

## Fixed Points Approach to Clustering \*

Alexander V. Genkin

Institute for Problems of  
Information Transmission

Ilya B. Muchnik

Boston University

**Abstract.** Assume that a dissimilarity measure between elements and subsets of the set being clustered is given. We define the transformation of the set of subsets under which each subset is transformed into the set of all elements whose dissimilarity to its is not greater than a given threshold. Then the cluster is defined as fixed point of this transformation. Three well-known clustering strategies are considered from this point of view: hierarchical clustering, graph-theoretic methods, and conceptual clustering. For hierarchical clustering generalizations are obtained that allow for overlapping clusters and/or clusters not forming a cover. Three properties of dissimilarity are introduced which guarantee the existence of fixed points for each threshold. We develop the relation to the theory of quasi-concave set functions, to help give an additional interpretation of clusters.

**Keywords.** Fixed points; Hierarchical clustering; Graph-theoretic clustering; Conceptual clustering; Overlapping clusters.

### 1. Introduction

A variety of clustering methods have already been developed, and most of them rely on a specific corresponding concept of cluster. These concepts are introduced in different terms, so that it is difficult to compare them with one another. Moreover, sometimes these concepts are given implicitly, particularly in algorithmic form. We suppose that such a situation reflects the real variety in humans' concepts of similarity, compactness, etc. Nevertheless we believe that a general point of view is highly desirable. It could produce the terminology to choose the appropriate clustering method.

---

We thank Boris Mirkin, Phipps Arabie, and anonymous referees for useful suggestions, constructive criticism, fruitful discussions and helpful comments and editing.

Authors' Addresses: A.V. Genkin, Institute for Problems of Information Transmission, Yermolovoj 19, Moscow, 101447, Russia. I.B. Muchnik, Boston University, College of Engineering, Biomolecular Center, 36 Cummington St., Boston, MA 02215, USA.

In this paper we try to advance in this direction. Three well-known clustering strategies are studied. These are hierarchical clustering (Jardine and Sibson 1971; Sneath and Sokal 1973), graph-theoretic methods (Hubert 1974; Matula 1977) and conceptual clustering (Michalski and Stepp 1982, 1983). We define clusters as fixed points of an appropriate transformation of the set of subsets and apply this definition to all three clustering strategies studied.

The definition of cluster proposed here is formulated according to a dissimilarity measure between elements and subsets of the set being clustered. We refer to it as an “*element-set*” type *dissimilarity measure (ESD)*. Measures of dissimilarity between elements are central in clustering, and measures between *subsets* are sometimes used as well, as in hierarchical clustering. The type of measure under consideration here has not yet been studied systematically, but it has found a number of applications. Recall, for example, a distance between an element and the (variously defined) centroid of cluster, exploited in ISODATA (Ball and Hall 1967) and a number of its followers. It plays the role of the “*element-set*” type measure, through it has the form of distance between two points.

A number of procedures for generation of ESD from more commonly used type of data has been proposed by Kuznetsov, Muchnik, Hincey, and Tchkuasely (1984) and by Kuznetsov, Muchnik, and Shvartser (1985). ESD can be generated from a dissimilarity matrix with the help of such simple operations as minimum, maximum, summation, etc. The multiple correlation coefficient provides an example of “*element-set*” type of similarity measure on the set of variables, which cannot be calculated from matrix of “*element-element*” type (one-mode, two-way, in the terminology of Tucker, 1964). It can be easily transformed into dissimilarity measure. There are some other examples of ESD’s in the present paper.

A number of methods of data analysis based on ESD have been proposed by Mulla (1976, 1977); Kuznetsov et al. (1984); Kuznetsov, Muchnik and Shvartser (1985); Muchnik, Chkuaseli and Shvartser (1986); Zaks and Muchnik (1989), in particular with monotone ESD (see below) and with passing relevance to clustering. However, we are presently concentrating on the specific concept of a cluster.

Given a finite set  $U$ , let an ESD be defined as real-valued function  $\pi(x, A)$ , where  $x \in U$ ,  $A \subseteq U$ . A subset  $A \subseteq U$  will be called a  $t$ -stable set, or a  $t$ -cluster, where  $t$  is a real value, iff the following conditions hold:

$$\pi(x, A) \leq t \text{ for all } x \in A, \quad (1.1a)$$

$$\pi(x, A) > t \text{ for all } x \in U - A, \quad (1.1b)$$

(where  $A - B$  means  $A \cap B^c$ ). Our first condition expresses the cluster's property of being compact while the second one expresses its property of being isolated. This definition seems rather simple and natural. Consider now a transformation of the set of all subsets of  $U$  defined as follows:

$$V_t(A) = \{x \in U : \pi(x, A) \leq t\}. \quad (1.2)$$

This transformation can be interpreted as mapping a subset  $A \subseteq U$  into the set of all elements similar to  $A$  in the sense of ESD. The set  $A$  is called a fixed point of the transformation  $V_t$  iff  $V_t(A) = A$ . Note now that the set is a  $t$ -cluster iff it is a fixed point of the mapping  $V_t$ . Throughout this paper, we use the terms “ $t$ -cluster,” “ $t$ -stable set,” “fixed point” interchangeably.

Mullat (1981) was the first to propose the fixed points transformations, similar to (1.2) in the continuous case, as a formal explication for the concept of the “area of condensation” of probability density. Transformation (1.2) was studied by Genkin, Zaks, and Muchnik (1988a, 1988b). In this paper we present the relevant results of these papers and continue the developments.

The first theoretical problem of the fixed points approach is the existence of  $t$ -clusters. Moreover, the definition of a  $t$ -cluster  $A$  allows  $A$  to be either the empty set or the whole of  $U$ . Thus, the existence of non-trivial  $t$ -clusters is also of interest. We state some sufficient conditions for ESD to have a  $t$ -cluster for each  $t$  and outline the conditions to have a non-trivial  $t$ -cluster where possible.

The fixed points approach concentrates on the properties of a particular cluster, not a partition or of some other kind of clustering as a whole. But the properties of the set of all  $t$ -clusters are of considerable importance. It is usually claimed they should form a partition. Sometimes-weaker claims are practical: to form a cover of  $U$  or to be mutually disjoint. We suppose it is reasonable to consider a still weaker claim: the  $t$ -clusters should not be subsets of one another. For each of these properties, we formulate conditions for ESD, which guarantee the properties to be true.

Another important problem is how  $t$ -clusters change when  $t$  changes. The growth of  $t$ -clusters when  $t$  becomes larger seems most sensible. We define a type of ESD such that each  $t$ -cluster has a superset  $t'$ -cluster for each  $t'$ , where  $t' > t$ . When  $t$ -clusters form a cover of  $U$  for each  $t$ , it corresponds to well-known dendrogram representation of hierarchical clustering.

Our paper contains six sections and a summary. Sections 2 and 3 are devoted to some theoretical aspects of  $t$ -clusters. Section 2 considers the problem of their existence. In section 3, monotone ESD's are studied. An ESD is monotone if it increases (or decreases) when its set-valued argument becomes larger. Such ESD's, together with related quasi-concave set functions, form a subject of monotone systems theory (Mullat 1976, 1977; Kuznetsov, Muchnik, and Shvartser 1985). We study  $t$ -clusters in the framework of this theory, which enables us to obtain a useful additional interpretation of  $t$ -clusters.

