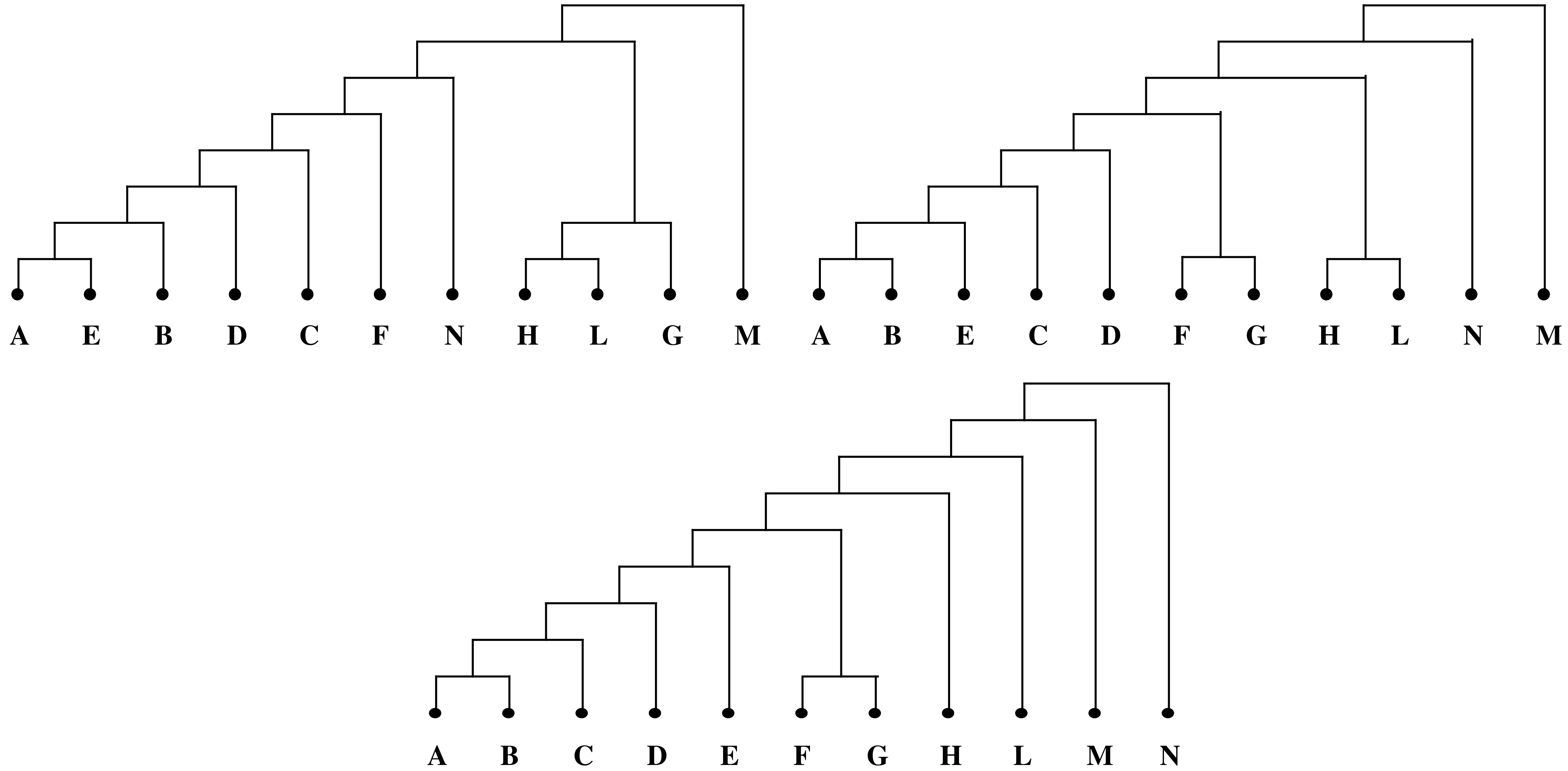
The fitting and consensus of closure systems

Bruno Leclerc

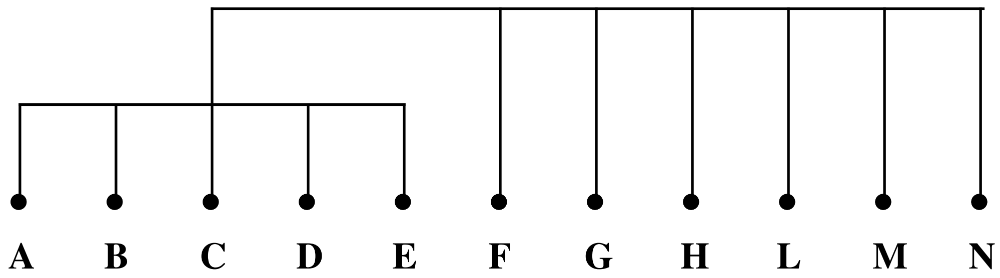
École des Hautes Études en Sciences Sociales
Centre d'Analyse et de Mathématique Sociales
54 boulevard Raspail
75270 Paris cedex 06
leclerc@ehess.fr

with Florent Domenach (Intercollege, Nicosia, Cyprus)

