Annu. Rev. Psychol. 1992. 43:169–203

COMBINATORIAL DATA ANALYSIS

Phipps Arabie

Graduate School of Management, Rutgers University, Newark, New Jersey 07102-1895

Lawrence J. Hubert

Department of Psychology, University of Illinois, Champaign Illinois 61820

Dedicated to Ledyard Tucker

KEY WORDS: seriation, clustering, partitioning, additive trees, network models

---

CONTENTS

0066-4308/92/0201-0169$02.00

INTRODUCTION

Although the *Annual Review of Psychology* periodically offers chapters on topics in quantitative methodology (e.g. L. V. Jones & Appelbaum's 1989 chapter on psychometric methods), it occasionally allows reviews to herald new subareas of intense development (e.g. Smith's 1976 chapter on analysis of qualitative data). The present chapter is of the latter kind. There is thus no strong consensus on the boundaries of our area or even on when coverage in a review chapter should begin. There is, however, one enduring certainty: We lack adequate space to discuss all the meritorious and relevant work.

Combinatorial data analysis (CDA) concerns the *arrangement* of objects for which relevant data have been obtained. Stated more explicitly, CDA is involved either with (*a*) the location of arrangements that are optimal for the representation of a given data set (and thus is usually operationalized using a specific loss-function that guides the combinatorial search defined by some set of constraints imposed by the particular representation chosen) or with (*b*) trying to determine in a confirmatory manner whether a specific object arrangement given *a priori* reflects the observed data. CDA does not (cf Rouanet et al 1986) postulate strong stochastic models based on specific unknown parametric structures as underlying a given data set. Although CDA might use or empirically construct various weighting functions, the weights so obtained are not intended to be interpreted as parameter estimates in some presumed stochastic model viewed in turn as responsible for generating the data. In CDA, manifest data are emphasized, and the traditional concern for an assumed relationship between the data and a restrictively parameterized stochastic population model is essentially ignored.

Methods of CDA are generally organized around (*a*) the types of combinatorial structures we might use to represent a given data set (e.g. Guénoche & Monjardet 1987, Hubert 1987) and (*b*) the classes of combinatorial optimization methods used in solving problems of actually locating optimal arrangements rather than the (confirmatory) evaluation of a given one. The latter emphasis on computation may seem unusual, but such staples of combinatorial structures as the "traveling salesman problem," the "minimum spanning tree," and "additive trees" have been used to connect seemingly unrelated techniques of discrete data analysis (see respectively Hubert & Baker 1978; Hubert 1974a; Carroll 1976); similarly, explicit optimization techniques like "branch-and-bound" and "dynamic programming" have been used to suggest a general approach to broad classes of data analysis tasks (see respectively Hand 1981; Hubert & Golledge 1981).

The need for new techniques to answer questions that, for computational ease, had often been approached using restrictive parametric models (even when their underlying assumptions may be very unrealistic for the problem under study) has not always been acknowledged. According to personal communications, Lerman had considerable difficulty getting his (1980) programmatic statement published, and Edgington (1980) found a generally unreceptive audience for the randomization emphasis in his text. But the Zeitgeist has caught up with and vindicated those authors, as judged by the acclaim for their continuing contributions to the area (e.g. Edgington 1987; Lerman 1988) and by special issues of quantitative journals devoted entirely to combinatorial data analysis (e.g. Leclerc & Monjardet 1987a,b,c).

Most of the papers cited here were published after 1974. We avoid certain research areas here either because they rely implicitly on parameter estimation (e.g. classification decisions based on discriminant analysis) or because they merit their own review chapters (e.g. the related topics briefly mentioned in the section on Rankings, Relations, and Partially Ordered Sets, below). Our terminology for models and types of data follows that of an earlier *Annual Review* chapter by Carroll & Arabie (1980:609–12). We make extensive use of Tucker's (1964) terminology, distinguishing between "ways" (a matrix that has rows and columns is two-way) and modes: If the two ways both correspond to the same set of entities, as in a proximity matrix of n stimuli by n stimuli, the data are two-way one-mode; if the rows and columns correspond to disjoint sets (e.g. subjects by attributes), the data are two-way two-mode. We note that this terminology has found its way into relevant book titles (e.g. one dedicated to Tucker and edited by Law et al 1984; Coppi & Bolasco 1989).

