1

then $l_k \geq l_{k+1}$ for all $k=1$, m-1. Consider the ordinal generalizations of correction procedures $F_j(x,y)$ proposed in [11]. When $m > 1$, possible corrections will be defined by parameter j:

$j=1$: $F_j(x,y) = l_m$, i.e., l_m is the minimal value of $L_v(x,y)$;

$j=2$: $F_j(x,y) = l_1$, i.e., l_1 is the maximal value of $L_v(x,y)$;

$j=3$: $F_j(x,y) = (\sum l_k)/m$, i.e., $F_j(x,y)$ is the mean of all values from $L_v(x,y)$;

$j=4$: $F_j(x,y) = l_k$, where $k \in \{1, …, m\}$ – parameter, F_2 is a special case of F_4;

$j=5$: $F_j(x,y) = median(L_v(x,y))$.

All correction procedures $F_j(x,y)$ for $j=1,…,5$ are invariant under numeration of objects and correction procedures $F_j(x,y)$ for $j=1,2,4,5$ are invariant under monotone transformations of similarity values.

6 Selection of Valuable Clusters

Considered clustering scheme for given values of parameters defines some hierarchical clustering procedure. Generally a hierarchy constructed by clustering procedure is considered as a sought similarity structure of data. If the goal of analysis is a search of a partition of data then in more traditional approach it is selected some level of hierarchy and clusters on this level define a partition of data. In naive approaches the number of clusters in partition a priori is fixed and the level of dendrogram is selected so that the corresponding partition contains desired number of clusters. The level-based approach to selection of clusters has a following weakness. Frequently, natural clusters existing in data are generated on different levels of hierarchy. For this reason on high levels of hierarchy the small natural clusters can disappear as result of union in large conglomerates. Correspondingly on small levels the large natural clusters can be separated on small non-natural fragments.

Another approaches extract from hierarchy "valuable" clusters, e.g., clusters existing on large number of levels or clusters constructed on high level of similarity (on small level of dissimilarity).

We use "structural" approach to selection of valuable clusters from hierarchy proposed in [4]. Suppose on some level of hierarchy two clusters A and B are joined together in cluster $C = A \cup B$. Then the importance m of these clusters is calculated as follows:

$$m(A) = m(B) = min(N_A, N_B),$$

where N_A and N_B are the numbers of objects in clusters A and B, respectively. We say that a cluster A is a "valuable" cluster if $m(A) \geq M$, where M is a given number greater than 1. The level M can be selected adaptively depending on the number of valuable clusters extracted for different values M. The reasons to consider such measure of importance of clusters are the following. Suppose $N_B < M$. It means that the set A is joined with "nonvaluable" amount of objects and hence A is still "in the process of formation of cluster." For this reason the cluster A, even if he has a large amount of objects, receives a small importance value. But if N_A, $N_B \geq M$, then we say that both A and B are "valuable" clusters. If more than two clusters are joined on some level then the importance value of all such clusters is determined by two of them which have maximal number of objects.

As usual, real data does not contain clear clusters. As in a desert, where sand-dunes have different forms and mutual locations and some small dunes can be considered as parts of large dunes or as separate dunes depending on the "definition" of the concept "dune," in considered parametric clustering scheme a change of parameters of clustering procedure will change a concept of "similarity" or "indistinguishability" and, hence, will cause the construction of slightly different hierarchical similarity structures. Analysis of hierarchical structures obtained by clustering procedure for different values of parameters can be used for extraction of all possible valuable clusters in data or for selection "stable" clusters presented in most of hierarchies. For example presented in Fig. 2 we can say that similarity structure of data contains clusters *{x,y,z}, {u,v,w,t}, {v,w,t}, {x,z}* and *{v,w}*.

Another approach to selection of "best clustering" of data uses distance measure between similarity relations. Change of parameters in clustering procedure can be used for selection of hierarchical structure corresponding to valued equivalence relation E on the output of clustering procedure with minimal distance from initial similarity relation S. For example presented in Fig. 2 such minimal distance has partition in Fig. 2b. Note that for crisp equivalence relations hierarchy of partitions is reduced to one partition.