In the next three sections, different cluster concepts are studied using the fixed points approach. In section 4, we study hierarchical clusters, which are usually presented in the form of dendrogram. Generalization of the dendrogram concept is proposed, permitting overlapping clusters and/or clusters not forming a cover, at each level of hierarchy. Necessary and sufficient conditions for the clusters to be disjoint as well as to form a cover are formulated.

Section 5 considers the graph-theoretical approach to clustering. Several clustering methods have been thus formulated (Hubert 1974; Matula 1977). We show that these methods may be represented in a  $t$ -cluster framework. Some of their properties are shown to be implied to  $t$ -cluster theory.

In section 6, the concept of a  $t$ -stable set is applied to conceptual clustering. There arises a problem which can be formulated as follows: given a family of subsets of a finite set generated under conceptual criteria, find a subfamily which will be as proximate as possible to being a partition. The proximity mentioned can be measured in different ways. We propose an objective function, the extremum of which can be interpreted as a  $t$ -stable set.

## 2. The Existence of $t$ -clusters

The problem of the existence of  $t$ -clusters has not been completely solved yet. We only have some sufficient conditions for ESD which guarantee the existence of  $t$ -clusters for any  $t$ . In this section, we formulate these conditions and consider their basic implications.

First we need some definitions. An ESD  $\pi$  is called *non-decreasing* (or *non-increasing*), iff  $A \subset B \subseteq U$  implies:

$$\pi(x, A) \leq \pi(x, B) \text{ (or } \pi(x, A) \geq \pi(x, B) \text{) for all } x \in U .$$

ESD is called *point-non-decreasing* (or *point-non-increasing*), iff the following holds:

$$\begin{aligned} \pi(x, A+x) \geq \pi(x, A-x) \text{ (or } \pi(x, A+x) \leq \pi(x, A-x) \text{)} \\ \text{for all } x \in U, A \subseteq U . \end{aligned}$$

(For brevity we write  $A+x$  and  $A-x$  instead of  $A \cup \{x\}$  and  $A \setminus \{x\}$  respectively).

An ESD which is both point-non-decreasing and point-non-increasing is called *point-independent*. Such ESD's are usually used to extend the functions, which are determined only for  $x \in A$  or only for  $x \in U \setminus A$ .

Now we are prepared to consider the following three existence conditions: non-increasing, intergability, ultrametricity.

### A. Non-increasing condition

Note that for non-increasing ESD, the transformation in (1.2) is isotone, i.e., if  $A \subset B \subseteq U$ , then  $V_t(A) \subseteq V_t(B)$  for each  $t$ . A partially ordered set is called a complete lattice if for each subset it possesses a greatest lower bound and a least upper bound (Birkhoff 1967, Ch.1, §4).

**Proposition 2.1.** *For non-increasing ESD a  $t$ -cluster exists for each  $t$ .*

*Proof.* Immediately follows from the Tarski theorem (see Birkhoff 1967, Ch.5, §3): an isotone transformation of a complete lattice has a fixed point.

**Proposition 2.2.** *(Genkin, Zaks, and Muchnik 1988b). The set of  $t$ -clusters of a non-increasing ESD for a given  $t$  forms a lattice.*

*Proof.* The fixed point terminology is used here for convenience. If  $A_1, A_2$  are fixed points of an isotone  $V$  and  $A = A_1 \cup A_2$ , then we have:  $V(A) \supseteq V(A_i) = A_i$ ,  $i = 1, 2$ . Thus  $V(A) \supseteq A$ ,  $V(V(A)) \supseteq V(A)$ , etc. As  $U$  is finite, this sequence converges to some fixed point  $B$ , and  $B \supseteq A$ . We shall prove, that  $B = \sup(A_1, A_2)$ . Indeed, let  $C$  be a fixed point and  $C \supseteq A$ . Then  $V(A) \subseteq V(C) = C$ ,  $V(V(A)) \subseteq C$ , etc., and at some iteration we obtain  $B \subseteq C$ . The existence of infimum is proved in the same way. ■

**Proposition 2.3.** *Let  $A$  be a  $t$ -cluster of a non-increasing ESD, and  $t' < t < t''$ . Then there exists a  $t'$ -cluster  $A'$  and a  $t''$ -cluster  $A''$  with the property:  $A' \subset A \subset A''$ .*

*Proof.* For each  $x \in A$  we have:  $\pi(x, A) \leq t < t''$ . If  $V$  is a transformation defined by (1.2) with  $t''$ , then  $V(A) \supseteq A$ . Consider a sequence  $A_0 = A, A_1, A_2, \dots$ , where  $A_{i+1} = V(A_i)$ ,  $i \geq 0$ . As  $V$  is isotone, we have  $A_{i+1} \supseteq A_i$  for all  $i \geq 0$ . As  $U$  is finite, the sequence converges to the fixed point  $A'' \supseteq A$ . By analogy, consider the decreasing sequence of sets to prove the existence of  $A'$ . ■

The empty set is a  $t$ -cluster for all  $t$  such that  $t < t' = \min_{x \in U} \pi(x, \emptyset)$ . The set  $U$  is a  $t$ -cluster for all  $t$  such that  $t \geq t'' = \max_{x \in U} \pi(x, U)$ . For each  $t$  such that  $t' < t < t''$ , a non-trivial  $t$ -cluster exists.

For example consider ESD:

$$\pi_\alpha(x, A) = \min_{y \in A} d_{xy}, \quad \pi_\alpha(x, \emptyset) = d_{xx},$$

where  $d_{xy}$  is a dissimilarity measure,  $x, y \in U$ ,  $A \subseteq U$ . Assume as usual  $d_{xy} = d_{yx}$  and  $d_{zz} \leq d_{xy}$  for all  $x, y, z \in U$ . This ESD is non-increasing by definition. To interpret  $t$ -clusters, consider a *threshold graph of level  $t$* :  $G_t = (U, E_t)$ ,  $E_t = \{(x, y) : x, y \in U, d_{xy} \leq t\}$ . Then the minimal nonempty  $t$ -clusters turn to be exactly all components of this graph. We show in section 5 that they define the *single-linkage* clustering method. Moreover, every  $t$ -cluster is the union of some of these components and vice versa. To obtain another example, assume  $U$  is the set of variables of a data matrix. Consider a regression problem with  $x$  as independent variable, and variables from  $A$  except  $x$  - as the independent variables. Then the ESD  $\pi(x, A) = 1 - R^2$  is non-increasing. In section 5, we consider one more example of non-increasing ESD, which helps define the weak  $k$ -linkage clustering method.

### B. Integrability

Given a set function  $f(A)$ ,  $A \subseteq U$ , define an ESD  $\pi$  with the equality:

$$\pi(x, A) = f(A + x) - f(A - x). \quad (2.1)$$

It is point-independent by definition. An ESD, which satisfies (2.1), with some appropriate set function  $f$ , is called *integrable*.

