UNIVERSAL ALGEBRA AND APPLICATIONS
BANACH CENTER PUBLICATIONS, VOLUME 9
PWN–POLISH SCIENTIFIC PUBLISHERS
WARSAW 1982

APPLICATIONS OF THE THEORY OF PARTIALLY ORDERED SETS TO CLUSTER ANALYSIS

M. F. JANOWITZ*

Department of Mathematics, University of Massachusetts, Amherst, MA 01003, U.S.A.

1. Introduction

A very broad order theoretic model for cluster analysis was introduced in [2] and [3], with special attention paid to the role of cluster analysis in connection with problems of biological classification. In the present paper this model will be further generalized so that it will encompass an even broader class of cluster methods. However, before proceeding, let us examine the role of such a model.

After all, cluster analysis already has a fairly well developed theory. Several books, and literally hundreds of papers have been written on the subject. People are able to analyze properties of various clustering techniques, compare them, and arrive at meaningful classifications of various sets of objects. What purpose can it possible serve for a lattice theorist to step in, abstract the subject to the point where it is no longer recognizable, apply some high powered mathematical machinery to this abstract version of cluster analysis, and finally reach some conclusions that were already common knowledge? Of course if this is all that could be accomplished, little purpose would be served by such a model. What then should a mathematical model be able to do?

1. It should provide an axiomatic framework in which various cluster methods may be analyzed and compared.

2. It should be broad enough to encompass a large number of commonly used clustering techniques.

3. It should possess enough structure so that meaningful mathematical techniques may be employed.

* Research supported in part by a grant from the Faculty Research Council of the University of Massachusetts.

4. It should provide results that are of some interest to people who actually use clustering techniques.

5. It should provide a structure in which one can separate conjecture from fact. One should be able to distingusih properties that *ought* to hold from those that *do* hold.

The model to be presented leads to some interesting problems in the theory of partially ordered sets. The question of how well it satisfies the above requirements must be left to the reader to decide.

2. Background material from the theory of partially ordered sets

A basic familiarity with the theory of posets, semilattices, and lattices will be assumed. Despite this, it will be convenient to specifically develop certain notions here. Unless otherwise specified, all partial orders will be dencted $\leqslant$, and the symbols $\vee$, $\wedge$ will be used for the join and meet operations where they are defined. The symbol 0 or 0_p (1 or 1_p) will be used to denote the least (greatest) element of the poset P, provided such an element is present.

Let P, Q be posets. A mapping $\varphi\colon P \to Q$ is called *isotone* if $p_1 \leqslant p_2$ in P implies that $\varphi(p_1) \leqslant \varphi(p_2)$ in Q; it is a *join homomorphism* if the existence of $p_1 \vee p_2$ in P implies that $\varphi(p_1 \vee p_2)$ acts as the join in Q of $\varphi(p_1), \varphi(p_2)$. To say that it is a *complete join homomorphism* is to say that:

(1) for M a nonempty subset of P, the existence of $\bigvee M$ in P implies that $\varphi(\bigvee M)$ acts as the join in Q of $\varphi(M)$;

(2) if P has a 0, then $\varphi(0)$ is the least element of Q. There are dual notions of a *meet homomorphism* and a *complete meet homomorphism*.

An *ideal* of a poset P is a nonempty subset I of P having the property that for each $x \in I$, $y \leqslant x$ implies that $y \in I$. An ideal of the form $(x] = \{y \in P\colon y \leqslant x\}$ is called a *principal ideal*. There are dual notions of a *filter* and a *principal filter*, the latter being denoted as $[x)$. It will prove convenient to classify mappings according to how they behave with respect to principal ideals and filters. The first result is

THEOREM 2.1 ([1], Theorem 2.1, p. 5). *Let P, Q be posets and $\varphi\colon P \to Q$. TAE:*

(1) *φ is isotone.*

(2) *The pre-image of every principal ideal of Q is either empty or an ideal of P.*

(3) *The pre-image of every principal filter Q is either empty or a filter of P.*

A mapping $\varphi\colon P \to Q$ is called *chain residuated* if the pre-image of every principal ideal of Q is the union of finitely many principal ideals of P; it is called *chain residual* if the corresponding condition holds for principal filters of Q. Let $\mathrm{CR}(P, Q)$ denote the set of chain residuated mappings of P into Q with $\mathrm{CR}^+(P, Q)$ the set of chain residual mappings.

THEOREM 2.2. *Let P, Q be posets and $C \in \mathrm{CR}^+(P, Q)$. Then:*

(1) *C is isotone.*

(2) *If Q has a largest element 1, then $1 \in$ range C.*

(3) *$x_\delta \downarrow x$ in P implies $C(x_\delta) \downarrow C(x)$ in Q.*

Proof. (1) This is immediate from Theorem 2.1.

(2) If $p \in C^{-1}([1])$, then $C(p) = 1$.

(3) To say that $x_\delta \downarrow x$ is to say that the indexing set $\{\delta\colon \delta \in D\}$ is directed, that $\delta_1 \leqslant \delta_2$ in D implies $x_{\delta_1} \geqslant x_{\delta_2}$, and that $x = \bigwedge_\delta x_\delta$. By (1), $C(x)$ is a lower bound for all $C(x_\delta)$. Suppose y is any other lower bound. Then for each $\delta \in D$, $x_\delta \in C^{-1}([y)) = [a_1) \cup [a_2) \cup \ldots \cup [a_k)$, where k is a positive integer. We may clearly choose an index δ_0 so that x_{δ_0} dominates a minimum number of these a_i's. Then for $\delta \geqslant \delta_0$, $x_\delta \geqslant a_i$ iff $x_{\delta_0} \geqslant a_i$. Thus there is an a_i that is a lower bound for $\{x_\delta\colon \delta \geqslant \delta_0\}$, so $a_i \leqslant x = \bigwedge \{x_\delta\colon \delta \geqslant \delta_0\}$. It is immediate that $y \leqslant C(a_i) \leqslant x$.

