Additive-Tree Representations

Hervé Abdi

*Université de Bourgogne & University of Texas at Dallas**

PURPOSE OF ADDITIVE-TREE REPRESENTATIONS

Additive-trees are used to represent objects as "leaves" on a tree, so that the distance on the tree between two leaves reflects the similarity between the objects. Formally, an observed similarity δ is represented by a tree-distance d. As such, additive-trees belong to the descriptive multivariate statistic tradition. Additive-tree representations useful in a wide variety of domains as illustrated by the following three examples of (1) filliation of manuscripts, (2) psychological similarity between animal terms, and (3) phylogenetic trees. These examples are briefly described in the next subsections.

Filiation of Manuscripts (Buneman, 1971)

This application was suggested (but not actually performed) by Buneman (1971) in a seminal paper that laid the foundations for additive-tree representations. Here, the objects to be represented were manuscript copies of the same text (e.g., as in the medieval tradition).

The problem is to infer from the set of texts the *"family tree"* that generated the variants (i.e., which manuscript is copied from which other one; are there lost copies? etc.). The similarity between texts can be defined for example as the number of "common errors", or more simply as the number of common words.

---

* Correspondence about this paper should be adressed to: Hervé Abdi, The University of Texas at Dallas, Program in Cognition, MS:GR.4.1., Richardson, TX75083-0688, USA. e-mail: `herve@utdallas.edu`. The author wishes to thank Sue Viscuso and Alice O'Toole for help and comments on previous drafts.

Ref: Abdi, H. (1990). Additive-tree representations. *Lecture Notes in Biomathematics*, **84**, 43–59.

Examples of filiation data are given in the volume edited by Hodson, Kendall & Tartu (1971) in which the Buneman paper has been published. Another example of this line of investigations can be found in Abdi (1985, 1989b).

Psychological similarity between animal terms

Henley (1969) conducted several experiments on semantic memory structure. She obtained from 21 subjects an estimation of the "subjective distance" between 12 animal terms. Subjects were asked to list from memory the animal names they knew.

For each animal pair the number of animals separating the pair was divided by the list length. For example, suppose that the "cow-mouse" pair was separated by 7 animals, and that the total number of animals given was 20, then the value attached to the pair "cow-mouse" will be 7/20=.35. This procedure was repeated for each subject. Then, the value for each pair was collapsed across subjects to obtain the average similarity. Abdi, Barthélemy & Luong (1984) used a tree (percentage of explained variance: 73) to represent this data matrix. The tree is displayed in Figure 1.

Phylogenetic trees

The literature concerning phylogenetic trees is very large; some influencial papers are Phipps (1971), Farris (1973), Waterman, Smith, Singh & Beyer (1977).

Biologists have often tried to describe the relation between contemporary species. The similarity between species is defined, for example, as the number of identical sequences for some protein or for the DNA, etc. The leaves represent the actual species, and (as in the filiation problem of Buneman) the interior vertices can be interpreted as "missing links" or common ancestors.

In general, biologists focus their interest more on the *shape* of the tree rather than on the *distance* between vertices of the tree, because it is more important in this context to assess the existence of common ancestors for some species rather than to suggest when the separation of the species did occur.

An example of phylogenetic tree is given in Figure 2 (cf. Dress & Krüger, 1987). Because different similarity measures can suggest different patterns of evolution one problem faced by biologists is to compare different trees obtained from a same set of objects. A related problem is to define a tree that expresses the consensus among different trees (cf. Bobisud & Bobisud; Robinson, 1971; Robinson & Fould, 1981).

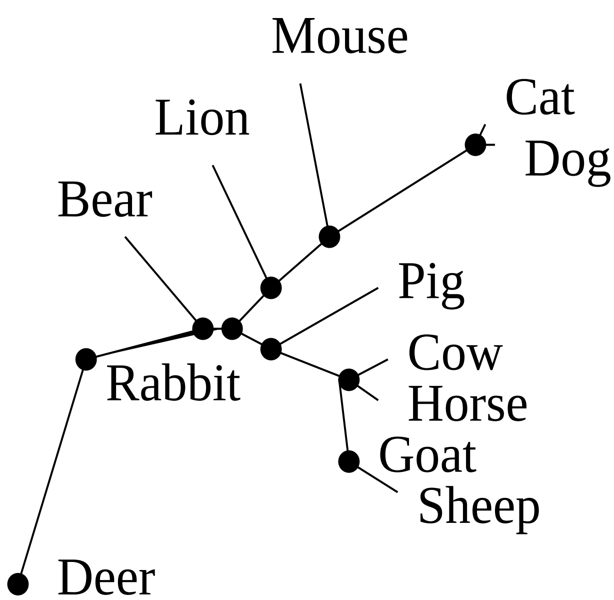

FIG. 1: Additive-tree representation of Henley's data (1969), from Abdi *et al* (1984)

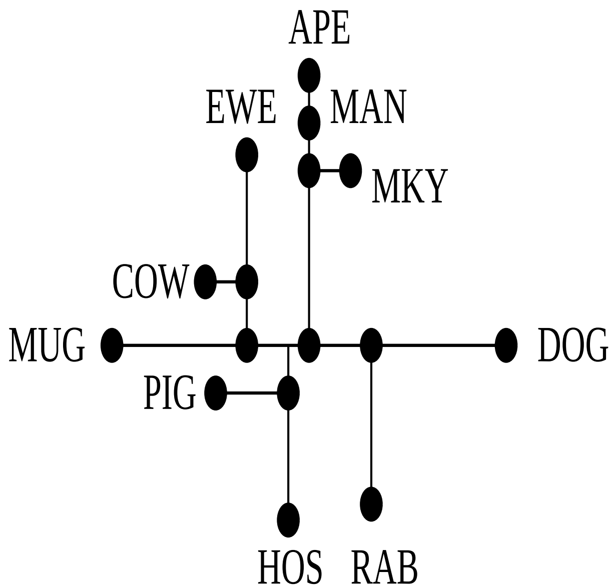

FIG. 2: A phylogenetic tree from Dress & Krüger (1987)