**Proposition 2.4.** *If an ESD is integrable, then there exists a  $t$ -cluster for each  $t$ .*

*Proof.* Let  $\pi$  satisfy (2.1) with the set function  $f$ . For  $t = 0$  one may rewrite the definition of  $t$ -cluster (1.1a,b) as follows:

$$\begin{aligned} f(A) &\leq f(A - x) \text{ for all } x \in A, \\ f(A) &< f(A + x) \text{ for all } x \in U \setminus A. \end{aligned}$$

As  $U$  is finite, there exists a subset  $A \subseteq U$ , which minimizes  $f$ , thus satisfying the first inequality. If the second inequality is not satisfied for some  $x$ , then  $f(A + x) = f(A)$ . Then consider  $A + x$ , and so on, until both inequalities are satisfied. For  $t \neq 0$  consider an ESD  $\pi'$  defines as:  $\pi'(x, A) = \pi(x, A) - t$ . It is clearly integrable with  $f'(A) = f(A) - t \cdot |A|$  (where  $|\cdot|$  denotes cardinality); consequently it possesses a 0-cluster, which is obviously a  $t$ -cluster of  $\pi$ . ■

We consider an example of integrable function in Section 6.

### C. Ultrametricity

ESD is called ultrametric iff it satisfies:

$$\pi(x, A+y) \leq \max \{ \pi(x, A), \pi(y, A) \} \quad (2.3)$$

for all  $x \in A$ ,  $y \in U$ ,  $A \subseteq U$ .

The term is chosen because (2.3) resembles the ultrametric inequality, which is central to hierarchical clustering. An ultrametric dissimilarity measure generates a dendrogram. In section 4, we shall see that an ultrametric ESD generates a generalized dendrogram. Note that a non-increasing ESD is certainly ultrametric. Thus, an ESD, which is ultrametric and point-non-increasing, may be called weak decreasing.

**Proposition 2.5.** *Let  $\pi(x, A)$  be a weak decreasing ESD. Then:*

- (i) *if  $\pi(x, A) \leq t$  for all  $x \in A$ , then there exists a  $t$ -cluster  $B \supseteq A$ ;*
- (ii) *if  $A_1$  is  $t_1$ -cluster and  $t_2 > t_1$ , then there exists a  $t_2$ -cluster  $A_2$  such that  $A_2 \supseteq A_1$ ;*
- (iii) *there exists a  $t$ -cluster for any  $t$ .*

*Proof.* To prove (i) find  $y \in U \setminus A$  such that  $\pi(y, A) \leq t$ . If there is no such  $y$ , then  $A$  itself is a  $t$ -cluster. If there is, then we infer that  $\pi(y, A+y) \leq t$ , since  $\pi$  is point-non-increasing, and  $\pi(y, A+y) \leq t$  for all  $x \in A$ , since  $\pi$  is ultrametric. Thus  $A' = A + y$  satisfies the condition of item (i). The approach taken for  $A$  works for  $A'$  as well, and so on. As  $U$  is finite, at some iteration, a  $t$ -cluster will be obtained.

To prove (ii), we note that because of the non-increasing property,  $\pi(x, A_1) \leq t_1 < t_2$  for all  $x \in A_1$ . Then (i) provides the poof.

Finally, we prove (iii). If  $\pi(x, \{x\}) \leq t$  for some  $x$ , then by (i) there exists a  $t$ -cluster  $A$  with  $\{x\} \subseteq A$ . Otherwise the empty set is a  $t$ -cluster. ■

If there exists an element  $x \in U$  such that  $\pi(x, \{x\}) < \pi(x, U)$ , then a non-trivial  $t$ -cluster exists, where  $t = \pi(x, \{x\})$ . Indeed, according to Proposition 2.5i, there exists a  $t$ -cluster  $A$  such that  $x \in A$ , and obviously  $A$  is not equal to  $U$ .

The above results are certainly true for non-increasing ESD also, but here  $t$ -clusters do not necessarily form a lattice.



As an example consider ESD defined by the formula

$$\pi_{\omega}(x, A) = \max_{y \in A+x} d_{xy},$$

where  $d_{xy}$  is again a dissimilarity measure. It is obviously non-decreasing and point-independent. Let us show that it is also ultrametric. Indeed, for  $x \in A$ :

$$\pi_{\omega}(x, A+y) = \max_{z \in A+y+x} d_{xz} = \max \left\{ \max_{z \in A+x} d_{zx}, d_{xy} \right\} = \max \left\{ \pi_{\omega}(x, A), d_{xy} \right\}.$$

To prove the statement we note that:

$$d_{xy} = d_{yx} \leq \max_{z \in A+y} d_{yz} = \pi_{\omega}(y, A).$$

Some more examples of ultrametric ESD's are considered in Section 5, where they help define the well-known strong  $k$ -linkage,  $k$ -overlap and  $k$ -clique clustering methods (Matula 1977).

### 3. Non-Decreasing ESD's and Quasi-Concave Set Functions

Monotone functions  $\pi(x, A)$ , where  $x \in U$ ,  $A \subseteq U$ , were studied by Mulla (1976,1977) and Kuznetsov, Muchnik, and Shvartser (1985). Their relation to quasi-concave set functions was developed there, forming the subject of so-called monotone system theory. We first review the main concepts of this theory, assuming the function  $\pi$  is ESD.

Given a non-decreasing ESD  $\pi$ , define a set function  $F$ :

$$F(A) = \min_{x \in A} \pi(x, A), \quad (3.1)$$

where  $A \subseteq U$ ,  $A \neq \emptyset$ . This function proves to be quasi-concave, i.e., for all  $A, B$  the inequality holds:

$$F(A \cup B) \geq \min(F(A), F(B)). \quad (3.2)$$

A subset of  $U$  that maximizes  $F$  is called a core. Because of quasi-concavity, there exists a *greatest core*  $\Gamma$ , which is the union of all cores. Moreover, a subset  $\Gamma'$ , which

contains all subsets  $A$  satisfying  $F(A) \geq F(\Gamma')$ , is called a *greatest quasi-core*. All greatest quasi-cores constitute a sequence  $\Gamma_1, \Gamma_2, \dots, \Gamma_p$  with the following properties:

$$\begin{aligned} \Gamma_1 \supset \Gamma_2 \supset \dots \supset \Gamma_p, \\ F(\Gamma_i) > F(\Gamma_{i-1}), \quad i = 2, \dots, p. \end{aligned}$$

An algorithm has been developed which produces a sequence of greatest quasi-cores, given a monotone function  $\pi$ . It requires  $O(|U|^2)$  evaluations of function  $\pi$ . Note that  $F$  in (3.1) depends only on the values of  $\pi$ , where  $x \in A$ . So in this section we usually assume ESD to be point-independent.

Monotone systems have different applications in data analysis (Kuznetsov et al. 1984; Kuznetsov, Muchnik, and Shvartser 1985; Zaks and Muchnik 1989). The concept similar to a  $t$ -cluster for the continuous case has been formulated by Mulla (1981).  $t$ -clusters for monotone ESD's were studied by Genkin, Zaks and Muchnik (1988a). Here we present the main results of the latter paper and then proceed with the new work.

As can be seen from (3.1), there exists numerous ESD's producing the same quasi-concave set function. The question arises: are the  $t$ -clusters of such ESD's identical? Or in other words, can the concept of  $t$ -cluster be formulated using these functions, without reference to ESD? As we shall see, the general answer is negative, but  $F$  imposes substantial limitations upon  $t$ -clusters.

For non-decreasing ESD,  $t$ -clusters may or may not exist, but if they do, some of their properties can be observed.