7 Example of Windham's Data

Clustering procedures with identity neighborhood functions f_1–f_3 will be illustrated here on Windham's data [13] shown in Fig. 3. Dissimilarity values (squares of distances) for these data are given in Appendix 2. A single linkage method will give a trivial clustering of these objects, i.e., all objects will be joined in one cluster or there will be 11 clusters containing one object. It is clear that similarity classes are absent in the given data but two clusters $\{1,2,3,4,5\}$ and $\{7,8,9,10,11\}$ can be considered as "natural." These clusters are constructed by relational clustering procedures based on procedures F_j with parameters h_3 and $j = 2$, $j = 4$ ($k = 2$), $j = 5$ for all values of parameter $p \in \{0.1, 0.2, \ldots, 1\}$, and for parameter $j = 3$ for almost all values of parameter p on higher levels of dendrogram.

In addition to "natural clusters," relational clustering procedures with parameter values $j = 1$, 3, 4 ($k > 2$) construct dendrograms with the following nontrivial clusters and partitions: $\{1,2,4\}$, $\{8,10,11\}$, $\{5,6,7\}$; $\{\{\{\{1,2,4\}, 3\},5\}, \{\{\{8,10,11\},9\},7\}\}$; $\{\{1,2,3,4\},\{5,6,7\}, \{8,9,10,11\}\}$. Invariance under numeration of objects of the constructed clusters can be easily seen from symmetry of data. These clusters describe the symmetric structure of the considered set of objects. Note that most of known clustering algorithms cannot extract such symmetric structure of data.
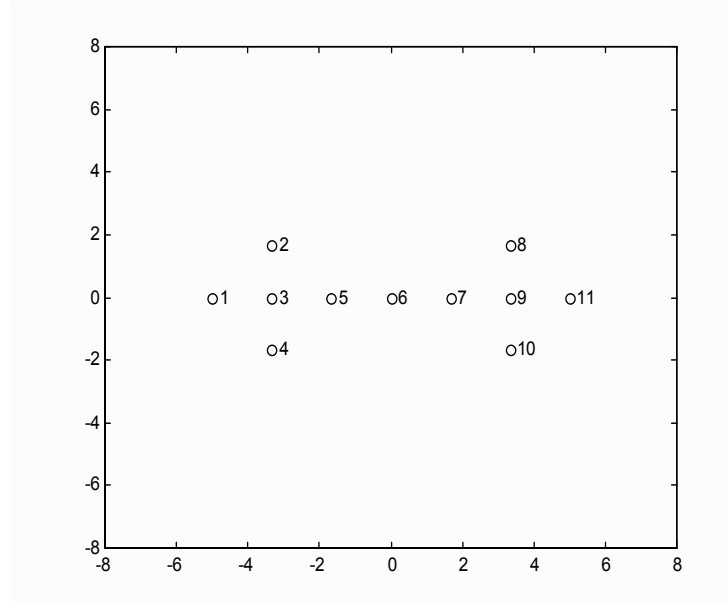

Fig. 3. Windham's data [13]

The distance *d(D,U)* between the initial dissimilarity function *D* corresponding to Windham's data and the ultrametrics *U* corresponding to dendrograms constructed by clustering algorithms was minimal for the ultrametric *U* corresponding to clustering into "natural clusters." As distances *d* we have used distances $d_s(S,T) = (\sum_x \sum_y |S(x,y)-T(x,y)|^s)^{1/s}$, with $s = 1$ and $s = 2$.