BASIC TREE-NOTIONS

Recall that distances between pairs is a function d that associates a positive real value to each pair of a given set, satisfying the following conditions:

$$
\begin{aligned}
&d_{i,j} \geq 0 \\
&d_{i,i} = 0 \\
&d_{i,j} = d_{j,i} \\
&d_{i,j} \leq d_{i,k} + d_{k,j} \qquad \text{(triangle inequality).}
\end{aligned}
$$

Four-points condition

A distance is a tree distance (i.e., derived from a tree), if and only if the so called four points condition is satisfied. This condition expresses that four points on a tree can always be labeled x, y, z, t such that:

$$d_{x,y} + d_{z,t} \leq (d_{x,z} + d_{y,t}) = (d_{x,t} + d_{y,z}) \ .$$

This condition has been discovered by several authors in different contexts (Zarestkii, 1965; Buneman, 1971; Patrinos & Hakimi, 1972; Dobson, 1974; etc.).

Note: the four-points condition clearly implies the triangle inequality. Intuitively, it seems clear that the four-points condition in turn is implied by the ultrametric inequality (for a proof, see Dobson, 1974). The ultrametric inequality is expressed as:

$$d_{x,y} \leq \max\{d_{x,z}, d_{z,y}\} \quad \forall\ x, y, z \ .$$

Thus, ultrametric trees represent a particular case of additive trees.

Strict and weak score, split and H-relations

The four-points condition can be expressed as a relation between pairs of vertices. An essential tool for understanding trees is the notion of *score* of a pair of vertices. When d is a tree-distance, then all 4-vertices can be labelled x, y, z, t such that either (a) or (b) hold:

$$(a) \qquad (d_{x,y} + d_{z,t}) < (d_{x,t} + d_{y,z}) = (d_{x,z} + d_{y,t})$$
$$(b) \qquad (d_{x,y} + d_{z,t}) = (d_{x,t} + d_{y,z}) = (d_{x,z} + d_{y,t})$$

Following Colonius & Schulze (1981), condition (a) is denoted $xy\mathcal{H}^*zt$ and condition (a) or (b) is denoted $xy\mathcal{H}zt$. An illustration of the conditions (and of their names is given in Figure 3)

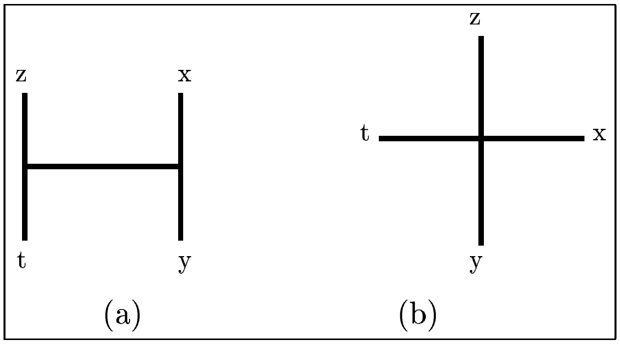

FIG 3: the two configurations of 4 points on a tree, the $\mathcal{H}^*$ and the $\mathcal{H}$ relations

The $\mathcal{H}$-relation can be expressed in different ways. When $xy\mathcal{H}^*zt$, the pair (x,y) is said to be *split* from the pair (z,t) (Buneman, 1971). In terms of tree structure, this means that by deleting one edge of the tree, two non-connected sub-trees are created. One tree contains the vertices x,y, the other one contains the vertices z,t. For example, if in Figure 3.a the internal edge is deleted, two trees will be created, the first one made of x,y, the second one made of z,t. These conditions can also be expressed in terms of a quaternary relation (i.e., a binary relation between pairs). Following Buneman (1971) and (Dobson, 1974), Colonius & Schulze (1981) developed this idea and characterized tree-distances from the (non-numerical) properties of the $\mathcal{H}^*$ relation as follow: for all x,y,z,t

i. $xy\mathcal{H}^*tz$ implies $tz\mathcal{H}^*xy$ and $yx\mathcal{H}^*tz$.
ii. $xy\mathcal{H}^*tz$ implies neither $xy\mathcal{H}^*ty$ nor $tx\mathcal{H}^*yz$ hold .
iii. $xy\mathcal{H}^*tz$ implies either $xy\mathcal{H}^*vz$ or $vx\mathcal{H}^*tz$ or both .
iv. $xy\mathcal{H}^*tz$ and $xy\mathcal{H}^*tv$ implies $xy\mathcal{H}^*vz$.

Actually, *iv* is implied by *i*, *ii*, and *iii* (Bandelt & Dress, 1986; see also Barthélemy & Luong, 1987).

The *strict score* of the pair (x, y) is the number of pairs (t, z) such that $xy\mathcal{H}^*tz$. The *loose score* of the pair (x, y) is the number of pairs (t, z) such that $xy\mathcal{H}tz$.

When d is a tree-distance, the tree structure can be recovered from the score matrix (cf. Buneman, 1971; Colonius & Schulze, 1981; Bandelt & Dress, 1986). This fundamental relation can be expressed in topological terms (i.e., in terms of "neighborness" or "neighborliness", cf. Bandelt & Dress, 1986; Barthélemy & Luong, 1987). In a sense, $xy\mathcal{H}^*tz$ expresses simultaneously that x and y are neighbors and that x, y are separated from z, t. In other words, x and y are farther from z and t than they are from each other. This notion is generalized by Barthélemy & Luong (1987) and by Barthélemy & Guénoche (1987). From that notion, Eastbrook, McMorris & Meacham (1985) and Day (1985) derive some indices used to compare trees, although some problems may occur when this neighboring relation is used within this framework (McMorris, 1985).

HISTORICAL SKETCH

Interest in additive-tree representations originated in several fields such as operational research and computer sciences, biology and psychology. The contribution of these fields is detailed in the following sections.

