Towards a Theory of Holistic Clustering

A.W.M. Dress
Bielefeld, Germany

No. 152 March, 1997

Abstract: In this note, cluster theory is presented from a rather abstract point of view, basic known results are reviewed from this view point, and some new results which motivated the proposed approach, as well as some new problems which naturally arise in this context, are presented.

Towards a Theory of Holistic Clustering

A.W.M. Dress
Bielefeld, Germany

Abstract: In this note, cluster theory is presented from a rather abstract point of view, basic known results are reviewed from this view point, and some new results which motivated the proposed approach, as well as some new problems which naturally arise in this context, are presented.

1 A Local-Global Approach to Clustering

Clustering procedures (as well as many other mathematical techniques) can be viewed as methods for extracting globally relevant features from locally distributed information. A rather natural, simple and sufficiently general conceptual framework for describing clustering procedures is, therefore, the following one: We assume that, for any given (generally finite) set X, we can form the set $Inf\,(X)$ comprising **all** possible structures defined on X which encode the information regarding X we are seeking: this could be the set of all tree structures or the set of all ultrametrics definable on X, or just the set $\mathcal{P}(\mathcal{P}(X))$ of all subsets of the power set $\mathcal{P}(X)$ of X. In addition, we assume that, for any pair (X, Y) consisting of a set X and a subset Y of X, information regarding X implies information regarding Y which is expressed in form of a map

$$res \;=\; res_{X \to Y} : Inf\,(X) \to Inf(Y) : i \mapsto i|_Y$$

called the *restriction map* (relative to X and Y), and we assume consistency of restriction by requiring that, for all $Z \subseteq Y \subseteq X$ and $i \in \mathrm{Inf}(X)$, we have

$$i|_Z = (i|_Y)|_Z \;.$$

Finally, we assume that there is a concept of *compatibility* of information, expressed for each set X as above in terms of a - not necessarily symmetric - binary relation

$$Cpb_X \subseteq Inf_X \times Inf_X$$

- with $(i,j) \in Cpb_X$ implying that given information "i", the information "j" is compatible with (or even implied by) "i". A natural requirement for these compatibility relations is, of course, that all these relations are *reflexive* and *hereditary* with respect to restriction, that is, one has $(i,i) \in Cpb_X$ for all X and $i \in Inf\,(X)$ as well as

$$(i|_Y, j|_Y) \in Cpb_Y$$

for all $Y \subseteq X$ and $(i,j) \in Cpb_X$. Another property one may or may not want to require is transitivity, that is that $(i,j),(j,k) \in Cpb_X$ for some X always implies $(i,k) \in Cpb_X$.

Now, locally distributed information can be expressed in terms of elements $i_Y \in Inf\,(Y)$ - with Y running through the set $\mathcal{Y} \subseteq \mathcal{P}(X)$ of all those - generally small - subsets Y of X for which certified information $i_Y \in Inf\,(Y)$ is available and, given such a family $(i_Y)_{Y\in\mathcal{Y}}$ of certified local information, we may ask for all elements $i_X \in Inf\,(X)$ with $(i_Y, i_X|_Y) \in Cpb\,(Y)$ for all $Y \in \mathcal{Y}$.

This scheme of formalising clustering procedures has the advantage that it allows for the separation of various stages of clustering, in particular it allows one to clearly separate the task of setting up an appropriate formal model for the kind of information we are seeking from (a) analysing the model from a mathematical point of view and (b) designing and analysing algorithms for actually computing globally relevant information from local data.

In this note, we'll study in particular the following two basic clustering models which we'll dub the *affine* and the *projective* clustering model: In the affine clustering model, the information we seek is gathered in terms of a collection $\mathcal{C}$ of *clusters*, that is, of subsets C of the set X of objects in question; so we have

$$Inf\,(X) = Cl^{aff}(X) := \mathcal{P}(\mathcal{P}(X)).$$

There are several ways to define, for any pair (X, Y) with $Y \subseteq X$, the associated restriction map. The most natural one probably is using intersection with Y, that is putting

$$res_{X \to Y}(\mathcal{C}) = del_{X \to Y} = del^{aff}_{X \to Y} := \{Y \cap C \mid C \in \mathcal{C}\}$$

for any $\mathcal{C} \subseteq \mathcal{P}(X)$, also sometimes called the *deletion operator* because elements outside Y are just neglected and, henceforth, deleted. Another, sometimes interesting and useful choice is using the *contraction operator*, defined by

$$contract_{X \to Y}(\mathcal{C}) = contract^{aff}_{X \to Y}(\mathcal{C}) := \{C \subseteq Y \mid C \cup (X - Y) \in \mathcal{C}\},$$

or its *dual*, the *focus operator*, defined by

$$focus_{X \to Y}(\mathcal{C}) = focus^{aff}_{X \to Y}(\mathcal{C}) := \{C \subseteq Y \mid C \in \mathcal{C}\}$$

which focuses attention to only those clusters $C \in \mathcal{C}$ which are already contained in Y. It is clear that these three operators satisfy the consistency condition described above. In this note, we'll only consider the deletion operator when dealing with Cl^{aff}.

Finally, at least when dealing with Cl^{aff}, we define the binary relation

$$Cpb_X \subseteq Inf\,(X) \times Inf\,(X)$$

by

$$Cpb_X = Cpb^{aff}_X := \{(\mathcal{C}, \mathcal{C}') \in Inf\,(X)^2 \mid \mathcal{C} \supseteq \mathcal{C}'\}.$$

The projective clustering model is defined in a rather similar way except that we replace the subsets C of X by the splits S of X, that is unordered pairs $S = \{A, B\}$ of subsets of X with $A \cup B = X$ and $A \cap B = \emptyset$ or - equivalently - equivalence relations $\overset{S}{\sim}$ defined on X with, at most, two distinct equivalence classes. So, with $\mathcal{S}p(X)$ denoting the set of all splits of X, we put

$$Cl^{proj}(X) := \mathcal{P}(\mathcal{S}p(X)),$$

we define

$$res_{X \to Y}(\mathcal{S}) = del_{X \to Y}(\mathcal{S}) = del^{proj}_{X \to Y}(\mathcal{S}) := \{\{A \cap Y, B \cap Y\} \mid \{A, B\} \in \mathcal{S}\}$$

for all $Y \subseteq X$ and $\mathcal{S} \subseteq \mathcal{S}p(X)$, and we define

$$Cpb_X^{proj} := \{(\mathcal{S}, \mathcal{S}') \in \mathcal{S}p(X)^2 \mid \mathcal{S} \supseteq \mathcal{S}'\}.$$

Of course, one could also define a restriction operator which is analogous simultaneously to the contraction and the focus operator by putting

$$contract_{X \to Y}(\mathcal{S}) = contract_{X \to Y}^{proj}(\mathcal{S}) := \{\{A, B\} \in \mathcal{S}p(Y) \mid \{A, B \cup (X - Y)\} \in \mathcal{S}\};$$

yet, we will not study this restriction operator in the present note and just mention in passing that in the special case $\#(X - Y) = 1$, both restriction operators, projective deletion and projective contraction, coincide.

Note also that in both set-ups the non-uniqueness of the solution i_X we are seeking for any given some family $(i_Y)_{Y \in \mathcal{Y}} (\mathcal{Y} \subseteq \mathcal{P}(X))$ is not a problem as - almost by definition - there exists always a unique largest element $\bar{i}_X \in Cl^{aff/proj}(X)$ such that i_X is a solution if and only if $(\bar{i}_X, i_X) \in Cpb^{aff/proj}(X)$.

These models formalise the following idea: Given a collection X of objects under consideration, we seek to specify "relevant" subsets C (or splits $S = \{A, B\}$) of X which group together elements of X which exhibit a certain degree of similarity relative to each other (or share enough common features) so as to distinguish them clearly, as a class C (or A, or B), from all elements outside this class. We assume further that for (at least some) small subsets Y of X, the data we can start with allow us to specify easily and directly those subsets (or splits) of Y which are considered to be relevant as far as only objects from Y are concerned, and we then ask for those subsets C (or splits S) of X which, when restricted to any such $Y \subseteq X$, produce only such subsets (or splits) of Y which have been specified before as being of some relevance relative to Y.

Of course, there are simple relations between affine and projective clustering: Given a system $\mathcal{C} \subseteq \mathcal{P}(X)$ of clusters, we can associate to it a system $\mathcal{S}$ of splits of X by putting

$$\mathcal{S} = \mathcal{S}(\mathcal{C}) := \{\{C, X - C\} \mid C \in \mathcal{C}\},$$

and given a system $\mathcal{S} \subseteq \mathcal{S}p(X)$ of splits of X, we can associate to it the system

$$\mathcal{C} = \mathcal{C}(\mathcal{S}) := \{C \subseteq X \,|\, \{C, X - C\} \in \mathcal{S}\}.$$

Clearly, we have

$$\mathcal{S}(\mathcal{C}(\mathcal{S})) = \mathcal{S}$$

and

$$\mathcal{C}(\mathcal{S}(\mathcal{C})) = \{C \subseteq X \,|\, C \in \mathcal{C} \text{ or } X - C \in \mathcal{S}\}.$$

Yet, there is a better way to relate affine and projective clustering which also explains why the terms *affine* and *projective* were suggested in this context: Given a set X as above and a system $\mathcal{C}$ of clusters $C \subseteq X$, we may add to X another *ideal* element *at infinity* denoted by $*$, and then form the split system $\mathcal{S}^*(\mathcal{C})$ of $X^* := X \cup \{*\}$, defined by

$$\mathcal{S}^*(\mathcal{C}) := \{\{C, X^* - C\} \,|\, C \in \mathcal{C}\},$$

while given a system $\mathcal{S}$ of splits of X^*, we may form the system $\mathcal{C}_*(\mathcal{S})$ of clusters in X defined by

$$\mathcal{C}_*(\mathcal{S}) := \{C \subseteq X \,|\, \{C, X^* - C\} \in \mathcal{S}\}.$$

Clearly, we have $\mathcal{S}^*(\mathcal{C}_*(\mathcal{S})) = \mathcal{S}$ for every $\mathcal{S} \subseteq \mathcal{S}p(X^*)$ as well as $\mathcal{C}_*(\mathcal{S}^*(\mathcal{C})) = \mathcal{C}$ for every $\mathcal{C} \subseteq \mathcal{P}(X)$. We'll also see later on that relevant properties of cluster systems $\mathcal{C} \subseteq \mathcal{P}(X)$ easily translate into corresponding properties of split systems $\mathcal{S} \subseteq \mathcal{S}p(X^*)$, and that similar results then hold in both situations. Moreover, as in geometry, it will turn out that while the affine version is more easily grasped and reflects the naive intuitive understanding of clustering, the projective version (which, from the affine point of view, consists in forgetting the special role of the point at infinity used in forming splits from clusters) allows one often more elegant formulations of theorems and proofs.