**Proposition 3.1.** *If  $A_i$  are  $t_i$ -clusters of non-decreasing ESD  $\pi$ ,  $i = 1, 2$ , and  $A_1 \supset A_2$ , then  $t_1 > t_2$ .*

*Proof.* Let  $x \in A_1 \setminus A_2$ . Then  $\pi(x, A_1) \leq t_1$ ,  $\pi(x, A_2) > t_2$ . But  $\pi(x, A_1) \geq \pi(x, A_2)$ , and we obtain  $t_1 > t_2$ . ■

**Corollary.**  *$t$ -clusters of non-decreasing ESD for given  $t$  are not subsets of one another.*

Equality (3.1) defines a mapping from the class defines a mapping from the class  $\mathcal{I}$  of non-decreasing ESD's to the class  $\mathcal{D}$  of set functions. Now we define mapping from  $\mathcal{D}$  to  $\mathcal{I}$  following Malishevsky (1986):

$$\pi(x, A) = \max \{ F(X) : x \in X \subseteq A + x \}. \quad (3.3)$$

By definition, the resulting  $\pi$  is point-independent and non-decreasing. Soon we shall see that it is also ultrametric.

**Lemma.** *If  $\pi$  is a non-decreasing ultrametric ESD and  $\pi(x, A) > \pi(y, A)$ ,  $x, y \in A$ ,  $A \subseteq U$ , then  $\pi(x, A) = \pi(x, A - y)$ .*

*Proof.* Rewrite  $\pi(x, A) = \pi(x, (A - y) + y)$ . Because of ultrametricity, either  $\pi(x, (A - y) + y) \leq \pi(x, A - y)$ ,  $\pi(x, (A - y) + y) \leq \pi(y, A - y) \leq \pi(y, A)$ . The latter inequality contradicts the condition of the Lemma, therefore the former one, together with the non-decreasing property gives the proof. ■

**Proposition 3.2.** *The point-independent non-decreasing ESD  $\pi$  satisfies (3.3) with some appropriate set function  $F$  iff it is ultrametric.*

*Proof.* Let (3.3) be true and  $x \in A$ ,  $y \in U - A$ . If  $\pi(x, A + y) = \pi(x, A)$  then the ultrametric inequality is satisfied. Otherwise  $\pi(x, A + y) > \pi(x, A)$ , and there exists  $X \subseteq A + y$  such that  $\pi(x, A + y) = F(X)$ . Then we have

$$\pi(x, A + y) = F(X) \leq \pi(y, X) \leq \pi(y, A + y) = \pi(y, A).$$

Now, let  $\pi$  be ultrametric and  $F(A)$  be defined by (3.1). Given  $x$  and  $A$ , where  $x \in A \subseteq U$ , there exists a greatest maximum of  $F$  in the right side of (3.3) because of quasi-concavity. Denote it by  $B$ . By non-decreasing property,  $\pi(x, A) \geq \pi(x, B) \geq F(B)$ . Since  $B$  is the greatest maximum of  $F$ ,  $F(B) > F(A)$ . Thus, we infer:  $\pi(x, A) > F(A)$ . Now consider the sequence  $A_0 = A, A_1, \dots, A_n = \emptyset$ , where  $A_{i+1} = A_i - y_i$ ,  $\pi(y_i, A_i) = F(A_i)$ ,  $i = 0, \dots, n - 1$ . Let  $A_k$  be the first member of this sequence with  $y_k \in B$ . Then  $B \subseteq A_k$  and  $F(B) \leq \pi(y_k, B) \leq \pi(y_k, A_k) = F(A_k)$ .

Inequality would contradict the definition of  $B$ ; thus,  $F(B) = F(A_k)$  and  $B = A_k$ . By the Lemma,  $\pi(x, A_i) = \pi(x, A_{i+1})$  for  $i = 0, \dots, k-1$ . Thus, we obtain  $\pi(x, B) = \pi(x, A)$ .

Suppose now that  $F(B) < \pi(x, B)$ . Then find  $r$  such that  $y_r = x$ . By Lemma,  $\pi(x, A_i) = \pi(x, A_{i+1})$  for  $i = k, \dots, r-1$ , and we obtain  $F(A_r) = \pi(x, A_r) = \pi(x, B) > F(B)$ , which contradicts  $B$ 's being a maximum. Thus,  $F(B) \geq \pi(x, B)$ . But  $F(B) \leq \pi(x, B)$  by definition of  $F$  in (3.1), and we obtain  $F(B) = \pi(x, B) = \pi(x, A)$ . ■

Ultrametric ESD's play an important role in the monotone system theory, as following proposition demonstrates.

**Proposition 3.3.** *Consider a class of all non-decreasing point-independent ESD's satisfying (3.1) with the given quasi-concave set function. There exists a unique minimal ESD, which is ultrametric and given by (3.3).*

*Sketch of proof.* Mapping defined by the expressions in (3.1) and (3.3) form a Galois correspondence (Birkhoff 1967, Ch 5, §8) between the class  $\Pi$  and  $\Phi$ . The images of the first mapping prove to be exactly all quasi-concave set functions. The images of the second mapping prove to be all ultrametric point-independent non-decreasing ESD's. ■

Now we are prepared to describe  $t$ -clusters using set functions. We call subset  $A$  the least quasi-core of level  $t$  of the quasi-concave set function  $F$ , if  $F(A) > t$ , and it contains no proper subset with such a property.

**Proposition 3.4.** Let  $F$  and  $\pi$  be respectively a quasi-concave set function and a non-decreasing ESD, satisfying (3.1). Let  $G_1, \dots, G_k$  be the least quasi-cores of level  $t$ . Then:

- (i) If a  $t$ -cluster exists, it contains none of  $G_i$ ,  $i = 1, \dots, k$ ;
- (ii) If  $\pi$  is ultrametric and point-independent, then  $t$ -clusters are exactly all maximal subsets of  $U$ , containing none of  $G_i$ ,  $i = 1, \dots, k$ .

*Proof.* We prove item (ii) first. Proposition 3.2 allows  $\pi$  to be represented by (3.3), and we rewrite the definition of a  $t$ -cluster in (1.1a,b) as follows:

$$\begin{aligned} \max \{ F(X) : x \in X \subseteq A \} &\leq t \text{ for all } x \in A, \\ \max \{ F(X) : y \in X \subseteq A+y \} &> t \text{ for all } y \in U \setminus A, \end{aligned}$$

or equivalently

$$\begin{aligned} \text{for all } X \subseteq A : F(X) &< t, \\ \text{for all } y \in U \setminus A \text{ there exists } X \subseteq A+y, &\text{ such that } F(X) > t. \end{aligned}$$

In other words, no least quasi-core is contained in  $A$ , but at least one of them is contained in each superset of  $A$ .

To prove item (i), assume  $\pi_0$  is defined by (3.3) with  $F$ . According to Proposition 3.3,  $\pi_0$  is ultrametric and point-independent, and  $\pi_0 \leq \pi$ . If  $A$  is a  $t$ -cluster of  $\pi$ , then  $\pi_0(x, A) \leq \pi(x, A) \leq t$  for all  $x \in A$ . By Proposition 2.5,  $\pi_0$  possesses a  $t$ -cluster  $B$  such that  $B \supseteq A$ . As we have just seen, none of  $G_i$ ,  $i = 1, \dots, k$ , is contained in  $B$ ; thus none is contained in  $A$ . ■

This proposition has a clear interpretation in cluster terms. A large value of  $F(A)$  shows that each element of  $A$  is dissimilar to  $A$ , see 3.1, so we may interpret  $A$  as a sparse set, and a  $t$ -cluster is a set countering no sparse subset.

