

On the consensus of closure systems Bruno Leclerc

▶ To cite this version:

Bruno Leclerc. On the consensus of closure systems. pp.11, 2004. hal-00018694

HAL Id: hal-00018694 https://hal.science/hal-00018694

Submitted on 7 Feb 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On the consensus of closure systems

Bruno Leclerc*

Abstract

The problem of aggregating a profile of closure systems into a consensus system has applications in domains such as social choice, data analysis, or clustering. We first briefly recall the results obtained by a lattice approach and observe that there is some need for a finer approach. Then, we develop some considerations based on implications and related notions, and present a uniqueness result. It appears to be a generalization of a previous result relevant from cluster analysis.

Key words: Closure, Implication, Lattice, Consensus

1 Introduction

We consider the problem of aggregating a profile (k-tuple) $F^* = (F_1, F_2, ..., F_k)$ of closure systems on a given finite set S into a consensus closure systems $F = c(F^*)$. The aim is, for instance, to find a structure on a set S described by variables of different types. Structural information (order, tree structure) provided by these variables may be totally or partially retained by a derived closure system (see examples in Section 2). Moreover, several consensus problems already studied in the literature are particular cases of the consensus of closure systems. A basic example is provided by hierarchical classification, where many works have followed those of Adams [Ada72] and Margush and McMorris [MM81] (see the survey [Lec98]).

Closure systems and their uses are presented in Section 2.1. Several equivalent structures are recalled in Section 2.2. Section 2.3 give elements about the involved lattice structures. Section 3 presents results provided by the particularization of general results on the consensus problem in lattices. An original approach based on implications is initiated in Section 4.

^{*} École des Hautes Études en Sciences Sociales, Centre d'Analyse et de Mathématique Sociales (CNRS UMR 8557), 54 bd Raspail, F-75270 Paris cedex 06, France, leclerc@ehess.fr

2 Closure systems

2.1 Definitions and uses

A *closure system* (abbreviated as CS) on a finite given set S is a set $F \subseteq P(S)$ of subsets of S satisfying the following two conditions:

(C1) $S \in \mathbb{F};$

(C2) $C, C' \in \mathbb{F} \Rightarrow C \cap C' \in \mathbb{F}.$

When considering classical types of preference or classification data describing a given set *S* of objects, one observes that, frequently, they naturally correspond to closure systems. A list, of course not limitative, of such situations is given in Table 1. A CS F is *nested* if it is linearly ordered by set inclusion: for all $F, F' \in F \Rightarrow F \cap F' \in \{F, F'\}$; it is a *tree of subsets* if, for all $F, F' \in F \Rightarrow F \cap F' \in \{\emptyset, F, F'\}$ (and *hierarchical* if, moreover, $\{s\} \in F$ for all $s \in S$); it is *distributive* if, for all $F, F' \in F \Rightarrow F \cap F' \in F$ and $F \cup F' \in F$.

Type of data	S endowed with a	Subsets of S	Type of closure system
Numerical, ordinal variable	Weak order W	Down-sets of <i>W</i>	Nested
Transitive preference relation	Preorder P	Down-sets of <i>P</i>	Distributive
Nominal variable	Partition Π	$S, \emptyset, $ and classes of Π	Tree of subsets of length 2
Taxonomy	Hierarchy H	Classes of H, and \varnothing	Hierarchical

Table 1. Types of data and related closure systems

2.2 Equivalent structures

Three notions are defined in this section. Together with CS's, they turn to be equivalent to each other. A *closure operator* φ is a mapping onto $\mathbb{P}(S)$ satisfying the properties of isotony (for all $A, B \subseteq S, A \subseteq B$ implies $\varphi(A) \subseteq \varphi(B)$, extensivity ((for all $A \subseteq S, A \subseteq \varphi(A)$ and idempotence (for all $A \subseteq S, \varphi(\varphi(A)) = \varphi(A)$. Then, the elements of the image $\mathbb{F}_{\varphi} = \varphi(\mathbb{P}(S))$ of $\mathbb{P}(S)$ by φ are the *closed* (by φ) sets, and \mathbb{F}_{φ} is a closure system on *S*. Conversely, the closure operator $\varphi_{\mathbb{F}}$ on $\mathbb{P}(S)$ is given by $\varphi_{\mathbb{F}}(A) = \bigcap \{F \in \mathbb{F}: A \subseteq F\}$.