There is a final type of mapping that will be of some interest. A mapping $\varphi\colon P \to Q$ is said to be *residuated* if the pre-image of every principal ideal of Q is a principal ideal of P; it is *residual* if the pre-image of every principal filter of Q is a principal filter of P. Let $\mathrm{Res}(P, Q)$ denote the set of residuated mappings of P into Q, and $\mathrm{Res}^+(P, Q)$ the set of residual maps. In case $P = Q$, we use the symbols $\mathrm{Res}(P)$ and $\mathrm{Res}^+(P)$ in place of $\mathrm{Res}(P, P)$, $\mathrm{Res}^+(P, P)$. In view of the fact that the theory of these mappings is well developed in [1], it will not be repeated here (see also [2] and [3]). It should be noted, however, that for every residuated mapping $\varphi\colon P \to Q$ there corresponds a unique residual mapping $\varphi^+\colon Q \to P$; these mappings are related by the equation

$$\varphi(p) \leqslant q \text{ if and only if } p \leqslant \varphi^+(q).$$

Notice that if P and Q are replaced by their duals, then φ^+ becomes residuated with φ the associated residual mapping. It will prove convenient to let φ^* denote the residuated mapping associated with the residual mapping φ.

To establish the reason for the terms "chain residuated" and "chain residual" we close this section by considering

THEOREM 2.3. *Let P, Q be posets. Assume that P has a largest element 1, Q has a smallest element 0, and that $Q \neq \{0\}$. A necessary and sufficient condition for each chain residuated mapping of P into Q to be residuated is that P be a chain.*

Proof. If P is a chain, it is clear that $\mathrm{CR}(P, Q) = \mathrm{Res}(P, Q)$. Suppose conversely that this condition holds. Let a, b denote arbitrary elements of P, and let q be a fixed nonzero element of Q. Define $\varphi\colon P \to Q$ by the rule $\varphi(p) = 0$ if $p \leqslant a$ or $p \leqslant b$, $\varphi(p) = q$ otherwise. Then $\varphi^{-1}(\{y\}) = (a] \cup (b]$ if $q \not\leqslant y$ and (1] if $q \leqslant y$. Thus $\varphi \in \mathrm{CR}(P, Q) = \mathrm{Res}(P, Q)$. Taking $p = \varphi^+(0)$, it is immediate that $(p] = (a] \cup (b]$, so $a \leqslant b$ or $b \leqslant a$. This shows that P is indeed a chain.

3. The order theoretic model

A description of the basic problem in cluster analysis appears in [2]. Essentially, we are given a finite set P whose cardinality is at least 3. The imput data may be regarded as a residual mapping $C\colon \boldsymbol{R}^+ \to \Sigma$, where $\boldsymbol{R}^+$ denotes the non-negative reals, and Σ the Boolean algebra of all subsets of the set of two element subsets of P. The output may be similarly regarded, so a cluster method may be defined to be a mapping $F\colon \mathrm{Res}^+(\boldsymbol{R}^+, \Sigma) \to \mathrm{Res}^+(\boldsymbol{R}^+, \Sigma)$. These ideas are generalized in [2] as follows: $\boldsymbol{R}^+$ is replaced by an arbitrary join semilattice L with 0; Σ is replaced by a pair M, N of posets, each of which has a unit element; a cluster method is then regarded as a mapping $F\colon \mathrm{Res}^+(L, M) \to \mathrm{Res}^+(L, N)$. There are a number of advantages to this procedure, some of which we now list:

1. It costs nothing, as one can always specialize the results back to the context in which they arose.

2. By viewing the theory in an abstract setting, one can obtain additional insight into what is happening.

3. Earlier models due to Jardine and Sibson [4] and Matula [5] may in a sense be subsumed by the new model.

4. One may view the output of a cluster method as a sequence of classifications of the set P, rather than as a sequence of reflexive symmetric relations on P. It turns out that this eliminates a great deal of confusion, especially when the output is to allow some overlap between clusters at a given level. An extremely lucid discussion of these issues occurs in [5].

Why then are we led to further generalize the model? The answer is both simple and illuminating. Let us return to the context in which the model arose and view a cluster method as a mapoing $F\colon \mathrm{Res}^+(L, \Sigma) \to \mathrm{Res}^+(L, \Sigma)$, where Σ is the Boolean algebra of all reflexive symmetric relations on the finite set P. A cluster method called *single linkage clustering* is one of the most commonly used techniques. It proceeds as follows: For each $S \in \Sigma$, let $\gamma(S)$ denote the equivalence relation generated by S. For $C \in \mathrm{Res}^+(L, \Sigma)$, take $F(C) = \gamma \circ C$. Notice though that by [2], Theorem 6.2, p. 67, this defines a cluster method iff L is a chain. Thus if one wants to have single linkage clustering available in situations where L is not a chain, the model must be generalized. What saves the day (see Lemma 4.1) is the fact that $F(C) \in \mathrm{CR}^+(L, \Sigma)$. This leads us to define a cluster method to be a mapping $F\colon \mathrm{Res}^+(L, \Sigma) \to \mathrm{CR}^+(L, \Sigma)$, and if cluster methods are to be composed we must take them to be mappings $F\colon \mathrm{CR}^+(L, \Sigma) \to \mathrm{CR}^+(L, \Sigma)$. To recapitulate, here is the final version of our model:

L is a join semilattice with 0 with $|L| > 1$;

M, N are posets with unit elements;

a *cluster method* is a mapping $F\colon \mathrm{CR}^+(L, M) \to \mathrm{CR}^+(L, N)$.