SERIATION

Methods of seriation are, in effect, techniques for the unidimensional scaling or sequencing of a set of objects along a continuum. In the last several decades, these and other strategies of analysis have been developed most aggressively by archaeologists whose frequent concern is with a two-mode matrix of artifacts by sites which is typically converted to a one-mode matrix amenable to nonmetric multidimensional scaling (MDS) techniques. A review of this methodology may be found in Carroll & Arabie (1980:617; also see Lerman 1981: Ch. 8; Pliner 1984; Halperin 1989). The inherent problem is combinatorial in nature and can be characterized via the ordering of the objects defined by the one mode. Because the ostensible task is to place objects along a continuum so as to optimize an objective function, the problem seems to be one of location estimation along the real line. It is thus susceptible to the usual type of gradient-based optimization techniques (e.g. Kruskal 1964a,b). Such intuition conflicts with widespread observed failures (e.g. Shepard 1974:378–79) of gradient-based MDS algorithms when unidimensional solutions are sought. De Leeuw & Heiser (1977:740), however, noted that the one-dimensional MDS problem was in fact one of combinatorial optimization and was thus reducible to the search for an ordering along a continuum, possibly with a secondary estimation of the actual coordinates if desired.

Hubert & Arabie (1986) demonstrated analytically why gradient-based MDS approaches fail for the unidimensional case. We have since provided a combinatorial algorithm that guarantees a global optimum and is computationally feasible for medium-sized (e.g. $n = 20$) proximity matrices (Hubert & Arabie 1986). The difficulties with location estimation based on the usual gradient methods are not restricted to the unidimensional case, and we have shown that gradient-based MDS techniques will fail for identical technical reasons when city-block representations are sought in two or more dimensions. As an alternative, combinatorial approaches to higher-dimensional city-block spaces are being developed (Hubert & Arabie 1988; Arabie et al 1989; Heiser 1989). These methods are in a tradition of combinatorial approaches to MDS (e.g. Hubert & Schultz 1976; Waller et al 1992).

CLUSTERING

Perhaps the most well-developed and commonly used form of combinatorial data analysis is clustering, which comprises those methods concerned in some way with the identification of homogeneous groups of objects, based on whatever data are available. The immensity of the literature on cluster analysis precludes our giving much space to applications; fortunately, the Classification Society of North America publishes an annual bibliography, *Classification Literature Automated Search Service,* based on citations to "classic" articles and books as compiled by the Institute for Scientific Information. The articles cited there are drawn primarily from the periodical literature on clustering and multidimensional scaling; there were 887 citations for 1990 (Volume 19; Day 1990).

Evidence of the activity propelling research in clustering is given by a selection of full-length texts devoted to the subject (Sneath & Sokal 1973; Bock 1974; Duran & Odell 1974; Hartigan 1975; Spaeth 1980, 1985; Gordon