Regarding the above set up, the following problems arise:

(1) Given some data regarding the elements of X, e.g. a - perhaps only partially known - (dis)similarity matrix, and some type of problem, e.g. the problem of (re)constructing from these data the topology of a (phylogenetic) tree, what is the

appropriate choice for the set $\mathcal{Y}$ of "small" subsets and which definition do we use to express the local information we have in terms of an element $i_Y \in Inf\,(Y) = Cl^{aff/proj}(Y)$?

(2) Given a family $(i_Y)_{Y\in\mathcal{Y}}$ of local data, what is the optimal algorithm for computing the associated unique largest solution

$$\bar{i}_X \in Inf\,(X) = Cl^{aff/proj}(X)?$$

(3) And, closely related to that question, are there *a priori* upper bounds regarding the number of clusters or splits in $\bar{i}_X$?

(4) In addition, the following questions are of interest also in the general context outlined above:

- given $\mathcal{Y} \subseteq \mathcal{P}(X)$ and $i \in Inf\,(X)$, what can be said about all $i_X \in Inf\,(X)$ with $(i|_Y,\ i_X|_Y) \in Cpb\,(Y)$ for all $Y \in \mathcal{Y}$? Obviously, this is the case for all $i_X \in Inf\,(X)$ with $(i, i_X) \in Cpb\,(X)$ whenever compatibility is hereditary with respect to restriction. So, the most basic question in this context is how to characterise those $\mathcal{Y} \subseteq \mathcal{P}(X)$ and $i \in Inf\,(X)$ for which this obviously sufficient condition is also necessary, that is, for which

$$\{i_X \in Inf\,(X) \,|\, (i, i_X) \in Cpb\,(X)\}$$

equals the set

$$\{i_X \in Inf\,(X) \,|\, (i|_Y, i_X|_Y) \in Cpb\,(Y) \text{ for all } Y \in \mathcal{Y}\}.$$

- given $\mathcal{Y} \subseteq \mathcal{P}(X)$ and a family of elements $i_Y \in Inf\,(Y)\ \ (Y \in \mathcal{Y})$, when does there exist some - or even a unique - element $i_X \in Inf\,(X)$ with $i_X|_Y = i_Y$ for all $Y \in \mathcal{Y}$?

In the following, we'll first collect some of what is already known regarding these questions for special choices of $\mathcal{Y} \subseteq \mathcal{P}(X)$ and of the family $(i_Y)_{Y\in\mathcal{Y}}$, and then we'll address question (3) in some more detail allowing $\mathcal{Y}$ and $(i_Y)_{Y\in\mathcal{Y}}$ to be almost arbitrary.

2 Some Previous Results

2.1 Hierarchies

According to well-established traditions (cf. [G87]), a *hierarchy* $\mathcal{C}$ defined on a set X - or, for short, an $X-$*hierarchy* - is defined to be a subset $\mathcal{C}$ of $\mathcal{P}(X)$ such that

$$C_1 \cap C_2 \in \{C_1, C_2, \emptyset\}$$

holds for all $C_1, C_2 \in \mathcal{C}$. For technical reasons, we require also that $\emptyset \in \mathcal{C}$ and $X \in \mathcal{C}$ should hold for any $X-$hierarchy $\mathcal{C}$ which then automatically implies also that $\mathcal{C}$ is closed with respect to intersection. Given a similarity measure s defined on X (that is, just a map $s : X \times X \to \mathbb{R}$ satisfying the symmetry condition $s(x,y) = s(y,x)$ for all $x, y \in X$), an $X-$hierarchy $\mathcal{C}_s$ can be associated with s according to the definition

$$\mathcal{C}_s := \{C \subseteq X \mid s(a,b) > s(a,c) \text{ for all } a, b \in C \text{ and } c \in X - C\}.$$

It is well known that, for any $X-$hierarchy $\mathcal{C} \subseteq \mathcal{P}(X)$, one can always find some similarity measure s defined on X with $\mathcal{C} = \mathcal{C}_s$ and that, with $n := \#X$, one always has $\#\mathcal{C} \leq 2n$. This can either be seen by induction, using the fact that for every maximal cluster $C_0 \in \mathcal{C}$ which is different from X, $\mathcal{C}$ decomposes into the three sets $\{X\}$, $\{C \in \mathcal{C} \mid C \subseteq C_0\}$ and $\{C \in \mathcal{C} \mid C \cap C_0 = \emptyset\}$ - with the last two having only the empty set $\emptyset \in \mathcal{P}(X)$ in common. Following an idea of Boris Mirkin (cf. [M97]), it can also be deduced as follows: Given an $X-$hierarchy $\mathcal{C}$ and a cluster $C \in \mathcal{C}$, let

$$V(C) \subseteq I_{\mathbb{R}}(X) := \{f : X \to \mathbb{R} \mid \sum_{x \in X} f(x) = 0\}$$

denote the real vector space of all maps f from X into $\mathbb{R}$ which vanish outside C, are orthogonal to every constant map (relative to the canonical inner product defined on $\mathbb{R}^X$), and are constant on every proper subcluster C' of C. It is easily seen that $V(C)$ is different from 0 if and only if C contains at least two distinct elements, and that any two such subspaces $V(C)$ and $V(C')$ are orthogonal to each other for any $C, C' \in \mathcal{C}$ with $C \neq C'$. Actually, a simple induction argument similar to the one used just above establishes that $I_{\mathbb{R}}(X)$ is the orthogonal direct sum of the spaces $V(C)$ $(C \in \mathcal{C})$. Yet, even without establishing this fact, the construction yields, for $n := \#X$ as above, the inequality

$$\#\mathcal{C} \leq \#\{C \in \mathcal{C} \mid \#C \leq 1\} + \dim_{\mathbb{R}} I_{\mathbb{R}}(X) \leq (1+n) + (n-1) = 2n$$

- and, hence, shows also that for an X−hierarchy $\mathcal{C} \subseteq \mathcal{P}(X)$ with $\#\mathcal{C} = 2n$ one has necessarily $\dim_{\mathbb{R}} V(C) = 1$ for each $C \in \mathcal{C}$ with $\#C > 1$.

All that can be put easily into the framework considered in the previous section: First, we note that a set system $\mathcal{C} \subseteq \mathcal{P}(X)$ with $\emptyset \in \mathcal{C}$ and $X \in \mathcal{C}$ is an X−hierarchy if and only if $\mathcal{C}|_Y = \{C \cap Y \mid C \in \mathcal{C}\} \subseteq \mathcal{P}(Y)$ is a Y−hierarchy for every $Y \subseteq X$ with $\#Y = 3$: indeed, if there would exist some subsets $C_1, C_2 \in \mathcal{C}$ with $\emptyset \neq C_1 \cap C_2 \neq C_i$ for $i = 1, 2$, then putting $Y := \{a, b, c\}$ with $a \in C_1 \cap C_2$, $b \in C_1 - (C_1 \cap C_2)$ and $c \in C_2 - (C_1 \cap C_2)$ produces a subset Y of cardinality 3 so that $\mathcal{C}|_Y$ is not a Y−hierarchy. Hence, putting

$$\mathcal{Y} := \mathcal{P}_{\leq 3}(X) = \{Y \subseteq X \mid \#Y \leq 3\}$$

and choosing, for each $Y \in \mathcal{Y}$, a Y−hierarchy $i_Y \in \mathcal{P}(\mathcal{P}(Y)) = Cl^{aff}(Y)$, the resulting set system

$$\bar{i}_X := \{C \subseteq X \mid C \cap Y \in i_Y \text{ for all } Y \in \mathcal{Y}\}$$

surely is an X−hierarchy. In addition, given either an X−hierarchy $\mathcal{C} \subseteq \mathcal{P}(X)$ or a similarity measure s defined on X, we may consider the Y−hierarchies

$$i_Y = i_Y^{\mathcal{C}} := \mathcal{C}|_Y = \{C \cap Y \mid C \in \mathcal{C}\} \in Cl^{aff}(Y)$$

or

$$i_Y = i_Y^{s} := \mathcal{C}_{s|_{Y \times Y}} \in Cl^{aff}(Y),$$

respectively, for every $Y \in \mathcal{Y}$. It is then easy to see that $\mathcal{C}$ (or $\mathcal{C}_s$) coincides with

$$\bar{i}_X := \{C \subseteq X \mid C \cap Y \in i_Y \text{ for all } Y \in \mathcal{Y}\} \in Cl^{aff}(X).$$