Though the theory will develop in this abstract setting, it will still be useful to think of M as representing the Boolean algebra of reflexive symmetric relations on a finite set P. Before examining the role of N, it will be necessary to present a definition. A *classification* of P ([5], p. 7) is a set $\mathscr{C}$ of subsets of P such that:

(1) $\bigcup \mathscr{C} = P$, and

(2) $C_1, C_2 \in \mathscr{C}$ with $C_1 \subseteq C_2 \Rightarrow C_1 = C_2$.

One then thinks of N as the set of classifications of P, partially ordered by the rule $\mathscr{C}_1 \leqslant \mathscr{C}_2$ iff $C_1 \in \mathscr{C}_1$ implies the existence of $C_2 \in \mathscr{C}_2$ such that $C_1 \subseteq C_2$.

There should be no confusion if this new model is referred to as *Model L*, where L denotes the underlying join semilattice in which all dissimilarities are measured. The earlier model that was developed in [2] and [3] could then be called the *Restricted Model L*, and a mapping $F\colon \mathrm{Res}^+(L, M) \to \mathrm{Res}^+(L, N)$ a *restricted cluster method.*

4. Flat cluster methods

As in [2], a cluster method F is called *flat* in case there is a mapping $\gamma\colon M \to N$ such that $F(C) = \gamma \circ C$ for all $C \in \mathrm{CR}^+(L, M)$. A mapping $\gamma\colon M \to N$ is called *L-flat* in case there is a cluster method F such that $F(C) = \gamma \circ C$ for all $C \in \mathrm{CR}^+(L, M)$; in other words, to say that γ is L-flat is to say that $C \in \mathrm{CR}^+(L, M)$ implies $\gamma \circ C \in \mathrm{CR}^+(L, N)$. This section is devoted to the solution of the analogues of the *First* and *Second Flat Cluster Problems* (See [2], § 6 and 7). This amounts to the characterization of flat cluster methods and L-flat mappings of M into N. We begin with the

First Flat Cluster Problem. *Let $\gamma\colon M \to N$ be given. What properties must γ have in order that there exist a cluster method F such that $F(C) = \gamma \circ C$ for all $C \in \mathrm{CR}^+(L, M)$?*

A useful result is provided by

LEMMA 4.1. *Assume that M is finite. A mapping $\gamma: M \to N$ is then chain residual iff it is isotone and maps 1_M into 1_N.*

Proof. By Theorem 2.2, every chain residual mapping of M into N has the specified properties. Suppose conversely that $\gamma: M \to N$ is isotone and maps 1_M into 1_N. For $n \in N$, Theorem 2.1 may be invoked to see that $\gamma^{-1}([n])$ is a filter of M; in that M is finite, $\gamma^{-1}([n])$ must even be the union of finitely many principal filters of M, whence γ is chain residual. The lemma now follows from the fact that the composition of chain residual mappings is chain residual.

THEOREM 4.2. (1) *If L is a two element chain, then a mapping $\gamma: M \to N$ is L-flat iff it maps 1_M into 1_N.*

(2) *If L is not a two element chain, and M is finite, then $\gamma: M \to N$ is L-flat iff it is isotone and maps 1_M into 1_N.*

Proof. (1) For arbitrary L, let $\gamma: M \to N$ be L-flat. The fact that $\gamma(1_M) = 1_N$ now follows as in the proof of [2], Lemma 6.1 (1), p. 65. If L is a two element chain, the converse is clear.

(2) If $\gamma: M \to N$ is isotone and maps 1_M into 1_N, then by Lemma 4.1, it is chain residual, and consequently L-flat. Suppose now that γ is L-flat. As in the proof of (1), $\gamma(1_M) = 1_N$. In order to show that γ is isotone we note that L must have a chain of the form $0 < h < k$. We must show that for $y < z < 1_M$ in M, $\gamma(y) \leqslant \gamma(z)$. So let us define $C: L \to M$ and $C^*: M \to L$ by the rules

$$C(h_1) = \begin{cases} 1_M, & h_1 \geqslant k, \\ z, & h_1 \geqslant h,\ h_1 \ngeqslant k, \\ y, & h_1 \ngeqslant h, \end{cases}$$

$$C^*(m) = \begin{cases} 0, & m \leqslant y, \\ h, & m \leqslant z,\ m \nleqslant y, \\ k, & m \nleqslant z. \end{cases}$$

Routine verificarion shows that C^* is residuated with C its associated residual mapping. The fact that $\gamma \circ C \in \mathrm{CR}^+(L, N)$ now shows that

$$\gamma(y) = (\gamma \circ C)(0) \leqslant (\gamma \circ C)(h) = \gamma(z),$$

so γ is indeed isotone.

COROLLARY 4.3. *If L is not a two element chain, then every flat L-cluster method is isotone.*

Next we turn to the

SECOND FLAT CLUSTER PROBLEM. *Suppose the cluster method F is given. What properties must F enjoy in order that it be flat?*

In connection with the solution to this problem, it will be useful to say that a cluster method F is *θ-compatible* for $\theta \in \mathrm{Res}^+(L)$ if $F(C \circ \theta) = [F(C)] \circ \theta$ for all $C \in \mathrm{CR}^+(L, M)$. In view of [2], Theorem 7.4, p. 68, we are led to conjecture that the answer will involve F being θ-compatible for *all* $\theta \in \mathrm{Res}^+(L)$. Indeed, we have

THEOREM 4.4. *A necessary and sufficient condition for a cluster method to be flat is that it be θ-compatible for every $\theta \in \mathrm{Res}^+(L)$.*

Proof. The proof of [2], Lemma 7.1, p. 68, also shows that F flat implies that F is θ-compatible for all $\theta \in \mathrm{Res}^+(L)$. We may therefore assume the indicated compatibility condition, and concentrate our efforts on showing F to be flat. The proof will be broken up into a number of steps.