8 Clustering Procedures with Nonidentity Neighborhood Functions

Nonidentity neighborhood functions f_1–f_3 in the considered scheme of clustering procedures may be used for the construction of clusters based on a perception "break bridges between clusters." This approach, from a certain point of view, is opposite to the approach "keep similarity classes" because some similarity clusters considered as "bridges" between natural clusters can be break down. Figure 4 shows an example of "butterfly" data, which may be considered as junction of two clusters forming the "wings" of a butterfly. Two central points, forming "similarity class," should belong to different clusters corresponding to the "wings." Clustering procedure
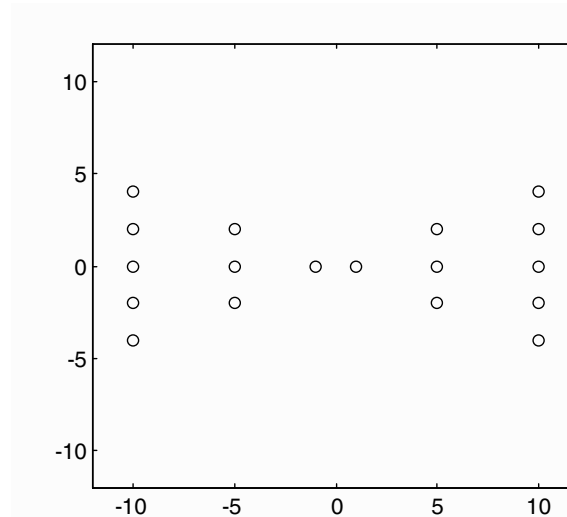

Fig. 4. "Butterfly" data with 18 points

should be able "to break" this "bridge" between clusters. Note that the distances between points increase when moving off origin of coordinates along x-axis. Coordinates of points are given in Appendix 3.

Relational clustering procedures with nonidentity neighborhood functions give possibility to classifying the "butterfly" data into two clusters. Euclidean distances between objects were considered as ordinal dissimilarity values and a clustering procedure with ordinal nonidentity functions f_2 and f_3 were used. These functions were defined by $f_2(D(x,y))=l_r$, $f_3(D(x,y))=l_q$, where l_r and l_q are the "rth" and "qth" dissimilarity values chosen from the ordered list of dissimilarity values $D(y,z)$, $D(x,z)$ that are greater than $D(x,y,)$ where $z \in X \backslash \{x,y\}$ for f_2 and $z \in V(x,y)$ for f_3. Similarity values can be obtained from dissimilarity values by $S(x,y) = max_{u,v}D(u,v) - D(x,y)$. The clustering into two wings on the higher levels of the dendrograms was obtained, for example, for the following values of parameters ($j = 1$ and h_3 were used):

(1) ($r = 1$, $q = 1,2$), ($r = 2$, $q = 1,2$), ($r = 3$, $q = 1,...,4$), $p = 0.1,...,0.5$ or $p = 0.1,...,0.7$;
(2) ($r = 3$, $q = 5,6$), $p = 0.1,...,1$.

As one can see for case (2) two wings were constructed for all values of parameter p (chosen with step 0.1), greater than 0.

9 Example of Clustering of Time Series

Consider application of relational clustering procedures to clustering time series of economic data. We use data from [31] which contain time series of World Per Capita Gross Domestic Product using Market Exchange Rates, 1980-2002. Time series used in example are presented in Fig. 5.

As dissimilarity measure between time series we used measure of local trend associations [12] calculated as $d(y,x) = 0.5(1-AM(y,x))$, where $AM(y,x) = max(AF_K(y,x))$, $K = \{2,3\}$. The values of this dissimilarity measuree used in clustering procedure are given in Appendix 4. We scanned similarity structures of these data by parametric clustering procedures with values of parameters: f_1, f_2, $f_3 = 1$, 1.5; $j = 1,2,3$; $p = 0, 0.1,...,1$. The minimal distance between initial similarity relation and final valued equivalence relation was obtained for the following values of parameters: $f_1, f_2, f_3 = 1.5$; $j = 1$; $p = 1$. The respective dendrogram is shown in Fig. 6.
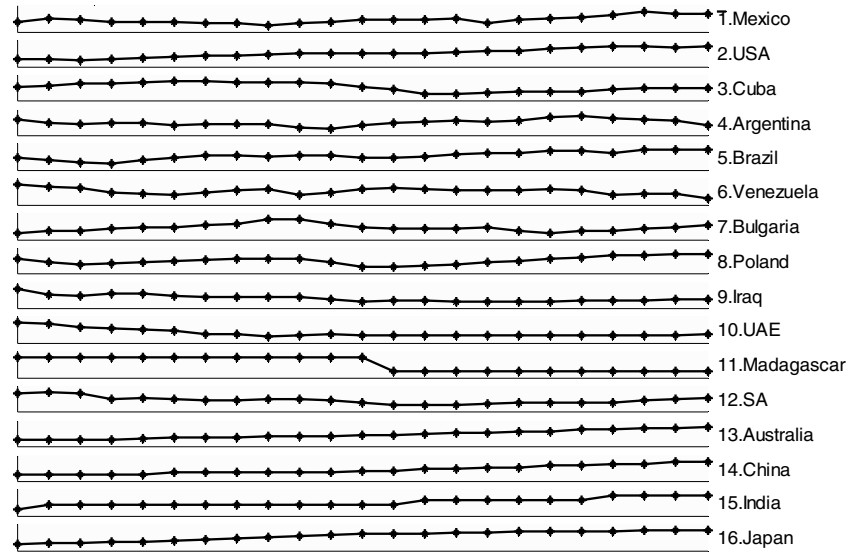
Fig. 5. Time series of world per capita gross domestic product using market exchange rates, 1980–2002