Operational research and computer sciences

Perhaps the first proof of the four-points condition is to be attributed to Zarestkii (1965). The condition is proved for a tree with a unitary valuation (i.e., all the edges have the same length). Smolenski (1963) proved the unicity of the tree associated with (unitary) tree-distance. However, because these papers were published in eastern journals, they were not widely known or accessible. Actually, the paper by Zarestkii, written in Russian, was known *via* a review of graph theory in the USSR (Turner & Kautz, 1970). Within this tradition, Boesch (1968), Simões-Pereira (1969) and later Patrinos & Hakimi (1972) characterized the tree-distance and, in the latter reference, characterized the four-points condition for the general case (i.e., "non-unitary"). These authors, incidently, did not mention Buneman (1971) but referred to Zarestkii (1965) and Simões-Pereira (1969), who in turn were not mentioned by Buneman (1971). In the meantime, tree-metrics have made their way into mathematics proper (cf. Dress, 1984; and the literature quoted there).

Biology

Buneman (1971) in his classic paper recalled that some authors (biologists, although not mentionned explicitly as such) have previously dealt with problems similar to the manuscript filiation problem. Specifically, he referred to work on phylogenetic trees and signaled that Cavalli-Forza & Edwards (1967) and Ecks & Dayhoff (1966) had described methods for finding an approximation of dissimilarity data by a tree distance. Fruitful research is still conducted in this field (cf. among others, Waterman, Smith, Singh & Beyer, 1978; Fitch, 1981; Meacham, 1987; Peacock, 1981).

As mentioned previously, a main problem in Biology is to integrate the results of different analyses. That is, to compare either different tree distances, or different trees obtained for one set of objects with different measures, methods or (dis)similarity indices. This problem led Phipps (1971) and Farris (1973) to define the tree-distance when the valuation of the edges is unitary (cf. previous section on Zarestkii). The problem of consensus between different trees, and the definition of a distance between trees is explicitly dealt with by Bobisud & Bobisud (1972). This tradition has been carried on by different authors (Robinson, 1971; Dobson, 1975; Robinson & Foulds, 1981). Currently, the problem of defining a consensus between trees remains open as different approaches can be proposed (cf. Day, 1985; Estbrook *el al.*, 1985; McMorris, 1985; Dress & Krüger, 1987).

The comparison of the approach taken in Biology with the approach taken in Psychology emphasizes the *rôle* of the field of application in the theoretical developments. That is, for biologists, the shape of the tree is of prime importance, while for psychologists the tree distance between leaves is the main concern (cf. the section on tree analysis as regression). As a consequence of this difference of emphasis, the very idea of *goodness of fit* differs among pratictioners. To caricature, for biologists a good tree recovers the shape of a possible "original tree," whereas for psychologists, a good tree preserves the original dissimilarity. Thus biologists favor measure of fit expressed in terms of quaternary relations (e.g., number of violations of the "natural" $\mathcal{H}$-relation by the tree). Psychologists, on the other hand, prefer measures like stress (due to the nonmetric multidimensional scaling tradition), or more adequately, the part of the variance of the original dissimilarity explained by the tree model (i.e., the squared correlation coefficient between the original distance and the tree-distance approximation). Hence, the notion of goodness of fit is far from universal in the tree world!

The idea of emphasizing a topological point of view in interpreting a tree is clearly akin to the biological point of view. However this idea was first proposed by mathematical psychologists Colonius & Schultz (1981) who placed themselves in the measurement theory tradition. This perspective has been expanded recently by Bandelt & Dress (1986) and Barthélemy & Luong (1987).

Psychology

In the psychological tradition, besides Buneman (1971), two papers (Carrol & Chang, 1973; Cunningham, 1974) have had a strong influence. As an oddity* these papers were actually "underground" papers. Precisely, they were given as talks in scientific meetings and never circulated (stricly speaking) as papers. However, these two papers aroused the interest of mathematical psychologists. Eventually, one of these papers was published as part of a broad paper that also explored "directed trees" (Cunningham, 1978).

Almost from the beginning of the interest for additive trees within psychology, two traditions were created.

The first tradition, following Carrol & Chang, 1973, comes from the "scaling" or psychometric tradition and considers a tree as a

*that would probably please people at the institute for scientific information.

more or less pratical *graphical* representation of (dis)similarity data. Consequently, the additive tree representations are contrasted with the more usual maps derived by nonmetric multidimensional scaling (cf. the review of Shepard, 1974, 1980; Carrol & Arabie, 1980). Here, the main problem is to decide when to favor tree representations over euclidean maps. Different criteria have been proposed, but stress or percentage of explained variance (i.e., square correlation between the original dissimilarity and the tree distance) are generally prefered. Pruzansky, Tversky & Carroll (1982) offer some guidelines that can help to decide, based on data properties (i.e., dispersion, skweness, proportion of elongated triangles, etc.), which representation is more appropriate for a given set of data. The main result of these investigations is that, on the whole, euclidean distance data show positive skewness, and tree distance data show negative skewness.

The second tradition is linked to cognitive psychology, specifically to work in semantic memory organization. Precisely, the empirical work of Rosch initiated this work (for reviews, see, for example, Rosch, 1975, 1983; Abdi, 1986a). Tversky (1977, Tversky & Gati, 1977) proposed a formalization of the notion of "psychological similarity" in parallel with his work on additive tree fitting with Sattah (Sattah & Tversky, 1977).

The theory was designed to give an account of several empirical observations, including the *typicality effect*. This term is used by psychologists to express that some members of a natural category are more representative of that category than are some of the other members. The classic example is that for the category "bird", canary is a better representative of the category than is penguin. As such, class elements are not equivalent and are often more or less representative of their category. Clearly, the standard ultrametric tree representation forces objects to be equivalent (i.e., in an ultrametric tree, all the elements of a class are at the same distance from the center of the class). As such these trees do not adequately represent data that have a gradient of representativity. The additive tree, with its different edges of varying lengths, seems more appropriate. Tversky (1977) went further than simply advocating tree representations as a tool. He showed that the additive tree distance is a particular case of a more general model of similarity named the *contrast model.* This theme is expanded further in this paper in the section on tree interpretation.