Step 1. Let $\bar{1}_M: L \to M$ be defined by $\bar{1}_M(h) = 1_M$ for all $h \in L$; $\bar{1}_N: L \to N$ is defined in a similar manner. Note that $\bar{1}_M \in \mathrm{Res}^+(L, M)$ and $\bar{1}_N \in \mathrm{Res}^+(L, N)$. We claim that $F(\bar{1}_M) = \bar{1}_N$. For suppose the range of $F(\bar{1}_M)$ contained an element $n \neq 1_N$. Then for some $h \in L$, $[F(\bar{1}_M)](h) = n$. Evidently h cannot be the largest element of L, so there is some element $k > h$. By [2], Lemma 7.2, p. 68, there exists $\delta \in \mathrm{Res}^+(L)$ with $\delta(h) = k$. Then $\bar{1}_M \circ \delta = \bar{1}_M$, so $F(\bar{1}_M) = F(\bar{1}_M \circ \delta) = F(\bar{1}_M) \circ \delta$; consequently, $n = [F(\bar{1}_M)](h) = [F(\bar{1}_M)](\delta(h)) = [F(\bar{1}_M)](k)$. Choosing k so that $[F(\bar{1}_M)](k) = 1_N$ will now produce a contradiction.

Step 2. Let C, $C' \in \mathrm{Res}^+(L, M)$, and suppose that C, C' each have range $\{m, 1_M\}$ for some fixed $m < 1_M$. We claim that $[F(C)](0) = [F(C')](0)$. To establish this, let $C^{-1}([1_M]) = [h)$ and $C'^{-1}([1_M]) = [k)$. Since neither C nor C' is $\bar{1}_M$, we must have $h, k > 0$. We may clearly assume that $h \nleqslant k$ and define δ as in [2], Lemma 7.3, p. 68. Then $C' = C \circ \delta$, so $F(C') = F(C \circ \delta) = F(C) \circ \delta$. Thus $[F(C')](0) = [F(C)](\delta(0)) = [F(C)](0)$.

Step 3. We are now ready to define a mapping $\gamma: M \to N$. We begin by defining $\gamma(1_M) = 1_N$. For $m < 1_M$, we choose $C \in \mathrm{Res}^+(L, M)$ so that its range is $\{m, 1_M\}$. Now take $\gamma(m) = [F(C)](0)$ and recall from Step 2 that this definition is independent of the choice of C. We have yet to show that $F(C) = \gamma \circ C$ for every $C \in \mathrm{CR}^+(L, M)$. We begin with a special case. For arbitrary $h \in L$, $[F(\bar{1}_M)](h) = \bar{1}_N(h) = 1_N = (\gamma \circ \bar{1}_M)(h)$ so the result holds for $\bar{1}_M$.

Step 4. Let $C \in \mathrm{Res}^+(L, M)$ have a 2 element range. By the very definition of γ, it must then be true that

$$[F(C)](0) = (\gamma \circ C)(0).$$

Step 5. *The result for $h > 0$.* Let $C \in \mathrm{CR}^+(L, M)$ with $C \neq \bar{1}_M$. If $m = C(h)$ for $h > 0$, we are to show that $\gamma(m) = [F(C)](h)$. If h is

the largest element of L, there is nothing to prove. Thus we may assume the existence of $k > h$ with $k \in C^{-1}([1_M])$. Then $0 < h < k$ and δ may be chosen as in [2], Lemma 7.2. It is easily verified that $C \circ \delta \in \mathrm{Res}^+(L, M)$ with range $\{m, 1_M\}$ or $\{1_M\}$. Using Step 4, it follows that

$$[F(C)](h) = [F(C)](\delta(0)) = [F(C \circ \delta)](0)$$
$$= (\gamma \circ C \circ \delta)(0) = (\gamma \circ C)(h) = \gamma(m).$$

Step 6. *The result for* $h = 0$. Let C be as in Step 5. Let $h_1 \in C^{-1}([1_M])$ and note that $h_1 > 0$ since $C \neq \bar{1}_M$. Suppose k can be chosen so that $k > 0$ and $h_1 \not\leqslant k$. If δ is defined as in [2], Lemma 7.3 (with h_1 taking the place of h), it follows that $C \circ \delta \in \mathrm{Res}^+(L, M)$ with range $\{C(0), 1_M\}$. By Step 4,

$$[F(C)](0) = [F(C)](\delta(0)) = [F(C \circ \delta)](0) = (\gamma \circ C \circ \delta)(0) = (\gamma \circ C)(0).$$

If no such k can be found, then $h_1 \leqslant k$ for all $k > 0$. It is immediate that $C \in \mathrm{Res}^+(L, M)$ with $C^{-1}([1_M]) = [h_1)$. The result of Step 4 may now be directly invoked.

The results provided by [2], Theorems 6.2 and 7.4 may now be viewed as providing the solution to the First and Second Flat Cluster Problems for *restricted* cluster methods.

5. Semiflat cluster methods

With each set $\mathscr{F}$ of cluster methods, one may associate the set of all residual maps θ on L for which each $F \in \mathscr{F}$ is θ-compatible. Furthermore, for each subset $\mathscr{S}$ of $\mathrm{Res}^+(L)$ one may associate those cluster methods F that are θ-compatible for each $\theta \in \mathscr{S}$. This procedure sets up a galois connection between the subsets of the set of all cluster methods and the subsets of $\mathrm{Res}^+(L)$. It would be interesting to determine the galois closed objects in this correspondence. The characterization of flat cluster methods was a start toward solving this problem. The current section is devoted to the study of a more general galois closed class of cluster methods.