1981; Lorr 1983; Murtagh 1985a; Godehardt 1988; Jain & Dubes 1988; Mandel 1988 [see Kamensky 1990]; McLachlan & Basford 1988) as well as a lengthy report commissioned by the US National Research Council (Panel on Discriminant Analysis, Classification, and Clustering 1988). This profusion of texts occasionally evinces a scholar's counterpart to Gresham's Law: The same publisher that took the English-language edition of Hartigan (1975) out of print released a second edition of Everitt (1980). The technique is mentioned in such potboilers as *Spycatcher* (Wright 1987:153), is central to *The Clustering of America* (Weiss 1988), and along with some other methodology suffers frequent misrepresentation in the literature on "artificial intelligence" (e.g. Denning 1989; and cf Dale's 1985 criticisms of Michalski & Stepp 1983). In addition to conference volumes largely devoted to clustering and closely related methods (e.g. Van Ryzin 1977; Felsenstein 1983a; Diday et al 1984; Diday et al 1986; Gaul & Schader 1986; Bock 1988), there have been some noteworthy reviews in such substantive areas as marketing (Wind 1978; Punj & Stewart 1983; Dickinson 1990; Beane & Ennis 1987; also see Green et al 1989: Section 7), experimental psychology (Luce & Krumhansl 1988; Shepard 1988), developmental psychology (K. Miller 1987), clinical psychology (Blashfield & Aldenderfer 1988), criminology (Brennan 1987), information retrieval (Willett 1988), biophysics (Hartigan 1973), image segmentation and related areas of computer science (Dubes & Jain 1979; Jain & Dubes 1988: Ch. 5) and phenetic taxonomy (Sokal 1986a, 1988a). More methodologically oriented review articles are cited below.

Formal Underpinnings

For many statisticians, the shady history of cluster analysis is an ongoing cause for suspicion. The early approaches to clustering (e.g. McQuitty 1960) were usually mere convenient algorithms devoid of any associated representational model or effort at optimizing a stated criterion. Subsequent remedies to this situation have taken two paths that with few exceptions (e.g. Ling 1973) are distinct.

The first is to note that the structures sought by many hierarchical clustering methods (whose output is often represented as an inverted tree diagram or dendrogram, as considered in Murtagh 1984a; see Gordon 1987a for an excellent, comprehensive review) conform to the ultrametric inequality

$$d_{ij} \leq \max(d_{ik}, d_{jk}) \text{ for all } k,$$

where i, j, and k are members of the stimulus set; d_{ij} is the distance between stimuli i and j predicted by the cluster analysis and corresponds to the observed dissimilarity measure that is input for the cluster analysis. The ultrametric inequality is to most forms of hierarchical cluster analysis what the triangle inequality is to two-way MDS (Shepard 1962a,b; Kruskal 1964a,b). That is, the predicted or reconstructed distances resulting from a relevant hierarchical cluster analysis conform to the ultrametric inequality, just as those from two-way MDS satisfy the triangle inequality. The ultrametric inequality was introduced independently into the literatures of biology, experimental psychology, and data analysis by C. J. Jardine et al (1967) and Johnson (1967; also see Hartigan 1967; and Lance & Williams 1967) in the same year; recognition of its importance in the physical sciences has been somewhat delayed (Rammal et al 1986) and apparently remains to be achieved in the neurosciences (e.g. Ambros-Ingerson et al 1990). A tradition of close inspection of its implications for clustering methods has led to a better understanding of how various hierarchical techniques are interrelated (Hubert & Baker 1976, 1979; Jambu 1978; Kim & Roush 1978; Leclerc 1979; Milligan 1979; Batagelj 1981; Leclerc 1981, 1986; Hubert 1983; Degens 1983, 1985; Herden 1984; Barthélemy & Guénoche 1988; Critchley & Van Cutsem 1989; Ohsumi & Nakamura 1989).

The second approach to providing a more defensible logical basis for clustering algorithms is to relate such algorithms to the vast literatures on graph theory (e.g. Monjardet 1978, 1981a; see reviews by Hubert 1974b and Guénoche & Monjardet 1987) and set and lattice theory (Barbut & Monjardet 1970a,b; Hubert 1977a; Janowitz 1978, 1979; Janowitz & Schweizer 1989). This strategy has yielded new insights about the results from cluster analyses (e.g. Ling 1975; Hubert & Baker 1976, 1979; Matula 1977; Frank 1978; Tarjan 1982; Godehardt 1988; Sriram 1990) and about the design of faster and more capacious algorithms (e.g. Hansen & Delattre 1978; Day 1984; Day & Edelsbrunner 1985; see reviews by Murtagh 1983, 1984b). Although much of this literature reaches a technical depth beyond the training of most psychologists (thus perhaps explaining why some of the field's most rudimentary aspects are continually reinvented—e.g. Cooke et al 1986), parts of it remain basic and highly applicable.

