PSYCHOMETRIKA—VOL. 41, NO. 4
DECEMBER, 1976

SPATIAL, NON-SPATIAL AND HYBRID MODELS FOR SCALING

J. DOUGLAS CARROLL

BELL LABORATORIES

In this paper, hierarchical and non-hierarchical tree structures are proposed as models of similarity data. Trees are viewed as intermediate between multidimensional scaling and simple clustering. Procedures are discussed for fitting both types of trees to data. The concept of multiple tree structures shows great promise for analyzing more complex data. Hybrid models in which multiple trees and other discrete structures are combined with continuous dimensions are discussed. Examples of the use of multiple tree structures and hybrid models are given. Extensions to the analysis of individual differences are suggested.

Key words: multidimensional scaling, hierarchical tree structures, clustering, geometric models, multivariate data analysis.

Roger Shepard's presidential address [Shepard, 1974] was a somewhat retrospective overview of multidimensional scaling; last year Kruskal [1976] emphasized ongoing research in the area. This address, although involving past, present, and future, will be somewhat more speculative, emphasizing possible future trends.

Spatial models of two-way multidimensional scaling are very well known. They assume that measures of similarity, dissimilarity, or to use Shepard's more general term, of proximity, relate in some simple way (e.g., linearly or monotonically) to distance in a postulated underlying metric space. This model has been extended to what has come to be called three-way or "individual differences" multidimensional scaling, in which parameters describing different individuals (or other data sources) are added to the parameters defining the multidimensional spatial representation of stimuli. The weighted Euclidean model that Jih-Jie Chang and I have called the INDSCAL model [Horan, 1968; Carroll & Chang, 1970], Tucker's [1972] three-mode scaling model, and Harshman's [Note 3] PARAFAC-2 model are three of

1976 Psychometric Society Presidential Address.
While people too numerous to list here have contributed ideas, inspiration, and other help, I particularly wish to acknowledge the contributions of Sandra Pruzansky, without whom this paper could not have been written. I would also like to acknowledge the past contributions of my long-time colleague Jih-Jie Chang, without whose help I probably would not have been *asked* to write it.

Requests for reprints should be sent to Dr. J. Douglas Carroll, Bell Laboratories, Murray Hill, New Jersey 07974.

the principal examples of such three-way MDS models. Of course there are many others, both actual and potential.

Nonspatial models are familiar in the guise of clustering, either simple or hierarchical. When we apply a clustering procedure to a matrix of proximities we are fitting a kind of nonspatial model. This can be either a set-theoretic model in the case of "simple" clustering, or a graph-theoretic model as in the tree structures of hierarchical clustering. In the set-theoretic models of simple clustering the underlying psychological model can be stated in terms of discrete attributes, rather than the more continuous or "dimensional" attributes usually associated with MDS.

Of course, as Torgerson [1965] pointed out, in a spatial model there is really no need to assume that all points in the space can correspond to actual stimuli. It would be perfectly consistent to postulate a spatial model, for example, in which each of the dimensions takes on only two, three, or some other finite or denumerably infinite number of values, so that the space, rather than being truly continuous, has "holes" that cannot be filled by actual stimulus objects.

In the final analysis, the choice of a model probably depends on parsimony. If the number of possible values for each dimension is large, or denumerably infinite, and the number of dimensions relatively small, a continuous spatial model should provide a quite satisfactory approximation. If, on the other hand, the number of possible values for each dimension is small (say two or three) and the number of dimensions large, a more discrete, nonspatial model may provide a more parsimonious representation.

In the model implicitly assumed by simple clustering, each stimulus is postulated to have one and only one of a number of such discrete attributes. These attributes are jointly exhaustive and mutually exclusive. In hierarchical clustering, this restriction is relaxed somewhat by allowing attributes that define overlapping sets. However, these sets define a strictly nested structure such that every pair of sets is either disjoint or else one of the two is a proper subset of the other. This may provide a good model of many linguistic situations, for example, where there are indeed nested sets of superordinate and subordinate concepts (e.g., living things, animal, dog, cocker spaniel, my cocker spaniel named "Sport", etc.).

John Hartigan [1967] was the first to suggest treating the problem of hierarchical clustering as one of fitting a certain geometric model—namely a rooted tree structure on which an ultrametric is defined—using combinatorial optimization techniques. Hartigan proposed an explicit algorithm aimed at optimizing a least squares criterion of fit between data "distances" and distances calculated from an ultrametric tree structure.

A few years later Jih-Jie Chang and I devised a procedure that generalized Hartigan's approach in two different ways [Carroll & Chang, 1973]. First, our procedure allowed some or all of the interior, or nonterminal, nodes of the tree to correspond to objects or stimuli, whereas Hartigan's procedure (and, indeed, most other hierarchical clustering procedures) restricted the objects or stimuli to *terminal* nodes. Second, we allowed a more flexible definition of the metric on the tree. In addition to the ultrametric, which amounts to associating what Johnson [1967] has called a "height" with each nonterminal node, and defining the distance between any two nodes as the height of their "lowest common ancestor" node (the lowest node on the [inverted] rooted tree at which the two paths upward from those two nodes meet), two other kinds of metrics were allowed. The first of these is a path length metric, in which lengths are associated with branches, or links, in the tree, and distance is simply the length of the (unique) path joining those two nodes of the tree. The second is a mixed case, in which "heights" are associated with nonterminal nodes *and* lengths with branches, while distance is defined as a sum of the path length and height of the lowest common ancestor node. (It might be noted that this last "mixed" metric can be meaningfully distinguished from the simpler path length metric *only* in the case in which some of the objects are at nonterminal nodes. In the more usual case in which all objects are at terminal nodes, with the interior nodes all being dummy or "invisible" nodes, both the ultrametric and this "mixed" metric are special cases of the path length metric. When some objects are at interior nodes the three models are all meaningfully different, although of course both the "unmixed" cases are special cases of the "mixed" one.)

We called these three types of metrics "nodes only" (corresponding to the ultrametric case), "branches only" (path length) and "nodes and branches" (mixed metric), because the first has parameters (heights) associated with (nonterminal) nodes only, the second has parameters (lengths) for branches (or links) only, and the third has parameters associated with both nodes and branches.

We applied this procedure to some data that Miller collected by use of a subjective sorting task [see Miller, 1969] on dissimilarities of names of body parts. Miller had intentionally selected these names to define a fairly clear-cut hierarchy based on inclusion relations. The best solution (by a modified least squares criterion utilizing a criterion of fit based on a pseudo F-statistic which takes number of parameters fit into account in a systematic way), is shown in Fig. 1. Ironically, after our effort at generalizing to the other two metrics, the ultrametric proved to provide the best fit to these data. The height values are shown in the boxes associated with nonterminal nodes. This was counter to our expectation that one of the other models would provide a better fit. This result may be due to the fact that the data were collected by a sorting task (the measure of dissimilarity being the number of people who sorted the two words into different categories). It is possible that the sorting task may lead to dissimilarities more or less satisfying the ultrametric inequality.

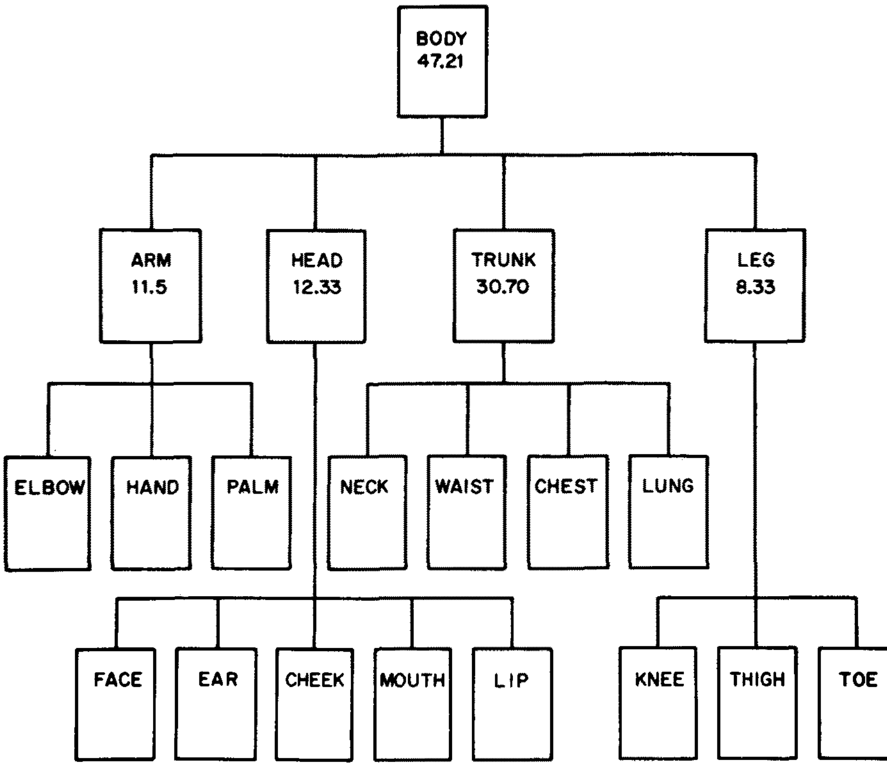


FIGURE 1

Hierarchical tree structure for Miller's "body parts" data. Numbers in boxes representing nonterminal nodes are estimated "heights" in "nodes only" or ultrametric model.