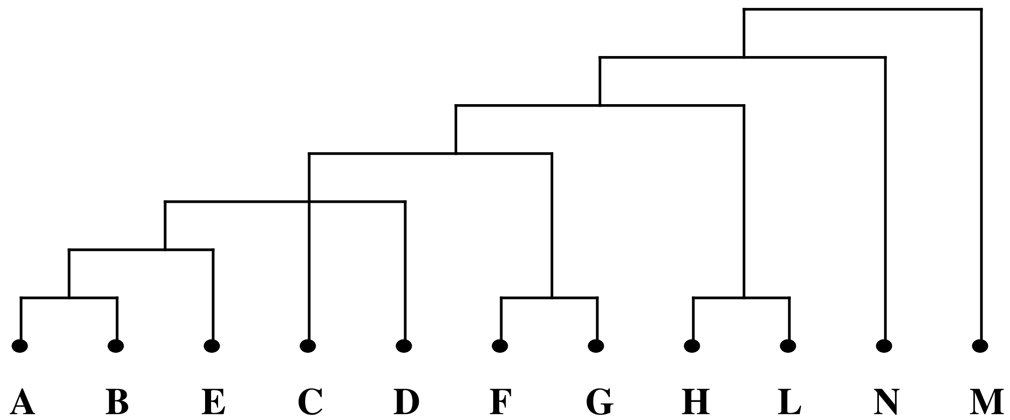
An example (from Day 1983, *Math. Biosciences*, after Johnson and Selander 1971, Schnell, Best and Kennedy 1978):
k = 3 trees on 11 species of kangaroo rats



Strict consensus (retain classes present in all trees):

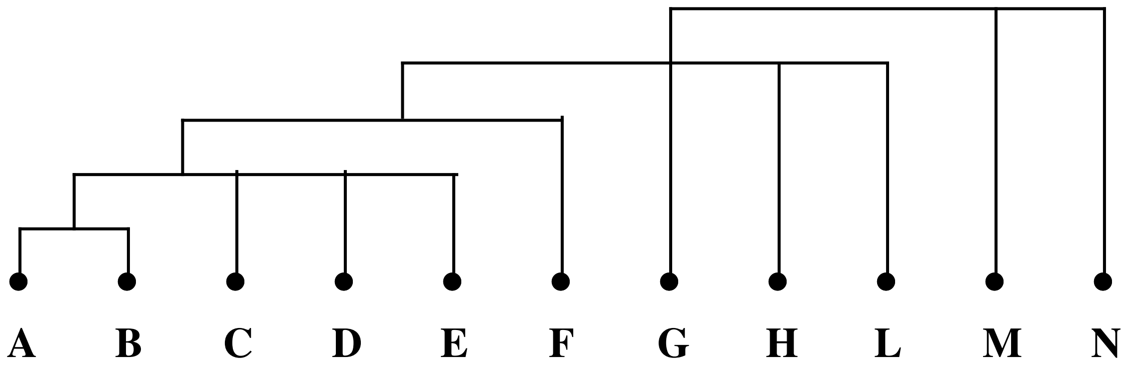


Quota rule consensus (retain classes present in σ trees): σ = 2

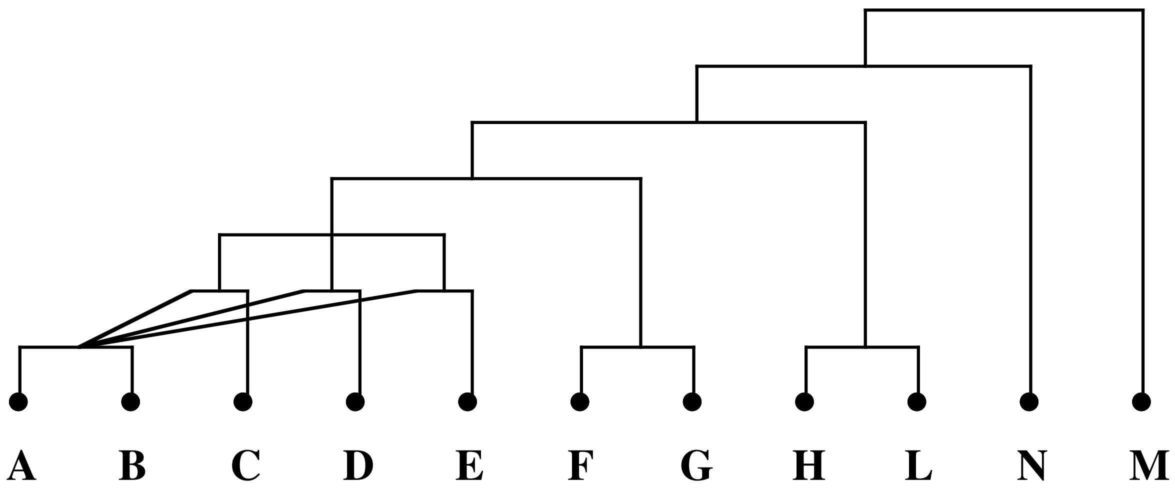


The output is a tree for $\sigma > k/2$ (Margush and McMorris 1981)

Adam's consensus



Frequent subclasses consensus



Closure system

Finite set S (objects to choose, to classify, …)

family $\mathcal{C} \subseteq 2^S$ *of subsets* of S satisfying:

(i) $S \in \mathcal{C}$ (universal set)

(ii) $C, C' \in \mathcal{C} \;\Rightarrow\; C \cap C' \in \mathcal{C}$

Then, $\mathcal{C}$ is a ***closure system*** (CS), or a *Moore family* on S.