This holds essentially by definition in case we start with a similarity measure s because, given some $C \subseteq X$, we clearly have $s(a,b) > s(a,c)$ for all $a, b \in C$ and $c \in X - C$ if and only if, for any $Y \in \mathcal{Y}$, the same inequality $s(a,b) > s(a,c)$ holds for all $a, b \in C \cap Y$ and

$c \in Y \backslash C$, that is, if and only if $C \cap Y \in C_{s|_{Y \times Y}}$ holds. If, on the other hand, we start with an $X-$hierarchy $\mathcal{C}$, then clearly $\mathcal{C} \subseteq \bar{i}_X$ holds by definition of $\bar{i}_X$, while $\bar{i}_X \subseteq \mathcal{C}$ - and, hence, equality - holds for the following reasons: first, given some $C \in \bar{i}_X$, we may assume w.l.o.g that $\emptyset \neq C \neq X$ holds, so we can find elements $a \in C$ and $b \in X - C$. Next, for any $x \in C$, we can find some $C' = C'_x \in \mathcal{C}$ with

$$\{a, x\} = C \cap \{a, b, x\} = C'_x \cap \{a, b, x\},$$

that is, with $a, x \in C'_x$ and $b \notin C'_x$. As $\mathcal{C}$ is a hierarchy and we have $a \in C'_x \cap C'_y$ for all $x, y \in C$, we must have $C'_x \subseteq C'_y$ or $C'_y \subseteq C'_x$ for all $x, y \in C$, so the set of these clusters is linearly ordered and, hence, contains a unique largest cluster $C' = C'(b) \in \mathcal{C}$ which still does not contain b, but - in view of $x \in C'_x \subseteq C'(b)$ - contains every element x from C; so, for every $b \in X - C$, we can find some cluster $C'(b) \in \mathcal{C}$ with

$$C \subseteq C'(b) \subseteq X - \{b\}.$$

Hence, we get

$$C = \bigcap_{b \in X - C} C'(b) \in \mathcal{C},$$

as claimed.

Finally, it can also be shown that whenever a $Y-$hierarchy $i_Y \subseteq \mathcal{P}(Y)$ is given for every $Y \in \mathcal{Y}$, the corresponding $X-$hierarchy

$$\bar{i}_X := \{C \subseteq X \mid C \cap Y \in i_Y \text{ for all } Y \in \mathcal{Y}\}$$

satisfies the relation

$$\bar{i}_X|_Y = i_Y$$

for every $Y \in \mathcal{Y}$ if and only if that holds for every subset $Z \subseteq X$ of cardinality at most 4 and the correspondingly defined $Z-$hierarchy

$$\bar{i}_Z := \{C \subseteq Z \mid C \cap Y \in i_Y \text{ for every } Y \in \mathcal{Y} \cap \mathcal{P}(Z)\}$$

(and every $Y \in \mathcal{Y} \cap \mathcal{P}(Z)$): It is obvious that this condition is necessary in view of $\bar{i}_X|_Z \subseteq \bar{i}_Z$ and, hence, $i_Y = \bar{i}_X|_Y = \bar{i}_X|_Z|_Y \subseteq \bar{i}_{Z|Y} \subseteq i_Y$.

To establish the converse, one may define

$$\langle a,b\rangle := \{c \in X \mid c \in \{a,b\} \text{ or } \{a,b\} \notin i_{\{a,b,c\}}\}$$

for all $a, b \in X$ (whether $a \neq b$ or $a = b$) and first note that $Y \in \mathcal{Y}$, $i_Y|'_Y \subseteq i_{Y'}$, for all $Y' \subseteq Y$, and $\langle a,b\rangle \in \bar{i}_X$ for all $a, b \in Y$ will imply $\bar{i}_X|_Y = i_Y$ because we surely have $\emptyset, Y \in \bar{i}_X|_Y$, while every other subset in i_Y is of the form $\{a,b\}$ for some $a, b \in Y$ in which case we have $\{a,b\} = \langle a,b\rangle \cap Y \in \bar{i}_X|_Y$ because $c \in Y - \{a,b\}$ implies $c \notin \langle a,b\rangle$ in view of $\{a,b\} \in i_Y$ and, therefore, $\{a,b\} \in i_Y|_{\{a,b,c\}} \subseteq i_{\{a,b,c\}}$.

Next, note that - vice versa - $\bar{i}_X|_Y = i_Y$ for all $Y \in \mathcal{Y}$ also implies $\langle a,b\rangle \in \bar{i}_X$ for all $a, b \in X$ because the smallest subset C in $\bar{i}_X$ containing a and b (that is, the intersection of all those subsets) surely must contain $\langle a,b\rangle$ in view of $\{a,b\} \subseteq C \cap \{a,b,c\} \in i_{\{a,b,c\}}$ for all $c \in X$ and, therefore, $c \in C$ for all $c \in X$ with $\{a,b\} \notin i_{\{a,b,c\}}$, while it cannot contain any c not in $\langle a,b\rangle$ because for any such c we have $c \notin \{a,b\}$ and $\{a,b\} \in i_{\{a,b,c\}} = \bar{i}_X|_{\{a,b,c\}}$, so there must exist some $C' \in \bar{i}_X$ with $C' \cap \{a,b,c\} = \{a,b\}$ which surely implies $C \subseteq C'$ as well as $c \notin C'$ and, hence, $c \notin C$, as claimed.

So, it remains to show that we have $\langle a,b\rangle \in \bar{i}_X$ for all $a, b \in X$ whenever we have $\langle a,b\rangle \cap Z \in \bar{i}_Z$ for all $Z \subseteq X$ with $a, b \in Z$ and $\#Z \leq 4$ or - equivalent ly - whenever we have $\langle a,b\rangle \cap \{b,c,d\} \in i_{\{b,c,d\}}$ for all $a, b, c, d \in X$. So, assume the latter, pick $a, b \in X$ and consider $\langle a,b\rangle \cap Y$ for some $Y \in \mathcal{Y}$. If $\#(\{a,b\} \cup Y) \leq 4$, then $\langle a,b\rangle \cap (\{a,b\} \cup Y) \in \bar{i}_{\{a,b\}\cup Y}$ by assumption, so we also clearly have

$$\langle a,b\rangle \cap Y = (\langle a,b\rangle \cap (\{a,b\} \cup Y)) \cap Y \in \bar{i}_{\{a,b\}\cup Y|Y} \subseteq i_Y.$$

Otherwise, we have $a \neq b$, $\#Y = 3$ and $a, b \notin Y$. Assume $\langle a,b\rangle \cap Y \notin i_Y$. Then there exist $x, y, z \in X$ with $x \in \langle a,b\rangle \cap Y$, $y \notin \langle a,b\rangle \cap Y$ and $Y = \{x,y,z\}$. From our assumption, we get $\{a,x\} = \langle a,b\rangle \cap \{a,x,y\} \in i_{\{a,x,y\}}$, that is $y \notin \langle a,x\rangle$. Hence, if also $z \notin \langle a,b\rangle \cap Y$, we can replace y by z in the above argument which leads to $z \notin \langle a,x\rangle$ and, hence,

$$\langle a,b\rangle \cap \{x,y,z\} = \{x\} = \langle a,x\rangle \cap \{x,y,z\rangle \in i_{\{x,y,z\}} = i_Y,$$

while, if $z \in \langle a,b\rangle \cap Y$, we can replace x by z in the above argument which leads to $y \notin \langle a,z\rangle$. Hence, as $z \in \langle a,x\rangle$ or $x \in \langle a,z\rangle$ because we cannot simultaneously have $\{a,x\} \in i_{\{a,x,z\}}$ **and** $\{a,z\} \in i_{\{a,x,z\}}$, we have either

$$\langle a,b\rangle \cap Y = \{x,z\} = \langle a,x\rangle \cap Y$$

or we have

$$\langle a,b\rangle \cap Y = \{x,z\} = \langle a,z\rangle \cap Y,$$

so - in any case - we have $\langle a,b\rangle \cap Y \in i_Y$, as claimed.

Another way to deduce the same result is also of some interest: given a family of hierarchies $i_Y(Y \in \mathcal{Y})$, define a tertiary relation $ab|c$ on X by

$$ab|c \Leftrightarrow c \neq a,b \text{ and } \{a,b\} \in i_{\{a,b,c\}}$$

$(a,b,c \in X)$. Clearly, one has

(H1) $ab|c \quad \Leftrightarrow \quad ba|c$

and

(H2) $ab|c \;\&\; ac|d \Rightarrow b \neq d$

for all $a,b,c,d \in X$, the latter because $b = d$ would imply $\#\{a,b,c\} = 3$ as well as well as $\{a,b\} \in i_{\{a,b,c\}}$ and $\{a,c\} \in i_{\{a,b,c\}}$ in contradiction to $\emptyset \neq \{a\} = \{a,b\} \cap \{a,c\} \neq \{a,b\}, \{a,c\}$. And it is also clear that for any ternary relation satisfying (H1) and (H2) the set system $\mathcal{C}$ consisting of all subsets $C \subseteq X$ with $ab|c$ for all $a,b \in C$ and $C \in X - C$ is a hierarchy.