With the help of the above Lemma, we obtain one more proposition that is important for the monotone system theory.

**Proposition 3.5.** *Let  $\pi$  be an ultrametric non-decreasing ESD. Let  $\Gamma_1, \dots, \Gamma_p$  be the sequence of greatest quasi-cores of the corresponding quasi-concave set function  $F$ , and  $\Gamma_{p+1} = \emptyset$ . Then:*

- (i)  $\pi(x, U) = \pi(x, \Gamma_i)$  for all  $x \in \Gamma_i$ ,  $i = 1, \dots, p$ ;
- (ii)  $\pi(x, \Gamma_i) = F(\Gamma_i)$  for all  $x \in \Gamma_i \setminus \Gamma_{i+1}$ ,  $i = 1, \dots, p$ .

*Proof.* Consider the sequence  $A_0 = U, A_1, \dots, A_n = \emptyset$ , where  $A_{j+1} = A_j - y_j$ ,  $\pi(y_j, A_j) = F(A_j)$ ,  $j = 0, \dots, n-1$ . Given  $i$ , let  $k$  be the smallest number satisfying:  $y_k \in \Gamma_i$ . Then  $\Gamma_i \subseteq A_k$  and

$$F(\Gamma_i) \leq \pi(y_k, \Gamma_i) \leq \pi(y_k, A_k) = F(A_k).$$

Inequality would disallow  $\Gamma_i$  as being the greatest quasi-core, thus,  $F(\Gamma_i) = F(A_k)$  and  $\Gamma_i = A_k$ . By Lemma,  $\pi(x, A_j) = \pi(x, A_{j+1})$  for  $j = 0, \dots, k-1$  for each  $x \in \Gamma_i$ , and we obtain (i).

To prove (ii), assume  $x \in \Gamma_i$ ,  $\pi(x, \Gamma_i) > F(\Gamma_i)$ , and  $x = y_r$  for some  $r$ . Then by Lemma,  $\pi(x, A_j) = \pi(x, A_{j+1})$  for  $j = k, \dots, r-1$ . Thus

$$F(A_r) = \pi(x, A_r) = \pi(x, \Gamma_i) > F(\Gamma_i),$$

and we conclude that  $A_r$  is contained in some greatest quasi-core  $\Gamma_m$  with  $m > i$ . ■

It follows immediately from the above proposition that in the ultrametric case, it takes only  $|U|$  calculations of  $\pi$  to find a sequence of greatest quasi-cores. Indeed  $\pi(x, U)$  for all  $x \in U$  contains full information.

For example consider  $t$ -clusters of ESD  $\pi_\omega$  introduced in Section 2. Each pair  $x, y$  such that  $d_{xy} > t$ , constitutes a least quasi-core. Then the set is a  $t$ -cluster iff it is a clique of the threshold graph  $G_t$ . In other words, the  $t$ -cluster is the maximal subset of  $U$ , where every two elements  $x$  and  $y$  satisfy the condition  $d_{xy} \leq t$ .

#### 4. Hierarchical Clustering and $t$ -clusters

The results of hierarchical clustering are often presented in the form a dendrogram, which is a sequence of increasingly refined partitions together with corresponding values of a dissimilarity measure (Jardine and Sibson 1971; Sneath and Sokal 1973). We review the necessary terms and then consider a more general concept.

A dendrogram is completely defined by an ultrametric dissimilarity measure. Given a set  $U$  of objects being clustered, a dissimilarity measure  $d_{xy}$  is called an ultrametric if the following conditions are satisfied for all  $x, y, z \in U$ :

- U1.  $d_{xy} \geq 0, d_{xx} = 0,$
- U2.  $d_{xy} = d_{yx},$
- U3.  $d_{yz} \leq \max(d_{xy}, d_{xz}).$

Given an ultrametric dissimilarity measure, clusters of level  $t$  for each  $t \geq 0$  are defined as subsets  $A \subseteq U$ , satisfying:

$$d_{x,y} \leq t \text{ for all } x, y \in A \quad (4.1a)$$

$$d_{x,y} > t \text{ for all } x \in A, y \in U \setminus A. \quad (4.1b)$$

The set of clusters for any given  $t$  forms a partition of  $U$ . For each cluster  $A$ , define  $t(A)$  as the minimal value of  $t$  such that  $A$  is the cluster of level  $t$ . When displaying a dendrogram, usually the value  $t(A)$  is displayed together with each  $A$ . A dendrogram possesses the following properties:

- D1. For any given  $t$ , clusters exist.
- D2. For any given  $t$ , clusters are mutually disjoint.
- D3. For any given  $t \geq 0$ , clusters form a cover of  $U$ , i.e., every object belongs to some cluster.
- D4. If  $A_1$  is a cluster of level  $t_1$ , and  $t_2 > t_1$ , then there exists a cluster  $A_2$  of level  $t_2$ , such that  $A_2 \supseteq A_1$ .

Property D1 is obviously the implication of D3, but we outline it for convenience. Properties D2 and D3 together imply that clusters form a partition of  $U$ , i.e., each object belongs to exactly one cluster. Now consider another property:

- GD. For any given  $t$ , clusters are not subsets of one another.

Obviously, D2 implies GD, and D1-D4 together imply the set of properties D1, GD and D4. That is the reason we call a set of subsets with the corresponding values of  $t$ , satisfying D1, GD and D4, a *generalized dendrogram*.

**Proposition 4.1.** *Given a non-decreasing weak decreasing ESD, its  $t$ -clusters together with their  $t$  values form a generalized dendrogram.*

*Proof.* D1 and D4 are true because of Proposition 2.5. GD is true because of the Corollary of Proposition 3.1. ■

A generalization is made in two directions. First,  $t$ -clusters for a given  $t$  may overlap when D2 is not true. Second, they may not form a cover of  $U$  when D3 is not true; i.e.,

there may be an object belonging to none of  $t$ -clusters. Here we formulate the necessary and sufficient conditions for  $t$ -clusters to be disjoint for each  $t$ , as well as the necessary and sufficient conditions for them to form a cover of  $U$  for each  $t$  large enough.

**Proposition 4.2.** *Given a non-decreasing weak decreasing ESD  $\pi$ , its  $t$ -clusters form a cover of the set  $U$  for all  $t \geq s$  iff  $\pi(x, \{x\}) \leq s$  for all  $x \in U$ .*

*Proof.* If  $\pi(x, \{x\}) \leq s$ , then for any  $t \geq s$  there exists a  $t$ -clusters containing  $\{x\}$ , because of Proposition 2.5. Let there now be  $\pi(x, \{x\}) > s$  for some  $x \in U$ . Then by the non-decreasing property we have  $\pi(x, A) > s$  for each  $A$  such that  $x \in A$ . Thus,  $x$  cannot belong to any  $s$ -cluster. ■

It is reasonable to assume the ESD to be non-negative. Then we should apply the above proposition with  $s = 0$  to obtain D3.

**Proposition 4.3.** *The condition*

$$\pi(x, A) \leq \max \left( \min_{y \in A} \pi(x, \{x, y\}), \max_{y \in A} \pi(y, a) \right) \quad (4.2)$$

for all  $A \subseteq U$ ,  $A \neq \emptyset$ ,  $x \in U \setminus A$ ,

is necessary and sufficient for  $t$ -clusters of non-decreasing, weak decreasing ESD to be disjoint for each  $t$ .