A *full implicational system*, denoted hereafter by $I, \rightarrow_I \text{ or simply} \rightarrow$, is a binary relation on $\mathbb{P}(S)$ satisfying the following conditions:

- (I1) $B \subseteq A$ implies $A \to B$;
- (I2) for any A, B, $C \subseteq S$, $A \rightarrow B$ and $B \rightarrow C$ imply $A \rightarrow B$;
- (I3) for any A, B, C, $D \subseteq S$, $A \to B$ and $C \to D$ imply $A \cup C \to B \cup D$.

An overhanging order on S is also a binary relation \times on $\mathbb{P}(S)$, now satisfying:

$$(O1) \qquad A \times B \Rightarrow A \subset B;$$

$$(O2) \qquad A \subset B \subset C \implies [A \times C \iff A \times B \text{ ou } B \times C];$$

$$(O3) \qquad A \times A \cup B \Rightarrow A \cap B \times B.$$

It follows from (01) and (02) that the relation \times is a strict order on P(S) (whereas \rightarrow is a preorder). The sets of, respectively, closure systems, closure operators, full implicational systems and overhanging orders on S are denoted, respectively, as **M**, **C**, **I** and **O**. They are related to each other by one-to-one correspondences. The equivalence between closure systems and operators has been recalled above. For a closure operator ϕ and its associated full implicational system \rightarrow and overhanging order \times , the first of the equivalences below is due to Armstrong [Arm74], and the second is given in [DL04a]:

$$A \to B \iff B \subseteq \varphi(A)$$

< n

$$A \times B \iff A \subset B \text{ and } \varphi(A) \subset \varphi(B)$$

There is an important literature, with meaningful results, on implications, due to their importance in domains such as logic, lattice theory, relational databases, knowledge representation, or latticial data analysis (see the survey [CM03]). Overhanging orders

take their origin in Adams ([Ada86]), where, named *nestings*, they were characterized in the particular case of hierarchies. Their generalization to all closure systems [DL04a] make them a further tool for the study of closure systems.

2.3. Lattices

The results of this section may be found in [CM03] and, for overhangings, in [DL03] and [DL04a]. First, each of the sets **M**, **C**, **I** and **O** is naturally ordered: **M**, **I** and **O** by inclusion, and **C** by the pointwise order: for $\varphi, \varphi' \in \mathbf{C}, \varphi \leq \varphi'$ means that $\varphi(A) \subseteq \varphi'(A)$ for any $A \subseteq S$. These orders are isomorphic or dually isomorphic:

$$\mathbb{F} \subseteq \mathbb{F}' \Longleftrightarrow \varphi' \leq \varphi \Longleftrightarrow I' \subseteq I \Longleftrightarrow \mathbb{C} \subseteq \mathbb{C}',$$

where φ , *I* and \oplus (resp. φ' , *I* and \oplus') are the closure operator, full implication system and overhanging relation associated to F (resp. to F ').

The sets **M** and **I** preserve set intersection, while **O** preserves set union and **C** pointwise intersection $(\phi \land \phi'(A) = \phi(A) \cap \phi'(A)$, for any $A \subseteq S$. The main correspondences between elements or operations in **M**, **C**, **I** and **O** are given in Table 2. Here, $(A] = \{B \subseteq S: B \subseteq A\}$ (prime ideal in P(S))and $[A] = \{B \subseteq S: A \subseteq B\}$ (prime filter).

From these observations, **M** and **I** are closure systems, respectively on $\mathbb{P}(S)$ and $(\mathbb{P}(S))^2$. The closure operator associated to **M** is denoted as Φ . It is well-known that, with the inclusion order, any closure system F on S is a lattice (F, \lor, \cap) with the meet $F \cap F'$ and the join $F \lor F' = \phi(F \cup F')$. If $F \subseteq F'$, F' covers F (denoted as $F \prec F'$) if $F \subseteq G \subseteq F'$ implies G = F or G = F'.