Next observe that in case

$$\langle a,b\rangle \cap \{b,c,d\} \in i_{\{b,c,d\}}$$

holds for all $a,b,c,d \in X$, one also has

(H3) $ab|c \;\&\; ac|d \Rightarrow ab|d$

as well as

(H4) $ab|d \;\&\; ac|d \Rightarrow bc|d$

for all $a,b,c,d \in X$. One can then show that there is a canonical 1-1 correspondence between (a) ternary relations defined on a finite set X satisfying (H1), (H2), (H3) and (H4) and (b) hierarchies $\mathcal{C} \subseteq \mathcal{P}(X)$, given by associating to any such relation - as above - the set system $\mathcal{C}$ consisting of all subsets $C \subseteq X$ with $ab|c$ for all $a, b \in C$ and $c \in X - C$, or - vice versa - to any $X-$hierarchy $\mathcal{C}$ the ternary relation defined by

$$ab|c \Leftrightarrow \text{ there exists some } C \in \mathcal{C} \text{ with } a, b \in C \text{ and } c \in X - C.$$

The fact that this sets up indeed a one-to-one correspondence is almost equivalent to the above-mentioned fact that, for any family $i_Y(Y \in \mathcal{Y})$ of $Y-$hierarchies, we have $\bar{i}_X|_Y = i_Y$ for all $Y \in \mathcal{Y}$ if and only we have $\bar{i}_Z|_Y = i_Y$ for all $Z \subseteq X$ with $\#Z \leq 4$ and all $Y \in \mathcal{Y} \cap \mathcal{P}(Z)$, so any one of these two facts can be deduced easily from the other one.

Altogether, we see that the theory of hierarchies - including the way they are related to similarity measures - fits perfectly well into the conceptual framework developed in the previous section.

2.2 Trees

Given a set X, a tree structure on X is a triple (V, E, ϕ) consisting of a set V - called the set of *vertices* or *nodes* of that tree structure - , a subset E of the set $\mathcal{P}_2(V) := \{e \subseteq V \,|\, \#e = 2\}$ of subsets of V of cardinality 2 - called the set of edges -, and a map $\phi : X \to V$ such that the pair (V, E) is a tree, that is, such that for any two vertices v, v' there exists one and only one pair $(d, p) = (d(v, v'), p(v, v'))$ consisting of a natural number

$$d = d(v, v') = d^E(v, v') \in \mathbb{N}_0$$

- called the (combinatorial or unweighted) *distance* between v and v' (relative to (V, E)) - and a sequence

$$p = p(v, v') = p^E(v, v') = (v_0, v_1, \ldots, v_d)$$

of length $d + 1$ of vertices

$$v_i = p_i(v, v') = p_i^E(v, v')$$

$(i = 0, \ldots, d)$ from V with $v_0 = v, v_d = v'$,

$$e_i = e_i(v, v') = e_i^E(v, v') := \{v_{i-i}, v_i\} \in E$$

for all $i = 1, \ldots, d$ as well as $v_{i-1} \neq v_{i+1}$ for all $i = 1, \ldots, d-1$, called the *path* from v to v' (in the tree (V, E)) (cf. [BD86]). It is easily seen that - with these notations - one must have $\#\{v_0, v_1, \ldots, v_d\} = d + 1$ and $\#\{e_1, \ldots, e_d\} = d$, that the set

$$\Delta(v, v') = \Delta_E(v, v') := \{e_1, \ldots, e_d\}$$

coincides with the intersection of all subsets E' of E for which there exist vertices $w_0 := v, w_1, \ldots, w_k := v'$ from V with $\{w_{i-1}, w_i\} \in E'$ for all $i = 1, \ldots, k$, and that for $v, v', v'' \in V$ one has $p_i(v, v') = p_i(v, v'')$ for some i (assumed to be at most equal to $min(d(v, v'), d(v, v''))$ so that both terms are defined) if and only if i is smaller than or equal to

$$d(v; v', v'') := \#(\Delta(v, v') \cap \Delta(v, v''))$$

and, hence,

$$\Delta(v', v'') = (\Delta(v, v') \cup \Delta(v, v'')) - (\Delta(v, v') \cap \Delta(v, v'')),$$

the symmetric difference between $\Delta(v, v')$ and $\Delta(v, v'')$ A tree structure (V, E, ϕ) defined on X is said to be an $X-$tree if (i) E coincides with its subset $\bigcup_{x,y \in X} \Delta(\phi(x), \phi(y))$ and if (ii) for every vertex $v \in V - \phi(X)$ there exist at least three distinct edges $e_1, e_2, e_3 \in E$ containing v. Equivalently, defining an equivalence relation "$\overset{e}{\sim}$" on X for each $e \in E$ by

$$x \overset{e}{\sim} y \Leftrightarrow e \notin \Delta(\phi(x), \phi(y)),$$

a tree structure (V, E, ϕ) defined on X is an $X-$tree if (i) none of these equivalence relations is trivial in which case the associated set of equivalence classes consists of exactly two (non-empty!) subsets of X and, hence, constitutes in particular a split $S_e \in \mathcal{S}p(X)$, and if (ii) for any two edges $e, e' \in E$ one has

$$S_e = S_{e'} \Leftrightarrow e = e'.$$

It is easily seen and well known that by associating to any tree structure (V, E, ϕ) defined on a set X the triple (V', E', ϕ') defined by

$$\begin{aligned} V' &:= \phi(X) \cup \{v \in V \,|\, 3 \leq \#\{S_e \,|\, v \in e \in E, \#S_e = 2\}\}, \\ E' &:= \Big\{\{v, w\} \in \mathcal{P}_2(V') \,|\, \text{there exist } e, f \in E \\ &\qquad \text{with } \ v \in e, w \in f \text{ and } S_e = S_f \neq \{X\}\Big\}, \end{aligned}$$

and

$$\phi' : X \to V' : x \mapsto \phi(x)$$

is an X–tree which is called the X– tree induced by the tree structure (V, E, ϕ). Clearly $(V, E, \phi) \doteq (V', E', \phi')$ if and only if (V, E, ϕ) is an X–tree.

It is also easily seen and well-known (see for instance [B71] and [BD86]) that, for any two X–structures (V_1, E_1, ϕ_1) and (V_2, E_2, ϕ_2), the following assertions are equivalent:

(A1) *The set* $\mathcal{S}(V_1, E_1, \phi_1) := \{S_{e_1} \,, | \, e_1 \in E_1, \#S_{e_1} = 2\}$ *coincides with the correspondingly defined set* $\mathcal{S}(V_2, E_2, \phi_2)$.

(A2) *There exist maps* $w_1 : E_1 \to \mathbb{R}_{>0}$ *and* $w_2 : E_2 \to \mathbb{R}_{>0}$ *such that for all* $x, y \in X$ *and* $\Delta_1 := \Delta_{E_1}(\phi_1(x), \phi_1(y)$, $\Delta_2 := \Delta_{E_2}(\phi_2(x), \phi_2(y)$ *one has*

$$\sum_{e_1 \in \Delta_1} w_1(e_1) = \sum_{e_2 \in \Delta_2} w_2(e_2).$$

(A3) *For the corresponding induced* X*–trees* (V_1', E_1', ϕ_1') *and* (V_2', E_2', ϕ_2'), *one has*

$$d^{E_1'}(\phi_1'(x), \phi_1'(y)) = d^{E_2'}(\phi_2'(x), \phi_2'(y))$$

for all $x, y \in X$.

(A4) *There exists one (and only one!) bijection* $\alpha : V_1' \xrightarrow{\sim} V_2'$ *between the sets of vertices of the corresponding induced* X*–trees such that* $E_2' = \{\alpha(e_1') \,|\, e_1' \in E_1'\}$ *and* $\phi_2' = \alpha \circ \phi_1'$.

If one and, hence, all of these assertions hold, the tree structures (V_1, E_1, ϕ_1) and (V_2, E_2, ϕ_2) are called *equivalent.*

Clearly, any tree structure (V, E, ϕ) defined on X is equivalent to its induced $X-$tree and that $X-$tree is determined uniquely up to canonical isomorphism by the equivalence class of (V, E, ϕ). It is also well-known that for any $X-$tree (V, E, ϕ) and any two edges $e, e' \in E$, the two associated splits S_e and $S'_e \in \mathcal{S}_p(X)$ are *compatible*, that is, the set $\{A \cap A' \mid A \in S_e, A' \in S_{e'}\}$ contains at most three non-empty subsets of X. And it has been established by P. Buneman (cf. [B71]) that, vice versa, for a finite set X and any set $\mathcal{S} \subseteq \mathcal{S}p(X)$ of splits $S = \{A, B\} \in \mathcal{S}p(X)$ which are pairwise compatible, that is, with

$$\#\{A \cap A' \mid A \in S, A' \in S', A \cap A' \neq \emptyset\} \leq 3$$

for all $S, S' \in \mathcal{S}$, there exists a tree structure (V, E, ϕ) for X with $\mathcal{S}(V, E, \phi) = \mathcal{S}$.

Actually, assuming without loss of generality $\{X, \emptyset\} \notin \mathcal{S}$, an $X-$tree (V, E, ϕ) with that property can be defined as follows: put

$$\begin{aligned} V &:= \{v : \mathcal{S} \to \mathcal{P}(X) \mid v(S) \in S \text{ for all } S \in \mathcal{S} \text{ and } v(S_1) \cap v(S_2) \neq \emptyset \text{ for all } S_1, S_2 \in \mathcal{S}\}, \\ E &:= \{\{v, w\} \in V \mid \#\{S \in \mathcal{S} \mid v(S) \neq w(S)\} = 1\}, \end{aligned}$$

and

$$\phi : X \to V : x \mapsto (v_x : \mathcal{S} \to \mathcal{P}(X) : S \mapsto S(x)),$$

where $S(x) \in S$ denotes that subset A of X contained - as an element - in S which contains x.