Except for Miller's body parts data, which was a deliberately designed illustration, we were not able to find another convincing case that requires stimulus objects to be placed at nonterminal nodes. (Cunningham has discussed a possible example, however, which will be mentioned later.) As a result, it seems wiser to fit a more general tree structure that puts all objects at terminal nodes, and if appropriate to infer that some of these objects should be at nonterminal nodes from special patterns of parameter values (e.g., certain branch lengths very close to zero). Furthermore, the combinatorial optimization approach turns out to be quite slow and cumbersome, highly subject to the discrete analogue of merely local optima, and generally rather inefficient. For these reasons we have turned to procedures based on mathematical programming concepts, in which the discrete optimization problem is translated into one of continuous optimization with certain inequality constraints. This almost always leads to trees of the more general type in which all objects are at terminal nodes. However, the procedures prove to be much more efficient and effective in general. They also lend themselves quite nicely to models involving *multiple* tree structures, and hybrid models involving mixtures of tree structures and continuous spatial structure, which will be discussed subsequently. First, though, I will mention some other developments involving nonspatial models.

ADCLUS, and Fitting of Single "Free Trees" to Dissimilarities Data

Roger Shepard [1974] has discussed the Shepard-Arabie [Note 6] ADCLUS procedure for extracting overlapping but not necessarily hierarchically organized clusters from proximities data. Basically the ADCLUS (for ADditive CLUSters) procedure attempts to approximate a similarity matrix by a positive linear combination of outer products of binary vectors. In English this means that each stimulus object either has or does not have each of a number of attributes (but these attributes need not be mutually exclusive). The similarity of two objects gets a positive increment (the amount defined by the weight associated with that particular attributes) if both of them have the attribute, but no increment if either one fails to have it. (Tversky, Note 7, has formulated a general theory that includes this and other discrete models, such as tree structures, as special cases). The procedure Shepard and Arabie have implemented approximates a least squares solution to the problem of simultaneously defining the optimal set of attributes and the optimal weights for the attributes so defined. Subsequently I will discuss a planned generalization of ADCLUS to a hybrid model of a certain type as well as its extension to three-way or individual differences data.

What I have been calling a tree with path length metric, or simply a path length tree, is sometimes called a "free tree" because, unlike an ultrametric tree, it has no natural "root" node. Since a path length tree has no root, it is not necessary to think of it as being vertically organized into a hierarchy; or, put differently, one can associate as many different hierarchies with a free tree as it has nodes (including both terminal and nonterminal nodes)—since every node can be the root node. Both Cunningham [Note 2] and Sattath and Tversky [Note 5] have devised algorithms for fitting free trees to data. Sattath and Tversky's method is a kind of natural generalization of the "pair group" method (e.g., single, average and complete linkage, also known to psychologists as the connectedness, average, and diameter methods, respectively) often used to generate hierarchical clustering solutions. Cunningham's method is somewhat closer in spirit to our own approach, as he attempts to approximate a least squares solution.

Cunningham's solution rests on a certain four-point condition [see, e.g., Buneman, 1971; Patrinos & Hakimi, 1972; or Dobson, 1974] that must be satisfied by path length distances. This condition is:

$$\text{if} \quad d_{ij} + d_{kl} \geq d_{ik} + d_{jl} \geq d_{jk} + d_{il}$$
$$\text{then} \quad d_{ij} + d_{kl} = d_{ik} + d_{jl}\ ; \tag{1}$$

i.e., the two largest sums of pairs of distances involving the subscripts i, j, k, and l must be equal. Cunningham's approach in effect assumes that the order of these sums of distances for the data will match (at least weakly) the order of the same sums in the optimal solution. This is a rather dubious assumption, unless the error variances are very small indeed.

Thus if δ_{ij} represents the data value (dissimilarity for the pair (i, j)), Cunningham imposes the linear constraint that

$$d_{ij} + d_{kl} = d_{ik} + d_{jl} ,$$

$$\text{(2)} \qquad \text{if} \quad \delta_{ij} + \delta_{kl} \geq \delta_{ik} + \delta_{jl} \geq \delta_{jk} + \delta_{il} .$$

One such linear constraint on the distances is defined for each tetrad of points. He then seeks the d's yielding a best least squares fit to the δ's, but subject to these linear constraints. This problem has a straightforward analytic solution, although such anomalies as negative distances can occur.

Cunningham's procedure works well for small data sets with relatively low rates of error. For larger data sets it quickly becomes unwieldy, since it requires inverting an $\binom{n}{2} \times \binom{n}{2}$ matrix. More critically, it begins to break down once the ratio of error variance to "true" variance exceeds about 10%. This is apparently because the constraints seemingly implied by the data are then not necessarily the constraints that hold for the optimal solution. In such cases the solution Cunningham's method obtains is frequently the rather uninteresting tree of the form shown in Fig. 2, with all branch lengths equal (so that all distances are equal to the same constant). A tree of this form is sometimes called a "star" by graph theorists.

It would appear that for such "noisy" data, a different approach is needed. Let us turn to a somewhat different approach to tree fitting which is better suited to such "noisy" situations.

Fitting of a Single Ultrametric Tree by Mathematical Programming Techniques

As pointed out by Johnson [1967], Hartigan [1967] and others, there is a homomorphism between distance matrices satisfying the ultrametric inequality and rooted trees. The ultrametric inequality simply states that

$$\text{(3)} \qquad d_{ik} \leq \max\,(d_{ij} , d_{jk}), \qquad \text{for all} \qquad i, j, k.$$

This is, of course, a much stronger condition than the more general triangle inequality, which states that

$$\text{(4)} \qquad d_{ik} \leq d_{ij} + d_{jk}, \qquad \text{for all} \qquad i, j, k.$$

The ultrametric inequality is stronger in part because the extent to which it is satisfied is invariant under strictly increasing monotonic function of the distances, whereas the triangle inequality can always be satisfied by applying

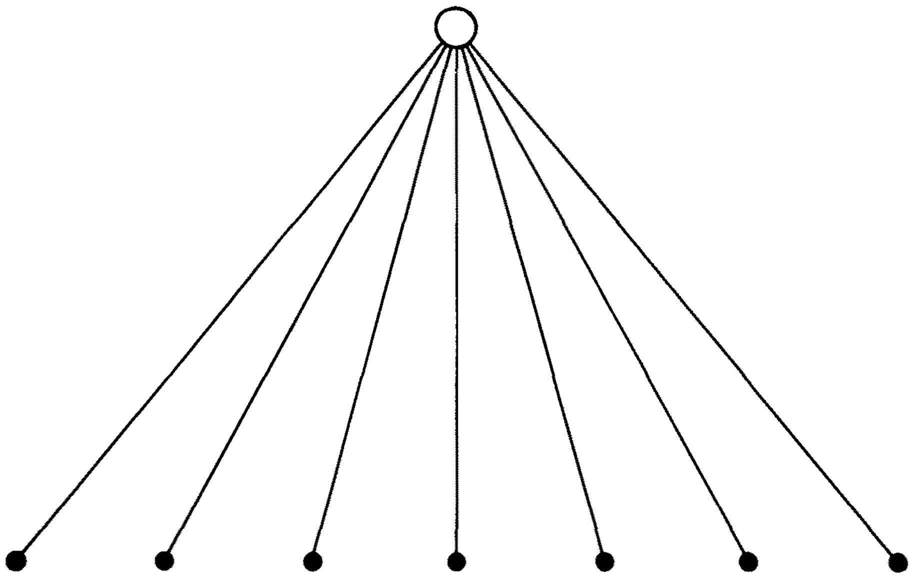


FIGURE 2
A degenerate tree, sometimes called a "star graph". If branch lengths are all equal all distances are constant; if unequal then distances are "additively decomposable".

the trivial monotonic function that adds a sufficiently large constant to all distances, or violated by subtracting a sufficiently large constant. Given a set of distances satisfying the ultrametric inequality (u.i.), the associated tree can easily be constructed (and height values defined). Given a tree, an infinite family of ultrametrics is defined; if the height values are specified, the particular ultrametric is uniquely specified.