An element J of F is *join irreducible* if $G \subseteq F$ and $J = \bigvee G$ imply $J \in G$; an equivalent property is that J covers exactly one element, denoted J⁻, of F. The set of all the join irreducibles is denoted by J. Setting $J(F) = \{J \in J: J \subseteq F\}$ for any $F \in F$, one has $F = \bigvee J(F)$ for all $F \in F$. A join irreducible is an *atom* if it covers the minimum element of F, and the lattice F is *atomistic* if all its join irreducibles are atoms. Similarly, an element M of F is *meet irreducible* if $G \subseteq F$ and $M = \bigcap G$ imply $M \in G$; equivalently, M is covered by exactly one element M^+ of F. For any $F \in F$, we have $F = \bigcap M(F)$, where M is the set of all the meet irreducibles of F and $M(F) = \{M \in M: F \subseteq M\}$.

М	С	I	0
P(S) (maximum)	$\varphi_{\min} = id_{P(S)}$ (minimum)	$\{(X, Y) \in \mathbb{P}(S)^2: Y \\ \subseteq X\} (\text{minimum})$	$\{(X, Y) \in \mathbb{P}(S)^2:$ $X \subset Y\}$ (maximum)
{ <i>S</i> } (minimum)	$\varphi_{\max}(A) = S, \text{ all } A$ $\subseteq S$ (maximum)	$\mathbb{P}(S)^2$ (maximum)	Ø (minimum)
join M∨M'	meet (pointwise intersection)	meet I∩I'	join Œ∪Œ'
meet M∩M'	join	join IvI'	meet Œ^Œ'
$\{S, A\}, A \subset S$ (join irreducible)	$\varphi(X) = S$ otherwise	$\mathbb{P}(A)^{2} \cup \{(X, Y) \in \mathbb{P}(S)^{2} \colon A \not\subseteq X\}$ (meet irreducible)	$\{(X, Y) \in (A] \times$ $(\mathbb{P}(S)-(A]): X \subset Y\}$ $(join irreducible)$
$\{X \subseteq S: A \subseteq X \Rightarrow s \\ \in X\}, A \subset S, s \in S-A$ (meet irreducible)	$\varphi(X) = X + s \text{ if } A \subseteq X$ $\varphi(X) = X \text{ otherwise}$ (join irreducible)	$\subseteq Y \text{ or } A \subseteq X, Y =$	$\{(X, Y) \in \mathbb{P}(S)^2: X$ $\subset Y\} - \{(X, Y) \in \mathbb{P}(S)^2: A \subseteq X, Y = X + s\}$ (meet irreducible)

Table 2. Correspondences between M, C, I and O

The lattice F is *lower semimodular* if, for every $F, F' \in F, F \prec F \lor F'$ and $F' \prec F \lor F'$ imply $F \cap F' \prec F$ and $F \cap F' \prec F$. The lattice F is *ranked* if it admits a numerical *rank function* r such that $F \prec F'$ implies r(F') = r(F) + 1. Lower semimodular lattices are ranked.

The lattice F is a *convex geometry* if it satisfies one of the following equivalent conditions (among many other characterizations [Mon90b]:

(CG1) For any $F \in F$, there is a unique minimal subset R of J such that $F = \lor R$;

(CG2) F is ranked with rank function r(F) = |J(F)|;

(CG3) F is lower semimodular with a rank function as in (CG2) above;

Since it is a closure system on P(S), the ordered set **M** is itself a lattice. This lattice is an atomistic convex geometry. For $F \in M$, we have $J(F) = \{\{A, S\}: A \neq S, A \in F\}$ and, so, |J(F)| = |F|-1.

3 Lattice consensus for closure systems

In this section, we consider the main consequences of the lattice structure of **M** for the consensus problem on closure systems, that is the aggregation of a profile $F^* = (F_1, F_2, ..., F_k)$ (of length k) of closure systems into a closure system $F = c(F^*)$. General results on the consensus problem in lattices may be found, among others, in [BM90b], [BJ91] and [Lec94]. Concerning closure systems, the results obtained in an axiomatic approach by Raderanirina [Rad01] (see also [MR04] about the related case of choice functions) are described in another contribution and not recalled here.