Hence, (equivalence classes of) tree structures defined on a finite set X can be identified with subsets $\mathcal{S}$ of the set $\mathcal{S}p(X)$ of all splits of X in which any two splits $S, S' \in \mathcal{S}$ are *compatible*, that is, satisfy the above condition

$$\#\{A \cap A' \mid A \in S, A' \in S', A \cap A' \neq \emptyset\} \leq 3,$$

and none coincides with $\{X, \emptyset\}$.

Consequently, also the theory of tree structures fits nicely into the framework proposed in the previous section: given a finite set X and a subset $\mathcal{S}$ of the set $\mathcal{S}p(X)$ of all splits of X, it is easy to see that $\mathcal{S}$ represents a tree structure of X - that is, one has

$$\#\{A \cap A' \mid A \in S, A' \in S', A \cap A' \neq \emptyset\} \leq 3$$

for all $S, S' \in \mathcal{S}$ - if and only if this holds for $\mathcal{S}|_Y = \{\{A \cap Y, B \cap Y\} \mid \{A, B\} \in \mathcal{S}\}$ for all $Y \subseteq X$ with $\#Y \leq 4$.

Vice versa, putting $\mathcal{Y} := \mathcal{P}_{\leq 4}(X) := \{Y \subseteq X \mid \#Y \leq 4\}$ and choosing, for each $Y \in \mathcal{Y}$, a set $i_Y \subseteq \mathcal{S}p(Y)$ of pairwise compatible splits - including, for the sake of technical simplicity - the trivial split $\{Y, \emptyset\}$ induced by the trivial equivalence relation, the associated element $\bar{i}_X \in Cl^{proj}(X)$ defined by

$$\bar{i}_X := \{\{A, B\} \in \mathcal{S}p(X) \mid \{A \cap Y, B \cap Y\} \in i_Y \text{ for all } Y \in \mathcal{Y}\}$$

always represents a tree structure on X.

It is also easy to see that in case $\#X = n + 1$, one has $\#\mathcal{S} \leq 2n$ for all subsets $\mathcal{S} \subseteq \mathcal{S}p(X)$ of pairwise compatible splits (including possibly the empty split), either by proving this directly or by - after choosing some $x_0 \in X$ to play the rôle of the "point at infinity" - replacing $\mathcal{S} \subseteq \mathcal{S}p(X)$ by

$$\mathcal{C}^{x_0}(\mathcal{S}) := \{A \subseteq X - \{x_0\} \mid \{A, X - A\} \in \mathcal{S}\}$$

and noting that $\mathcal{C}^{x_0}(\mathcal{S}) \cup \{\emptyset, X - \{x_0\}\}$ is necessarily an $(X - \{x_0\})$ - hierarchy whenever any two splits in $\mathcal{S}$ are compatible - actually, any two splits in $\mathcal{S}$ are pairwise compatible if and only if $\mathcal{C}^{x}(\mathcal{S}) \cup \{\emptyset, X - \{x\}\}$ is an $(X - \{x\})$ - hierarchy for every $x \in X$.

To construct an X-tree from local data in this way, the decision about which tree structure to choose for any given small subset $Y \in \mathcal{Y}$ can be based on whatever creed one adheres to: given a distance $d : X \times X \to \mathbb{R}$ defined on X, one might - for any $Y \in \mathcal{Y}$ - take all splits $\{A, B\}$ of Y with

$$d(a, a') + d(b, b') < d(a, b) + d(a', b')$$

for all $a, a' \in A$ and $b, b' \in B$ in which case the corresponding family $\bar{i}_X$ of splits of X is exactly the set $\mathcal{S}_{Buneman}(d)$ of all splits $\{A, B\}$ of X satisfying exactly the same requirement. As indicated by our notation, this set of splits and the corresponding tree structure was considered already by P. Buneman in the paper [B71] mentioned already above which, by the way, was devoted to archeological classification.

If X is a set of sequences $x = (x_1, \dots, x_k)$ whose entries x_i all come from some alphabet $\mathcal{A}$ on which a metric $d_{\mathcal{A}}$ is defined (e.g. the trivial metric $d_{\mathcal{A}}(a, b) := 1 - \delta_{ab}$), there are many alternatives to define $i_Y \in \mathcal{S}p(Y)$ for all $Y \in \mathcal{Y}$: If all sequences are of the same length k, one can define a metric d on X by

$$d(x, y) := \sum_{i=1}^{k} d_{\mathcal{A}}(x_i, y_i)$$

for all $x = (x_1, \dots, x_k), y = (y_1, \dots, y_k) \in X$ and then proceed as above. One may also explore, for each $Y \in \mathcal{Y}$, the very few possible tree structures definable on Y and then decide for the most parsimonious tree structure as the appropriate local input. In general, one will first have to align the sequences according to some alignment score, and one might then use just that score (or a distance measure derived from it) for constructing the Buneman splits (to which end one may actually restrict oneself to pairwise alignment). Instead, one might try to simultaneously construct, for each $Y \in \mathcal{Y}$, the tree structure as well as the alignment - again, say, by exploring all possible tree structures - and then invoke a parsimony or a maximum likelihood principle.

For real data sets, it might actually be advisable to explore all or, at least, quite a few of these alternatives as only those splits can be trusted as being reliable which are observed in many of the resulting $X-$trees; - actually, an extensive literature exists regarding how to construct a *consensus* tree structure from many given ones (cf. for example [DM85]) which could also be evoked at that stage, even though in most practical cases - at least, when it comes to problems in phylogeny - there will be only a few doubtful splits which should rather be discussed individually, taking into account all sorts of arguments and not exclusively only those which are based on formal tree-construction and/or consensus procedures.

To conclude this subsection, we just mention that in analogy to the affine case, given a family $i_Y \in Cl^{proj}(Y)$ $(Y \in \mathcal{Y} = \mathcal{P}_{\leq 4}(X))$ of tree structures, there exists a tree structure $i_X \in Cl^{proj}(X)$ with $i_X|_Y = i_Y$ for all $Y \in \mathcal{Y}$ if and only if this holds for any subset Z of X of cardinality at most 5, that is, if and only if for any subset Z of X with $\#Z \leq 5$ there exist a tree structure $i_Z \in Cl^{proj}(Z)$ with $i_Z|_Y = i_Y$ for all $Y \in \mathcal{Y} \cap \mathcal{P}(Z)$

(cf. [BD86]).

2.3 Weak Hierarchies and Weakly Compatible Split Systems

Next, I want to point out that also the theory of weak hierarchies and weakly compatible split systems as developed in [BD89] and [BD92] fits nicely into the above framework. According to [BD89], a *weak hierarchy* $\mathcal{C}$ defined on a set X is a subset of $\mathcal{P}(X)$ such that $C_1 \cap C_2 \cap C_3 \in \{C_1 \cap C_2,\ C_2 \cap C_3,\ C_3 \cap C_1\}$ holds for all $C_1, C_2, C_3 \in \mathcal{C}$.

It follows that a subset $\mathcal{C}$ of $\mathcal{P}(X)$ is a weak hierarchy defined on X if and only if $\mathcal{C}|_Y := \{C \cap Y \mid C \in \mathcal{C}\}$ is a weak hierarchy defined on Y for every $Y \in \mathcal{Y} := \mathcal{P}_{\leq 3}(X)$ which in turn is the case if and only if $\mathcal{C}|_Y$ does not contain all three subsets of Y of cardinality 2 for every $Y \subseteq X$ of cardinality 3. Hence, given a weak hierarchy $i_Y \subseteq \mathcal{P}(Y)$ for every $Y \in \mathcal{Y}$, the corresponding subset $\bar{i}_X = \{C \subseteq X \mid C \cap Y \in i_Y \text{ for } Y \in \mathcal{Y}\} \subseteq \mathcal{P}(X)$ of $\mathcal{P}(X)$ will always be a weak hierarchy, too. Moreover, it follows easily from adapting the first (rather trivial) part of an argument presented in **2.1** to this situation, that now we have $res_{X \to Y}(\bar{i}_X) = i_Y$ for all $Y \in \mathcal{Y}$ in case (i) every weak hierarchy i_Y contains Y and the empty set among its clusters and is closed with respect to intersection and (ii) we have $res_{Z \to Y}(\bar{i}_Z) = i_Y$ for all $Z \subseteq X$ of cardinality at most 5 and all $Y \in \mathcal{Y} \cap \mathcal{P}(Z)$ (with $\bar{i}_Z := \{C \subseteq Z \mid C \cap Y \in i_Y \text{ for all } Y \in \mathcal{Y} \cap \mathcal{P}(Z)\}$, of course).