The associated closure operator $\varphi_{\mathcal{C}}$ on 2^S:

$$\varphi_{\mathcal{C}}(A) = \bigcap\{C \in \mathcal{C}: A \subseteq C\}$$

Example: after using a classification procedure on a set S of objects to classify, one often gets a *set of classes* $\mathcal{C}$ satisfying (i), (ii) and:

(iii) $s \in S \;\Rightarrow\; \{s\} \in \mathcal{C}$ (individual classes)

Elements of a class $C \in \mathcal{C}$ ought to be *similar* or *sharing common properties*

Lattice structure of a closure system $\boldsymbol{C}$ on S, ordered by inclusion

for $C, C' \in \boldsymbol{C}$,

meet $\quad C \cap C'$

join $\quad C \vee C' = \varphi_{\boldsymbol{C}}(C \cup C')$

- covering relation p

- join-irreducible $J \in \mathcal{J}_{\boldsymbol{C}}$. For any $C \in \boldsymbol{C}$,

$$C = \vee\{J \in \mathcal{J}_{\boldsymbol{C}}: J \subseteq C\} = \vee\mathcal{J}(C)$$

(full join irreducible representation)

- meet-irreducible $M \in \mathcal{M}_{\boldsymbol{C}}$. For any $C \in \boldsymbol{C}$,

$$C = \cap\{M \in \mathcal{M}_{\boldsymbol{C}}: C \subseteq M\} = \cap\mathcal{M}(C)$$

Types of closure systems

Distributive CS: $C, C' \in \mathcal{C} \Rightarrow C \cup C' \in \mathcal{C}$,

Tree of subsets: $C, C' \in \mathcal{C} \Rightarrow C \cap C' \in \{\emptyset, C, C'\}$,

(a tree completed with the empty set)

Nested CS: $C, C' \in \mathcal{C} \Rightarrow C \cap C' \in \{C, C'\}$.

(both tree and distributive)

Convex geometry: every element of $\mathcal{C}$ has a unique irredundant (minimal) join-irreducible representation.

Combinatorial geometry (matroid) and so on…

Obtaining closure systems (1)

Data

Type of variable v	**Structure of domain D of v**	**Subsets of S**	**Type of closure system**
Numerical, ordinal	Linear order	$\{s \in S: v(s) \leq \alpha\}, \alpha \in D$	Nested
		Intervals of D	Convex geometry
Nominal	Finite set $D = \{v_1, \ldots, v_k\}$	$\{s \in S: v(s) = v_i\}$	Tree of subsets
Multicriterion evaluation	Product of linear orders	$\{s \in S: v(s) \leq \alpha\}, \alpha \in D$	Distributive
Taxonomic	Rooted tree	$\{s \in S: v(s) \leq \alpha\}, \alpha \in D$	Tree of subsets

Obtaining closure systems (2)

Choice models

W complete ordering (weak order) on S

for $s \in S$, $Ws = \{s' \in S\colon (s', s) \in S\}$ (elements at least as good as s),

then, $\{Ws\colon s \in S\}$ is a nesting family on S

Classification models

Hierarchy $\mathcal{H}$ on S,

(H1) $S \in \mathcal{H}$,

(H2) $s \in S \;\Rightarrow\; \{s\} \in \mathcal{H}$,

(H3) $H, H' \in \mathcal{H} \;\Rightarrow\; H \cap H' \in \{\emptyset, H, H'\}$,

then $\mathcal{H} \cup \{\emptyset\}$ is a *hierarchical classification system*.

Others: *pyramids*, *weak hierarchies*, …

(Galois) lattices

Databases, Association rules mining,…

The lattice structure of $\mathbf{M}$

Let $\mathbf{M}$ be the set of all closure systems on S;

- $2^S \in \mathbf{M}$,
- for C, $C' \in \mathbf{M}$, $C \cap C' \in \mathbf{M}$

So, $\mathbf{M}$ is a closure system on 2^S.

- For any family $\mathcal{F}$ of subsets of S, is there is a smallest CS $\Phi(\mathcal{F})$ including $\mathcal{F}$ (make all intersections of subsets of $\mathcal{F}$ comprising $S = \cap\varnothing$)
- Join-irreducibles of $\mathbf{M}$ are closure systems $\{A, S\}$, with a unique (proper) closed subset $A \subset S$.
  Then, for $C \in \mathbf{M}$,

$$C = \Phi(\mathcal{F}) \iff \mathcal{M}_C \subseteq \mathcal{F}$$

- So, $\mathbf{M}$ is a ***convex geometry*** (lower locally distributive) on 2^S.

Consensus of closure systems

searching a consensus function f

$$\mathbf{M}^k \xrightarrow{f} \mathbf{M}$$

(aggregation of a *profile* $C^* = (C_1, C_2, \ldots, C_k) \in \mathbf{M}^k$ of CS's into a *unique* CS)

So, we can apply results on the consensus problem

- in lattices (Monjardet 1990, Barthélemy and Janowitz 1991, L. 1994, and others)

- particularly, in convex geometries (Raderanirina 2001, L. 2003)

Median consensus

Given a metric d on $\mathbf{M}$, find a ***median*** $C^\mu \in \mathbf{M}$ such that

$$\rho(C^\mu, C^*) = \sum_{1\leq i\leq k} d(C^\mu, C_i) \rightarrow \min$$

- often difficult to compute,
- not necessarily unique,
- satisfies Young's ***consistency***: for $C^* \in \mathbf{M}^k$, $C'^* \in \mathbf{M}^{k'}$,

$$\mu(C^*)\cap\mu(C'^*) \neq \emptyset \Rightarrow \mu(C^*C'^*) = \mu(C^*)\cap\mu(C'^*),$$

where $\mu(C^*)$ is the set of the medians of C^*
$C^*C'^* \in \mathbf{M}^{k+k'}$ is the concatenation of C^* and C'^*.

- Problem: do medians satisfy

$$\bigcap_{1\leq i\leq k} C_i \subseteq C^\mu$$

(a *unanimity property*: does C^μ preserves those closed sets present in all C_i's)
?