As in [3], a cluster method F is called *semiflat* in case there is a family $(\gamma_m)_{m \in M}$ of mappings of M into N such that $F(C) = \gamma_{C(0)} \circ C$ for all $C \in \mathrm{CR}^+(L, M)$; while a family $(\gamma_m)_{m \in M}$ of mappings of M into N is called *L-semiflat* in case there is a cluster method F for which $F(C) = \gamma_{C(0)} \circ C$ for each $C \in \mathrm{CR}^+(L, M)$ — in other words, if $C \in \mathrm{CR}^+(L, M)$ implies that $\gamma_{C(0)} \circ C \in \mathrm{CR}^+(L, N)$. As in [3], p. 49, it costs nothing to assume that for $x \not\geqslant m$, $\gamma_m(x) = \gamma_m(m)$, so this we now do. The next lemma is then an analogue of [3], Lemma 2.1, p. 49, with a proof so similar that it will be omitted.

LEMMA 5.1. *Let* $(\gamma_m)_{m \in M}$ *be an* L*-semiflat family. Then*

(1) $\gamma_m(1_M) = 1_N$ *for each* $m \in M$.

(2) $\gamma_{1_M}(m) = 1_N$ *for all* $m \in M$.

(3) *If* L *is not a 2 element chain, then* $m \leqslant m'$ *in* M *implies* $\gamma_m(m) \leqslant \gamma_m(m')$ *in* N.

In view of the above lemma, we agree that L shall not be a 2 element chain, and that the family $(\gamma_m)_{m \in M}$ of mappings of M into N will be called *potentially* L*-semiflat* in case:

(1) $\gamma_m(x) = \gamma_m(m)$ for all $x \not\geqslant m$;

(2) $\gamma_m(1_M) = 1_N$ for each $m \in M$;

(3) $\gamma_{1_M}(m) = 1_N$ for all $m \in M$;

(4) $\gamma_m(m) \leqslant \gamma_m(m')$ whenever $m \leqslant m'$ in M.

The content of Lemma 5.1 is that for L not a 2 element chain, *every* L-semiflat family is potentially L-semiflat. We turn now to the analogue of the First Semiflat Cluster Problem ([3], p. 48):

Find necessary and sufficient conditions for a given family of mappings of M *into* N *to be* L*-semiflat.*

The next theorem should be compared with the result provided by [3], Theorem 2.4, p. 52.

THEOREM 5.2. *Assume* M *is finite. Then*

(1) *If* L *has height* 3, *a family* $(\gamma_m)_{m \in M}$ *is* L*-semiflat iff it is potentially* L*-semiflat.*

(2) *If* L *has height greater than* 3, *a family* $(\gamma_m)_{m \in M}$ *is* L*-semiflat iff it is a potentially* L*-semiflat family of isotone mappings.*

Proof. (1) Let $(\gamma_m)_{m \in M}$ be a potentially L-semiflat family of mappings of M into N, and let $C \in \mathrm{CR}^+(L, M)$. We must show that with $m = C(0)$, $\gamma_m \circ C \in \mathrm{CR}^+(L, N)$. Accordingly, let $n \in N$ and consider $\gamma_m^{-1}([n)) = \{m_1, m_2, \ldots, m_k\}$, which is finite since M is finite. For each index i, $C^{-1}([m_i))$ is the union of finitely many principal filters of L.

Case 1. *For some index* i, $C^{-1}([m_i)) = L$. Then $m = C(0) \geqslant m_i$, so $\gamma_m(m) = \gamma_m(m_i) \geqslant n$ and $\gamma_m^{-1}([n)) = M$. It follows that $(\gamma_m \circ C)^{-1}([n)) = C^{-1}(M) = L$.

Case 2. *For each index* i, $C^{-1}([m_i)) \neq L$. The principal filters of L are all of the form L, $\{1_L\}$ or $\{h, 1_L\}$ for some h with $0 < h < 1_L$. It follows that each $C^{-1}([m_i))$ is a finite subset of $L \setminus \{0\}$, and the same is true of $C^{-1}(\{m_i\})$. Hence $C^{-1}(\{m_1, m_2, \ldots, m_k\})$ is a finite subset of $L \setminus \{0\}$. In that $\gamma_m(1_M) = 1_N$, we know that $1_M = m_j$ for some index j, whence $1_L \in C^{-1}(\{m_j\})$. In summary, we have that $(\gamma_m \circ C)^{-1}([n))$ is the union of a finite collection of principal filters.

(2) Assume now that L has height greater than 3 and that $(\gamma_m)_{m\in M}$ is L-semiflat. We must show each γ_m to be isotone. If $s<t$ in M and $m\not\leqslant s$, then $\gamma_m(s)=\gamma_m(m)\leqslant\gamma_m(t)$: if $t=1_M$, we may invoke Lemma 5.1 to obtain $\gamma_m(s)\leqslant\gamma_m(t)$. This leaves us with the case where $m<s<t<1_M$. Using the fact that L height greater than 3, we choose a chain $0<h<k<h_1$ of elements of L, and define $C\colon L\to M$ by the rule

$$C(y)=\begin{cases}1_M, & y\geqslant h_1,\\ t, & y\geqslant k,\ y\not\geqslant h_1,\\ s, & y\geqslant h,\ y\not\geqslant k,\\ m, & y\not\geqslant h.\end{cases}$$

As in the proof of [3], Lemma 6.1, p. 68, we have $C\in\mathrm{Res}^+(L,M)$. It follows that

$$\gamma_m(s)=(\gamma_m\circ C)(h)\leqslant(\gamma_m\circ C)(k)=\gamma_m(t).$$

Now assume $(\gamma_m)_{m\in M}$ is a potentially L-semiflat family of isotone mappings. By Lemma 4.1, each γ_m is chain residual. Hence $C\in\mathrm{CR}^+(L,M)$ implies $\gamma_{C(0)}\circ C\in\mathrm{CR}^+(L,N)$.