That it is not enough to require condition (ii) for all subsets $Z \subseteq X$ of cardinality at most 4 can be deduced from the following simple (counter)example:
put $X := \{1, 2, 3, 4, 5\}$, put $i_Y := \mathcal{P}(Y)$ if $\mathcal{Y} \leq 2$, and in case $\#Y = 3$ put

$$i_Y := \begin{cases} \mathcal{P}(Y) \backslash \{\{1,2\},\{3,4\}\} & \text{if } 5 \notin Y, \\ \{Y\} \cup \mathcal{P}_{\leq 1}(Y) & \text{if } Y = \{3,4,5\}, \\ \mathcal{P}(Y) - \{\{a,5\} \mid a \in Y - \{5\}\} & \text{else.} \end{cases}$$

It is easy to see that this implies

$$\begin{aligned} \bar{i}_{\{1,2,3,4\}} &= \{\{1,2,3,4\}\} \cup \mathcal{P}_{\leq 2}(\{1,2,3,4\}) - \{\{1,2\},\{3,4\}\}, \\ \bar{i}_{\{1,2,a,5\}} &= \{\{1,2,a,5\},\{1,2,a\},\{1,a\},\{2,a\}\} \cup \mathcal{P}_{\leq 1}(\{1,2,a,5\}) \end{aligned}$$

for $a = 3, 4$ and

$$\bar{i}_{\{a,3,4,5\}} = \{\{a,3,4,5\},\{a,3\},\{a,4\}\} \cup \mathcal{P}_{\leq 1}(\{a,3,4,5\})$$

for $a = 1, 2$ from which formulae one can easily conclude that condition (ii) is fulfilled in this example for all $Z \subseteq X$ with $\#Z \leq 4$. Yet, there can't be a subset $C \subseteq X$ with $C \in \bar{\imath}_X$ (that is, with $C \cap Y \in i_Y$ for all $Y \subseteq X$ with $\#Y \leq 3$) as well as $C \cap \{1, 2, 5\} = \{1, 2\}$ because any $C \in \bar{\imath}_X$ with $\{1, 2\} \subseteq C$ must intersect $\{1, 2, a\}$ $(a \in \{3, 4\})$ in the only subset in $i_{\{1,2,a\}}$ containing $\{1, 2\}$ which is $\{1, 2, a\}$ itself, so we must have $3, 4 \in C$ which implies that C must intersect $\{3, 4, 5\}$ in the only subset in $i_{\{3,4,5\}}$ containing $\{3, 4\}$ which is $\{3, 4, 5\}$ itself. So, we must have $5 \in C$ and, hence, $C \cap \{1, 2, 5\} \neq \{1, 2\}$, as claimed.

It is easy to see that one has $\#\mathcal{C} \leq \#\mathcal{P}_{\leq 2}(X) = \binom{n}{2} + \binom{n}{1} + \binom{n}{0}$ for every weak hierarchy defined on a set X of cardinality n: Indeed this follows easily from the fact that for any non-empty cluster C in a weak hierarchy $\mathcal{C}$ there exist $a, b \in C$ with $C \subseteq C'$ for all $C' \in \mathcal{C}$ with $a, b \in C'$ because otherwise there would exist a smallest subset T of C of cardinality > 2 and $C \subseteq C'$ for every $C' \in \mathcal{C}$ with $T \subseteq C'$, so for any three distinct elements $a_1, a_2, a_3 \in T$ there would exist, for each $i \in \{1, 2, 3\}$, some cluster $C_i \in \mathcal{C}$ with $a_i \notin C_i$ and $T - \{a_i\} \subseteq C_i$, in contradiction to $C_1 \cap C_2 \cap C_3 \in \{C_1 \cap C_2,\ C_2 \cap C_3,\ C_3 \cap C_1\}$ for all $C_1, C_2, C_3 \in \mathcal{C}$.

Next, it is obvious that for any weak hierarchy $\mathcal{C} \subseteq \mathcal{P}(X)$ we have $\mathcal{C} \subseteq \bar{\imath}_X$ for the weak hierarchy $\bar{\imath}_X$ associated with the family $i_Y := \mathcal{C}|_Y$ $(Y \in \mathcal{Y})$, and that equality implies that $\mathcal{C}$ is closed with respect to intersection provided that that holds for all $i_Y\,(Y \in \mathcal{Y})$. More precisely (see below), it can be shown that $\bar{\imath}_X$ is always contained in the smallest intersection-closed subset $\hat{\mathcal{C}}$ of $\mathcal{P}(X)$ containing $\{X\} \cup \mathcal{C}$ (which - for a weak hierarchy $\mathcal{C}$ - is easily seen to coincide with $\{X\} \cup \{C_1 \cap C_2 \,|\, C_1, C_2 \in \mathcal{C}\}$) and, hence, that $\hat{\mathcal{C}}$ always coincides with the weak hierarchy $\bar{\jmath}_X$ associated with the family $j_Y := \hat{\imath}_Y$ (with $\hat{\imath}_Y$, of course, denoting the smallest intersection-closed subset of $\mathcal{P}(Y)$ containing $\{Y\} \cup i_Y$). Note also that i_Y will always coincide with $\hat{\imath}_Y = j_Y$ provided $\mathcal{C}$ contains X as well as every subset $C \subseteq X$ of cardinality ≤ 1; so in this case, we will always have $\bar{\imath}_X = \hat{\mathcal{C}}$.

Finally, given a symmetric map $s : X^2 \to \mathbb{R}$, the set system

$$\mathcal{H}_s = \mathcal{H}_s(C) := \{C \subseteq X \mid s(a,b) > min(s(a,c)s(b,c)) \text{ for all } a,b \in C \text{ and } c \in X - C\}$$

always is an intersection-closed weak hierarchy (which was the starting point for their discussion in [BD89]). Clearly, $\mathcal{H}_s$ is the weak hierarchy $\bar{i}_X$ associated with the family of weak hierarchies $i_Y := \mathcal{H}_s(C|_{Y\times Y})$ $(Y \in \mathcal{Y})$. So, as in the case of hierarchies, there is a simple way to construct local information in the form of weak hierarchies from other local information (here given in the form of the map s) and to derive the desired global information directly from these locally defined weak hierarchies without further recourse to any other form of local information from which these locally defined weak hierarchies might have been deduced.

Similarly (cf. [BD92]), weakly compatible split systems $\mathcal{S}$ are defined to be those subsets $\mathcal{S}$ of $\mathcal{S}p(X)$ for which no three splits

$$S_1 = \{A_1, B_1\},\ S_2 = \{A_2, B_2\},\ S_3 \in \{A_3, B_3\} \in \mathcal{S}$$

with $A_1 \cap A_2 \cap A_3 \neq \emptyset$, $A_1 \cap B_2 \cap B_3 \neq \emptyset$, $B_1 \cap A_2 \cap B_3 \neq \emptyset$ and $B_1 \cap B_2 \cap A_3 \neq \emptyset$ exist or, equivalenty, for which

$$\mathcal{C}^x(\mathcal{S}) := \{A \subseteq X - \{x\} \mid \{A, X - A\} \in \mathcal{S}\}$$

is a weak hierarchy defined on $X - \{x\}$, for every $x \in X$. So, we have $\#\mathcal{S} \leq \binom{n}{2} + \binom{n}{1} + \binom{n}{0} = \binom{n+1}{2} + 1 = \binom{\#X}{2} + 1$ for every family $\mathcal{S}$ of weakly compatible splits (including possibly the trivial split $\{X, \phi\}$) defined on a set X of cardinality $n + 1$.

It follows easily from the results in [BD92] and [BD93] - or from the results regarding weak hierarchies just reported above and the relation between (affine) weak hierarchies and (projective) weakly compatible split systems - that also the theory of weakly compatible split systems fits excellently into the conceptual framework developed in §1. In particular, such split systems can also be viewed as resulting from the proposed "standard" procedure of extracting globally relevant from locally distributed information, provided that that locally distributed information conforms to some rather mild and easily specified requirements.

A certain generalization of weak hierarchies was discussed in [BD94], where a collection $\mathcal{C}$ of subsets of X was defined to be a *weak hierarchy of breadth at most* k if for all $C_1, C_2, \ldots, C_{k+1} \in \mathcal{C}$ one has

$$C_1 \cap C_2 \cap \cdots \cap C_{k+1} \in \{\bigcap_{i \neq j} C_i \mid j = 1, \ldots, k+1\}$$

or, equivalently, if there exist no clusters $C_1, \ldots, C_{k+1} \in \mathcal{C}$ and elements $x_1, \ldots, x_{k+1} \in X$ with $x_i \in C_j$ if and only if $i \neq j$. Clearly, $\mathcal{C} \subseteq \mathcal{P}(X)$ is a weak hierarchy of breadth at most k if and only if there exists at least one subset of cardinality k in any subset $Y \subseteq X$ of cardinality $k+1$ which is not contained in $\mathcal{C}|_Y = \{C \cap Y \mid C \in \mathcal{C}\}$ - that is, if and only if $\mathcal{C}|_Y$ is a weak hierarchy of breadth at most k for every $Y \in \mathcal{P}_{\leq k+1}(X)$. Equivalently, $\mathcal{C} \subseteq \mathcal{P}(X)$ is a weak hierarchy of breadth at most k if and only if one has

$$\mathcal{P}(Y) = \{C_1 \cap C_2 \cap \cdots \cap C_\ell \cap Y \mid \ell \in \mathbb{N}_0,\ C_1, \ldots, C_\ell \in \mathcal{C}\}$$

for some $Y \subseteq X$ only if $\#Y \leq k$. As above, this implies easily

$$\#\mathcal{C} \leq \#\mathcal{P}_{\leq k}(X) = \binom{n}{k} + \binom{n}{k-1} + \binom{n}{k-2} + \cdots + \binom{n}{0}$$

for any such $\mathcal{C} \subseteq \mathcal{P}(X)$ (and $n := \#X$) as any $C \in \mathcal{C}$ necessarily contains a subset T of cardinality at most k with $C \subseteq C'$ for every $C' \in \mathcal{C}$ with $T \subseteq C'$.