Two classical metrics on a lattice

- MPL metric ∂:

$\partial(C, C')$ is the minimum path length in the covering graph ($\mathbf{M}$, p)

- (Generalized) symmetric difference metric δ:

$$\delta(C, C') = |\, \mathcal{J}_C \,\Delta\, \mathcal{J}_{C'} \,| = |\, C \,\Delta\, C' \,|$$

- Since $\mathbf{M}$ is a convex geometry, $\partial = \delta$,

a characterization of LLD lattices (L. 2003)

Federation consensus rules and quota rules

Federation on $K = \{1, \ldots, k\}$: inclusion monotone family $\mathcal{K}$ of subsets of K:

$$[L \in \mathcal{K}, L' \supseteq L] \Rightarrow [L' \in \mathcal{K}]$$

Federation consensus function $c_{\mathcal{K}}$ on $\mathbf{M}$:

$$c_{\mathcal{K}}(C^*) = \vee_{L \in \mathcal{K}} (\cap_{i \in L} C_i)$$

Include:

Oligarchic consensus functions: $\mathcal{K} = \{L \subseteq K: L \supseteq I\}$ for a fixed $I \subseteq K$.

$$c_{\mathcal{K}}(C^*) = \cap_{i \in I} C_i,$$

Quota rules: with $1 \leq q \leq k$ (majority rule: $q > k/2$)

$$c_q(C^*) = \Phi(\mathcal{A}_q),$$

where $\mathcal{K} = \{L \subseteq K: |L| \geq q\}$, for a fixed q,

$\mathcal{A}_q$ is the set of closed sets present in at least q elements of the profile

Results in the lattice M

- Properties of c_q:

Unanimity: $\cap_{1 \leq i \leq k} C_i \subseteq c_q(C^*)$;

Isotony: $C_i \subseteq C'_i$ for all $i = 1, \ldots, k \Rightarrow c_q(C^*) \subseteq c_q(C'^*)$.

In ***convex geometries***, quota rules share *consistency* with the median procedure (L. 2003). Consider a relative frequency $\alpha \in [0, 1[$:

$$c_\alpha(C^*) = c_\alpha(C'^*) = C \quad \Rightarrow \quad c_\alpha(C^* C'^*) = C$$

Sketched proof. From $\mathcal{M}(C) \subseteq \mathcal{A}_q \subseteq C$, $\mathcal{M}(C') \subseteq \mathcal{A}'_q \subseteq C'$, and standard properties of frequencies:

for any $C \subset S$, $\min(\gamma(C, C^*), \gamma(C, C'^*)) \leq \gamma(C, C^*C'^*) \leq \max(\gamma(C, C^*), \gamma(C, C'^*))$,

one gets $\mathcal{M}(C) \subseteq \mathcal{A}_q(C^*C'^*) \subseteq C$

This property is not true, e.g., in the partition lattice (Barthélemy and L. 1995).
Problem : does it characterize LLD ones ?

Weak majorities and medians

For any median C^μ,

$$C^\mu \subseteq c_{k/2}(C^*),$$

that is, $C^\mu \subseteq \boldsymbol{\Phi}(\mathcal{A}_{k/2})$,

where $\mathcal{A}_{k/2}$ is a set of closed sets present in at least half of the elements of the profile.

Any closed set of a median CS is an intersection of "majority closed sets".

Consequence: if such closed sets do not exist (but S), the trivial closure system $\{S\}$ is the unique median of C^*.

Axiomatic results

A consensus rule f: $\mathbf{M}^k \rightarrow \mathbf{M}$ satisfies *unanimity* and is *neutral monotonic*: for all A, $B \subset S$, C^*, $C'^* \in \mathbf{M}^k$,

$$\{i: A \in C_i\} \subseteq \{i: B \in C'_i\} \Rightarrow [A \in f(C^*) \Rightarrow B \in f(C'^*)]$$

if and only if it is *oligarchic* (Raderanirina 2001, Monjardet and Raderanirina 2004, by particularization of Monjardet 1990)

with many related results on special cases of closure systems, choice functions, …

- the above result applies to the unanimity rule c_k.

Discussion

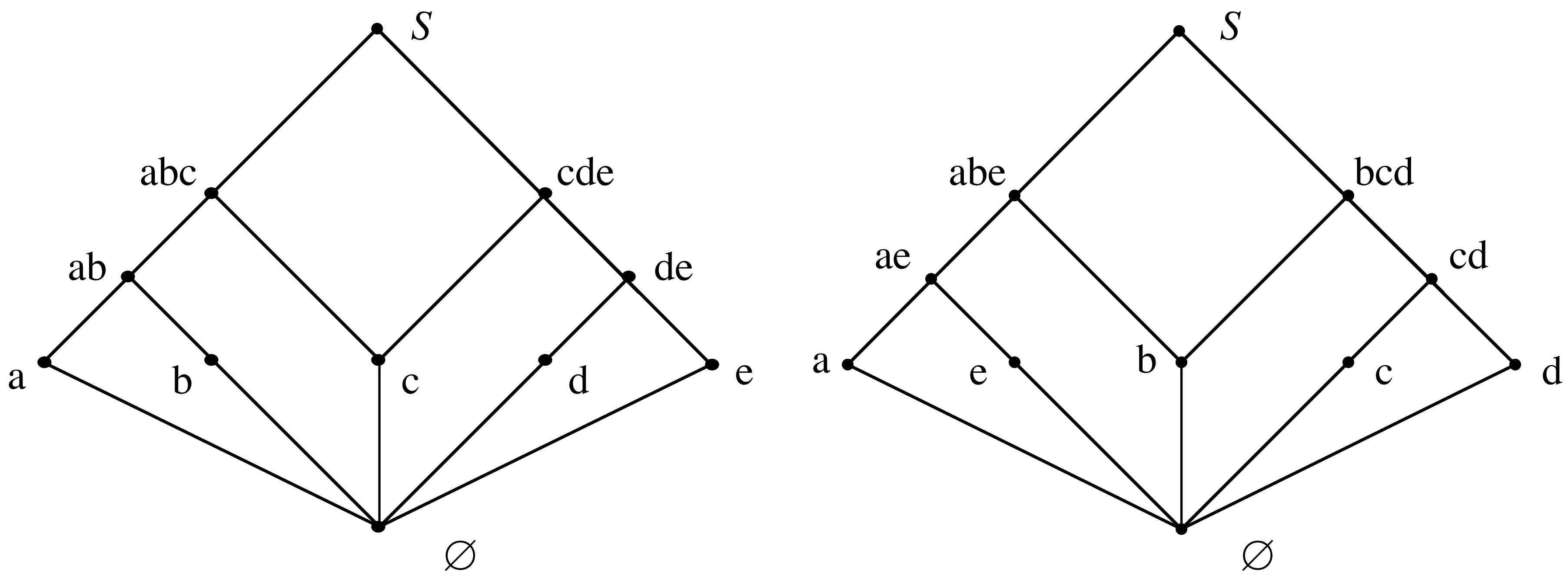
Significant results were obtained, especially for quota rules (including majority rule)