Now we dispose of the

Second Semiflat Cluster Problem. *What properties must a cluster method F have in order that there exist a family $(\gamma_m)_{m\in M}$ of mappings of M into such that $F(C)=\gamma_{C(0)}\circ C$ for every $C\in\mathrm{CR}^+(L,M)$?*

In view of Theorem 4.4 and [3], Theorem 3.8, p. 57, we seek a solution that involves compatibility of F with respect to a certain family of residual mappings on L. To begin with, we note that the proof of [3], Lemma 3.1, p. 54, also produces

Lemma 5.3. *Every semiflat cluster method F is θ-compatible for all $\theta\in\mathrm{Res}^+(L)$ such that $\theta(0)=0$.*

Assume now that $F\colon\mathrm{CR}^+(L,M)\to\mathrm{CR}^+(L,N)$ has the property that:

I. If L does not have a largest element, then F is θ-compatible for all $\theta\in\mathrm{Res}^+(L)$ that are *decreasing* in that $\theta(h)\leqslant h$ for all $h\in L$.

II. If L does have a largest element, then F is θ-compatible for all $\theta\in\mathrm{Res}^+(L)$ for which $\theta(0)=0$.

Our goal is to show that F is semiflat. The proof will be broken up into a series of lemmas. For ease of reference, the proof of each of the next 3 lemmas will be divided into 2 cases. In Case I we shall assume that L does not have a largest element, and in Case II that it does.

Lemma 5.4. $F(\bar{1}_M)=\bar{1}_N$.

Proof. Case I. Let $h>0$ and choose a sequence $0<h<k<h_1$ of elements of L. By [3], Lemma 3.2, p. 54, there is a decreasing $\delta\in\mathrm{Res}^+(L)$ such that $\delta(h)=0$. If h is chosen so that $[F(\bar{1}_M)](h)=1_N$, one then has that

$$1_N=[F(\bar{1}_M)](h)=[F(\bar{1}_M)]\big(\delta(h)\big)=[F(\bar{1}_M)](0).$$

Case II. Let $0<h<1_L$. Taking $\delta(y)=0$ for $y\neq 1_L$, $\delta(1_L)=1_L$, one has $\delta\in\mathrm{Res}^+(L)$ with $\delta(h)=0$. Hence

$$[F(\bar{1}_M)](h)=[F(\bar{1}_M)]\big(\delta(h)\big)=[F(\bar{1}_M)](0).$$

On the other hand, we may take $\delta(y)=1_L$ for $y\geqslant h$, $\delta(y)=0$ for $y\not\geqslant h$ to get

$$[F(\bar{1}_M)](h)=[F(\bar{1}_M)]\big(\delta(h)\big)=[F(\bar{1}_M)](1_L)=1_N.$$

Lemma 5.5. *Let $m<1_M$ in M and C, $D\in\mathrm{Res}^+(L,M)$. Then*

(1) *If the range of C is $\{m,1_M\}$ with $C^{-1}\big([1_M)\big)=[h)$, $F(C)\in\mathrm{Res}^+(L,N)$ and has range $\{n,1_N\}$ for some $n\leqslant 1_N$ in N. Furthermore, if $n<1_N$, then $[F(C)]^{-1}\big([1_N)\big)=[h)$.*

(2) *If C, D each have range $\{m,1_M\}$, then $[F(C)](0)=[F(D)](0)$.*

Proof. (1). *Case* I. Choose $h_2>h_1>h$ so that $[F(C)](h_1)=1_N$. Now define $\delta\colon L\to L$ by

$$\delta(y)=\begin{cases}y, & y\geqslant h_2,\\ h, & y\geqslant h,\ y\not\geqslant h_2,\\ 0, & y\not\geqslant h.\end{cases}$$

Then δ is a decreasing residual mapping on L. Since $C=C\circ\delta$, we have $F(C)=F(C)\circ\delta$. Hence

$$[F(C)](y)=\begin{cases}1_N, & y\geqslant h,\\ [F(C)](0), & y\not\geqslant h.\end{cases}$$

Case II. The mapping a_h^+ given by $a_h^+(y)=1_L$ for $y\geqslant h$ and 0 for $y\not\geqslant h$ has the property that $C=C\circ a_h^+$ with $a_h^+(0)=0$. Hence $F(C)=F(C)\circ a_h^+$, so again $[F(C)](y)=1_N$ or $[F(C)](0)$.

(2) Let $C^{-1}\big([1_M)\big)=[h)$, $D^{-1}\big([1_M)\big)=[k)$ and note that $h,k>0$. We may clearly assume that $h\neq k$, since otherwise $C=D$.

Case I. If $h<k$, choose $h_1>k$ and use [3], Lemma 3.2, p. 54, to get a decreasing residual mapping δ such that $\delta(k)=h$ and $\delta(y)\geqslant h$ iff $y\geqslant k$. Then $D=C\circ\delta$ shows that $F(D)=F(C)\circ\delta$ and consequently,

$$[F(D)](0)=[F(C)\circ\delta](0)=[F(C)](0).$$

If $h\not\leqslant k$, choose $h_1>h\vee k$ and apply the above argument to the pair $h,h\vee k$ as well as to the pair $k,h\vee k$.

Case II. Recalling that $a_k^+(y)=1_L$ for $y\geqslant k$ and 0 for $y\not\geqslant k$, we have $D=C\circ a_k^+$, so

$$[F(D)](0)=[F(C)\circ a_k^+](0)=[F(C)](0).$$

LEMMA 5.6. *Let* C, $D \in \mathrm{Res}^+(L, M)$ *each have range* $\{m, s, 1_M\}$ *with* $m < s < 1_M$. *If* $C^{-1}([s]) = [h]$ *and* $D^{-1}([s]) = [k]$, *then* $[F(C)](h) = [F(D)](k)$.