*Proof.* To prove sufficiency, assume (4.2) is true, but that there exist overlapping  $t$ -clusters  $A$  and  $B$ . Let  $y \in A \cap B$ ,  $x \in B \setminus A$ . Then  $\pi(x, \{x, y\}) \leq \pi(x, B) \leq t$ , and by applying (4.2) we conclude that  $\pi(x, A) \leq t$ , which contradicts  $A$ 's being  $t$ -cluster.

To prove necessity, assume (4.2) is not true, and there exist  $A \subseteq U$ ,  $y \in A$ ,  $x \in U \setminus A$ , where the following holds:

$$\pi(x, \{x, y\}) \leq t, \max_{z \in A} \pi(z, A) \leq t, \pi(x, A) > t.$$

By Proposition 2.5, there exists a  $t$ -cluster  $A'$  such that  $\pi(y, \{y\}) \leq \pi(y, A) \leq t$  and  $\pi(x, \{y\}) \leq \pi(x, \{x, y\}) \leq t$ . Then by ultrametricity we obtain:

$$\pi(y, \{x, y\}) \leq \max(\pi(y, \{y\}), \pi(x, \{y\})) \leq t.$$



Applying Proposition 2.5, we conclude that there exists a  $t$ -cluster  $B$  with  $\{x, y\} \subseteq B$ . But  $x \in U \setminus A'$ , because  $\pi(x, A') \geq \pi(x, A) > t$ . Thus, we obtain overlapping  $t$ -clusters  $A'$  and  $B$ , where  $y \in A' \cap B$ ,  $x \in B \setminus A'$ . ■

The inequality (4.2) seems to be rather difficult to interpret. The following proposition presents a more simple necessary condition. We introduce an auxiliary dissimilarity measure:

$$d_{xy}^\pi = \max(\pi(x, \{x, y\}), \pi(y, \{x, y\})) \quad (4.3)$$

for all  $x, y \in U$ .

**Proposition 4.4.** *Let ESD  $\pi$  be non-decreasing and weak decreasing and let  $t$ -clusters be disjoint for each  $t$ . Then the dissimilarity defined by (4.3) satisfies the ultrametric inequality, i.e., for each  $x, y, z \in U$ :*

$$d_{xz}^\pi \leq \max(d_{xy}^\pi, d_{yz}^\pi).$$

*Proof.* Suppose  $d_{xy}^\pi \leq t$ ,  $d_{yz}^\pi \leq t$  for some  $x, y, z \in U$  and some number  $t$ . Because of Proposition 4.3, the condition (4.2) is true. Applying this condition with  $A = \{y, z\}$  we obtain:

$$\begin{aligned} \pi(x, \{y, z\}) &\leq \left( \min_{v \in \{y, z\}} \pi(x, \{v, x\}), \max_{v \in \{y, z\}} \pi(v, \{y, z\}) \right) \leq \\ &\leq \max(d_{xy}^\pi, d_{yz}^\pi) \leq t. \end{aligned}$$

From non-decreasing property, we conclude that  $\pi(x, \{z\}) \leq t$ . The non-decreasing and weak decreasing properties together imply the ESD is point-independent, so  $\pi(x, \{x, z\}) = \pi(x, \{z\}) \leq t$ . In the same way, we can prove:  $\pi(z, \{x, z\}) \leq t$  and thus we obtain  $d_{xz}^\pi \leq t$ . ■

This proposition shows that when  $t$ -clusters are disjoint, they may be interpreted as clusters of  $d_{xy}^\pi$  in the hierarchical clustering sense; see (4.1a,b). On the other hand, let a dissimilarity measure satisfying U1-U3 be given. Then the ESD

$$\pi(x, A) = \max_{y \in A \cup x} d_{xy},$$

is non-decreasing and weak decreasing (see example at the end of Section 3 for the proof). It is easy to see that the generalized dendrogram generated by this ESD is equal to the standard dendrogram generated by the dissimilarity measure.

## 5. Graph-Theoretic Clustering Methods and $t$ -clusters

In this section we consider some clustering techniques based on graph-theoretic concepts (Hubert 1974; Matula 1977), using  $t$ -clusters. We demonstrate that they provide examples of generalized dendrogram, which is studied in the previous section.

Conditions D1 and D4 will always apply throughout this section because of the ESD's being weak decreasing; see Proposition 2.5. These conditions are certainly also true when ESD is non-increasing. To obtain GD, non-decreasing property is sufficient because of the Corollary of Proposition 3.1. This possibility is used in Section 4. Otherwise, we can take into consideration only all minimal or all maximal  $t$ -clusters to guarantee the condition GD is true.

Assume a dissimilarity measure  $d_{xy}$  on finite set  $U$ , where  $d_{xy} = d_{yx}$  for all  $x, y \in U$ ,  $x \neq y$ . This measure can also be thought of as a complete graph with vertices set  $U$  and edges marked with values of  $d_{xy}$ . We denote by  $\alpha$  and  $n$  respectively the maximal and minimal values of  $d_{xy}$  for all  $x, y \in U$ ,  $x \neq y$ .

Consider the threshold graph of level  $t$ :  $G_t = (U, E_t)$ , as defined earlier. By  $G_t[A]$  we denote the subgraph of  $G_t$  induced by the vertex subset  $A \subseteq U$ . Now we introduce some functions to help define graph-theoretic clustering methods. For  $A \subseteq U$ ,  $x, y \in A$ ,  $x \neq y$ , let:

$\sigma_\lambda(x, y, A, t)$  - the number of edge-disjoint parts of the graph  $G_t[A]$ , joining the vertices  $x$  and  $y$ ;

$\sigma_\varepsilon(x, y, A, t)$  - the number of vertex-disjoint (except for endpoints) path of  $G_t[A]$ , joining the vertices  $x$  and  $y$ .

Given these two functions, we define one more function, where a dot in subscript substitutes for the letter  $\lambda$  or  $\varepsilon$ ;  $k$  is an integer,  $k \leq |U| - 2$ ;  $A \subseteq U$ ,  $x, y \in A$ :

$$\rho_{\bullet k}(x, y, A) = \begin{cases} n, & \text{if } x = y, \\ \min \{t : \sigma_\bullet(x, y, A, t) \geq k\}, & \text{if } x \neq y \text{ and } |A| \geq k + 2, \\ \alpha + 1, & \text{otherwise.} \end{cases}$$