It is easy to see that an equivalent statement of the u.i. is that all triangles are isosceles, with the two longest sides equal. This can be stated as

$$(5) \qquad d_{ij} \geq d_{jk} \geq d_{ik} \langle = \rangle d_{ik} = d_{jk}, \quad \text{for all} \quad i, j, k.$$

Pruzansky and I have devised an approach, based on mathematical programming techniques, to solve for ultrametric distances that provide a best fit, in a least squares sense, to a given matrix of dissimilarities. Details of this algorithm are contained in a recent paper by Carroll and Pruzansky [Note 1]. It is based on a "penalty function" approach, in which a penalty is added to the usual least squares loss function. This penalty is directly proportional to a quantitative measure of the degree to which the u.i. is violated. The weight attached to this penalty function, relative to that attached to the loss function itself, is gradually increased until the u.i. is essentially perfectly satisfied, but in such a way that the loss function is at least at a constrained local optimum. More importantly, it might be stated that our current method is the best we have developed so far. We expect further numerical experimentation to improve this procedure quite radically.

Fitting of Multiple Tree Structures Via Alternating Least Squares

There are many sets of proximities data that are not well represented by either simple or hierarchical clusterings, as has already been noted. One alternative model that has already been mentioned is the ADCLUS model and method of Shepard and Arabie, in which proximities data are assumed to arise from discrete attributes that define overlapping but nonhierarchically organized sets. However, in such a case it may be possible to organize the attributes into two or more separate hierarchies. Each of these separate hierarchies could represent, for example, an organized family of subordinate and superordinate concepts of the kind we have already discussed. For example, in the case of animal names one might imagine one hierarchical conceptual schema based on the phylogenetic scale, and another based on function or relationship to man (e.g., tame vs. wild, with tame animals further broken down into pets, work animals, and animals raised for food, pets further broken down into house pets vs. outdoor pets, and so on). Two such conceptual hierarchical structures would obviously be far from independent of one another—whether or not an animal is a pet, for example, is hardly independent of the phylogenetic classification of the animal-but they could be sufficiently distinct that an appropriate technique could pull them apart. Such multiple hierarchies in data may often be obscured in standard clustering analyses, because of the possibly high degree of correlation among separate structures. (I hasten to say that we have not yet applied the technique to be described here to data on animal names, so we do not really know if the results would in fact come out the way suggested in this hypothetical example.)

Apparently a method is needed in this case for describing data by a model entailing *multiple* tree structures—a multidimensional generalization of the single tree structure, as it were. I have whimsically suggested calling this approach "multiarboreal scaling", but multiple tree structures will probably have to do.

One plausible way for multiple tree structures to arise from data amalgamated over subjects (which is the nature of much two-way proximities data) is to assume that each subject uses only a single tree, but different subjects choose from some relatively small number of different trees. In this case the relative saliences of the trees could be thought to relate to the proportion of subjects using each.

One might ask rhetorically, however, if is it plausible to assume genuine *multiple* tree structures underlying proximities data for a *single* subject. I, for one, would answer that question affirmatively. After all, most of us accept the idea quite readily that a single subject's similarities judgments, for example, can be based on more than one dimension. Why then, couldn't they as readily be based on more than one tree structure—or, perhaps more properly, on internal processes or structures that are *isomorphic* to structures of this kind? One can easily imagine, for example, hierarchically organized list structures in memory that could be isomorphic to multiple tree structures of this type.

Pruzansky and I have generalized the approach to tree fitting described earlier to include fitting of *multiple* tree structures to dissimilarities data. We use a numerical strategy that has come to be called alternating least squares, a name suggested by de Leeuw [see, e.g., de Leeuw, Young & Takane, 1976]. This general approach, which was originally called NIPALS (*N*onlinear *I*terative *PA*rtial *L*east *S*quares) by Wold [1966], attempts to partition the total parameter set into subsets of parameters such that, if one fixes all but one of these subjects, there is an *exact* least squares solution for the remaining set. This is what might be called the *conditional* least squares problem. The NIPALS or ALS (for *A*lternating *L*east *S*quares) strategy was used by Chang and I in devising the CANDECOMP procedure that provides the numerical basis for analysis in terms of the INDSCAL model.

In the case of multiple tree structure the model can be stated as

$$\Delta \cong D_1 + D_2 + \cdots + D_m , \tag{6}$$

where Δ is a matrix of dissimilarities and each D_i is a distance matrix satisfying the ultrametric inequality. The symbol $\cong$ can be taken as meaning "equals, except for (unspecified) error terms", or ambiguously, as implying that, given the fixed term on the left side, we seek least squares estimates of the parameters specified on the right.

The conditional least squares problem that must be solved in the present case can easily be defined by noting that

$$\Delta_l^* \cong \Delta - \underline{\sum_{i \neq l}^{m} D_i} \equiv D_l \tag{7}$$

This means that we may solve the conditional least squares problem by applying our mathematical programming algorithm for least squares fitting of a single ultrametric tree to the Δ_l^* matrix. The ALS procedure, then, consists of successively reestimating each of the D_i matrices iteratively in this fashion until convergence occurs, as indicated by no further change in the D_i matrices, or in the (least squares) measure of fit.

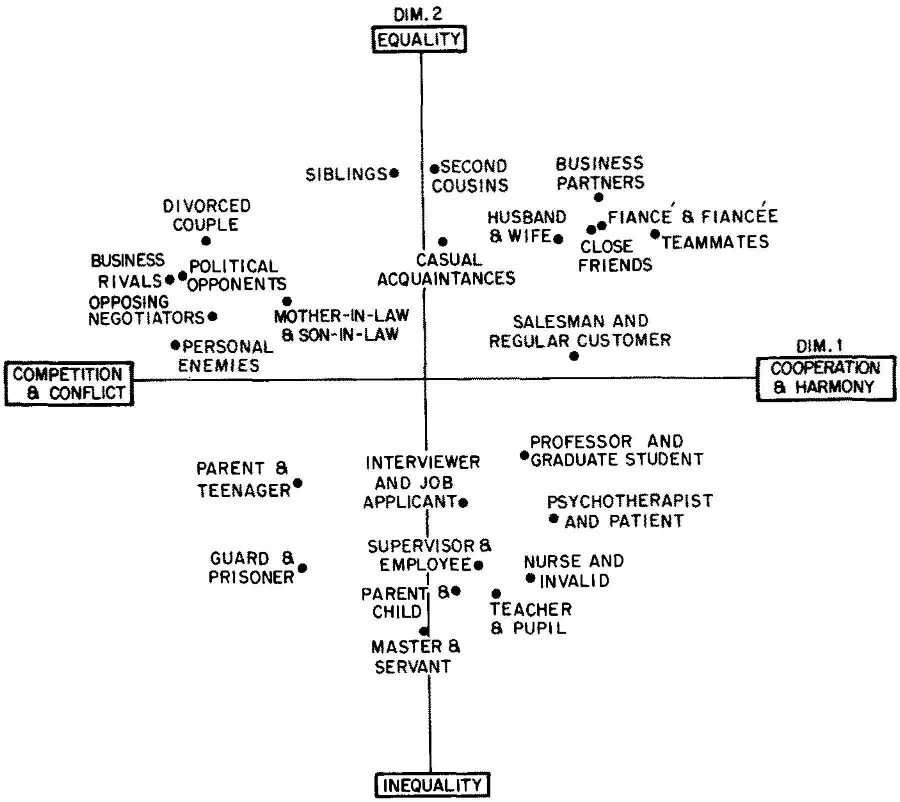
Earlier I suggested that methods of analyzing data in terms of some of these models were still in the realm of future development. In the case of multiple tree models, however, I believe that we have seen the future, and it works—pretty well anyhow (although there still are some problems of slow convergence and local minima). In some cases we had to try many different starting configurations and use quite a lot of computer time to get what seem to be reasonably good solutions. We do feel, however, that we are on the threshold of solving a number of the numerical problems associated with this method, and thus getting to the point where we can routinely expect to attain quite good solutions without expenditure of undue computer time and money. A large part of the solution appears to be in how we define the starting values for the mathematical programming component of the overall algorithm. We have recently made some fortunate discoveries that hopefully will yield more effective ways of doing this.

Some Applications of Multiple (Ultrametric) Tree Structure Fitting to Proximities Data

The procedure has been applied to the Miller-Nicely data on confusions of consonant phonemes [Carroll & Pruzansky, Note 1]. We found that two trees fit these data very well, accounting for about 99% of the variance in the data. The first tree corresponded essentially to the structure derived by use of standard hierarchical clustering techniques, and could be interpreted in terms of voicing and nasality. (It might also be mentioned that the single tree derived in a one-tree solution also corresponded essentially to this first tree, and accounted by itself for 94% of the variance. I stress, however, that it is not generally the case that the single tree from a one-tree solution will be one of the trees in a two-tree solution. If, as suggested earlier, there were two highly correlated tree structures that were about equally salient in the subjects' judgments, the best single tree would generally represent some kind of compromise between the two. A two-tree solution would, of course, pull the two apart, so to speak. Similar statements could be made for two versus three tree solutions, and so on.)