Note also (cf. [BD94]) that a collection $\mathcal{C} \subseteq \mathcal{P}(X)$ of subsets of X is a weak hierarchy of breadth at most k if and only if, for any map $s : \mathcal{C} \to \mathbb{R}$ with $s(C) \leq s(C')$ for all $C, C' \in \mathcal{S}$ with $C \supseteq C'$, one has $\#A \leq k$ for any subset $A \subseteq X$ with

$$max(s(C) \mid A \subseteq C \in \mathcal{C}) < max(s(C') \mid A - \{a\} \subseteq C' \in \mathcal{C})$$

for all $a \in A$.

And it is also easy to see that given a weak hierarchy $i_Y \subseteq \mathcal{P}(Y)$ of breadth at most k for any subset Y of X of cardinality at most $k+1$ which is closed with respect to intersection and contains Y as well as the empty set $\emptyset$, we have $res_{X \to Y}(\bar{i}_X) = i_Y$ for all $Y \in \mathcal{P}_{\leq k+1}(X)$ if and only if we have $res_{Z \to Y}(\bar{i}_Z) = i_Y$ for all $Z \subseteq X$ of cardinality at most $2k+1$ and all $Y \in \mathcal{Y} \cap \mathcal{P}(Z)$ (with $\bar{i}_Z$ defined relative to the i_Y for all Z in X just as above, of course).

And finally, given a weak hierarchy $\mathcal{C} \subseteq \mathcal{P}(X)$ of breadth at most k, we always have $\mathcal{C} \subseteq \bar{\imath}_X \subseteq \hat{\mathcal{C}}$ for the weak hierarchy $\bar{\imath}_X$ of breadth at most k associated with the family $i_Y := \mathcal{C}|_Y$ $(Y \in \mathcal{P}_{\leq k+1}(X))$ and the smallest intersection-closed subset $\hat{\mathcal{C}}$ of $\mathcal{P}(X)$ containing $\{X\} \cup \mathcal{C}$ which in this case coincides with

$$\{X\} \cup \{C_1 \cap \cdots \cap C_k | C_1, \cdots, C_k \in \mathcal{C}\}.$$

Moreover, $\bar{\imath}_X = \hat{\mathcal{C}}$ holds if and only if i_Y contains Y and is closed with respect to pairwise intersection for each $Y \in \mathcal{P}_{\leq k+1}(X))$ which in turn is surely the case if X and all subsets of cardinality at most $k-1$ belong to $\mathcal{C}$.

To prove these statements, it is enough to show that $\mathcal{C} = \bar{\imath}_X$ holds if $\mathcal{C}$ coincides with $\hat{\mathcal{C}}$. Otherwise, there would exist some subset $A \subseteq X$ with $A \in \bar{\imath}_X \backslash \mathcal{C}$ and, hence, there would also exist some smallest subset Z of X with $A \cap Z \notin \mathcal{C}|_Z$. It follows that for any $z \in Z$, there would exist some $C_z \in \mathcal{C}$ with $A \cap (Z - \{z\}) = C_z \cap (Z - \{z\})$ as well as $A \cap Z \neq C_z \cap Z$ and, hence, with

$$C_z \cap Z = \begin{cases} (A \cap Z) - \{z\} \text{ if } z \in A \cap Z, \\ (A \cap Z) \cup \{z\} \text{ else.} \end{cases}$$

While the first assertion implies $\#(A \cap Z) \leq k$ in view of the assumption that $\mathcal{C}$ is a weak hierarchy of breadth at most k, the second assertion implies $(Z \backslash A) \leq 1$ because $z_1, z_2 \in Z \backslash A$ and $z_1 \neq z_2$ would imply $C_{z_1} \cap C_{z_2} \cap Z = A \cap Z$ in contradiction to $C_{z_1} \cap C_{z_2} \in \mathcal{C}$ and our choice of A and Z. So, we would have $\#Z \leq k+1$, now in contradiction to the fact that this implies $A \cap Z \in i_Z = \mathcal{C}|_Z$ for all $A \in \bar{\imath}_X$ in viewe of the definition of $\bar{\imath}_X$.

3 Some Upper Bounds Regarding More General Cluster Systems

Finally, we want to justify the rather general framework for clustering theory presented in the first section by establishing the following rather general result which implies most of the inequalities mentioned above:

Theorem. *Given a collection $\mathcal{C} \subseteq \mathcal{P}(X)$ of subsets of a finite set X and a simplicial complex $\mathcal{X} \subseteq \mathcal{P}(X)$ of subsets of X (that is, a collection of subsets of X which is closed relative to taking subsets, that is, $A \subseteq B$ and $B \in \mathcal{X}$ implies $A \in \mathcal{X}$), such that every subset A of X with $\mathcal{P}(A) = \{A \cap C \mid C \in \mathcal{C}\}$ is contained in $\mathcal{X}$, then we have*

$$\#\mathcal{C} \leq \#\mathcal{X}.$$

Proof: For each $C \subseteq X$, consider the map

$$\chi_C : \mathcal{X} \to \mathbb{R} : A \mapsto (-1)^{\#A \cap C}.$$

All we need to show is that the maps χ_C with C in $\mathcal{C}$ are linearly independent which we do by induction on $n := \#X$. Clearly, our assumption holds in case $n = 0$. So assume that we have coefficients $r_C \in \mathbb{R}$ $(C \subseteq X)$ with $r_C = 0$ for $C \notin \mathcal{C}$ and $\sum_{C \subseteq X} r_C \chi_C(A) = 0$ for all $A \in \mathcal{X}$. Choose some arbitrary $x \in X$, consider $\mathcal{C}_x := \{C \backslash \{x\} \mid C \in \mathcal{C}\} \subseteq X - \{x\}$ and note that $\{Y \cap C \mid C \in \mathcal{C}_x\} = \{Y \cap C \mid C \in \mathcal{C}\}$ for all $Y \subseteq X - \{x\}$, so we have $Y \in \mathcal{X}_x := \{A \subseteq X - \{x\} \mid A \in \mathcal{X}\}$ for all $Y \subseteq X - \{x\}$ with $\{Y \cap C \mid C \in \mathcal{C}_x\} = \mathcal{P}(Y)$. As $\sum_{C \subseteq X} r_C \chi_C(A) = 0$ for all $A \in \mathcal{X}$ implies $\sum_{C \subseteq X - \{x\}} (r_C + r_{C \cup \{x\}}) \chi_C(A)$ for all $A \in \mathcal{X}_x$ and as $r_C + r_{C \cup \{x\}} = 0$ whenever $C \notin \mathcal{C}_x$, our induction hypothesis implies

$$r_C = -r_{C \cup \{x\}}$$

for all $x \in X$ and $C \subseteq X - \{x\}$. Hence, if $r_{C_0} \neq 0$ for one subset $C_0 \subseteq X$, a simple induction argument based on the cardinality of $(C \backslash C_0) \cup (C_0 \backslash C)$ would imply $r_C \neq 0$ for all $C \subseteq X$ and therefore $\mathcal{C} = \mathcal{P}(X)$ which in turn would imply $\mathcal{X} = \mathcal{P}(X)$ and, therefore, $\sum_{C \in \mathcal{C}} r_C \chi_C(X) = 0$ with

$$r_C = (-1)^{\#C} r_\emptyset \neq 0$$

for all $C \subseteq X$ in contradiction to

$$\sum_{C \in \mathcal{C}} r_C \chi_C(X) = \sum_{C \in \mathcal{C}} r_\emptyset \cdot (-1)^{\#C} \cdot (-1)^{\#(C \cap X)} = 2^n \cdot r_\emptyset \neq 0. \qquad \square$$

This result - and its "projective" analogue - clearly presents a far-reaching generalization of (some of) the results recalled in the previous section. It shows that we should be able to compute in reasonable time the cluster systems $\bar{i}_X$ related to systems of "local" information $i_Y (Y \in \mathcal{Y})$ provided the simplicial complex $\mathcal{X}$ consisting of all $A \subseteq X$ with $\bar{i}_A = \mathcal{P}(A)$ is of a reasonable size (with $\bar{i}_A := \{A' \subseteq A \mid A' \cap Y \in i_Y \text{ for all } Y \in \mathcal{Y}\}$, as above).

We will not work out the consequences of this result here. Rather, we refer to forthcoming papers with H.J. Bandelt, V. Chepoi and J. Koolen where in particular those *extremal* collections $\mathcal{C}$ of subsets of X will be studied for which $\#\mathcal{C} = \#\mathcal{X}$ holds for the simplicial complex $\mathcal{X} = \mathcal{X}(\mathcal{C})$ which consists of all subsets $Y \subseteq X$ with $\mathcal{P}(Y) = \{C \cap Y \mid C \in \mathcal{C}\}$, - just noting that, in case a collection $\mathcal{C}$ of subsets of X is closed with respect to intersection and contains X, it satisfies the condition $\#\mathcal{C} = \#\mathcal{X}(\mathcal{C})$ if and only if it is a convex geometry as defined in [ED85] (or - equivalently - an anti matroid).

Rather, we close with the following observation: As much stronger bounds hold for hierarchies than for weak hierarchies, one might be tempted to believe that given a subset $i_Y \subseteq \mathcal{P}(Y)$ for, say, each $Y \in \mathcal{P}_{\leq 3}(X)$ with $\# i_Y \leq \#\mathcal{P}(Y) - 2$ for each $Y \in \mathcal{P}_3(X)$, the cardinality of the resulting set $\bar{i}_X := \{A \subseteq X \mid A \cap Y \in i_Y \text{ for all } Y \in \mathcal{P}_{\leq 3}(X)\}$ should also be considerably smaller than that of $\mathcal{P}_{\leq 2}(X)$, the upper bound we get immediately from the above theorem. Yet the example $X := \{1, 2, \dots, 2n\}$ and

$$i_Y := \begin{cases} \mathcal{P}(Y) \text{ if } \#Y \leq 2 \\ \{\emptyset\} \cup \{\{i,j\} \subseteq Y \mid i,j \leq n \text{ or } n+1 \leq i,j\} \text{ if } \#Y = 3 \end{cases}$$

which leads to

$$\bar{i}_X = \{\emptyset\} \cup \{\{i,j\} \subseteq X \mid i,j \leq n \text{ or } n+1 \leq i,j\}$$

shows that such an expectation would not be justified.