For example, a topic of considerable economic importance in graph theory is the *minimum spanning tree* (MST) problem, in which one employs a graph whose nodes correspond to stimuli and whose edges represent possible links, with weights typically used to predict or reconstruct the empirical dissimilarities data. The objective is to find that tree spanning the graph (so that there is a path between each pair of nodes, but without any cycles) for which the sum of the edge weights is a minimum/maximum for dis/similarities. Solving the MST problem is formally equivalent to performing single-link clustering (Gower & Ross 1969), and the connection between clustering and spanning trees has proven substantively useful (Hubert 1974a; Murtagh 1985a: Ch. 4). Although it is generally assumed that interest in the MST first arose in engineering (e.g. in the layout of telephone, powerline, and other types of networks), Graham & Hell's (1985) laudable and comprehensive history of the problem indicates that work by the anthropologist Czekanowski (1909) enabled Florek et al (1951) to devise single-link clustering before the appearance of the paper (Kruskal 1956) most heavily cited in the English literature on MST algorithms relevant to engineering. Thus, a formal problem first recognized in the behavioral sciences became one of enormous practical importance.

Lest our summary of these developments in clustering seem too optimistic, we should note the absence of progress in a few areas. For example, N. Jardine & Sibson (1971) introduced the useful distinction between methods of clustering versus algorithms for implementing them: "It is easy to show that the single-link method, for example, can be implemented by a divisive algorithm, an agglomerative algorithm, or by an algorithm which belongs to neither category" (p. 42). While this sensible recommendation has been endorsed by various leaders in the field (e.g. Rohlf 1982:267), there are still many authors whose writing would be much clearer if the distinction were respected in their papers. Another regrettable tendency is the occasional resurgence of a nostalgic preference for algorithms supported neither by optimization nor models (e.g. Whaley 1982:173) versus the more current approach outlined above.

We now consider some of the areas in which clustering has seen greatest development in recent years.

Clustering of Binary Data

As the limiting case of discreteness, binary (0, 1) data have often claimed a special place in the discussion of numerous forms of data analysis (Cox 1970; Tucker 1983), and clustering is no exception. Such input data are usually two-mode, and a selected list of clustering methods developed specifically for them includes Lerman et al (1980), Buser & Baroni-Urbani (1982), Brossier (1984), Guénoche (1985), Cliff et al (1986), Muchnik et al (1986), Govaert (1988), Barthélemy (1989), Li & Dubes (1989), and Mkhadri (1989). An especially elegant and model-based method is that of De Boeck & Rosenberg (1988; also see Van Mechelen & De Boeck 1990).

A closely related problem of data analysis, often omitted in surveys of methodology because of its nonprobabilistic basis, is the following: Devise a set of binary vectors or *keys* for optimally and parsimoniously classifying a set of objects, each represented itself as a binary vector (as in a two-mode matrix of objects by attributes). Results on this important problem are found largely in the biological literature (Barnett & Gower 1971; Gower 1973, 1974; Gower & Payne 1975; Payne & Preece 1980; Sackin & Sneath 1988).

Measures of Association or Dissimilarity Coefficients

Many methods of clustering (especially hierarchical ones) require two-way *one*-mode data, in the form of matrices variously gauging direct judgments of perceived similarity, brand switching among products, confusions, correlations, etc. But the data as they occur are often two-way *two*-mode (e.g. objects by attributes). As a step in preprocessing such data *prior* to performing a cluster analysis, the conversion from two- to one-mode data is such a common problem that it usually receives its own chapter in texts on clustering, and it is also relevant to MDS (Shepard 1972; Coxon 1982: Ch. 2) and related techniques of data analysis (see Gower 1985). Although the topic has traditionally seen greatest emphasis in biology (see references in Rao 1982), psychologists concerned with such substantive issues as content analysis (Krippendorf 1980, 1987), interrater agreement (Hubert 1977b; Hubert & Golledge 1983; Popping 1988), information retrieval (W. P. Jones & Furnas 1987), and choice processes (Doignon et al 1986) have recently begun contributing to the literature.