Now we introduce one more function, where  $A \subseteq U$ ,  $x, y \in A$ :

$$\rho_{\omega k}(x, y, A) = \begin{cases} n, & \text{if } x = y, \\ d_{x,y}, & \text{if } x \neq y \text{ and } |A| \geq k + 1, \\ X + 1, & \text{otherwise.} \end{cases}$$

and consider all three functions  $\rho_{\bullet k}$ , where a dot in a subscript substitutes for one of letter  $\lambda$ ,  $\varrho$ ,  $\omega$ . These functions can be interpreted as conditional dissimilarities, meaning that conditions are determined by the subset  $A$ . Obviously, they possess the following properties:

$$\rho_{\bullet k}(x, y, A) = \rho_{\bullet k}(y, x, A), \quad (5.1)$$

$$\text{if } A \subset A', \text{ then } \rho_{\bullet k}(x, y, A) \geq \rho_{\bullet k}(x, y, A'), \quad (5.2)$$

for all  $x, y \in U$ ,  $A, A' \subseteq U$ , and any integer  $k$ . Now everything is ready for defining ESD's:

$$\pi_{\bullet k}(x, A) = \max_{y \in A+x} \rho_{\bullet k}(x, y, A+x),$$

which are point-independent by definition. With the help of properties (5.1) and (5.2) we shall demonstrate that they are also ultrametric. Indeed, for  $x \in A$ :

$$\begin{aligned} \pi_{\bullet k}(x, A+y) &= \max_{u \in A+y} \rho_{\bullet k}(x, u, A+x+y) = \\ &= \max \left( \max_{u \in A} \rho_{\bullet k}(x, u, A+y), \rho_{\bullet k}(x, y, A+y) \right). \end{aligned}$$

Applying (5.2) to the first item in brackets, we see it is not greater than  $\pi_{\bullet k}(x, A)$ . Applying (5.1) to the second one, we get:

$$\begin{aligned} \rho_{\bullet k}(x, y, A+y) &= \rho_{\bullet k}(y, x, A+y) \leq \\ &\leq \max_{u \in A+y} \rho_{\bullet k}(y, u, A+y) = \pi_{\bullet k}(y, A). \end{aligned}$$

Now conditions D1 and D4 are satisfied because of Proposition 2.5. Taking into consideration only maximal  $t$ -clusters for each  $t$ , we satisfy the condition GD, and thus obtain a generalized dendrogram. Applying Proposition 4.2 with  $s = n$ , we conclude that our clusters form a cover at each level of generalized dendrogram.

The  $t$ -clusters of  $\pi_{\lambda k}$  and  $\pi_{\alpha k}$  are subsets such that any pair of vertices of the induced subgraph of  $G_t$  is joined by at least  $k$  edge-disjoint (respectively, vertex-disjoint except endpoints) paths, and no vertex may be added without losing this property. They are called  $k$ -components and  $k$ -blocks. In both cases we choose only the maximal subsets with such a property, and obtain what Matula (1977) called *strong  $k$ -linkage and  $k$ -overlap clustering methods*.  $t$ -clusters of  $\pi_{\omega k}$  are cliques with more than  $k$  elements, called  $k$ -cliques. They define the  $k$ -cliques clustering method. When  $k = 1$ , we obtain  $\pi_{\omega 1} = \pi_{\omega}$ ; the ESD  $\pi_{\omega}$  was introduced in Section 2 and studied in Section 3.

Consider now a function:

$$\tau(x, A, t) = \begin{cases} \text{degree of } x \text{ in } G_t[A], & \text{if this graph is connected,} \\ 0, & \text{otherwise.} \end{cases}$$

and an ESD:

$$\pi_{\delta k} = \begin{cases} \min \{t : \tau(x, A, t) \geq k\}, & \text{if } |A| \geq k, \\ x + 1 & \text{otherwise.} \end{cases}$$

It is obviously non-increasing, as is the ESD  $\pi_{\alpha}$  introduced in Section 2.  $t$ -clusters of  $\pi_{\delta k}$  are subsets such that any vertex of the corresponding induced subgraph is incident to at least  $k$  vertices, and no vertex may be added without losing this property. They are called  $k$ -bonds.  $t$ -clusters of  $\pi_{\alpha}$  satisfy this property: any pair of vertices of the corresponding induced subgraph is linked by a path, and no vertex may be added without losing this property. Such a subgraph is called a *component*.

To obtain a generalized dendrogram, we consider only minimal  $t$ -clusters. Thus, we have defined the *weak  $k$ -linkage* and the *single-linkage clustering methods*. In the latter case, the generalized dendrogram turns out to be a standard dendrogram. Indeed, it is easy to see that the conditions of Propositions 4.2 and 4.3 are true.

ESD allows us to generate a typical hierarchical agglomerative clustering procedure if we define a set-to-set dissimilarity measure  $D(\cdot, \cdot)$  as follows:

$$D(A, B) = \max \left\{ \max_{x \in A} \pi(x, B), \max_{y \in B} \pi(y, A) \right\}. \quad (5.3)$$

It is easy to see that  $\pi_\alpha$  and  $\pi_\omega$  lead to the *single-* and *complete-linkage* clustering respectively. To obtain more examples we define the *direct r-dissimilarity*  $\rho_{Dr}(x, y, A)$  as minimal dissimilarity value, say  $t$ , such that  $x$  and  $y$  may be connected by a sequence of at most  $r - 1$  intermediate elements of  $A$ , where each adjacent pair in the sequence has a dissimilarity value less than or equal to  $t$ . The *indirect r-dissimilarity*  $\rho_{Ir}(x, y)$  is the same but intermediate elements are chosen arbitrary from  $U$ . Since  $\rho_{Dr}$  satisfies (5.1) and (5.2), the ESD

$$\pi_{Dr}(x, A) = \max_{y \in A+x} \rho_{Dr}(x, y, A+x)$$

is ultrametric. The following ESD is obviously non-decreasing:

$$\pi_{Ir}(x, A) = \max_{y \in A} \rho_{Ir}(x, y).$$

Some hierarchical clustering methods may be defined through the complete-linkage with an adequate dissimilarity (or similarity) measure. It applies to *indirect*, but not to *direct r-diameter clustering method*, as stated by Hubert and Baker (1977). With the help of ESD's  $\pi_{Ir}$  and  $\pi_{Dr}$  we obtain both through (5.3).

## 6. Conceptual Clustering and $t$ -clusters

The idea of conceptual clustering (Michalski and Stepp 1982, 1983) treats a cluster not as a simply a subset of a given set of objects, but as a specific concept. Each cluster represents a certain generalized description of a corresponding subcategory of objects. These descriptions are conjunctive concepts, involving attributes of objects, and thus are supposed to have a clear interpretation.

Implementation of conceptual clustering should involve at least two kinds of subprocedures: generation of candidate concepts and choosing an optimal subset of them. These subprocedures may be combined in different ways. The second subprocedure is our main concern. Let a family of candidate concepts be generated. The problem is choosing a subfamily of it, which produces a partition of the set of objects if possible or in some sense

an “approximate partition” otherwise. We are going to demonstrate that this problem can be adequately formulated as finding a  $t$ -stable subset of concepts.

Each candidate concept corresponds to some subset of the set of objects  $\Omega$  and is supposed to be dense or compact enough in the “conceptual sense” (it is not necessary to define it precisely here). The problem is, given a family  $U$  of such subsets, choose a subfamily which is as proximate as possible to being a partition of  $\Omega$ .