In fact, semantic memory has been a fruitful source of inspiration for psychologists. For example, Schulze & Colonius (1979) and Colonius & Schulze (1981) were originally interested in the exploration of the semantic meaning of verbs. To do so, they used directly the quaternary $\mathcal{H}$-relation and asked subjects to group quadruplets of verbs in pairs (cf. the IVb quadrant of Coombs's *theory of data*, 1964). This task gave the impetus for their later characterization of the tree distance in terms of the $\mathcal{H}$-relation.

ALGORITHMS FOR ADDITIVE TREE APPROXIMATION

The algorithms for tree approximation can be roughly divided in two large families.

Algorithms of the first family proceed by estimating the tree structure, and then adjusting for the length of the edges in order to fit the "original distance". The structure of the tree is determined, in general, using the score matrix (e.g., Sattah & Tversky, 1977). The score matrix can be recomputed after each iteration (as in Sattah & Tversky), or approximated (for example by using a "quasi single link" method on the score matrix as in Abdi *et al*, 1984; Barthélemy & Luong, 1985; Abdi, Guénoche & Luong, 1988; Abdi, 1989). The edge length can be evaluated either by the least squares method or by diverse geometrical means.

The algorithms of the second family (Cunningham, 1978; Carrol & Pruzansky, 1980; De Soete, 1980; Roux, 1986; Brossier, 1987) proceed by trying to find a best tree distance (in some pre-defined sense) approximating the original distance or dissimilarity and then by constructing the tree associated with the tree distance. This second step is indeed straightforward, but the first is strongly dependent on the criterion choosen to evaluate "best approximation". Cunnigham (1978) and De Soete (1980) use a least squares criterion. Alternatively, Roux (1986) minimizes the number of quadruplets that disagree with condition (a) of section II. Following suggestions from Carroll (1976), several authors [Carroll & Pruzansky (1980), Carroll, Clark & DeSarbo (1985), Brossier (1986), Brossier & Calvé, 1986)] use the property that an additive tree distance is decomposable into the sum of an ultrametric distance and a distance to center (i.e., a "star distance"). These authors estimate separately the two components, and reconstitute the tree distance by the sum of its two components.

A detailed description of tree approximation algorithms can be found in Luong (1987), along with comparisons relative to complexity, accuracy, etc. Some comparisons on small data sets ($n = 7$) can be found in Guénoche (1987). Finally, a general framework is given in Barthélemy & Guénoche (1988).

HOW TO ANALYZE A TREE THE "TVERSKY WAY*"

Tversky (1977) developed a general approach to similarity called the *contrast model*. Each stimulus (say, $a, b, \dots$) is associated with a set of features (denoted $A, B, \dots$). The similarity from a to b is defined as a function of the three sets: $A \cap B$, A/B and B/A.

In sum:

$$s(a,b) = F(A \cap B,\ A/B,\ B/A).$$

If some additional conditions are imposed (namely: monotonicity, independence, along with the two "technical conditions" of solvability and invariance†), then the similarity between a and b can be expressed as:

$$S(a,b) = \alpha f(A \cap B) - \beta f(A/B) - \gamma f(B/A)$$

where f is an isotonic function (or an interval scale function, cf. Krantz *et al.*, 1971) and α, β, γ are positive constants.

* This section and the following are adapted from Abdi, 1986b.

† Actually, it is possible to derive the contrast model from a slightly different set of axioms, as shown by Osherson (1987).

Moreover, when $\beta = \gamma$, and f is additive [i.e., $f(A \cup B = f(A) + f(B) - f(A \cap B)$], then there exists a measure g such that (cf. Sattath, 1976):

$$S(a, b) = \lambda - g(A/B) - g(B/A) = \lambda - g(A \Delta B)$$

with λ a positive constant, and $A \Delta B$ being the set symmetric difference between A and B (i.e., the number of features that belong to either a or b but not to both). Now, if the features follow a tree model (that is if three stimuli can always be named a, b, c with $A \cap B = A \cap C \subset B \cap C$), this property allows a very convenient tree representation of the similarities. If the stimuli are leaves on the tree, and the edges are appropriatly valued, then the tree distance from a to b will be

$$d(a, b) = g(A/B) + g(B/A) = g(A \Delta B).$$

This is equivalent to defining the distance in terms of distinctive features.

This expression of the distance on the tree in terms of distinctive features can be used to estimate the features composing the stimuli. To see how this is done, it is convenient to introduce three notions:

- the *median* of a tree.
- the *eccentricity* of a vertex (eccentricity is to be taken here as "distance from a center" according to its etymology).
- the *intersection vertex* of two vertices.

The *median* of a tree is the vertex that minimizes the sum of the distances to the set of the vertices. The *eccentricity* of vertex a, denoted $e(a)$ is defined as the distance from a to the median of the tree. The *intersection vertex* of vertices a and b is the vertex with minimal eccentricity situated on the path from a to b. The reason for this naming will become obvious later on. Call a, b two vertices and x their intersection vertex, then the set components of the contrast model are obtained by:

$$g(A/B) = d_{a,x} = e(a) - e(x)$$
$$g(B/A) = d_{b,x} = e(b) - e(x)$$
$$g(A \cap B) = e(x)$$

Note, incidentally, that $e(a) = g(A)$ can be seen as a measure of the overall saliency of stimlus a.

Thus, the additive tree representation can be used as a tool to recover *a posteriori* the features (and their weights) composing a set of stimuli from a distance matrix. In particular, a tree can give the number of distinctive features common to every pair of stimuli, and decompose each stimulus into (weighted) features so that the distance on the tree between stimuli is computed simply as a distance between (weighted) features. To do so, it is sufficient to use the "city-block distance", which can be interpreted as a generalization of the symmetric difference distance (for more details see Abdi, 1985). In this sense, a tree can be seen as a regression model, or a factorial method. All these notions, hopefully will be made clear by a example.