Proof. Assume first that $h = k$. Let $C^{-1}([1_M]) = [h_1]$ and $D^{-1}([1_M]) = [k_1]$. If $k_1 > h_1$, we note that $h_1 > h > 0$ so by [3], Lemma 3.2, p. 54, we may define a residual mapping δ on L by

$$\delta(y) = \begin{cases} y, & y \geqslant k_1, \\ h, & y \geqslant h,\ y \not\geqslant k_1, \\ 0, & y \not\geqslant h. \end{cases}$$

Then δ is decreasing and $D = C \circ \delta$. Consequently,

$$[F(D)](h) = [F(C) \circ \delta](h) = [F(C)](h).$$

For the general situation, the above argument may be applied to the pair $h_1, h_1 \vee k_1$ as well as to the pair $k_1, h_1 \vee k_1$.

We now consider the situation where $h \neq k$. By what has just transpired, we may assume that $h_1 = k_1$.

Case I. If $h < k$, note that $0 < h < k < h_1$. Defining δ as in [3], Lemma 3.2, p. 54, we have δ decreasing, $D = C \circ \delta$, and

$$[F(D)](k) = [F(C) \circ \delta](k) = [F(C)](h).$$

In general, this argument is applied to the pair $h, h \vee k$ and to the pair $k, h \vee k$ with h_1 chosen so that $h_1 > h \vee k$.

Case II. We may again assume that $h_1 = k_1$ and use [3], Lemma 3.2, p. 54, to define $\delta \in \mathrm{Res}^+(L)$ so that $\delta(0) = 0$, $\delta(k) = h$, and $\delta(y) \geqslant h$ iff $y \geqslant k$. We then proceed as in Case I.

LEMMA 5.7. *Let* $C \in \mathrm{CR}^+(L, M)$. *If* $C(h) = 1_M$, *then* $[F(C)](h) = 1_N$.

Proof. If h is the largest element of L, there is nothing to prove. If not, we can find an $h_1 > h$ in L; indeed, h_1 may be chosen so that $h_1 > h$ and $[F(C)](h_1) = 1_N$. We now define $\delta\colon L \to L$ by

$$\delta(y) = \begin{cases} y, & y \geqslant h_1, \\ h, & y \geqslant h,\ y \not\geqslant h_1, \\ 0, & y \not\geqslant h. \end{cases}$$

Then $\delta \in \mathrm{Res}^+(L)$ is decreasing, and if $D = C \circ \delta$, we have

$$D(y) = \begin{cases} 1_M, & y \geqslant h, \\ C(0), & y \not\geqslant h. \end{cases}$$

Thus $D \in \mathrm{Res}^+(L, M)$. Now if $C(0) = 1_M$, we could invoke Lemma 5.4; otherwise, we have $D^{-1}([1_M]) = [h]$. By Lemma 5.5, $[F(D)](h) = 1_N$, so $[F(C)](h) = [F(C) \circ \delta](h) = [F(C \circ \delta)](h) = [F(D)](h) = 1_N$.

Finally we are ready to define a family $(\gamma_m)_{m \in M}$ of mappings of M into N. We begin by taking $\gamma_{1_M}(r) = 1_N$ for all $r \in M$. Suppose $m < 1_M$. We choose $C \in \mathrm{Res}^+(L, M)$ to have range $\{m, 1_M\}$ and define $\gamma_m(m) = [F(C)](0)$. By Lemma 5.5, this definition is independent of the choice of C. For $m < r < 1_M$, we choose $C \in \mathrm{Res}^+(L, M)$ so that the range of C is $\{m, r, 1_M\}$. We then let $\gamma_m(r) = [F(C)](h)$, where $[h] = C^{-1}([r])$. By Lemma 5.6 this definition is independent of the choice of C. Finally, γ_m is extended to all of M by defining $\gamma_m(1_M) = 1_N$ and $\gamma_m(r) = \gamma_m(m)$ if $r \not> m$.

LEMMA 5.8. *For* $C \in \mathrm{CR}^+(L, M)$, $F(C) = \gamma_{C(0)} \circ C$.

Proof. First of all, since $F(\bar{1}_M) = \bar{1}_N$, we have for all $h \in L$ that $[F(\bar{1}_M)](h) = \bar{1}_N(h) = 1_N = (\gamma_{1_M} \circ \bar{1}_M)(h)$, so the result is true for $C = \bar{1}_M$. Now choose $C \in \mathrm{CR}^+(L, M)$ and assume that $C \neq \bar{1}_M$. Let $m = C(0)$ and $r = C(h)$ for some fixed $h \in L$. Choose $h_1 \geqslant h$ so that $C(h_1) = 1_N$ and consider the following cases:

Case 1. $r = 1_M$. Then by Lemma 5.7 $[F(C)](h) = 1_N = \gamma_m(r)$.

THEOREM 5.9. *Let* M, N *be partially ordered sets with* 1. *For a mapping* $F\colon \mathrm{CR}^+(L, M) \to \mathrm{CR}^+(L, N)$, (1) $\Leftrightarrow$ (2) $\Rightarrow$ (3). *If* L *does not have a largest element, then also* (3) $\Rightarrow$ (1).

(1) F *is semiflat.*

(2) F *is* θ*-compatible for all* $\theta \in \mathrm{Res}^+(L)$ *such that* $\theta(0) = 0$.

(3) F *is* θ*-compatible for all decreasing* $\theta \in \mathrm{Res}^+(L)$.

The above theorem completes the solution of the Second Semiflat Cluster Problem for the present version of Model L, while [3], Theorem 3.8, p. 57, may be regarded as a solution for *restricted* cluster methods.

6. Relation to other models

An extremely rich mathematical model for cluster analysis is developed in great detail in [4] by N. Jardine and R. Sibson. Its precise relationship to Model L is explored in [2], and for that reason will not be discussed here. A more recent graph theoretic model has been introduced by D. Matula [5], and it is this model that will now be discussed. In the interest of readability, it will be convenient to briefly introduce the model before showing its relation to Model L.