6.Venezuela
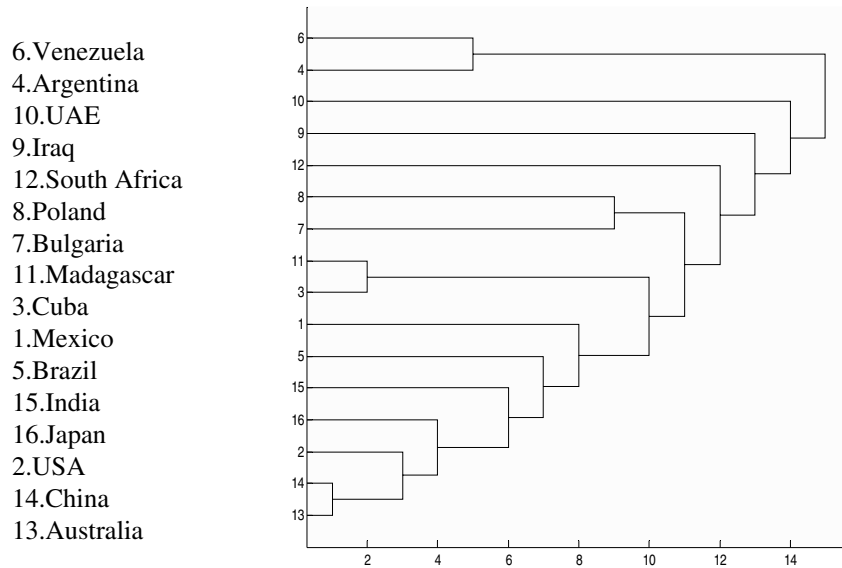4.Argentina
10.UAE
9.Iraq
12.South Africa
8.Poland
7.Bulgaria
11.Madagascar
3.Cuba
1.Mexico
5.Brazil
15.India
16.Japan
2.USA
14.China
13.Australia

Fig. 6. Clustering of time series for parameters: $f_1, f_2, f_3 = 1.5$; $j = 1$; $p = 1$

9.Iraq
6.Venezuela
4.Argentina
12.South Africa
10.UAE
7.Bulgaria
11.Madagascar
3.Cuba
1.Mexico
15.India
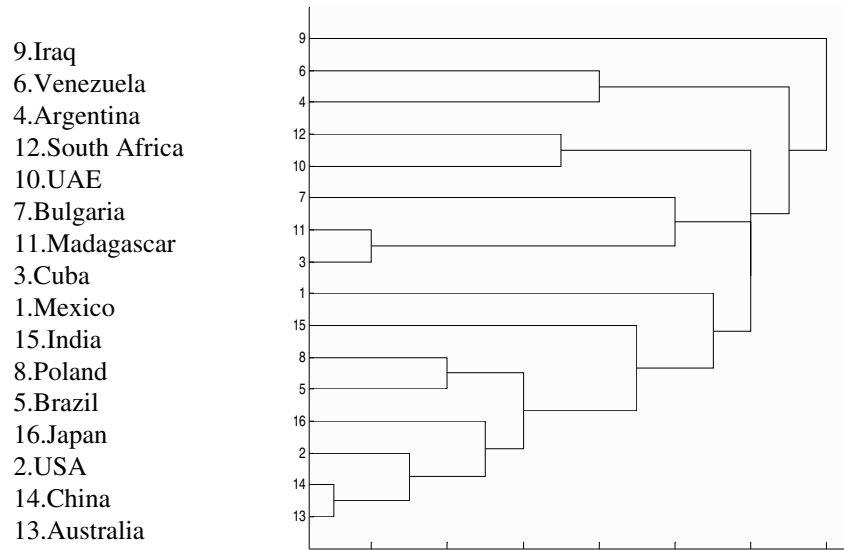8.Poland
5.Brazil
16.Japan
2.USA
14.China
13.Australia