3.1. A property of quota rules

A *federation* on *K* is a family K of subsets of $K = \{1, ..., k\}$ satisfying the monotonicity property: $[L \in K, L' \supseteq L] \Rightarrow [L' \in K]$. Then, the federation consensus function c_K on **M** is associated to K by $c_K(F^*) = \bigvee_{L \in K} (\bigcap_{i \in L} F_i)$. Such consensus function includes the *quota rules*, where $K = \{L \subseteq K: |L| \ge q\}$, for a fixed number $q, 0 \le q \le k$. The quota rule c_q is equivalently defined as:

$$c_q(\mathbb{F}^*) = \mathbf{\Phi}(\mathbb{A}_q),$$

where $\mathbb{A}_q = \{A \subset S: |\{i \in K: A \in \mathbb{F}_i\}| \ge q\}$, the set of all proper subsets of *S* appearing in at least *q* elements of \mathbb{F}^* , and Φ is the operator mentioned in Section 2.3: $\Phi(\mathbb{A}_q)$ is the smallest closure system including \mathbb{A}_q . For q = k/2, $c_q(\mathbb{F}^*) = m(\mathbb{F}^*)$ is the so-called (weak) *majority rule* and, for q = k, it is the *unanimity rule* $u(\mathbb{F}^*)$.

Quota rules have good properties in any lattice structure, for instance:

- Unanimity : for any $F \in \mathbf{M}$, $c_q(F, F, \dots, F) = F$;
- *Isotony*: for any F* = (F₁, F₂, ..., F_k), F'* = (F'₁, F'₂, ..., F'_k), profiles of M, F_i, ⊆ F'_i: for all i = 1, ..., k implies c_q(F*) ⊆ c_q(F'*).

The next property of consistency type (see Section 3.2) is not general (for instance it is not true in partition lattices [BL95]) but holds in the so-called LLD lattices [Lec03],

which include convex geometries; it implies unanimity. In what follows, the profile $F^*F'^*$ is just the concatenation of profiles F^* and F'^* , which are not required to have the same length.

Proposition 3.1. Let \mathbb{F}^* and \mathbb{F}'^* be two profiles of \mathbb{M} . If $c_q(\mathbb{F}^*) = c_q(\mathbb{F}'^*) = \mathbb{F}$, then $c_q(\mathbb{F}^*\mathbb{F}'^*) = \mathbb{F}$.

3.2. Bounds on medians

For a metric approach of the consensus in **M**, we first have to define metrics. For that, we just follow [BM81] and [Lec94]. A real function v on **M** such as $F \subseteq F'$ implies v(F) < v(F') is a *lower valuation* if it satisfies one the following two equivalent properties:

(LV1) For all $s, t \in L$ such that $s \lor t$ exists, $v(s) + v(t) \le v(s \lor t) + v(s \land t)$;

(LV2) The real function d_v defined on \mathbf{M}^2 by the following formula is a metric on \mathbf{M} :

$$d_{\mathcal{V}}(\mathbb{F}, \mathbb{F}') = \mathcal{V}(\mathbb{F}) + \mathcal{V}(\mathbb{F}') - 2\mathcal{V}(\mathbb{F} \cap \mathbb{F}').$$

A characteristic property of lower semimodular semilattices is that their rank functions are lower valuations. So, taking property (CG2) into account, the *rank metric* is obtained taking v(F) = |F| and, so, $d_v(F, F') = \partial(F, F') = |F\Delta F'|$, where Δ is the symmetric difference on subsets. The equality between the rank and the symmetric difference metric is characteristic of convex geometries or of close structures [Lec03] and is a reason to focuse on that metric.

Given the metric ∂ , the *median* consensus procedure consists of searching for the *medians* of the profile F*, that is the elements F^µ of **M** minimizing the *remoteness* $\rho(F^{µ}, F^*) = \sum_{1 \le i \le k} \partial(F^{µ}, F_i)$ (see [BM81]). If $\mu(F^*)$ is the set of all the medians of the profile F*, the median procedure has a *consistency* type property (YL78), described as follows.