A *graph* $G = (V, E)$ may be defined to be a nonvoid *vertex* set $V = \{o_1, o_2, \ldots, o_n\}$ and an edge set $E = \{e_1, e_2, \ldots, e_m\}$ where $m \leqslant n(n-1)/2$ and each edge e_k is a distinct pair $o_i o_j$ of elements of V, denoted $e_k = o_i o_j$. If $m = n(n-1)/2$, the graph G is called *complete*. If the set E of edges is linearly ordered, then G is an *ordered graph*. A *proximity graph* $P = (V, E)$ is then an ordered graph where $V = \{o_1, o_2, \ldots, o_n\}$ is a set of

objects to be classified and $E = \{e_1, e_2, \ldots, e_m\}$ is a set of object pairs called *links*, and the order relations on the links is determined by some sort of measure of dissimilarity. To illustrate this concretely, set $V = \{w, x, y, z\}$ and let d be given as follows:

d	w	x	y	z
w		2	1	3
x			5	2
y				3.5
z				

Here d: $V \times V \to \boldsymbol{R}^+$ has the property that the higher the value of $d(a, b)$, the more dissimilar is the pair (a, b). This yields a complete proximity graph $G = (V, E)$, where E has the ordering

$$wy < wx \equiv xz < wz < yz < xy .$$

Notice that $wx \equiv xz$ signifies that $d(w, x) = d(x, z)$. The *splitting levels* of the proximity graph $P = (V, E)$ are the levels $s = 0$, $s = m = |E|$ and all s for which $e_s < e_{s+1}$. In the example at hand, the splitting levels are $0, 1, 3, 4, 5, 6$.

The *classifications* of V are partially ordered as in Section 3. A *stratified clustering* of V is taken to be a sequence $S = (L_0, L_1, \ldots, L_k)$ of classifications, where

(i) $L_0 = \{\{o_1\}, \{o_2\}, \ldots, \{o_n\}\}$,

(ii) For $i < k$, $L_i \leqslant L_{i+1}$.

Matula defines a *level clustering method* to be a mapping γ of the class $\mathscr{P}$ of proximity graphs into the class of classifications of the object sets of the corresponding proximity graphs, and a *stratified clustering method* as a mapping σ of proximity graphs into stratified clusterings; in other words, both $\gamma(G)$ and $\sigma(G)$ are classifications of the object set V of the proximity graph $G = (V, E)$. A condition ([5], pp. 7–8) is then added to guarantee that the cluster methods do not depend upon the nature of the objects being clustered, but depend instead only upon the ordinal proximity relations. Finally, Matula defines a *graphical* level clustering method as a level clustering method that ignores the ordering of the set E of links of $P = (V, E)$.

Let us now see how all of this fits into Model L. Let V be a set of objects to be classified with $|V| = n \geqslant 3$. Take $M =$ the Boolean algebra of all reflexive symmetric relations on V, $N =$ the poset of classifications of V, and $L = \boldsymbol{R}^+$. If we adopt Matula's convention for the splitting levels of a proximity graph, agreeing to take $C(0) = 0_M$, we can associate a unique element of $\mathrm{Res}^+(L, M)$ with each proximity graph $G = (V, E)$. A level clustering method may then be viewed as the transformation of an

increasing sequence $m_0 = 0_M < m_1 < m_2 < \ldots < m_k = 1_M$ of elements of M into an element of N; a stratified clustering method becomes a transformation of the indicated sequence of M into a sequence $n_0 = 0_N < n_1 < \ldots < n_t$ of elements of N. If necessary, we may adjoin $n_{t+1} = 1_N$ to this sequence, so that it may be associated with a unique element of $\mathrm{Res}^+(L, N)$, whose splitting levels are $0, 1, 2, \ldots, t+1$. Finally, a graphical level clustering method is simply a mapping of M into N. Matula's threshold stratified clustering methods then turn out to be flat cluster methods, and his ordinal cluster methods are equivalent to a certain type of monotone equivariant method (see [2]). A definition of these terms may be found in [5], p. 9.

In closing, we mention some of the differences between the models. First of all, Matula allows an imput to be a proximity graph that is not complete. This amounts to considering partially defined residual mappings — a feature that has not been incorporated into Model L. Secondly, his outputs do not necessarily group all of the members of the object set into a single classification at the final stage of the clustering process. But this is done for technical reasons only, and one can simply add 1 more classification to the output in order to achieve this condition. Thirdly, his splitting levels are always nonnegative integers, whereas in Model L, any member of L may serve as a splitting level. The final, and perhaps the most important, difference between the models is in their viewpoints. The graph theoretic nature of Matula's model makes it natural to consider clustering techniques that are defined in terms of various connectivity and separation properties of graphs, while the order theoretic flavor of Model L seems to lead one to quite different types of methods. Despite these differences, the cluster methods studied by Matula in [5] may for the most part be viewed as monotone equivariant methods within Model L, with $L = \boldsymbol{R}^+$.

References

[1] T. S. Blyth and M. P. Janowitz, *Residuation theory*, Pergamon Press, 1972.

[2] M. F. Janowitz, *An order theoretic model for cluster analysis*, SIAM J. Appl. Math. 34 (1978), 55–72.

[3] —, *Semiflat L-cluster methods*, Discrete Math. 21 (1978), 47–60.

[4] N. Jardine and R. Sibson, *Mathematical taxonomy*, Wiley, 1973.

[5] D. W. Matula, *Graph theoretic techniques for cluster analysis algorithms*, Southern Methodist University, Tech. Report CS 76015 (1976).

Presented to the Semester
Universal Algebra and Applications
(February 15 – June 9, 1978)