The second tree, which of course does not emerge at all in a clustering solution, seemed to relate to place of articulation. One interesting point that is consistent with the earlier suggestion about the possible relationship between ADCLUS and multiple tree structures is that almost every cluster found by Shepard and Arabie in their ADCLUS solution for the same data is represented as a node in one of the two trees.

Another multiple tree analysis, not previously reported in print, was of some data collected by Wish, Kaplan and Deutsch [1973] on various kinds of dyadic relationships. Wish et al. asked subjects to judge, on a 9-point rating scale, the similarity among each pair of dyadic relationships. (A typical judgment would involve rating the similarity of the relationship between, say, "mother-in-law and son-in-law" and that between "personal enemies.") Wish [in press] did many analyses, including an INDSCAL analysis of the three-way array of data for all subjects. The results of this analysis are shown in Figs. 3 and 4, which display the (unrotated) 1—2 and 3–4 planes of the INDSCAL stimulus space, respectively. As can be seen, Wish et al. interpreted Dimension 1 as "cooperation and harmony vs. com-

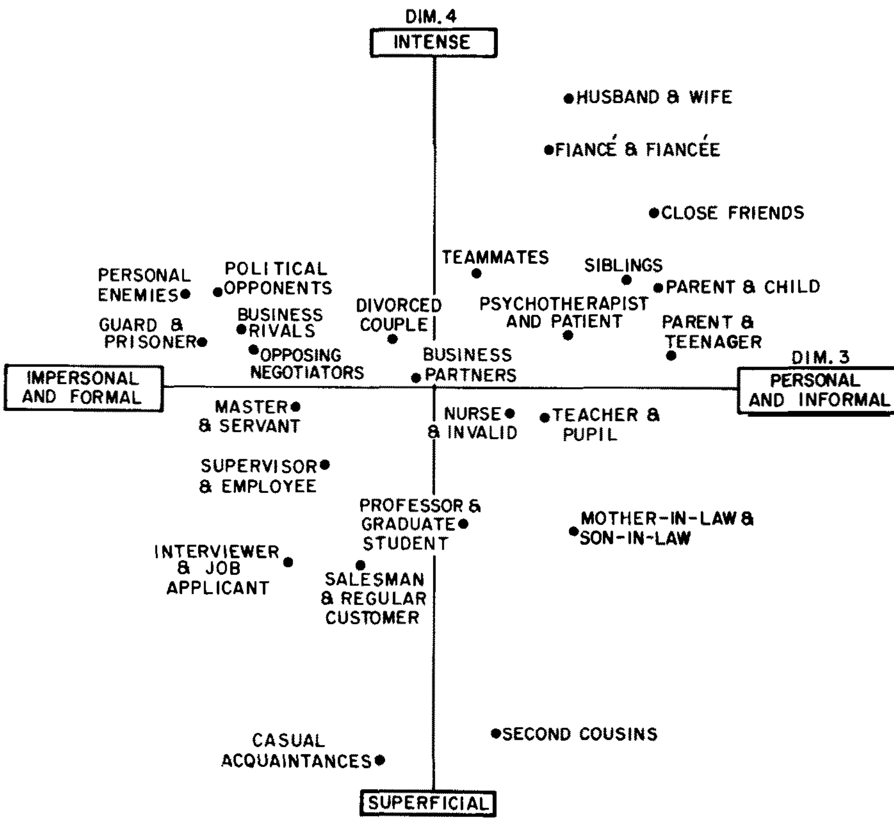


SIMILARITY RATINGS DATA:
DIMENSIONS 1 AND 2

FIGURE 3
Plane defined by dimensions one and two of group stimulus space from INDSCAL analysis of Wish et al. interpersonal relations data.

petition and conflict", Dimension 2 as "equality vs. inequality", Dimension 3 as "personal and informal vs. impersonal and formal" and Dimension 4 as "intense vs. superficial". These interpretations were supported by evidence based on rating-scale judgments.

Our two-tree solution for this same data (averaged, in this case, over subjects) resulted in structures that seemed to conform remarkably well with these INDSCAL results. In fact, there seems to be a direct relation between each of the trees and one of the two planes of the four dimensional stimulus space from INDSCAL. The major branch of the first tree (see Fig. 5), appears to divide the cooperative or only mildly competitive from the highly competitive relations ("mother-in-law and son-in-law", "second cousins" and "casual acquaintances" not quite fitting into this scheme). There



SIMILARITY RATINGS DATA:

DIMENSIONS 3 AND 4

FIGURE 4

Plane defined by dimensions three and four of group stimulus space from INDSCAL analysis of Wish et al. interpersonal relations data.

appears to be a mild anomaly in the location of "guard and prisoner", which is in the less competitive branch. However, looking at the first INDSCAL dimension suggests that this is in fact consistent with that solution—"guard and prisoner" is at almost exactly the same position on this dimension as is "parent and teenager". (In fact, "parent and teenager" and "guard and prisoner" are almost identically positioned on three of four of the INDSCAL dimensions, the exception being Dimension 3 "personal and informal" versus "impersonal and formal", on which they differ radically, being at opposite extremes. Perhaps this is telling us something important about intergenerational relationships!)

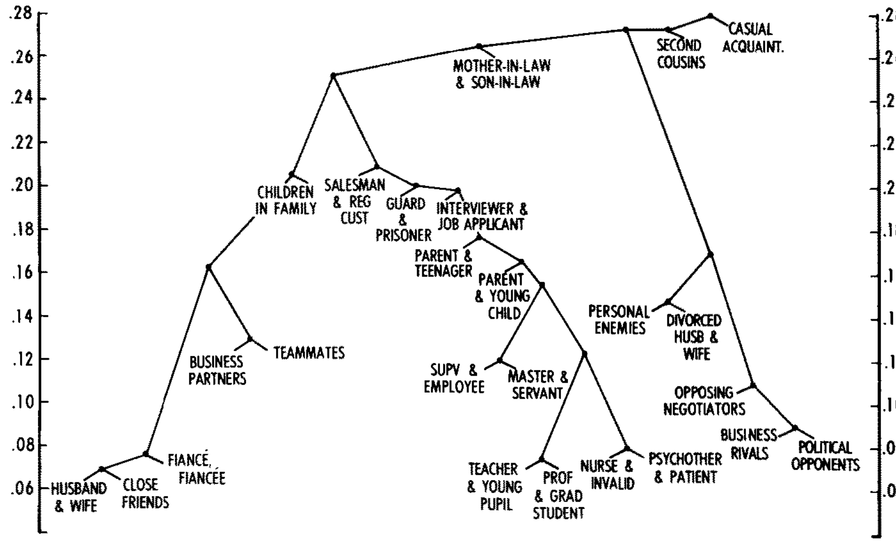


FIGURE 5

First tree structure from "two tree" solution derived from Wish et al. interpersonal relations data (averaged over subjects).

Within the cooperative or less competitive branch, we see a further division between the more-or-less symmetric or equal relations (e.g., "close friends," "husband and wife," "teammates") and the non-symmetric or unequal ones (e.g., "professor and graduate student", "nurse and invalid" and "interviewer and job applicant"). There is no comparable division among the highly competitive relations, but this is not surprising. An inspection of the particular relations that are high in the competitive dimension reveals that they are all more or less equal or symmetric, as evidenced by the fact that they are grouped very close together on Dimension 2. Perhaps this is a matter of stimulus selection, or perhaps highly competitive relations are by nature symmetric or equal. To put it differently, unequals cannot compete, for the ascendancy of the dominant one of the pair makes competition impossible. (Perhaps this also partly explains the seemingly anomalous position of "guard and prisoner".) Perhaps a better way to characterize this first tree is to say that it exhibits three main branches, corresponding to "equal and cooperative", "equal and competitive", and "unequal".

The tree structure also includes some things that are not so obviously present in the dimensional representation. The existence of nodes specific to the relations, "teacher and young pupil" and "professor and graduate student" for example, or to the two relations "business rivals" and "political opponents" (just to pick two of a number that could be mentioned) suggest attributes unique to these relations which would be very difficult to capture in a small dimensional spatial model, but are easily incorporated into a tree structure representation.