Instead, given a simplicial complex $\mathcal{X}$ of subsets of X, one might define a system $\mathcal{C}$ of subsets of X to be an $\mathcal{X}$-hierarchy if $C_1 \cap C_2 \in \mathcal{X}$ holds for all $C_1, C_2 \in \mathcal{C}$ for which neither $C_1 \subseteq C_2$ nor $C_2 \subseteq C_1$ holds. Clearly, if $\mathcal{C}$ is an $\mathcal{X}$-hierarchy, then so is any subset of $\mathcal{C}$ as well as the set $\mathcal{C} \cup \{X\} \cup \mathcal{X}^*$ with $\mathcal{X}^*$ defined by

$$\mathcal{X}^* := \{A \subseteq X \mid \text{every proper subset of } A \text{ belongs to } \mathcal{X}\};$$

so in particular, the smallest subset $\hat{\mathcal{C}}$ of $\mathcal{P}(X)$ containing a given $\mathcal{X}$-hierarchy $\mathcal{C}$ and being closed with respect to intersection is always an $\mathcal{X}$-hierarchy. It is also clear that hierarchies are just $\{\emptyset\}$-hierarchies. And it follows easily from the above result regarding hierarchies that we have

$$\#(\mathcal{C}\backslash\mathcal{X}^*) < \#\mathcal{A}$$

for the set

$$\mathcal{A} := \{A \subseteq X \mid A \notin \mathcal{X} \text{ and } A - \{a\} \in X \text{ for all but (at most) one } a \in A\},$$

because associating to any $C \in \mathcal{C}\backslash\mathcal{X}^*$ the subset $\mathcal{A}(C) := \mathcal{P}(C) \cap \mathcal{A}$ of $\mathcal{A}$ produces an $\mathcal{A}$-hierarchy $\mathcal{C}' := \{\mathcal{A}(C) \mid C \in \mathcal{C}\}$ whose cardinaslity $\#\mathcal{C}'$ coincides with that of $\mathcal{C}\backslash\mathcal{X}^*$ in view of $C = \bigcup_{A \in \mathcal{A}(C)} A$ for all $C \in \mathcal{C}\backslash\mathcal{X}^*$, which consists of subsets of $\mathcal{A}$ each containing at least two distinct elements from $\mathcal{A}$.

Of course, the hierarchy $\mathcal{C}'$ is in general far from being binary and, hence, its cardinality will be *considerably* smaller than that of $\mathcal{A}$. Yet, the example $X := \mathbb{P}_d(\mathbb{F}_2) := \mathbb{F}_2^{d+1} - \{0\}$, the d-dimensional projective space over $\mathbb{F}_2$, and $\mathcal{C} := \{U - \{0\} \mid U \text{ a subspace of } \mathbb{F}_2^{d+1} \text{ of dimension } 2\}$, the set of lines in $\mathbb{P}_d(\mathbb{F}_2)$, provides an example of an $\mathcal{X}$-hierarchy for $\mathcal{X} := \mathbb{P}_{\leq 1}(X)$ of cardinality $(2^{d+1}-1)(2^d-1)/3$ defined on a set of cardinality $2^{d+1}-1$ which shows that at least the order of magnitude of $\mathcal{C}\backslash\mathcal{X}^*$ is described correctly by the above bound. Still, it is probably quite an interesting problem to study the extremal $\mathcal{X}$-hierarchies $\mathcal{C}$ (that is, those $\mathcal{X}$-hierarchies $\mathcal{C}$ which have the largest possible cardinality among all $\mathcal{X}$-hierarchies) in some detail, - at least in the case of "k-hierarchies", that is the $\mathcal{P}_{\leq k}(X)$-hierarchies, for which the above - and surely improvable - bound gives

$$\#(\mathcal{C}\backslash\mathcal{P}_{\leq k+1}(X)) < \binom{n}{k+1}.$$

Still more generally, for any two simplicial complexes $\mathcal{X}_1$ and $\mathcal{X}_2$ consisting of subsets of X, we may define a cluster system $\mathcal{C}$ contained in $\mathcal{P}(X)$ to be an $(\mathcal{X}_1, \mathcal{X}_2)$-hierarchy if and only if for all $k \in \mathbb{N}$ and $C_1, \ldots, C_k \in \mathcal{C}$ we have (non-exclusively) either (i) $C_1 \cap \cdots \cap C_k \in \mathcal{X}_1$ or (ii) $C_1 \cap \cdots \cap C_k = C_1 \cap \cdots \cap C_{i-1} \cap C_{i+1} \cap \cdots \cap C_k$ for some $i \in \{1, \ldots, k\}$ or (iii) $\{a_1, \ldots, a_k\} \in \mathcal{X}_2$ for all $a_1, \ldots, a_k \in X$ with $a_i \in C_j$ if and only if $i \neq j$, for all $i, j = 1, \ldots, k$. In particular, we may define $\mathcal{C}$ to be a (k, ℓ)-hierarchy for any two integers $k, \ell \geq -1$ if and only $\mathcal{C}$ is an $(\mathcal{P}_{\leq k}(X),\ \mathcal{P}_{\leq \ell}(X))$-hierarchy, that is if and only if for all $C_1, \ldots, C_{\ell+1} \in \mathcal{C}$ we have $\#(C_1 \cap C_2 \cap \cdots C_{\ell+1}) \leq k$ or $C_1 \cap C_2 \cap \cdots \cap C_{\ell+1} = C_1 \cap \cdots \cap C_{i-1} \cap C_{i+1} \cap \cdots \cap C_{\ell+1}$ for some $i \in \{1, \ldots, \ell+1\}$. Using this terminology, it is easy to see that a hierarchy $\mathcal{C}$ as defined in **2.1** is just a $(0,1)$-hierarchy, while an $\mathcal{X}$-hierarchy is an $(\mathcal{X}, \mathcal{P}_{\leq 1}(X))$-hierarchy, a weak hierarchy is a $(-1, 2)$-hierarchy, and a weak hierarchy of breadth at most ℓ is a $(-1, \ell)$-hierarchy. Note also that the almost obvious fact that every hierarchy is a weak hierarchy (which actually - ten years ago - presented the motivation for naming them that way) now generalises to the simple lemma that every (k, ℓ)-hierarchy is a $(k-1, \ell+1)$-hierarchy.

It is left to the interested reader to establish that a cluster system $\mathcal{C} \subseteq \mathcal{P}(X)$ is a (k, ℓ)-hierarchy if and only if $\mathcal{C}|_Y$ contains at most ℓ subsets of cardinality $k + \ell + 1$ for any subset Y of X of cardinality $k + \ell + 2$, to find useful upper bounds regarding the number of clusters in an $(\mathcal{X}_1, \mathcal{X}_2)$-hierarchy $\mathcal{C}$ by using the theorems proved above or to search for even better bounds as well as to translate all that from the affine to the projective case. All that I wanted to establish (and hope to have established by now) is that viewing clustering techniques from the point of view proposed in the first section of this note, does not only allow us one to put a large body of known results into a uniform conceptual framework but also leads to a considerable number of new and interesting results and data-analysis tools.

Acknowledgement

This paper was written while the author was hosted by the Biomathematics Research Centre at the Mathematics Department, University of Canterbury, New Zealand, which is gratefully acknowledged. Many thanks go in particular to Ann Tindall for her careful typesetting and to Mike Steel for many useful comments. Many thanks go also to Boris Mirkin for his steadfast encouragements to actually work out and put down the ideas presented here without which encouragments this note would probably never have been written.

References

[B71] P. Buneman, *The recovery of trees from measures of dissimilarity*, In: Mathematics in the Archaeological and Historical Sciences, F.Hodson, D. Kendall, and P. Tautu (Ed.s) Edinburgh University Press, Edinburgh, 1971, pp.387-395.

[G87] A.D. Gordon, (1987) *A review of hierarchical classification*, J. Royal Statistical Society Ser. A 150, 119-137.

[DM85] W.H.E. Day and F.R. McMorris, (1985), A formalization of consensus index methods, *Bull. Math. Biol.* **47**, 215-229.

[ED85] P.H. Edelman and R. Jamison, The Theory of Convex Geometries, *Geom. Dedicata* **21** (1985), 247-274.

[BD86] H.-J. Bandelt and A. Dress, *Reconstructing the shape of a tree from observed dissimilarity data.* Advances in Applied Mathematics, Vol. 7 (1986), pp. 309-343.

[BD92] H.-J. Bandelt and A. Dress, *A canonical decomposition theory for metrics on a finite set*, Advances in Mathematics, 92(1992), pp.47-105.

[BD94] H.-J. Bandelt and A. Dress, An order theoretic framework for overlapping clustering, *Discrete Mathematics*, **136**, (1994), 21-37.

[BD93] H.-J. Bandelt and A. Dress, A relational approach to split decomposition; *Information and Classification*, O. Opitz, B. Lausen, R. Klar editors, Berlin; Springer-Verlag, (1993), 123-131.

[M97] B. Mirkin, this volume