Let F* and F'* be two profiles of M. If $\mu(F^*) \cap \mu(F'^*) \neq \emptyset$, then $\mu(F^*F'^*) = \mu(F^*) \cap \mu(F'^*)$. Set $\mathbf{J}_m = \{\{A, S\}: A \in \mathbb{A}_{k/2}\}$ (the set of majority atoms). From results in [Lec94], every median M of F* is the join of some subset of \mathbf{J}_m . It then follows that every median CS is included into the majority rule one:

Theorem 3.2. For any profile \mathbb{F}^* and for any median \mathbb{F}^{μ} of \mathbb{F}^* , the inclusion $\mathbb{F}^{\mu} \subseteq m(\mathbb{F}^*)$ holds.

4 The fitting of overhangings

The results of the previous section (and those of the same type mentioned there) only take into account the presence or absence of the same closed set in enough (oligarchies, majorities) or all (unanimity) elements of the profile. It was observed in the literature that this strong limitation may prevent one to recognize actual common features. This criticism, essentially pointed out in the classification context, remains valid for general closure systems [DL04b]. Moreover, consensus systems based on closet sets may frequently be trivial. For instance, if there does not exist any majority non-trivial closed set, then, the majority rule (and unique median) is the trivial closure system reduced to $\{S\}$. Adams [Ada86] presented a consensus method (for hierarchies) able to retain common features even in such cases. It is based on overhanging orders (and, then, on implications). Here we initiate the same approach for closure systems of any type.

We state here a very general uniqueness result. Let Ξ be a binary relation on $\mathbb{P}(S)$, with the only assumption that $(A, B) \in \Xi$ implies $A \subset B$. Consider the following two properties for a closure system F, with associated closure operator φ and overhanging relation \mathbb{E} :

(A Ξ 1) $\Xi \subseteq \oplus$; (preservation of Ξ)

(qualified overhangings)

(A Ξ 2) for all $M \in M_{\mathbb{F}}, (M, M^+) \in \Xi$.

Note that (A Ξ 2) is a weakening of the converse of (A Ξ 1) since, by definition, any pair (M, M^+) belongs to \times . Only those special pairs are required to already belong to the relation Ξ .

Theorem 4.1. If both F and F' satisfy Conditions (A Ξ 1) and (A Ξ 2), then F = F'.

Proof. Observe first that the maximum set S is in both F and F'. If $F \neq F'$, The symmetric difference $F\Delta F'$ is not empty. Let F be a maximal element of $F\Delta F'$. This subset F is not equal to S and it may be assumed without loss of generality that F belongs to F (and, so, not to F'). If F was not a meet-irreducible element of F, it would be an intersection of meet-irreducibles, all belonging to both F and F' and, so, F would belong to F'.

Thus, *F* is a meet-irreducible, covered by a unique element F^+ of F, with $F^+ \in F'$. By (A $\equiv 2$), $(F, F^+) \in \Xi$ and, by (A $\equiv 1$), $F \oplus F^+$. Set $F' = \varphi'(F)$. We have $F \subset F'$, since $F \notin F'$, and $F' \oplus F'$, since $F' = \varphi'(F) = \varphi'(F') \subset \varphi'(F^+) = F^+$. But, according to the hypotheses, $F \subset F'$ implies $F' \in F$, with $F \subset F' \subset F^+$, a contradiction with the

hypothesis that F^+ covers F in F.

The following question then arises: given a binary relation Ξ on P(S) (implying strict inclusion), does it exist an overhanging relation \mathbb{E} satisfying conditions (A Ξ 1) and (A Ξ 2). Adams provides a positive answer in the case of a profile of hierarchies, and with $\Xi = \bigcap_{i \in K} \mathbb{E}_i$ where $\mathbb{E}_1, \mathbb{E}_2, ..., \mathbb{E}_k$ are the overhanging orders associated to the elements of the profile. In the general case, one can consider any convenient combination of $\mathbb{E}_1, \mathbb{E}_2, ..., \mathbb{E}_k$. For instances, $\Xi = \bigcap_{i \in K} \mathbb{E}_i$ corresponds to a kind of unanimity rule on overhangings, and $\Xi = \bigcup_{L \subseteq K, 2|L| > k} \bigcap_{i \in L} \mathbb{E}_i$ to a majority rule.