In an analogous way, the second tree, shown in Fig. 6, captures much of the structure in Dimensions 3 and 4. In this case the main branch seems to be correlated with Dimension 3. The left branch contains the personal and informal relations, while the right branch contains those that are more impersonal and formal. Each of these then can be seen to split into intense vs. superficial. Of the personal and informal relations, "husband and wife" and "fiance and fiancee" are examples of intense relations, while "casual acquaintances" and "second cousins" are more superficial ones. Of the impersonal and formal relations, "guard and prisoner" and "political opponents" are examples of intense relations, while "interviewer and job applicant" and "salesman and regular customer" are more superficial. Again, there is a structure in the tree that is hard to represent spatially, and, of course, there are ways in which the tree is inconsistent with the configuration in that plane.

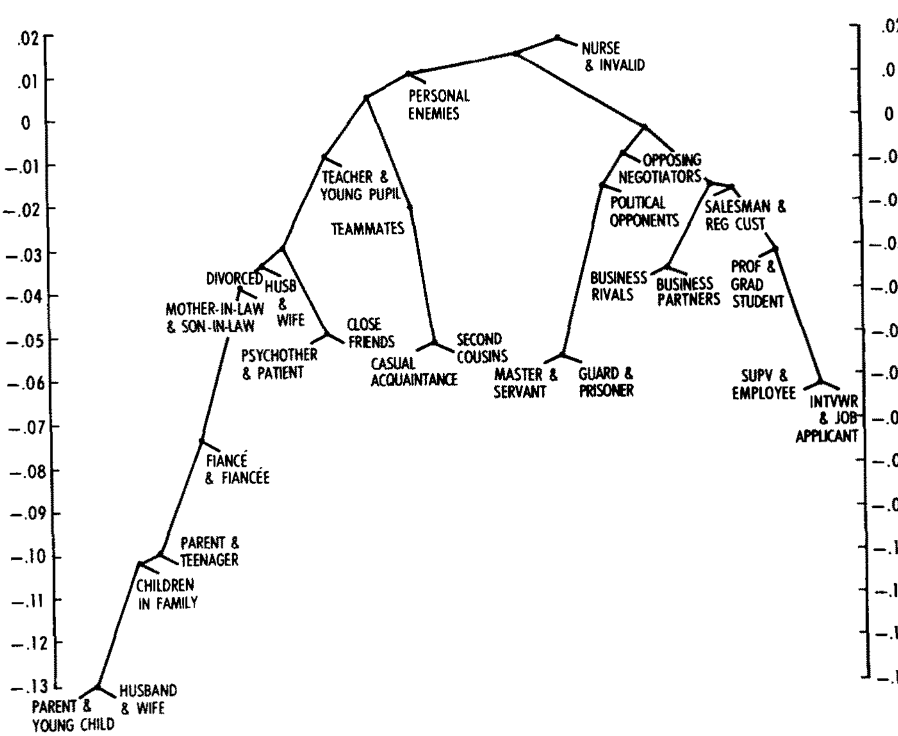


FIGURE 6

Second tree structure from Wish et al. interpersonal relations data (averaged over subjects).

Another way to see the relation between these two trees and the two planes of the INDSCAL space is to try to "map out" the contours for clusters defined by each of the two trees on the plane we have identified as corresponding to that tree. Figs. 7 and 8 show that we can do this reasonably well, although not as well as we might like. But if these contour maps are not satisfying, consider trying to map contours from Tree 2 on the 1–2 plane and from Tree 1 on the 3–4 plane. I tried this, as an exercise, and quickly discovered it to be an exercise in complete futility! As a result of these and other analyses, I have formulated a crude conjecture: roughly speaking, a single tree is worth two dimensions. This means, informally, that a tree contains about the same amount of information, generally speaking, as a two-dimensional space or plane. Obviously there are exceptions to this. For example, a linear ordering is a special case of a tree, but corresponds to a perfectly one-dimen-

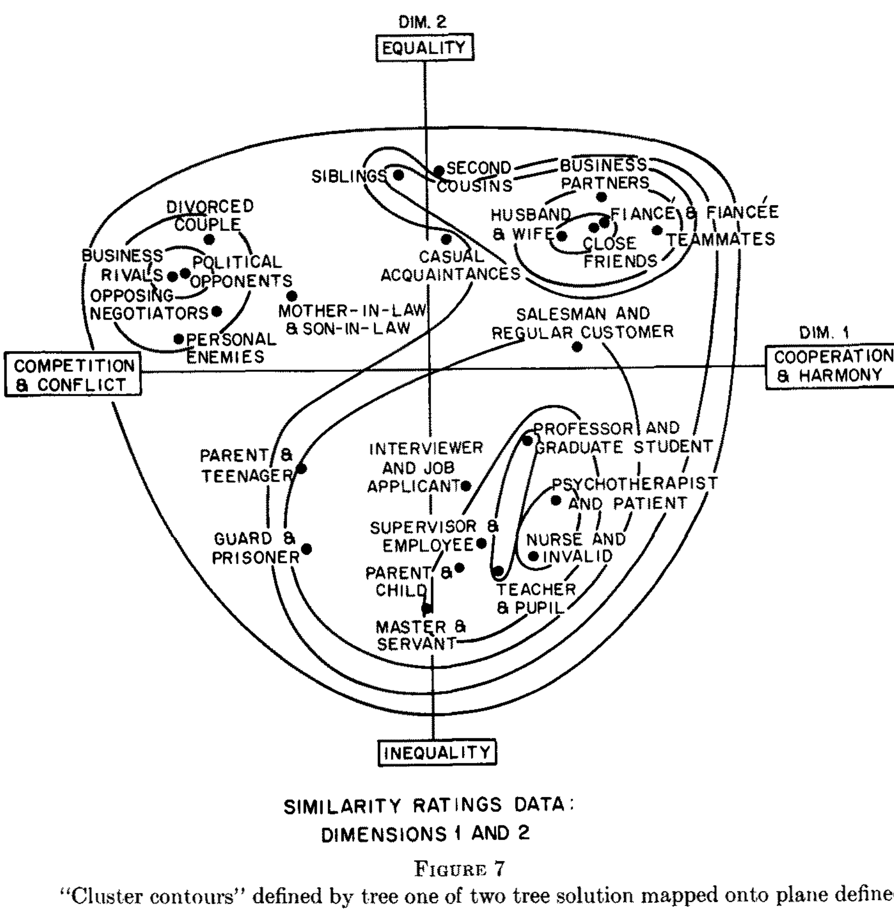


FIGURE 7

"Cluster contours" defined by tree one of two tree solution mapped onto plane defined by first two INDSCAL dimensions. (Wish et al. interpersonal relations data).

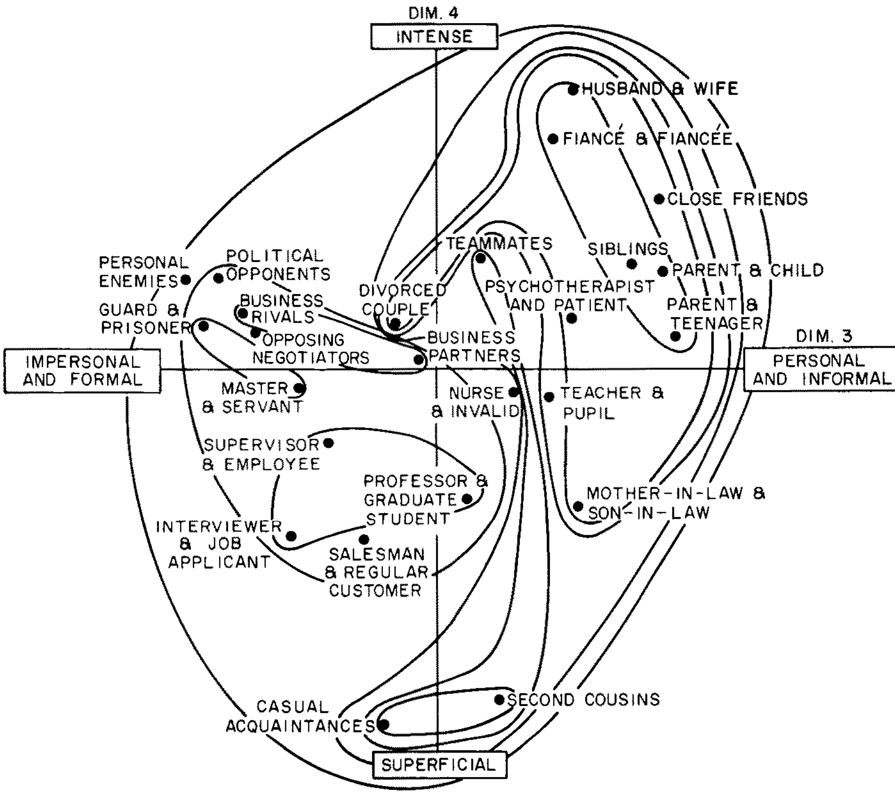


FIGURE 8

"Cluster contours" defined by tree two of two tree solution mapped onto plane defined by first two INDSCAL dimensions; third and fourth INDSCAL dimensions defined by tree two of two tree solution mapped onto plane defined by first two INDSCAL dimensions (Wish et al. interpersonal relations data).

sional structure. Also, as I have suggested, a tree can capture information that would require very many more dimensions in a spatial representation, although each such dimension or attribute may serve only to distinguish one or two stimuli from the others. Probably the right way to put it is that a general tree contains about the same *amount* of information as a two-dimensional structure, but it will generally be somewhat *different* information.