Such considerations as the scale type of the data—whether they are binary, more generally discrete, or continuous—have long been paid considerable attention, especially in biology. Gower (1971) and Lerman (1987), for example, devised general coefficients of similarity allowing for data where the "variables" (as in a two-mode matrix of objects by variables) were "mixed" among those different classes.

The case of binary data is of special interest. If we consider two binary-valued vectors **x** and **y**, then the element-by-element matches are of the four types labeled a, b, c, and d in Figure 1. For example, if **x**= (1, 1, 0, 0) and **y**= (1, 0, 1, 0), then the first entries (1, 1) in each vector are an a-type pair, the second (1, 0) are a b-type pair, etc. An endless number of coefficients of agreement can be written as a function of those four types; for example, Pearson product-moment correlation is given by $(ad - bc)/[(a + b)(a + c)(b + d)(c + d)]^{.5}$. Cheetham & Hazel (1969) were among the first to catalog the various coefficients published and based on the format of Figure 1, and their list had fewer than 25 entries, whereas Hubálek's (1982) had 43. The framework of Figure 1 is also useful for comparing pairs of partitions, as considered below in the section on Assessing and Comparing Structures, where a state of "1" corresponds to a pair of objects appearing in the same equivalence class or cluster in a partition, and "0" otherwise. The coefficients then can be used to gauge relatedness of pairs of partitions.

Given the plethora of such coefficients, several strategies of research have evolved to answer data analysts' questions concerning which coefficient(s) to use. For example, in research somewhat more relevant to MDS and related spatial models than to clustering, Gower (1971, 1986a,b), Critchley (1986b), Fichet (1986), Gower & Legendre (1986), and Zegers (1986) have studied those coefficients leading to one-mode matrices allowing the corresponding stimuli to be embedded perfectly in Euclidean spaces (see Heiser 1986 and Gower 1986b for a current summary of the main issues) by the fitting of various spatial models. W. P. Jones & Furnas (1987) have taken another type

		y	
		1	0
x	1	a	b
	0	c	d

Figure 1 The four possible patterns resulting from matching elements of two binary-valued vectors **x** and **y**.

of geometric approach, leading to sophisticated visual portrayals of differential performance of such coefficients in the practice of information retrieval.

Another line of research has sought to fortify these descriptive statistics to sustain inferential statistical tests. This formidable task faces the immediate obstacle that many of the measures, as initially proposed, are not even bounded by the familiar limits of $[-1, +1]$ or $[0, 1]$. Determining the maximum value of such coefficients for a given set of data is often a computationally difficult problem of combinatorial optimization (cf Hubert & Arabie 1985:199). Impressive advances on this general problem have been made by Lerman (1983a,b, 1987), Giakoumakis & Monjardet (1987a,b), and Lerman & Peter (1988). For coefficients most commonly used in empirical studies, some important distributional results have recently been reported (Heltshe 1988, Snijders et al 1990).

It is not surprising that when confronted with such an abundance of coefficients, various investigators have taken *a priori* approaches, including emphasis upon patterns of sensitivity to certain aspects of data (Faith 1984), admissibility conditions (Hubálek 1982, Vegelius & Janson 1982, Gower & Legendre 1986, Zegers 1986), and formal axioms (Baulieu 1989). As an exemplar of the last approach, Baroni-Urbani & Buser (1976) began with a set of substantively motivated axioms and then derived a new measure satisfying them; the authors also simulated their coefficient's numerical behavior.

While some practitioners would no doubt agree with Proverbs 20:23 that "Divers weights [i.e. measures] are an abomination unto the Lord," we cannot improve upon the advice of Weisberg (1974:1652–1653): ". . . I would contend that analysts frequently should not seek a single measure and will never find a perfect measure. Different measures exist because there are different concepts to measure. . . . It is time to stop acting embarrassed about the supposed surplus of measures and instead make the fullest possible use of their diversity."