GOOD GUYS AND BAD GUYS: A TREE ANALYSIS

In a recent paper, Goldstein, Chance & Gilbert (1984) made a new contribution to the topic of "implicit physiognomy". Implicit physiognomy means that observers agree among themselves and find it meaningful to attribute personality traits, intentions, occupations, etc. merely by looking at a face or a photograph of a face. In the Goldstein *et al* study, subjects were presented with five arrays each composed of 20 photographs of white middle-aged men taken from a casting directory, and asked to find among the 20 portraits of each array three bad guys (mass murderer, armed robber, rapist) and three good guys (medical doctor, clergyman, engineer). The data were analyzed *via* 30 chi-square tests (one for each of the six "occupations" of the 5 arrays). Because 27 of these 30 tests were statistically significant at the .05 level, it was concluded that there was indeed a clear consensual agreement among the subjects. As an illustration, Goldstein *et al* gave a contingency table corresponding to the results obtained from 58 subjects with the third array they used (i.e., they gave the number of subjects that assigned a given occupation to each portrait from array #3). These results are summarized below. Only the analysis on the occupations is reported here. The analysis proceeds in two steps. First a distance matrix between occupations will be computed. Second, a tree will be computed to fit these data.

portraits	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*	*K*	*L*	*M*	*N*	*O*	*P*	*Q*	*R*	*S*	*T*
Mass Murderer	11	6	3	2	0	0	0	18	2	6	5	1	1	0	0	1	1	0	0	1
Armed Robber	3	4	0	4	0	4	0	13	0	6	4	1	7	1	0	3	0	5	2	1
Rapist	5	2	1	7	4	5	0	4	1	18	0	0	1	1	0	8	0	1	0	0
Medical Doctor	0	1	2	0	2	1	12	1	0	0	6	8	0	2	4	2	5	2	4	6
Clergyman	0	1	5	0	1	1	6	0	2	0	4	1	1	0	6	2	20	1	3	4
Engineer	0	1	1	0	1	3	6	0	1	0	1	5	3	4	6	5	6	2	4	9

TABLE 1: Association between 20 faces and 6 "occupations". At the intersection between one row and one column, the number of subjects that selected the occupation for the face is given (data from Goldstein *et al.,* 1984).

Distance matrix between occupations

The strong connection between the tree distance and the city block metric justifies using the city block metric to compute distance between occupations. Precisely, if $k_{i,j}$ denotes the number of subjects that assigned face i to occupation j, then the distance between two occupations j and j' is computed by:

$$d_{j,j'} = \sum_i |k_{i,j} - k_{i,j'}|$$

	Doctor	*Engineer*	*Clergyman*	*Murderer*	*Robber*	*Rapist*
Doctor	0	34	46	90	84	96
Engineer		0	44	98	80	86
Clergyman			0	86	88	98
Murderer				0	44	70
Robber					0	58
Rapist						0

TABLE 2: City block distance computed from Table 1

	Doctor	*Engineer*	*Clergyman*	*Murderer*	*Robber*	*Rapist*
Doctor	0	34.00	46.00	91.25	82.75	94.00
Engineer		0	44.00	89.25	80.75	92.00
Clergyman			0	93.25	84.75	96.00
Murderer				0	44.00	68.00
Robber					0	59.75
Rapist						0

TABLE 3: Tree distance approximation of distance of Table 2

Table 2 gives the city block distance matrix, Table 3 gives the tree distance matrix, and Figure 4 displays the tree. The squared correlation coefficient between the original matrix and the tree distance matrix is .972. It is denoted by the letter τ, and is taken as an index of goodness of fit. Its value indicates a fairly good fit between the data and a tree-model.

Name of the features			**Occupations**			
	Doctor	*Engineer*	*Clergyman*	*Murderer*	*Robber*	*Rapist*
Doctor	18.00					
Engineer		16.00				
Clergyman			24.00			
Murderer				26.25		
Robber					17.75	
Rapist						35.50
Engi. & Doct.	4.00	4.00				
Engi. & Doct. & Cler.	18.25	18.25	18.25			
Murd. & Robb.				6.50	6.50	
Murd. & Robb. & Rapi.				18.25	18.25	18.25
Sum=eccentricity	40.25	38.25	42.25	51.00	42.50	53.75

TABLE 4: Tree reconstitution of the specific and common features

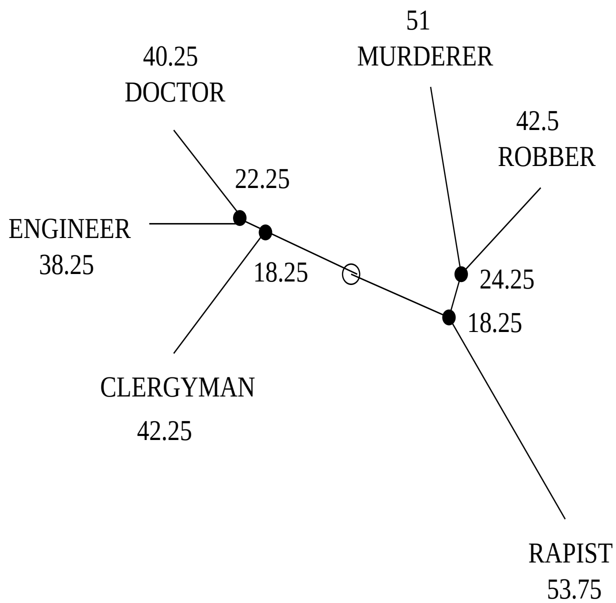

FIG. 4: Additive tree of the occupation distance matrix from Table 2. The number near each vertex is the eccentricity of the vertex.