One bit of evidence in favor of my conjecture is that a general tree with path length metric (or a "free tree") has about the same number of independent parameters (branch lengths) as does a two-dimensional spatial configuration. The number is about $2(n - 1)$ in each case. (One could argue that this figure should be $2n - 3$ for the Euclidean case, but then it would be $2(n - 1) = 2n - 2$ for a non-Euclidean two space.) Of course, this assumes a fixed tree, whereas in the case of fitting tree structures the topology of the tree is optimized as well as the set of parameter values.

Another reason for my conjecture is the frequency with which it is possible to map cluster contours from a hierarchical clustering onto two-dimensional MDS configurations, even when there is clear evidence that more than two dimensions are necessary to account for the data. This fact would be explained if we simply assume that a single hierarchical tree extracted by clustering methods is in fact accounting for about the same information in the data as a two-dimensional MDS solution.

Holman [1972] has presented a theoretical objection to my conjecture by showing that distances from a general ultrametric tree on n objects cannot be accounted for perfectly by a Euclidean space of less than $n - 1$ dimensions. However, it seems to me that this theorem of Holman's is really saying something comparable to the fact that distances derived, say, from an r-dimensional city block space (or a space with any non-Euclidean metric) cannot be perfectly accounted for in an r-dimensional Euclidean space (nor, indeed, in any Euclidean space, no matter how large the dimensionality). While this is true, the Euclidean approximation may account for very nearly all the variance in the distances, and extract practically the same information as is contained in the city block configuration, although with some slight distortions. Of course, the real situation is probably one in which *neither* model is perfectly accurate, but each is accounting for about the same amount of structure, although each will favor certain kinds of structure over others. I am increasingly inclined to think of tree structures and spatial structures not so much as competing models as complementary ones, each of which captures certain aspects of a reality which is probably in fact much more complex than either model alone. For many purposes, of course, it may be justifiable to deliberately simplify reality in order to gain greater insight into major features of the data. This notion would probably appeal more, of course, to psychometricians. Mathematical psychologists may be more inclined to think in terms of critical experiments for distinguishing between discrete and continuous structure underlying judgmental behavior. Such experimental criteria would be extremely valuable, if available, to psychometricians as well.

Single and Multiple Tree Structures with Path Length Metric

Without going into detail, we may note that a very simple device allows us, in effect, to assume a model entailing single or multiple tree structures with path length metric (i.e., "free trees") rather than single or multiple ultrametric trees. This amounts to adding to the multiple ultrametric tree structure model an additional component which can be described as a symmetric matrix whose off diagonal entries are perfectly additively decomposable. That is, a matrix

$$C \equiv ||c_{ij}||,$$

where

$$c_{ij} = c_{ji} = a_i + a_j, \quad \text{for all} \quad i \neq j, \tag{8}$$

and

$$c_{ii} = 0, \quad \text{for all} \quad i.$$

Details of this are given in Carroll & Pruzansky [Note 1] where it is shown that a distance matrix from a path length tree can always be decomposed into the sum of a distance matrix from an ultrametric tree plus a matrix C of the form defined above. (This result was based in part on an unpublished theorem due to S. J. Farris.) One practical consequence of this is that it may be difficult to distinguish between ultrametric trees and path length or free trees, particularly when one realizes that the C matrix mentioned above is itself the distance matrix resulting from a "free tree" of a very special type, namely one that has exactly one interior node, with all terminal nodes (corresponding to stimulus objects) having branches to that single node. A tree of this kind is illustrated in Fig. 2, which was referred to earlier as a "star", except that in this case the branch lengths are not necessarily all equal. In fact, some of the branch lengths may be negative, so that this additive part may not correspond at all to a "real" tree with positive branch lengths. However, since there is more than one way to decompose a path length tree into an ultrametric tree plus an additively decomposable part (there are, in fact, as many ways as the tree has nodes), it may be that one way will correspond to an additive part that is all positive. Sattath and Tversky [Note 5] discuss conditions for this to be the case. There is a quite plausible error theory that would give rise to a matrix C of this additively decomposable form, and thus have the effect of making the two resulting statistical models—one based on ultrametric trees and the other on free trees—completely indistinguishable.

Hybrid Models

Degerman [1970] proposed the first formal hybrid model combining elements of continuous dimensional structure and of discrete class-like structure. Degerman described a scheme for rotating a high dimensional MDS solution to find subspaces in which there was class-like rather than continuous variation. Since then, much has been said about such mixed or hybrid models. but little has been done about them.

One possible hybrid model can be formulated by further generalizing the multiple tree structure model we have proposed to include a continuous spatial component in addition to the tree structure components. To go back to our animal name example, we might postulate, in addition to the two hierarchical structures already mentioned, continuous dimensions of the kind best captured in spatial models. Obvious examples in the case of animals would be such dimensions as size, ferocity, color (which itself is multidimensional) and so on.

The multiple tree structure model has been generalized in precisely this direction. The model, formally, can be expressed as

$$\underline{\Delta} \cong D_1 + D_2 + \cdots + D_m + D_{E_r}^{\ 2}, \tag{9}$$

where D_1 through D_m are distance matrices arising from tree structures (either ultrametric trees of free trees) and $D_{E_r}^{\ 2}$ is a matrix of *squared* distances arising from an r-dimensional Euclidean space. (The reason for adding squared rather than first power Euclidean distances is a technical one having to do largely with mathematical tractability, and it will not be dealt with here.) In effect, to estimate this additional continuous component we simply add an extra phase to our alternating least squares algorithm that derives conditional least squares estimates of these components. Again, technical details are in the Carroll and Pruzansky [Note 1] paper.

An illustration comes from a study by Rosenberg and Kim [1975], who obtained data based on subjective sortings of kinship terms. Rosenberg and Kim were quite perplexed by the fact that standard hierarchical clustering, when applied to these data, resulted in a structure that made a great deal of sense but failed to be sensitive to a very critical dimension they knew was in the data, namely sex. While a majority of subjects in effect ignored sex in making their sortings (for example, always sorting "mother" and "father" together, and "son" and "daughter" together), a minority of the subjects clearly used that dimension as the basis for sorting, which could be seen from simple inspection of the raw data. Somehow the structure used by the majority of the subjects completely dominated and masked that used by this minority. MDS analysis had also shown clearly the presence of the sex dimension. We first thought a two-tree structure model would be appropriate, and indeed it did capture the essential features of the data quite nicely. The first tree, shown in Fig. 9, corresponds very well to a standard anthropological model—the Romney-D'Andrade model [Romney & D'Andrade, 1964]—for kinship terms. The major branch distinguishes between the "direct" relatives—lineals (who are in the same line of descent) and siblings—and the "collaterals"—the colineals (e.g., uncle, aunt, nephew, niece) and ablineals (cousins). Within each branch there is a further breakdown based on what might be called "absolute generation" (i.e., generational distance from "ego", or oneself). Within the "directs", for example, we have a node representing siblings who are of the same generation as ego; then a node representing children (daughter, son) or parents (father, mother), i.e., lineals one generation removed from ego; and then a node representing those lineals two generations