A limitation:
Quota rules, and related methods only take into account presence or absence of closed sets in a significant number (oligarchies, majorities) or in all (unanimity) elements of the profile:

- Small q: lack of significance of the consensus

- Consensus closed sets vanish when q increases. Unless the elements of the profile C^* are close to each other, $c_q(C^*)$ may become trivial

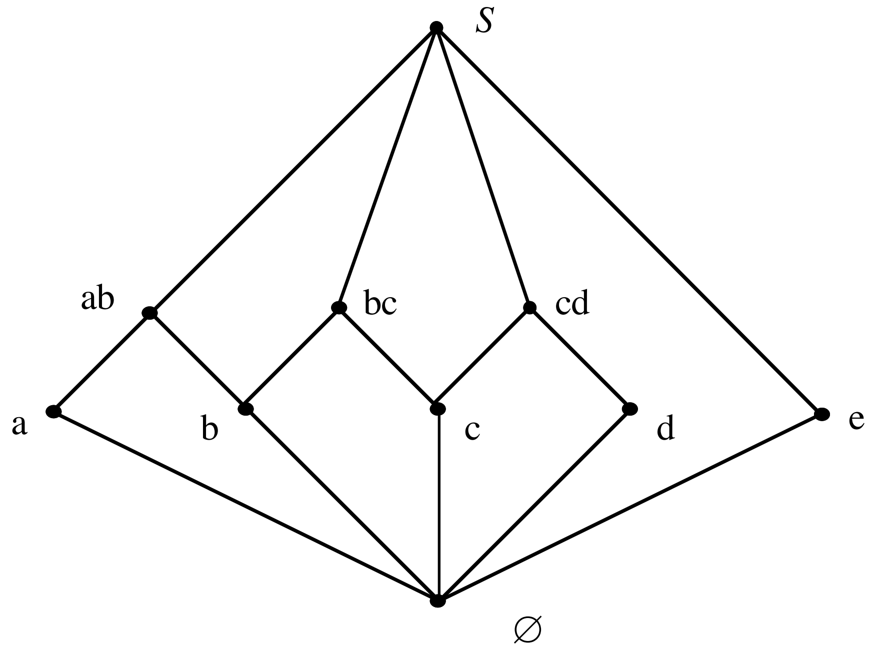
- Actual common features not recognized: see the 2-profile below



No common non trivial closet set

Common association of: ab, bc, cd

A possible consensus closure system for σ = 2 (unanimity on nestings):



For a finer approach, we consider ***implications*** and their ***overhangings*** variant

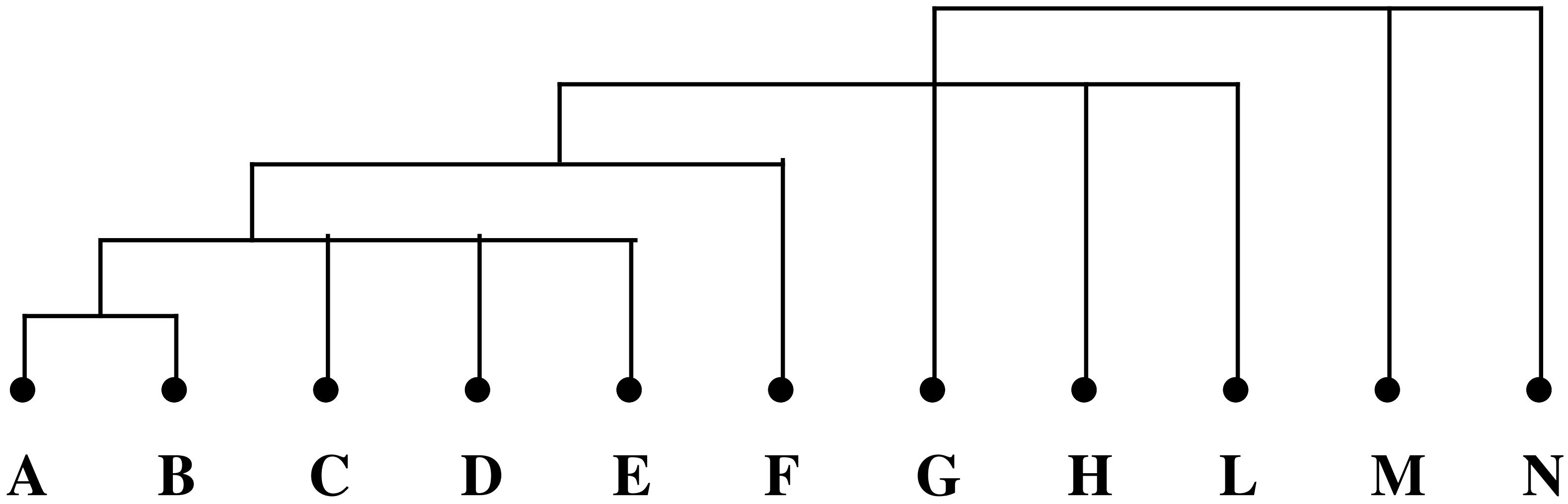
Adams' intersection rule (1972, 1986) for the consensus of classification trees:

- $S \in a(\mathcal{C}^*)$

Let C be an obtained class,

- select the maximal C'_i 's in $\mathcal{C}_i$ s.t. $C'_i \subset C$,
- For a tuple $(C'_1, C'_2, \ldots, C'_k)$, set $C' = \cap_{1 \leq i \leq k} C'_i \in a(\mathcal{C}^*)$,

and iterate…



Adam's Theorem

Let us associate to a tree $\mathcal{H}$ its *nesting order* Œ on 2^S:

$$A \text{ Œ } B \text{ if } A \subset B \text{ and } H_A \subset H_B$$

H_A is the smaller class in $\mathcal{H}$ including A

Given a profile $\mathcal{H}^* = (\mathcal{H}_1, \mathcal{H}_2, \ldots, \mathcal{H}_k)$ of hierarchical classification systems with overhangings/nesting orders $Œ_1, Œ_2, \ldots, Œ_k$

Adams' tree is the unique tree $\mathcal{H}$ (with nesting order Œ) s.t.:

(A1) $\bigcap_{1 \leq i \leq k} Œ_i \subseteq Œ$ (preservation of unanimity)

(A2) $H, H' \in \mathcal{H}$ and $H \subseteq H'$ imply $(H, H') \in \bigcap_{1 \leq i \leq k} Œ_i$ (qualified nestings)

Discussion (2)

Quota rules:

- sensitive to noising (existence of common classes required)
- take frequencies into account

Adams rule:

- able to provide new classes (based on common *subclasses*)
- does not take frequencies into account

• what about other closure systems than trees?
• other frequencies than unanimity ?

Implication relation (of a closure system)

A binary relation $\rightarrow$ on 2^S: $A \rightarrow B$ if every closed set containing A also contains B

Characterization (*complete* – or *full* – *implication system* CIS, Armstrong 1974):

(I1) $B \subseteq A \Rightarrow A \rightarrow B$,

(I2) $A \rightarrow B$ and $B \rightarrow C \Rightarrow A \rightarrow C$,

(I3) $A \rightarrow B$ and $C \rightarrow D \Rightarrow A \cup C \rightarrow B \cup D$.

Important literature (databases, lattice or symbolic data analysis, data mining,…), with strong results (existence of a canonical implication basis, Maier 1983, Guigues and Duquenne 1986)

Survey by Caspard and Monjardet (2003)

Overhanging/nesting order (of a closure system)

A binary relation Œ on 2^S: A Œ B if $A \subset B$ and not $A \rightarrow B$

(there exists a closed set containing A and not B)

Example: Adams' nestings for hierarchical CS's

Characterization (Domenach and L. 2003):

(O1) A Œ $B \Rightarrow A \subset B$,

(O2) $A \subset B \subset C \Rightarrow$ [A Œ $C \Longleftrightarrow A$ Œ B or B Œ C],

(O3) A Œ $A \cup B \Rightarrow A \cap B$ Œ B.

From (O1) and(O2), Œ is a *strict order* on 2^S.

Cryptomorphisms… Four isomorphic or dually isomorphic lattices

M set of all closure systems on S, **C** set of all closure operators on 2^S,
I set of all complete implication systems on S,
O set of all complete overhanging orders on S, among others…

M	**C**	**I**	**O**
2^S (maximum)	$\varphi_{min} = id_{2^S}$ (minimum)	$\{(X, Y) \in (2^S)^2: Y \subseteq X\}$ (minimum)	$\{(X, Y) \in (2^S)^2: X \subset Y\}$ (maximum)
$\{S\}$ (minimum)	$\varphi_{max}(A) = S$ (maximum)	$(2^S)^2$ (maximum)	$\varnothing$ (minimum)
join $\mathcal{M} \vee \mathcal{M}'$	meet (pointwise intersection)	meet $\mathcal{I} \cap \mathcal{I}'$	join Œ$\cup$Œ'
meet $\mathcal{M} \cap \mathcal{M}'$	join	join $\mathcal{I} \vee \mathcal{I}'$	meet Œ$\wedge$Œ'
$\{S, A\}$, $A \subset S$ (join irreducible)	$\varphi(X) = A$ if $X \subseteq A$; $\varphi(X) = S$ otherwise (meet irred.)	$(2^A)^2 \cup \{(X,Y) \in (2^S)^2: A \subseteq X\}$ (meet irred.)	$\{(X,Y) \in (A] \times (2^S - (A]): X \subset Y\}$ (join irred.)
$\{X \subseteq S: A \subseteq X \Rightarrow s \in X\}$, $A \subset S$, $s \in S\text{-}A$ (meet irred.)	$\varphi(X) = X+s$ if $A \subseteq X$ $\varphi(X) = X$ otherwise (join irred.)	$\{(X, Y) \in (2^S)^2: X \subseteq Y$ or $A \subseteq X$, $Y = X+s\}$ (join irred.)	$\{(X,Y) \in (2^S)^2: X \subset Y\} - \{(X,Y) \in (2^S)^2: A \subseteq X, Y = X+s\}$ (meet irred.)

Overhanging orders (special cases)

(Domenach and L. 2004-2007…)

- Classification systems: (O1), (O2), (O3) and

(OE) $\emptyset$ Œ $\{s\}$ for any $s \in S$,

(OS) $A \notin \{\emptyset, \{s\}\} \Rightarrow \{s\}$ Œ $A \cup \{s\}$, for all $s \in S$.

- Nested families: (O1), (O2) and

(ON) A Œ C and B Œ $C \Rightarrow A \cup B$ Œ C.

- Trees of subsets: (O1), (O2), either (ON) or (OE) and

(OT) A Œ C and B Œ $C \Rightarrow A \cup B$ Œ C or $A \cap B = \emptyset$ (Adams' axiom) .