Consider a Boolean matrix  $\|a_{x\omega}\|$  with the row index set  $U$  and the column index set  $\Omega$ , where  $a_{x\omega} = 1$  if  $\omega \in x$ , and  $a_{x\omega} = 0$  otherwise. To estimate a subfamily  $A \subseteq U$  suppose we add an element  $x \in U \setminus A$  to it. Doing so is appropriate if  $x$  covers some objects not covered by any member of  $A$ , but not if  $x$  covers objects that are already covered, because it creates overlapping clusters. Thus the utility of joining  $x$  to  $A$  may be estimated by the function:

$$\pi(x, A) = \sum_{\omega \in x} \delta \left( \sum_{y \in A-x} a_{y\omega} \right),$$

where a function  $\delta(n)$  is decreasing, positive for  $n = 0$ , and negative for  $n \geq 1$ . Obviously,  $\pi$  is non-decreasing, but it is also integrable. Indeed, consider a function  $\phi$  defined as follows:

$$\phi(0) = 0, \quad \phi(n) = \sum_{i=1}^{n-1} \delta(i) \text{ for } n \geq 1.$$

Now  $\pi$  satisfies (2.1) with the set function:

$$f(A) = \sum_{\omega \in \Omega} \phi \left( \sum_{x \in A} a_{x\omega} \right).$$

Our problem of choosing the approximate partition may now be formulated in the following extremal form:

$$\max \{ f(A) - t \cdot |A| : A \subseteq U \},$$

or equivalently:

$$\min \{ -f(A) + t \cdot |A| : A \subseteq U \}, \quad (6.1)$$

where the term  $t \cdot |A|$  plays the role of a penalty for too many clusters, or for too fine a clustering.

By Proposition 2.4, the minima of (6.1) are exactly the  $t$ -clusters of  $\pi$ . Moreover, as  $\pi$  is non-decreasing, (6.1) constitutes the problem of minimizing the submodular set function, as is shown in Nemhauser, Wolsey, and Fisher (1978). Some fast approximate algorithms for such a problem can be found there. For exact algorithm, which is optimal in “the worst case,” see Genkin and Muchnik (1990).

## 7. Summary

The fixed point concept provides a basis for a rather general definition of a cluster, called  $t$ -cluster, or a  $t$ -stable set. The definition of  $t$ -cluster is based on a dissimilarity measure between elements and subsets of the set being clustered, called the “element-set” type dissimilarity measure (ESD). Some basic mathematical properties of  $t$ -clusters are shown, where the most important are the existence problem and the relation to monotone system theory.

Our results are applied to different clustering strategies, which turn out to be particular cases of the theory developed. Hierarchical clusters produced by agglomerative procedures are  $t$ -clusters with monotone and ultrametric ESD. Such ESD’s produce clustering structures, which are more general than hierarchical ones in that the overlapping clusters and elements belonging to no clusters are permitted.

A set of clustering strategies based on graph-theoretic concepts is considered. An appropriate definition of ESD permits each of them to be formulated through the  $t$ -cluster concept.

Conceptual clustering gives quite another application of our theory. The problem is to choose the subfamily of a family of subsets of a finite set, which is approximate partition of this set. This problem is formulated as finding a  $t$ -stable subset in the family just mentioned.

## References

- BALL, G.H., and HALL, D.J. (1967), "A clustering Technique for Summarizing Multivariate Data," *Behavioral Sciences*, 12, 153-155.
- BIRKHOFF, G. (1967), *Lattice Theory*, Providence, RI: American Mathematical Society.
- GENKIN, A.V., and MUCHNIK, I.B. (1990), "Optimal Algorithm for Maximization of Submodular Functions," *Automation and Remote Control*, 51, 1121-1128.
- GENKIN, A.V., ZAKS, Yu.M., and MUCHNIK, I.B. (1998a), "The Structure of Final Sets of Counter-monotone System and Algorithms for its Analysis," in *Methods and Algorithms for Empirical Data Analysis*, Moscow: Institute of Control Sciences, 35-42 (in Russian).
- GENKIN, A.V., ZAKS, Yu.M., and MUCHNIK, I.B. (1998b), "Comparing the Structures of Monotone and Counter-monotone Systems," in *Methods and Algorithms for Empirical Data Analysis*, Moscow: Institute of Control Sciences, 42-48 (in Russian).
- HUBERT, L.J. (1974), "Some applications of Graph Theory to Clustering," *Psychometrika*, 39, 283-309.
- HUBERT, L.J., and BAKER, F.B. (1977), "An empirical Comparison of Baseline Models for Goodness-of-fit in r-diameter Hierarchical Clustering," in *Classification and Clustering*, Ed., J. Van Ryzin, New York: Academic Press, 131-153.
- JARDINE, N., and SIBSON, R. (1971), *Mathematical Taxonomy*, 131-153.
- KUZNETSOV, E.N., MUCHNIK, I.B., HENCEY, G., and TCHKUASELY, N.F. (1984), "Monotone Systems on Data Matrices," *Kozlemyek*, 31, 153-158 (in Russian).
- KUZNETSOV, E.N., MUCHNIK, I.B., and SHVARTSER, L.V. (1985), "Monotone Systems and their Properties," *Non-numerical Information Analysis in Social Studies*, Moscow: Nauka (in Russian), <http://www.data laundering.com/download/monosysp.pdf> .
- MALISHEVSKY, A.V. (1986), "About the Properties of the Ordinal Set Functions," in *All-Union Seminar for Optimization and its Applications*, Dushanbe, 151-152 (in Russian), <http://www.data laundering.com/download/order.pdf> .
- MATULA, D.W. (1977), "Graph Theoretic Techniques for Cluster Analysis Algorithms," in *Classification and Clustering*, Ed., J. Van Ryzin, New York: academic Press, 95-129.
- MICHALSKI, R.S., and STEPP, R. (1982), "Revealing Conceptual Structure in Data by Inductive Inference," in *Machine Intelligence 10*, Eds., J.E. Hayes, D.Michie and Y-H. Pao, New York: Wiley, 173-196.
- MICHALSKI, R.S., and STEPP, R. (1983), "Learning from Observation," in *Machine learning: An Artificial Intelligence Approach*, Eds., R.S. Michalsli, J.C. Carbonell and T.M. Mitchell, Palo Alto, CA: Tioga, 163-190.
- MUCHNIK, I.B., CHKUASELLI, N.F., and SHVARTSER, L.V. (1986), "Linguistic Analysis of 0-1 Matrices using Monotone Systems," *Automation and Remote Control*, 47, 553-560, <http://www.data laundering.com/download/linguist.pdf> .
- MULLAT, J.E. (1976, 1977), "Extremal Subsystems of Monotonic Systems, I,II,III," *Automation and Remote Control*, 1976, 37, 758-766, 37, 1286-1294; 1977, 38, 89-96, <http://www.data laundering.com/mono/extremal.htm> .
- MULLAT, J.E. (1981), "Contramotonic Systems in the Analysis of the Structure of Multivariate Distributions," *Automation and Remote Control*, 42, 986-993, <http://www.data laundering.com/download/contra.pdf> .
- NEMHAUSER, G.L., WOLSEY, L.A., and FISHER, M.L. (1978), "An analysis of Approximations for Maximizing Submodular Set Functions-I," *Mathematical Programming*, 14, 265-294.
- SNEATH, P.H.A., and SOKAL, R.R. (1973), *Numerical Taxonomy*, San Francisco; Freeman.
- TUCKER, L.R. (1964), "The Extention of Factor Analysis to Three-Dimensional Matrices," in *Contribution to Mathematical Psychology*, Eds., N. Frederiksen and H. Gulliksen, New York: Holt, Reinhart, and Winston, 109-127.
- ZAKS, Yu.M., and MUCHNIK, I.B., (1989), "Incomplete Classification of a Finite Set of Objects Using Monotone Systems," *Automation and Remote Control*, 50, 553-560, <http://www.data laundering.com/download/incomple.pdf> .