5 Conclusion

The last section provides a framework for the consensus of closure systems. One of the main questions is to recognize the binary relations Ξ on $\mathbb{P}(S)$ for which an overhanging order \mathbb{E} satisfying (A Ξ 1) and (A Ξ 2) exists. For instance, setting $\Xi = \bigcup_{L \subseteq K, \ 2|L| > k} \bigcap_{i \in L} \mathbb{E}_i$ accounts for the fact that a CS appears several times in a profile, contrary to intersection rules. Algorithmic issues are very important, since overhanging relations are very big objects.

References

[Ada72] E.N. Adams III (1972), Consensus Techniques and the Comparison of Taxonomic Trees, *Syst. Zool.* 21, 390-397.

[Ada86] Adams III, E.N., 1986. N-trees as nestings: complexity, similarity and consensus. Journal of Classification 3, 299–317.

[Arm74] Armstrong, W.W., 1974. Dependency structures of data base relationships. Information Processing 74, 580–583.

[BJ91] J.P. Barthélemy, M.F. Janowitz (1991), A formal theory of consensus, *SIAM J. Discr. Math.* 4, 305-322.

[BL95] J.P. Barthélemy, B. Leclerc (1995), The median procedure for partitions, *in* I.J. Cox, P. Hansen, and B. Julesz, eds., *Partitioning data sets*, *DIMACS Series in Discrete Mathematics and Theoretical Computer Science 19*, Amer. Math. Soc.,

Providence, RI, pp. 3-34.

[BM81] J.P. Barthélemy, B. Monjardet (1981), The Median Procedure in Cluster Analysis and Social Choice Theory, *Math. Soc. Sci.* 1, 235-268.

[CM03] Caspard, N., Monjardet, B., 2003. The lattices of Moore families and closure operators on a finite set: a survey. Discrete Applied Math., 127 (2), 241–269.

[DM03] W.H.E. Day, F.R. McMorris (2003), *Axiomatic Consensus Theory in Group Choice and Biomathematics*, Frontiers in applied mathematics **29**, SIAM, Philadelphia.

[DL03] F. Domenach, B. Leclerc (2003), The structure of the overhanging relations associated with some types of closure systems, submitted.

[DL04a] F. Domenach, B. Leclerc (2004), Closure systems, implicational systems, overhanging relations and the case of hierarchical classification, *Mathematical Social Sciences*, 47 (3), 349-366.

[DL04b] F. Domenach, B. Leclerc (2004), Consensus of classification systems, with Adams' results revisited, *in* D. Banks, L. House, F.R. McMorris, P. Arabie, W. Gaul (editors), *Classification, Clustering and Data Mining Applications*, Berlin, Springer, pp. 417-428.

[Lec94] B. Leclerc (1994), Medians for weight metrics in the covering graphs of semilattices, *Discrete Applied Math.* 49, 281-297.

[Lec98] B. Leclerc (1998), Consensus of classifications: the case of trees, in Advances in Data Science and Classification (A. Rizzi, M. Vichi, H.-H. Bock, eds), Studies in Classi-fication, Data Analysis and Knowledge Organization, Berlin, Springer-Verlag, 81-90.

[Lec03] Leclerc, B., 2003. The median procedure in the semilattice of orders. Discrete Applied Math., 127 (2), 285–302.

[MM81] T. Margush, F.R. Mc Morris (1981), Consensus *n*-trees, *Bull. Mathematical Biology* 43 (2), 239-244.

[Mon90b] B. Monjardet (1990a), Comments on R.P. Dilworth's Lattices with unique irreducible decompositions, in K. Bogart, R. Freese, J. Kung, eds. The Dilworth theorems; Selected works of Robert P. Dilworth, Birkhaüser, pp. 192-201.

[MR04] B. Monjardet, V. Raderanirina (2004), Lattices of choice functions and consensus problems, *Social Choice and Welfare*, to appear.

[Rad01] Raderanirina, V., 2001. Treillis et agrégation de familles de Moore et de fonctions de choix, Ph.D thesis, Université Paris 1, Paris.

[YL78] H.P. Young, A. Levenglick (1978), A Consistent Extension of Condorcet's Election Principle, *SIAM Journal on Applied Mathematics* 35, 285-300.