Fig. 7. Clustering of time series for parameters: $f_1, f_2, f_3 = 1; j = 2; p = 0.2,...,1$

If we consider clusters containing at least two objects (see Sect. 6) as "valuable" then we will obtain the following valuable clusters: $C_1=$ {13.Australia, 14.China, 2.USA, 16.Japan, 15.India, 5.Brazil, 1.Mexico}, $C_2=$ {3.Cuba, 11.Madagascar}, $C_3=$ {7.Bulgaria, 8.Poland}, $C_4=$ {6.Venezuela, 4.Argentina}, $C_5= C_1 \cup C_2$, $C_6= X \backslash C_4$. Analysis of similarity structures generated for other values of parameters gives also, for example, the following valuable clusters: {13.Australia, 14.China, 2.USA, 16.Japan}, {3.Cuba, 11.Madagascar, 7.Bulgaria}, {12.South Africa, 10.United Arab Emirates}, {5.Brazil, 8.Poland}. Figure 7 gives another example of hierarchical clustering of data obtained by clustering procedures for parameter values $f_1, f_2, f_3 = 1; j = 2; p = 0.2,...,1$.

10 Conclusion

In this chapter we studied a general scheme of clustering procedures based on correction and transitive closure of the given similarity relation. The main properties of these procedures are invariance under numeration of objects and invariance under monotone transformations of similarity

values. Another important property of general clustering scheme is its parametric definition. Different choice of these parameters gives possibility to change a meaning of "indistinguishability" of objects and to analyze possible similarity structures of data, in an exploratory manner. Such analysis gives possibility to analyze similarity structures of data from different points of view. Several of the considered parametric clustering procedures are based on perceptions "keep similarity classes" and "break bridges between clusters." Several propositions and testing examples illustrate the properties of the clustering scheme. The proposed scheme can be extended in several directions, for example a new type of correction procedure can be proposed. In [10] it was considered such extension of clustering scheme based on perception "break bridges between clusters."

Acknowledgments

References

1. Barthelemy J.P. & Guenoche A. (1991). Trees and Similarity Representations. Chichester: Wiley
2. Batyrshin I.Z. (1980). Clustering based on fuzzy similarity relations. In: *Third Workshop "Control with Presence of Fuzzy Categories"*, Perm, Russia, pp. 25–27 (in Russian)
3. Batyrshin I.Z. (1982). *Methods of Systems Analysis Based on Valued Relations*. PhD Thesis. Moscow Power Engineering Institute (in Russian)
4. Batyrshin I.Z. & Shuster V.A. (1984). The structure of semantic spaces of verbal estimates of actions. *Acta et Commentationes Universitas Tartuensis, Transactions on Artificial Intelligence, Principle Questions of Knowledge Theory*, Tartu, 688, 20–38 (in Russian)
5. Batyrshin I.Z. (1985). About approximation task in a partially ordered set. In: *Mathematical Methods of Optimization and Control in Systems*. Kalinin: Kalinin State University, pp. 50–56
6. Batyrshin I. (1994). Errors of type 2 in cluster analysis and invariant cluster procedures based on similarity relations. In: *Application of Fuzzy Systems, ICAFS-94* (Ed. by R. Aliev and R. Kenarangui). Iran: University Press of Tabriz, pp. 374–378

7. Batyrshin I.Z. & Khabibulin R.Ph. (1995). Attribution of pseudonymous works of literature based on invariant relational clustering algorithms. In: *Computational Linguistics and its Applications*, Proceedings of the International Workshop, Kazan, pp. 43–55 (in Russian)

8. Batyrshin I. & Khabibulin R. (1998). On invariance of clustering procedures. *Journal of Fuzzy Mathematics*, 6(3), 721–733