	Doctor	*Engineer*	*Clergyman*	*Murderer*	*Robber*	*Rapist*	**Sum**
Doctor	–	22.25	18.25	0	0	0	*40.50*
Engineer	22.25	–	18.25	0	0	0	*40.50*
Clergyman	18.25	18.25	–	0	0	0	*36.50*
Murderer	0	0	0	–	24.75	18.25	*43.00*
Robber	0	0	0	24.75	–	18.25	*43.00*
Rapist	0	0	0	18.25	18.25	–	*36.50*

TABLE 5: Tree reconstitution of the common features

The tree shows that the subjects clearly separated the "bad guys" from the "good guys". Moreover, the stereotypes of the good guys are less differentiated than the stereotypes of the bad guys. In particular, the rapist is the most stereotyped occupation, while the engineer is the least stereotyped. These conclusions are supported by table 5 where the eccentricity of the occupations is decomposed into specific and common weighted features. Thus, the tree analysis can be seen as a variety of the classical factorial analysis. As pointed out previously, this "canonical weighted features matrix" is equivalent to the tree-distance matrix when the computed distance is the city block. Note that the features are labelled in agreement with the contrast model interpretation of the tree.

WHEN TO USE A TREE? DIRECTIONS FOR USE....

Perhaps due to some halo effect, it is sometimes thought that the prescriptions for additive tree representation parallel those for nonmetric multidimensional scaling. It should be emphasized that additive tree are meaningful only for data that are invariant by linear transformation (i.e., interval scale measurement, cf. Suppes & Zinnes, 1963; Roberts, 1979). By contrast, nonmetric dimensional scaling deals with data matrices that are supposed to be invariant by monotonic transformation. An equivalent way of expressing this condition with the measurement theory vocabulary is that the data are supposed to be measured on an ordinal scale.

An example from De Soete (1983) clearly confirms that additive trees are non meaningful for data measured on a ordinal scale. The two dissimilarity matrices a and b from Table 6 are equivalent for the set of monotonic transformations (b is obtained by changing in a, 3 to 6, and 9 to 10):

	x	y	z	t
x	0	3	7	8
y		0	8	9
z			0	11

(a)

	x	y	z	t
x	0	6	7	8
y		0	8	10
z			0	11

(b)

TABLE 6: two dissimilarity matrices obtained by a monotonic transformation.

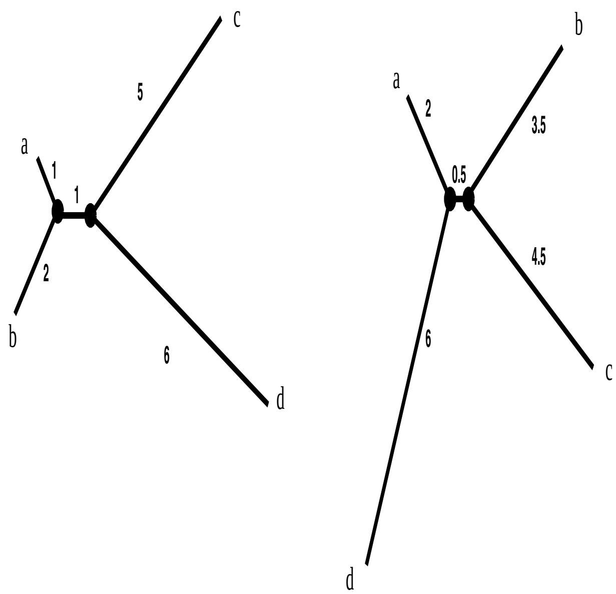

FIG. 5: Additive tree representation of matrix (a) and (b) from Table 6.

Unfortunately, as shown in Figure 5, the trees obtained from these two matrices are topologically different, which means that tree representations are not invariant for monotonic transformations. As a consequence, the application domain of additive trees is much more restricted than the domain of multidimensional scaling. Actually, it can be shown that additive tree representations are meaningful only for an *interval scale* measurement (i.e., additive-tree representations are invariant for the set of linear transformations, see Brossier, 1985; or Barthélemy & Guénoche, 1985).

CONCLUSION

Although this introduction to additive tree representations is an incomplete one, hopefully the presentation of examples were sufficient to enable the reader see the forest for the tree. The reference section serves as a guide for readers interesting in learning more about the topic. Some recent developments should also be alluded to here. In particular, the trees described in this paper were built from complete 2-way square matrices. But trees can also be computed from rectangular matrices (Furnas, 1980; Brossier, 1986), three ways matrices (Carrol, Clark, DeSarbo, 1984) or incomplete matrices (De Soete, 1984). An other line of research has been pioneered by Corter & Tversky (1987), who tried to relax the tree constraint in order to represent proximity data by an extended tree. These extended trees generalize traditional trees by including marked segments that correspond to overlapping clusters. Closely related to this approach is the notion of weak hiereachies recently proposed by Bandelt & Dress (1988).

Hence trees are still growing and bearing diverse fruits.

REFERENCES

Abdi, H. (1985)—Représentations arborées de l'information verbatim. In J., Wittwer (Ed.): *Psycholinguistique textuelle*. Paris: Bulletin de Psychologie.

Abdi, H. (1986a)—La mémoire sémantique, une fille de l'intelligence artificielle et de la psychologie. In C., Bonnet, J.C., Hoc, G., Tiberghien (Eds.): *Psychologie, intelligence artificielle et automatique*. Bruxelles: Mardaga.

Abdi, H. (1986b)—Faces, prototypes, and additive tree representations. In H.D., Ellis, M.A., Jeeves, F., Newcombe, A., Young (Eds.): *Aspects of face processing*. Dordrecht: Nijhoff.

Abdi, H. (1989a)—ABD-TREE: A complete APPLE 2 set of programs for additive-tree representations. In X. Luong (Ed.): *Additive-tree representations of textual data*. Nice: CUMFID.

Abdi, H. (1989b)—Additive-tree representation of verbatim memory. *CUMFID*, **16**, 99–124.

Abdi., H., Barthélemy, J.P., Luong, X. (1984)—Tree representations of associative structures in semantic and episodic memory research. In E., Degreef, J. Van Huggenhaut (Eds.): *Trends in mathematical psychology*. New York: Elsevier.