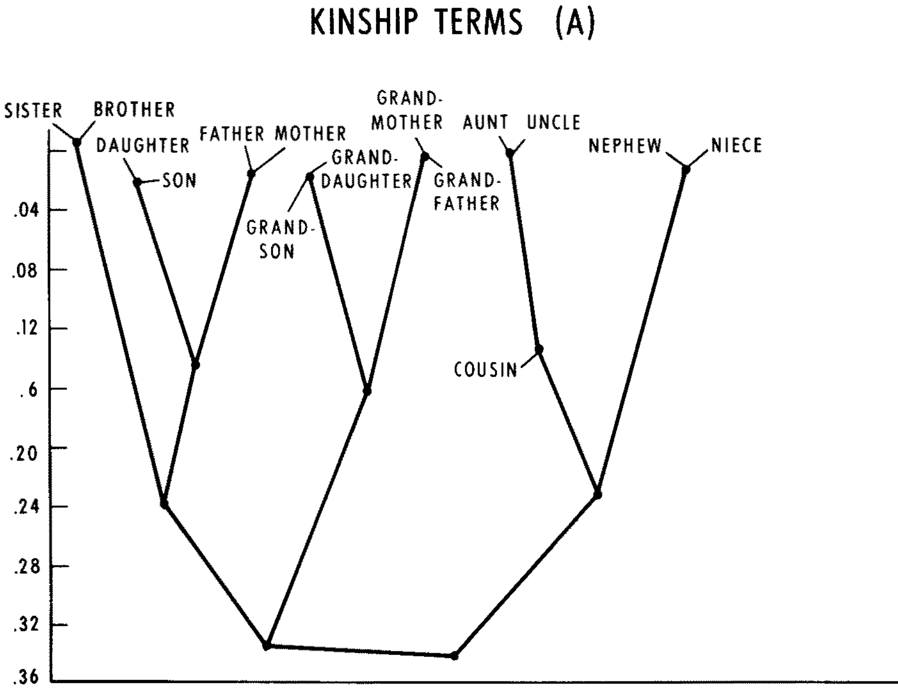


FIGURE 9
First tree from a "two tree" solution for Rosenberg and Kim kinship data (amalgamated over subjects).

removed from ego (grandchildren and grandparents). The collateral branch splits into those one generation above ego (aunt, uncle) those one generation below ego (nephew, niece) with the term at the same generational level (cousin) in some sense "between" those two nodes. Sex is nowhere in evidence as a basis for this tree. The analogous kinterms of opposite sex are always together at the same node.

The second tree, seen in Fig. 10, has its main division based on sex—all the male terms at one node and all the female at another, with cousin, the only genderless term, in the middle. The structure within each of these branches seems to have something to do with generation, but it is difficult to decipher.

We then decided to try our hybrid model, thinking that perhaps a single tree structure plus one dimension (presumably "sex") would provide a more parsimonious account of these data. As it happens, however, the data suggested *two* dimensions, in addition to the tree. The second dimension, while accounting for only a very small proportion of variance, did appear to be very regular, and also appears to be relatively interpretable. The single

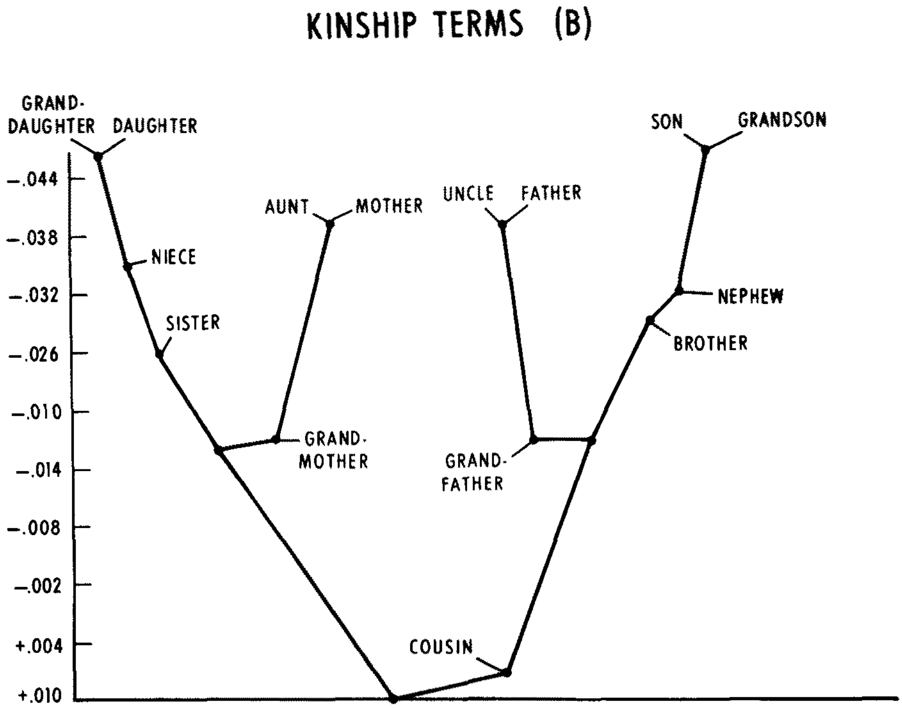


FIGURE 10
Second tree from a "two tree" solution for Rosenberg and Kim kinship data (amalgamated over subjects).

tree, in this case, was essentially identical to the first tree of our two-tree solution, and so will not be discussed. The two-dimensional spatial component is shown in the next figure. Clearly the first dimension is sex—the male terms on the left, female terms on the right, and the genderless cousin in the middle. The second dimension appears in this case to be a kind of "folded" generation dimension, but different from the absolute generation dimension (which amounts to folding the generation dimension around ego—or the same generation as itself). This dimension appears to be the generation dimension folded somewhere between one generation earlier (e.g., father) and two generations earlier (e.g., grandfather). While this would far from perfectly account for this second dimension, it does give a reasonably good account of it. Wish has suggested "dominance" as a reasonable interpretation of this dimension. Presumably, in western cultures at least, parents tend to be more dominant than grandparents, for example. Pruzansky and Kaplan suggested "dependence" as possibly a better name, since grandparents are certainly more likely to be dependent, economically and otherwise, than are parents, although they may not be any less dominant. Skeptics, of course, may suggest that

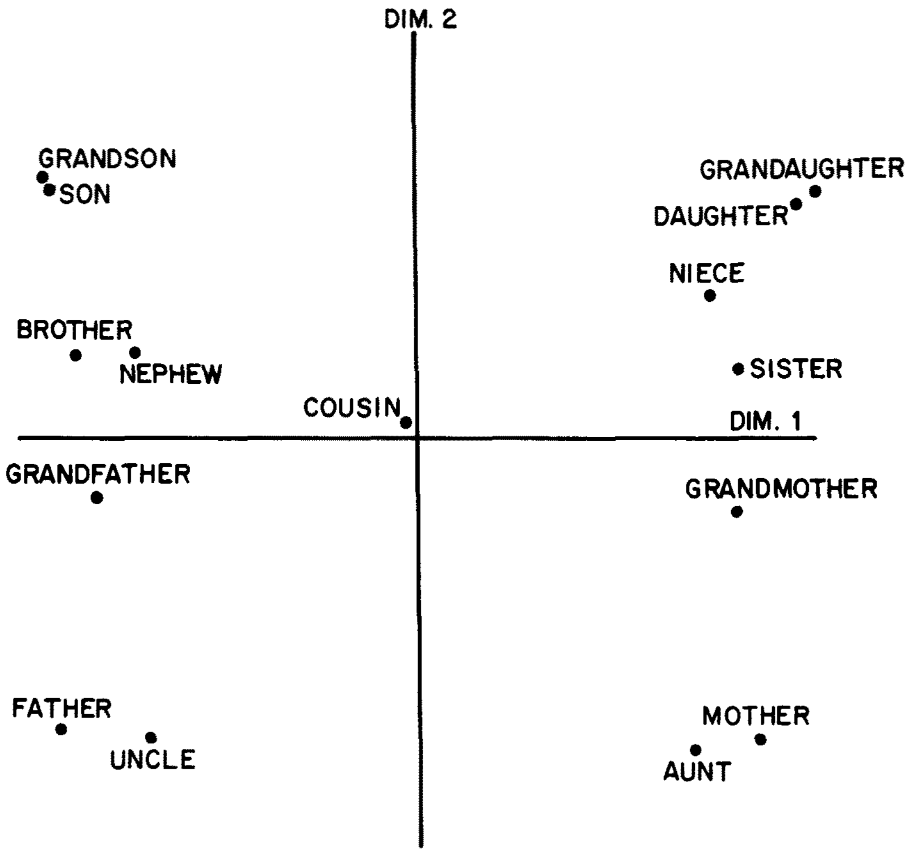


FIGURE 11

Two dimensional Euclidean spatial component from an analysis of Rosenberg-Kim kinship data in terms of a hybrid model assuming a simgle tree structure plus two dimensions. Tree structure component was essentially identical with first tree from "two tree" solution, shown in Figure 9.

this merely proves how adept psychologists can be at interpreting almost anything! In contrast with these "psychological" interpretations, Romney is firmly convinced that this dimension is what he and D'Andrade (who are both anthropologists) called "reciprocity", with the "juniors" above, the "seniors" below, and those at ego's generation level in the middle. Thus the dimension has at least four interpretations! Seriously, these are four different ways of describing the same thing—from four somewhat different perspectives (that might be called formal analytical, social psychological, sociological-economic, and formal linguistic-anthropological.) The four interpretations are not necessarily in conflict; rather, they should probably be viewed as complementary. To put it differently, one may plausibly and without logical contradiction assert that all four are correct!*

Our work on hybrid models is just getting underway, but the initial results justify continued effort.