9. Batyrshin I., Khabibulin R., Fatkullina R. (1996). Application of fuzzy relational clustering algorithms to ecological data. In: *ICAFS-96, Second International Conference on Application of Fuzzy Systems and Soft Computing* (Ed. by R.A. Aliev et al.). Siegen, Germany, pp. 115–117

10. Batyrshin I. & Klimova A. (2002). New invariant relational clustering procedures. In: *Proceedings of East West Fuzzy Colloquium 2002, 10th Zittau Fuzzy Colloquium,* Zittau, Germany, pp. 264–269

11. Batyrshin I. & Rudas T. (2000). Invariant clustering procedures based on corrections of similarities. In: *Proceedings of East West Fuzzy Colloquium*, Zittau, Germany, pp. 302–309

12. Batyrshin I., Herrera-Avelar R., Sheremetov L., Panova A. Moving approximation transform and local trend associations in time series data bases. In this book

13. Bezdek J.C. (1990). A note on two clustering algorithms for relational network data. SPIE, Vol. 1293, *Applications of Artificial Intelligence*, VIII, 268–277

14. Birkhoff G. (1967). *Lattice theory*. Providence, RI: American Mathematical Society

15. Duda R.O. & Hart P.E. (1973). *Pattern Classification and Scene Analysis*. New York: Wiley

16. Dunn J.C. (1974). A graph-theoretic analysis of pattern classification via Tamura's fuzzy relation. *IEEE Transaction on Systems, Man and Cybernetics*, SMC-4, 310–313

17. Hartigan J.A. (1967). Representation of similarity matrices by trees. *Journal of the American Statististical Association*, 62, 1140–1158

18. Hubert L.J. (1973). Monotone invariant clustering procedures. *Psychometrica,* 38(1), 47–62

19. Jambu M. (1978). *Classification automatique pour l'analyse des donnees*. Paris, France: Dunod

20. Jardine C.J., Jardine N., Sibson R. (1967). The structure and construction of taxonomic hierarchies. *Mathematical Biosci*ences, 1, 173–179

21. Jardine N. & Sibson R. (1971). Mathematical taxonomy. London: Wiley

22. Johnson S.C. (1967). Hierarchical clustering schemes. *Psychometrika*, 32(3), 241–254

23. Lance G.N. & Williams W.T. (1969). A general theory of classificatory sorting strategies. I. Hierarchical systems. *The Computer Journal*, 9(4), 373–380

24. Matula D.W. (1977). Graph theoretic techniques for cluster analysis algorithms. In: *Classification and Clustering* (Ed. by J. Van Ryzin)*.* New York: Academic, pp. 95–129
25. Naessens H., De Meyer H., De Baets B. (1999). Novel algorithms for the computation of transitive closures and openings of proximity relations. In: *Proceedings of EUROFUSE-SIC*'99, pp. 200–203
26. Sokal R.R. (1977). Clustering and classification: background and current directions. In: *Classification and Clustering* (Ed. by J. Van Ryzin)*.* New York: Academic, pp. 1–15
27. Swamy M.N.S. & Thulasiraman K. (1981). *Graphs, Networks, and Algorithms*. New York: Wiley
28. Tamura S., Higuchi S., Tanaka K. (1971). Pattern classification based on fuzzy relations. *IEEE Transaction on Systems, Man and Cybernetics*, SMC-1, 61–66
29. Young M.R. & DeSarbo W.S. (1995). A parametric procedure for ultrametric tree estimation from conditional rank order proximity data. *Psychometrica*, 60(1), 47–75
30. Zadeh L.A. (1973). Similarity relations and fuzzy orderings. *Information Sciences*, 3, 177–200
31. International Gross Domestic Product, Population, and General Conversion Factors Information. http://www.eia.doe.gov/emeu/international/other.html. Energy Information Administration. Official Energy Statistics from the U.S. Government

Appendix 1

Table 1. Distances between seven points symmetrically located on a circle (Fig. 1)

	a	b	c	d	e	f	g
a	0	0.8678	1.5637	1.9499	1.9499	1.5637	0.8678
b	0.8678	0	0.8678	1.5637	1.9499	1.9499	1.5637
c	1.5637	0.8678	0	0.8678	1.5637	1.9499	1.9499
d	1.9499	1.5637	0.8678	0	0.8678	1.5637	1.9499
e	1.9499	1.9499	1.5637	0.8678	0	0.8678	1.5637
f	1.5637	1.9499	1.9499	1.5637	0.8678	0	0.8678
g	0.8678	1.5637	1.9499	1.9499	1.5637	0.8678	0