Bandelt, H.J., Dress, A. (1986)—Reconstructing the shape of a tree from observed dissimilarity data. Preprint to appear in: *Advances in Applied Mathematics*.

Bandelt, H.J., Dress, A. (1988)—*Weak hierachies associated with similarity measures.* Unpublished manuscript. Bielefeld University.

Barthélemy, J.P., Guénoche, A. (1988)—*Arbres et représentations des proximités.* Paris: Masson.

Barthélemy, J.P., Luong, X. (1986)—Représentations arborées des mesures de dissimilarité. *Statistiques et Analyse des Données*, **5**, 20–41.

Barthélemy, J.P., Luong, X. (1987)—Sur la topologie d'un arbre phylogénétique: aspects théoriques, algorithmes et applications à l'analyse des données textuelles. *Mathématiques et Sciences Humaines*, **100**, 57–80.

Bobisud, H.M., Bobisud, A. (1972)—A metric for classification. *Taxonomy*, **21**, 607–613.

Boesch, F.T. (1968)—Properties of the distance matrix of a tree. *Quaterly of Applied Mathematics*, **16**, 607-9.

Boorman, S.A., Olivier, D.C. (1973)—Metrics on spaces of finite trees. *Journal of Mathematical Psychology*, **10**, 26-59.

Brossier, G. (1985)—Approximation des dissimilarités par des arbres additifs. *Mathématiques et Sciences Humaines*, **91**, 5–21.

Brossier, G., LeCalvé, G. (1986)—Analyse des dissimilarités sous l'éclairage $\sqrt{D}$, application à la recherche d'arbres additifs optimaux. In D., Diday *et al.* (Eds.): *Data analysis and informatics, IV.* Amsterdam: North Holland.

Buneman, P. (1971)—The recovery of trees from measures of dissimilarity. In J. Hodson *et al* (Eds.): *Mathematics in the archaelogical and historical sciences.* Edinbhurg. Edinbhurg University Press.

Buneman, P. (1974)—A note on the metric properties of trees. *Journal of Combinatorial Theory*, **17**, 48–50.

Carroll, J.D., Arabie, P (1980)—Multidimensional Scaling. *Annual Review of Psychology*, **31**, 607–49.

Carrol, J.D., Chang, J.J. (1973)—A method for fitting a class of hierarchical tree structure models to dissimilarity data and its application to some body parts data of Miller's. In *Proceedings of the 81th Annual Convention of the American Psychological Association*, **8**, 1097–8.

Carroll, J.D., Clark, L.A., Desarbo, W.S. (1984)—The representation of three-way proximity data by single and multiple tree structure models. *Journal of Classification*, **1**, 25–74.

Carrol, J.D., Pruzansky, S. (1975)—*Fitting of hierarchical tree structure models.* Paper presented at the US-Japan seminar on Multidimensional Scaling. San-

Diego.

Cavalli-Sforza, L.L., Edwards, A.W.F. (1967)—Philogenetic analysis models and estimation procedures. *American Journal of Human Genetic*, **19**, 233–57.

Chaiken, S., Dewdney, A.K., Slater, P.D. (1983)—An optimal diagonal tree code. *SIAM Journal of Applied Discrete Mathematics*, **4**, 424-9.

Colonius, H., Schulze, H.H. (1979)—Repräsentation nichnümerischer Anlichkeitsdaten durch Baumstrukturen. *Psychologische Beitrage*, **21**, 98-111.

Colonius, H., Schulze, H.H. (1981)—Tree structuress for proximity data. *British Journal of Mathematical and Statistical Psychology*, **34**, 167–80.

Corter, J.E., Tversky, A. (1986)—Extended similarity tree. *Psychometrika*, **51**, 429–451.

Coombs, C.H. (1964)—*A theory of data*. New York: Wiley.

Coxon, A.P.M. (1982)—*The user's guide to Multidimensional scaling*. London: Heinemann.

Culler, M. (1987)—Group actions on R-trees. *Proceedings of the London Mathematical Society*, 55.

Cunningham, J.P. (1974)—*Finding an optimal tree realization of a proximity matrix*. Paper presented at the Mathematical Psychology Meeting: Ann Arbor.

Cunningham, J.P. (1978)—Free trees and bidirectional trees as representations of psychological distances. *Journal of Mathematical Psychology*, **17**, 165-88.

Day, W.H.E. (1985)—*Analysis of quartet dissimilarity measures between undirected phylogenetic trees*. Research report #CRM-1315. Department of Computer sciences, Montréal (Canada).

De Soete, G. (1983a)—Are non-metric additive tree representations of numerical proximity data meaningful? *Quality & Quantity*, **17**, 475–8.

De Soete, G. (1983b)—A least-square algorithm for fitting additive trees to proximity data. *Psychometrika*, **48**, 621-6.

De Soete, G. (1984)—Additive-tree representations of incomplete dissimilarity data. *Quality & Quantity*, **8**, 387–97.

Dewdney, A.K. (1979)—Diagonal tree codes. *Information & Control*, **40**, 234–9.

Dobson, J. (1974)—Unrooted tree for numerical taxonomy. *Journal of applied Probability*, **11**, 32–42.

Dress, A. (1984)—Trees, tight extensions of metric spaces, and cohomological dimensions of certain groups. *Advances in Mathematics*, **53**, 321–402.

Dress, A., Kruger M. (1987)—Parsimonius phylogenetic trees in metric spaces and simulated annealing. *Advances in Applied Mathematics*, **8**, 8–37.

Eck, R.V., Dayhoff, M.O. (1966)—*Atlas of protein sequence and structure.* New York: National Biomedical Research Foundation.

Estabrook, G.F., McMorris, F., Meacham, C. (1985)—Comparison of undirected phylogenetic trees based on subtree of four evolutionary units. *Systematic Zoology*, *22*, 193–200.

Farris, J.S. (1973)—On comparing the shapes of taxonomic trees, *Systematic Zoology*, **2**, 50-4.

Fitch, W.M. (1981)—A non-sequential method for constructing trees and hierarchical classifications. *Journal of Molecular Evolution*, **18**, 30-7.