- Distributive CS: (O1), (O2) and

(OD) $s \in S$, $A \subseteq S$, and $\{a\}$ Œ $\{a, s\}$ for any $a \in A \Longleftrightarrow A$ Œ $A \cup \{s\}$.

- Convex geometries: (O1), (O2), (O3), (OE) and

(OC) $A \cup B \subseteq C$, $A \cap B$ Œ $C \Rightarrow A$ Œ C or B Œ C.

Fitting overhangings: a dual closure

Data: a binary relation R on 2^S,   with $(A, B) \in R$ implies $A \subset B$,

Problem: find an overhanging approximation of R.

An obvious solution: since

- $\mathbf{O}$ is $\cup$-stable,
- The empty relation $\varnothing$ is the minimum of $\mathbf{O}$,

there is a *dual closure operator* ω on $2^{(2^S)^2}$

$$\omega(R) = \cup\{Œ \in \mathbf{O}: Œ \subseteq R\}.$$

Getting $R \subseteq \omega(R)$, while there are reasons to prefer an approximation "from the top".

Fitting overhangings: a uniqueness result (Domenach and L. 2004)

Given a binary relation R on 2^S, with $(A, B) \in R$ implies $A \subset B$,
there is ***at most one*** *closure system* $\mathcal{C}$ (with overhanging order Œ) satisfying:

(AR1) $R \subseteq$ Œ (preservation of R)

(AR2) For any **meet-irreducible** M of $\mathcal{C}$, $(M, M^+) \in R$ (qualified overhangings)

Remark: (A2) is a (very) partial converse of (A1)

Proof. Assume that both $\mathcal{C}$ and $\mathcal{C}'$ satisfy (AR1) and (AR2). Observe first that S belongs to $\mathcal{C}$ and $\mathcal{C}'$. If the symmetric difference $\mathcal{C}\Delta\mathcal{C}'$ is not empty, let C be a maximal element of $\mathcal{C}\Delta\mathcal{C}'$. Assume without loss of generality that C belongs to $\mathcal{C}$. If C was not a meet-irreducible $\mathcal{C}$, it would be an intersection of meet-irreducibles, all belonging to both $\mathcal{C}$ and $\mathcal{C}'$ and, so, C would belong to $\mathcal{C}'$, and not to $\mathcal{C}\Delta\mathcal{C}'$.
Thus, C is covered by a unique element C^+ of $\mathcal{C}$, with $C^+ \in \mathcal{C}'$. By (AR2), the pair (C, C^+) belongs to R and, by (AR1), C Œ' C^+. Set $C' = \varphi'(C)$. We have $C \subset C'$, since $C \notin \mathcal{C}'$, and C' Œ' C^+, since $C' = \varphi'(C) = \varphi'(C') \subset \varphi'(C^+) = C^+$. But $C \subset C'$ implies $C' \in \mathcal{C}$, with $C \subset C' \subset C^+$, a contradiction with the hypothesis that C^+ covers C in $\mathcal{C}$.

Adams theorem:

- hierarchical case
- $R = \bigcap_{1 \leq i \leq k} Œ_i$
- axiom (AR2) is weaker than particularized (A2)

and existence guaranteed by Adams algorithm!

The solution for (A*R*1) and (A*R*2) does not always exist

Example 1: $R = \varnothing$, solution $\mathcal{C} = \{S\}$

Example 2: $R = \{(A, S)\}$, with $A \subset S$, solution $\mathcal{C} = \{A, S\}$

Example 3: $R = \{(A, B)\}$, with $A \subset B \subset S$, no $\mathcal{C}$ satisfying (AR1) and (AR2)

Properties

- If R satisfies Conditions (O1) and (O2), then there exists a closure system $\mathcal{C}$ satisfying Conditions (AR1) and (AR2).

- Approximation "from the top": if Œ satisfies Conditions (AR1) and (AR2), then, for any overhanging order Œ',

$$R \subseteq Œ' \subseteq Œ \text{ implies } Œ' = Œ.$$

What about the consensus case?

A profile $\mathcal{C}^* = (\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_k)$ of closure systems

A minimal frequency requested on nestings (fixed $\sigma \leq k$)

Set $R = \bigcup_{I \subseteq K,\, |I| \geq \sigma} \bigcap_{1 \leq i \leq k} Œ_i$ (then, $\omega(R)$ corresponds to $c_q(\mathcal{C}^*)$)

- Adams' intersection method: trees, $\sigma = k$.

In terms of overhangings

(FO) for all $A, B \subseteq X$, $|\{i \in K : A \text{ Œ}_i B\}| \geq p$ implies $A \text{ Œ } B$,

(frequent overhangings preservation)

(QO) for all $M \in \mathcal{M}(C)$, $|\{i \in K : M \text{ Œ}_i M^+\}| \geq p$.

(qualified overhangings)

In terms of implications

(FI) for all $A, B \subseteq X$, $A \to B$ implies $|\{i \in K : A \to_i B\}| \geq k - p$,