Appendix 2

Table 2. Dissimilarity values for Windham's data (Fig. 3)

n/n	1	2	3	4	5	6	7	8	9	10	11
1	0	6	3	6	11	25	44	72	69	72	100
2	6	0	3	11	6	14	28	56	47	44	72
3	3	3	0	3	3	11	25	47	44	47	69
4	6	11	3	0	6	14	28	44	47	56	72
5	11	6	3	6	0	3	11	28	25	28	44
6	25	14	11	14	3	0	3	14	11	14	25
7	44	28	25	28	11	3	0	6	3	6	11
8	72	56	47	44	28	14	6	0	3	11	6
9	69	47	44	47	25	11	3	3	0	3	3
10	72	44	47	56	28	14	6	11	3	0	6
11	100	72	69	72	44	25	11	6	3	6	0

Appendix 3

Table 3. Coordinates of "Butterfly" data (Fig. 4)

x	−10	−10	−10	−10	−10	−5	−5	−5	−1
y	4	2	0	−2	−4	2	0	−2	0
x	1	5	5	5	10	10	10	10	10
y	0	2	0	−2	4	2	0	−2	−4

Appendix 4

Table 4. Dissimilarity values between time series (Fig. 5)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	0,000	0,387	0,553	0,437	0,537	0,532	0,731	0,601	0,677	0,345	0,549	0,424	0,310	0,315	0,283	0,404
2	0,387	0,000	0,450	0,457	0,225	0,681	0,383	0,206	0,507	0,667	0,577	0,531	0,066	0,168	0,320	0,184
3	0,553	0,450	0,000	0,770	0,431	0,706	0,282	0,301	0,565	0,547	0,155	0,317	0,484	0,573	0,573	0,475
4	0,437	0,457	0,770	0,000	0,479	0,238	0,762	0,468	0,372	0,540	0,724	0,684	0,554	0,451	0,602	0,584
5	0,537	0,225	0,431	0,479	0,000	0,547	0,378	0,174	0,458	0,545	0,462	0,413	0,270	0,272	0,596	0,376
6	0,532	0,681	0,706	0,238	0,547	0,000	0,723	0,677	0,612	0,429	0,594	0,468	0,630	0,547	0,698	0,530
7	0,731	0,383	0,282	0,762	0,378	0,723	0,000	0,239	0,405	0,708	0,338	0,385	0,433	0,546	0,445	0,465
8	0,601	0,206	0,301	0,468	0,174	0,677	0,239	0,000	0,223	0,525	0,402	0,313	0,301	0,359	0,438	0,546
9	0,677	0,507	0,565	0,372	0,458	0,612	0,405	0,223	0,000	0,298	0,585	0,588	0,630	0,602	0,695	0,714
10	0,345	0,667	0,547	0,540	0,545	0,429	0,708	0,525	0,298	0,000	0,433	0,232	0,618	0,598	0,614	0,695
11	0,549	0,577	0,155	0,724	0,462	0,594	0,338	0,402	0,585	0,433	0,000	0,295	0,607	0,689	0,658	0,565
12	0,424	0,531	0,317	0,684	0,413	0,468	0,385	0,313	0,588	0,232	0,295	0,000	0,369	0,500	0,450	0,597
13	0,310	0,066	0,484	0,554	0,270	0,630	0,433	0,301	0,630	0,618	0,607	0,369	0,000	0,056	0,266	0,180
14	0,315	0,168	0,573	0,451	0,272	0,547	0,546	0,359	0,602	0,598	0,689	0,500	0,056	0,000	0,261	0,262
15	0,283	0,320	0,573	0,602	0,596	0,698	0,445	0,438	0,695	0,614	0,658	0,450	0,266	0,261	0,000	0,433
16	0,404	0,184	0,475	0,584	0,376	0,530	0,465	0,546	0,714	0,695	0,565	0,597	0,180	0,262	0,433	0,000