Furnas, G.W. (1980)—*Objects and their features: the metric representation of two-class data.* Thesis, Standford University.

Gati, I., Tversky, A. (1984)—Weighting common and distinctive features in perceptual and conceptual judgements. *Cognitive Psychology*, **16**, 341–70.

Guénoche, A. (1987)—Etude comparative de cinq algorithmes d'approximation des dissimilarités par des arbres à distance additives. Preprint to appear in: *Mathématiques et Sciences Humaines.*

Hodson, F.R., Kendall, D.G., Tautu, P. (Eds.) (1971)—Mathematics in the Archaelogical and Historical Sciences. Edinbhurg: Edinbhurg University Press.

Hakimi, S.L., Yau, S.S. (1964)—Distance matrix of a graph and its realizibility. *Quaterly of Applied Mathematics*, **22**, 305–17.

Hartigan, J.A. (1967)—Representations of similarity matrices by trees. *Journal of the American Statistical Association*, **62**, 1140–58.

Hartigan, J.A. (1975)—*Clustering Algorithm.* New York: Wiley.

Henley, N.M. (1969)—A psychological study of the semantics of animal terms. *Journal of Verbal Learning and Verbal Behavior*, **8**, 176–84.

Krantz, D.H., Luce, R.D., Suppes, P., Tversky, A. (1971)—*Foundations of measurement.* New York: Academic Press.

Luce, R.D., Galanter, E. (1963)—Psychophysical scaling. In R.D. Luce, R.R. Luce, E. Galanter (Eds.): *Handbook of Mathematical Psychology.* New York: Wiley.

Luong, X. (1983)—Voisinage lâche, score et famille scorante. Multigraph, Besançon (France).

Luong, X. (1988)—*Méthodes en analyse arborée: algorithmes et applications.* Thesis: Paris V University.

McMorris, F.R. (1987)—Axioms for consensus functions on undirected phylogenetic trees. *Mathematical Biosciences*, **74**, 17–21.

Meacham, C.A. (1981a)—A manual method for character compatibility analysis. *Taxon*, **30**, 591–600.

Meacham, C.A. (1981b)—A probability measure for character compatibility, *Mathematical Biosciences*, **57**, 1–18.

Osherson, D.N. (1987)—New axioms for the contrast model of similarity. *Journal of Mathematical Psychology*, **31**, 93–103.

Patrinos, A.N., Hakimi, S.L. (1972)—The distance matrix of a graph and its tree realization. *Quaterly of Applied Mathematics*, **30**, 255-69.

Phipps, J.B. (1971)—Dendogram Topology. *Systematic Zoology*, **20**, 306–8.

Pruzansky, S., Tversky, A., Carrol, J.D. (1982)—Spatial *versus* tree representations of proximity data. *Psychometrika*, **47**, 3-24.

Roberts, F.S. (1979)—*Measurement theory*. New York: Addison-Wesley.

Robinson, D.F. (1971)—Comparison of labeled trees of valency three. *Journal of Combinatorial Theory (B Series)*, **11**, 105–119.

Robinson, D.F., Foulds, L.R. (1981)—Comparison of phyologenetic trees. *Mathematical Biosciences*, **53**, 131-47.

Rosch, E. (1973)—On the internal structure of perceptual and semantic categories. In T.E., Moore (Ed.): *Cognitive development and the acquisition of language.* New York: Academic Press.

Rosch, E. (1975)—Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, **104**, 192-233.

Rosch, E. (1983)—Prototype classification and logical classification: the two systems. In E.K. Scholnik (Ed.): *New trends in conceptual representations.* Hillsdale: Erlbaum.

Roux, M. (1986)—Représentation d'une dissimilarité par un arbre aux arêtes additives. In D. Diday *et al* (Eds.): *Data Analysis and Informatics IV.* Amsterdam: North Holland.

Sattah, S., Tversky, A. (1977)—Additive similarity trees. *Psychometrika*, **42**, 319-45.

Shepard, R.N. (1974)—Representation of structure in similarity data: problems and prospects. *Psychometrika*, **39**, 373-421.

Shepard, R.N. (1980)—Multidimensional scaling, tree-fitting and clustering. *Science*, **210**, 390-8.

Shepard, R.N., Romney, A.K., Nerlove, S.B. (Eds.)—*Multidimensional Scaling (2 Vol.).* New York: Academic Press.

Simoes-Prereira, J.M.S. (1967)—A note on the tree-realizability of a distance matrix. *Journal of Combinatorial Theory*, **6**, 303-10.

Smoleskii, Y.A. (1963)—A method for linear recording of graphs. *USSR Computional Mathematics & Mathematical Physics*, **2**, 396–7.

Suppes, P., Zinnes, J.L. (1963)—Basic measurement theory. In R.D., Luce, R.R. Bush, E. Galanter (Eds.): *Handbook of Mathematical Psychology*. New York: Wiley.

Turner, J., Kautz, W.H. (1970)—A survey of progress in graph theory in the Soviet Union. *SIAM Review*, **12**, 17.

Tversky, A. (1977)—Features of similarity. *Psychological Review*, **84**, 327–52.

Tversky, A., Gati, I. (1978)—Studies of similarity. In E. Rosch, B.B. Llyod (Eds.): *Cognition and categorization*. Hillsdale: Erlbaum.

Tversky, A., Gati, I. (1982)—Similarity, Separability and the triangle inequality. *Psychological Review*, **89**, 123–54.

Waterman, M.S., Smith, T.F. (1978)—On the similarity of dendograms. *Journal of Theoretical Biology*, **73**, 789–800.

Waterman, M.S., Smith, T.F., Singh, M., Beyer, W.A. (1977)—Additive evolutionary trees. *Journal of theoretical Biology*, **64**, 199–213.

Zaretskii, K. (1965)—Constructing a tree on the basis of a set of distance between the hanging vertices (in Russian). *Uspekhi Mat. Nauk.*, **20**, 90–2.