Other Models for the Future

The kind of "hybrid" model discussed above treats the discrete components and their continuous spatial components as completely independent. The dimensions cut across all subgroups or clusters defined by different nodes of the tree or trees. Another kind of hybrid model would assume that these are not at all independent, but rather that some dimensions are relevant only to objects within a particular group or class and not to objects in other classes. A typical example would be the dimension of size, which may be relevant to such classes of objects as animals, furniture and buildings, but not to entities such as ideas or emotions. Similarly, the color of an electron or the weight of an argument are undefined (except in some vague metaphoric way). Arabie and I are now jointly working on a generalization of the ADCLUS procedure (using a different algorithm based on mathematical programming and alternating least squares approaches) that will allow for such a definition of dimensions that are relevant only to objects in the same class or group. Of course, one might also include in such a hybrid model some dimensions that "cut across" all classes. The exciting prospect is that current development of both computer hardware and numerical optimization software enable us to implement programs that permit fitting such highly complex models. The only limiting factor, ultimately, may be our own imaginations.

Another direction of future exploration which seems potentially fruitful

*As a final note on interpretation of this somewhat anomalous second dimension, I followed a suggestion of Rosenberg's to "look at the data", which he then kindly provided me. I looked, and found, indeed, that there were three subjects (out of 85) whose sortings were perfectly accounted for by this dimension, in the sense that one could divide that dimension into contiguous segments, with each segment corresponding to a grouping of the kinship terms formed by that subject. For example, one subject simply sorted the terms into only two groups—one containing "granddaughter," "grandson," "daughter," "son" and "niece" (??), and the other containing all the other terms. Splitting Dimension 2 between "niece" and "nephew" would give precisely this partition of the terms (this in fact may account for "niece" and "nephew" having slightly different positions on that dimension). The second subject partitioned the terms into three groups, and the third into five which could be defined by splitting Dimension 2 into three or five segments respectively. Thus there is no question that this dimension is systematically related to the data of these three subjects, although it may still not be obvious exactly what it "means". (All three of these sortings could just as well be accounted for by an *unfolded* generation dimension, incidentally, but with "niece" and "nephew" in slightly different positions to account for the first subject mentioned above. This fits well with the "reciprocity" notion, except that "reciprocity" would be binary, or possibly ternary, rather than continuous. The fact that this particular "folding" of the generation dimension in fact emerged presumbly may have to do with small but systematic trends in the data of other subjects; for example, a slight tendency to sort grandparents somewhat more frequently than parents with younger generation terms, which fits with the "dominance" or "dependence" hypothesis.)

is the inclusion of individual differences in some of the nonspatial and hybrid models I have discussed here. Arabie and I are planning to extend ADCLUS to individual differences, with individual subject weights for the discrete attributes providing, in many ways, a discrete analogue to the INDSCAL model. In the case of tree structures and multiple tree structures, the obvious generalization to individual differences would involve assuming that different individuals base their judgments on the same family of trees, but with different sets of parameters (node "heights" and/or branch lengths) for different individuals. In the hybrid case, individual subject weights, à la INDSCAL, can also be introduced. Pruzansky and I already have formulated algorithm for such extensions.

But we must walk before we run. There are still many unsolved numerical problems in the two-way (non-individual differences case). It would seem that these must be solved before we extend these tree structure and hybrid models to the three-way case.

In summary, my perception—or apperception—of the future of multidimensional scaling sees it extending to include many new and different models of all shapes and kinds; spatial, non-spatial and hybrid, in two-way, three-way and perhaps even multi-way versions. I think sophisticated use of the full power of the high speed digital computer will eventually free us almost totally from having to restrict ourselves to models that are mathematically simple and tractable, and allow us to pursue instead those models that most effectively mimic what is going on inside the head. Since what is going on inside the head is likely to be complex, and is equally likely to have both discrete and continuous aspects, I believe the models we pursue must also be complex, and have both discrete and continuous components. We have begun moving in that direction, but there's still a long way to go.

REFERENCE NOTES

1. Carroll, J. D., and Pruzansky, S. Fitting of hierarchical tree structure (HTS) models, mixtures of HTS models, and hybrid models, via mathematical programming and alternating least squares. Presented at the *U.S.-Japan Seminar* on Theory, Methods, and Applications of Multidimensional Scaling and related techniques, San Diego, Aug. 1975. Paper in informally published Proceedings of U. S.-Japan Seminar, available on request.
2. Cunningham, J. P. On finding an optimal tree realization of a proximity matrix. Paper presented at the Mathematical Psychology meeting, Ann Arbor, Michigan, August 29, 1974.
3. Harshman, R. A. PARAFAC2: Mathematical and technical notes (Working Papers in Phonetics No. 22). Los Angeles: University of California, March, 1972.
4. Kruskal, Joseph B. Some advances in parametric mapping. Psychometric Society Presidential Address, presented at joint meeting of Psychometric Society and Classification Society, Iowa City, Iowa, April, 1975.
5. Sattath, S. and Tversky, A. Additive similarity trees. Unpublished manuscript, Hebrew University, 1976.
6. Shepard, R. N. and Arabie, P. Additive cluster analysis of similarity data. Paper

presented at the U.S.-Japan Seminar on Theory, Methods, and Applications of Multidimensional Scaling and Related Techniques, San Diego, 1975. Paper in informally published Proceedings of U.S.-Japan Seminar.

7. Tversky, A. Features of similarity. Unpublished manuscript, Hebrew University, Jerusalem, Israel, 1976.

REFERENCES

Buneman, P. The recovery of trees from measures of dissimilarity. In F. R. Hodson' D. G. Kendall and P. Tautu (Eds.), *Mathematics in the archaeological and histroical sciences*. Edinburgh: Edinburgh University Press, 1971, 387–395.

Carroll, J. D., and Chang, J. J. Analysis of individual differences in multidimensional scaling via an N-way generalization of "Eckart–Young" decomposition. *Psychometrika*, 1970, *35*, 283–319.

Carroll, J. D. and Chang, J. J. A method for fitting a class of hierarchical tree structure models to dissimilarities data and its application to some "body parts" data of Miller's. *Proceedings of the 81st Annual Convention of the American Psychological Association*, 1973, *8*, 1097–1098.

Degerman, R. Multidimensional analysis of complex structure: Mixtures of class and quantitative variation. *Psychometrika*, 1970, *35*, 475–491.

de Leeuw, J., Young, F. W. and Takane, Y. Additive structure in qualitative data: An alternating least squares method with optimal scaling features. *Psychometrika*, 1976, 000–000.

Dobson, J. Unrooted trees for numerical taxonomy. *Journal of Applied Probability*, 1974, *11*, 32–42.

Hartigan, J. A. Representation of similarity matrices by trees. *Journal of the American Statistical Association*, 1967, *62*, 1140–1158.

Holman, E. W. The relation between hierarchical and Euclidean models for psychological distances. *Psychometrika*, 1972, *37*, 417–423.

Horan, C. B. Multidimensional scaling: Combining observations when individuals have different perceptual structures. *Psychometrika*, 1960, *34*, 139–165.

Johnson, S. C. Hierarchical clustering schemes. *Psychometrika*, 1967, *32*, 241–254.

Miller, G. A. A psychological method to investigate verbal concepts. *Journal of Mathematical Psychology*, 1969, *6*, 169–191.

Patrinos, A. N. and Hakimi, S. L. The distance matrix of a graph and its tree realization. *Quarterly of Applied Mathematics*, 1972, *30*, 255–269.

Romney, A. K. and D'Andrade, R. G. Cognitive aspects of English kin terms. *American Anthropologist*, 1964, *66*, No. 3, part 2 (special publication), 146–170.

Rosenberg, S. and Kim, M. P. The method of sorting as a data-gathering procedure in multivariate research. *Multivariate Behavioral Research*, 1975, *10*, 489–502.

Shepard, R. N. Representation of structure in similarity data: Problems and prospects. *Psychometrika*, 1974, *39*, 373–421.

Torgerson, W. S. Multidimensional scaling of similarity. *Psychometrika*, 1965, *30*, 379–393.

Tucker, L. R. Relations between multidimensional scaling and three-mode factor analysis. *Psychometrika*, 1972, *37*, 3–27.

Wish, M., Kaplan, S. J. and Deutsch, M. Dimensions of interpersonal relations: Preliminary results. *Proceedings of the 81st Annual Convention of the American Psychological Association*, 1973, *8*, 179–180.

Wish, M. Comparisons among multidimensional structures of interpersonal relations. *Multivariate Behavioral Research*, in press.

Wold, H. Estimation of principal components and related models by iterative least squares. In P. R. Krishnaiah (Ed.), *Multivariate Analysis*. New York: Academic Press, 1966.