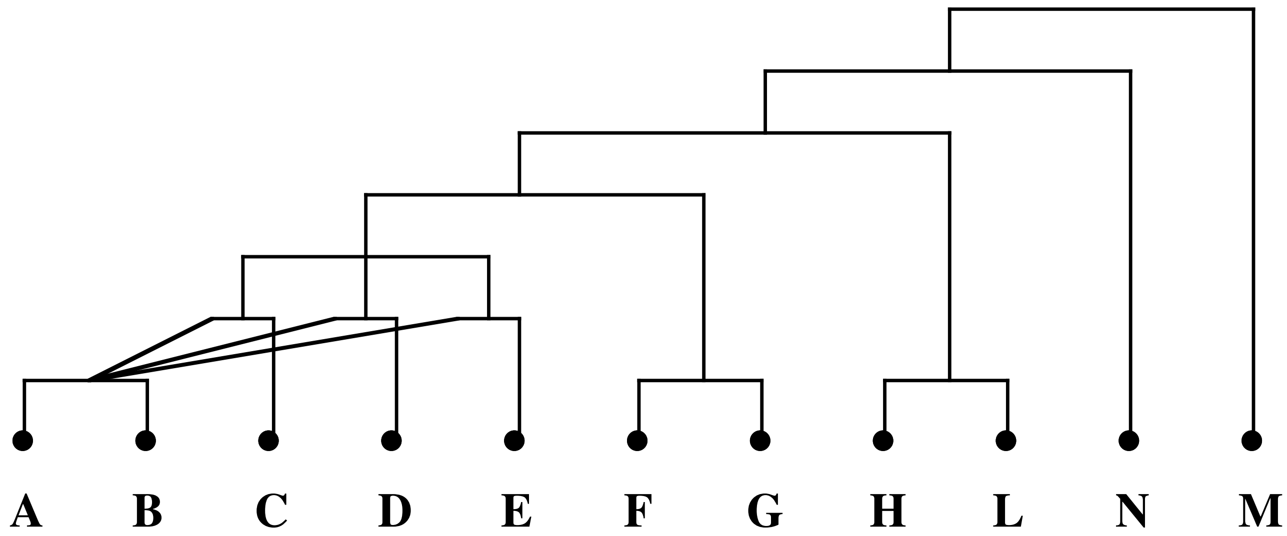
(frequent implications preservation)

(UI) for all $M \in \mathcal{M}(C)$, $|\{i \in K : M \to_i M^+\}| < k - p$.

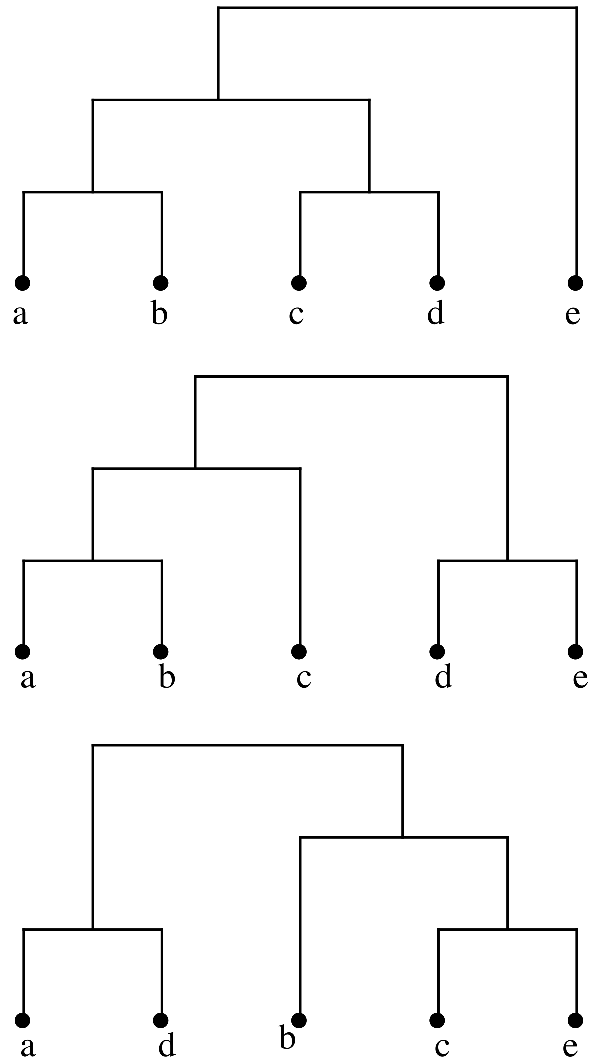
(disqualified implications)

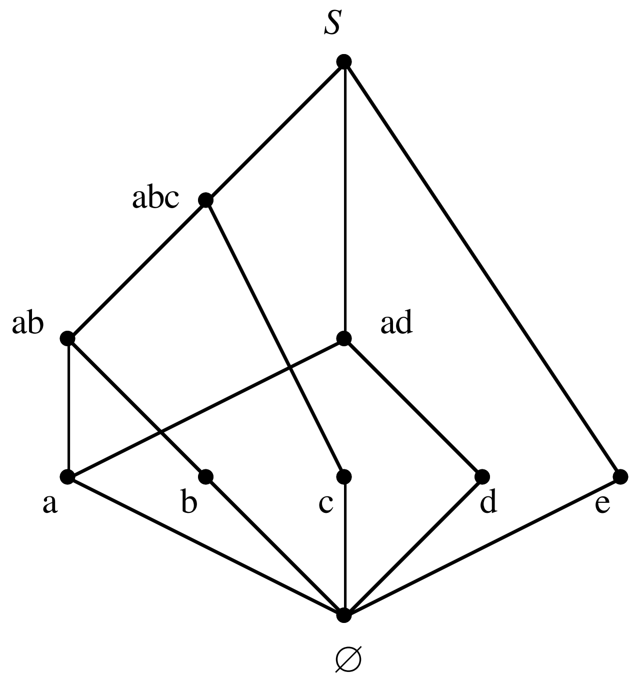
Back to kangaroo rats

$\sigma = 2$



- includes the majority classes
- brings further ones : ABC, ABD, with reasons to distinguish them from larger groups
- no longer a tree


a b c d e
a b c d e
a d b c e


S
abc
ab
ad
a
b
c
d
e
∅

Conjecture: for a relation $R = \cup_{I\subseteq K,\ |I|\geq\sigma} \cap_{1\leq i\leq k} \text{Œ}_i$,

there always exists a closure system satisfying Conditions (AR1) and (AR2).

Two kinds of problems

- Possibility results and algorithms
- Impossibility results