Mixture Models

No overview of recent developments in clustering would be complete without consideration of mixture models, in which an underlying continuum is assumed to give rise to distinct but continuous clusters or subpopulations. Sampling from such a space gives rise to a "mixture" from the clusters and to the problem of estimating parameters characterizing those clusters. Because this aspect of clustering relies heavily on parameter estimation (particularly by maximum likelihood), it is somewhat outside the declared scope of our chapter and will therefore receive only cursory treatment. Following in the tradition of J. H. Wolfe (1970), recent advances have appeared at an increasing rate (Hartigan 1975: Ch. 5; Marriott 1982; McLachlan 1982; Meehl & Golden 1982; Basford & McLachlan 1985; Windham 1985, 1987; Bryant 1988; Ganesalingam 1989; Govaert 1989), culminating in McLachlan & Basford's (1988) laudable *Mixture Models: Inference and Applications to Clustering* (see reviews by Windham 1988 and Morin 1990).

Overlapping Clustering

In the "modern" era of clustering, the first formalized approach to overlapping, instead of hierarchical or partitioning, clustering (N. Jardine & Sibson 1968) occasioned extensive rehearsals (in the form of algorithmic developments by Cole & Wishart 1970; Rohlf 1974, 1975), but to date only one performance (published analysis of empirical data with interpretation: Morgan 1973).

More recent times have been conducive to the developments of models, associated algorithms, and substantive applications. The ADCLUS model (Shepard & Arabie 1979), devised for fitting to a single (two-way one-mode) proximity matrix, has seen extensions to the three-way ("individual differences" or INDCLUS; Carroll & Arabie 1983) case, as well as other developments (DeSarbo 1982; Corter & Tversky 1986; Bandelt & Dress 1989), including linkage to latent class analysis (Grover & Srinivasan 1987). An important theoretical derivation of the relation between common and distinctive feature models, as represented by fitting the ADCLUS model, is given by Sattath & Tversky (1987). Algorithmic and software developments include those of Arabie & Carroll (1980a,b), DeSarbo (1982), Carroll & Arabie (1982, 1983), Hojo (1983), Mirkin (1987, 1989a,b, 1990), Imaizumi & Okada (1990).

For a list of published applications of overlapping clustering, see Arabie et al (1987a:57, 63; 1987b), as well as Faith (1985), Mirkin (1986, 1987, 1989b, 1990), Arabie & Maschmeyer (1988), Sabucedo et al (1990), and Walker (1989).

Partitioning

As an alternative both to hierarchical and to overlapping clustering, partitioning approaches assign each object to exactly one cluster. Described generically, the objective is to maximize similarity/cohesiveness/homogeneity within each cluster while maximizing heterogeneity among clusters. While the importance of partitioning approaches to data analysis has long been recognized in the behavioral and biological sciences (MacQueen 1967; Lee 1980; Lee & MacQueen 1980), it has recently enjoyed great emphasis in economics as well, in such applications as facility location (Klastorin 1982) and especially in the layout of computer circuitry (Barnes 1982). Because the number of distinct partitions of n objects into m clusters increases approximately as $m^n/m!$ (the printers' demons have conspired so successfully against this expression that its denominator is incomplete in Duran & Odell 1974:41 and missing altogether in Hartigan 1975:130), attempting to find a globally optimum solution (regardless of the measure of goodness-of-fit employed) is usually not computationally feasible.

Thus, a wide variety of heuristic approaches (capably reviewed both by Belbin 1987 and by Jain & Dubes 1988:89–117) have been developed to find local optima. Hartigan (1975:102) summarized differences among approaches as stemming from "(i) the starting clusters, (ii) the movement rule [i.e. transferring objects among clusters], and (iii) the updating [of goodness-of-fit] rule." In addition, the measure of goodness-of-fit should be consistent with the scale type of the data (see Hartigan 1975: Ch. 4, 6, 7). Not surprisingly, the scale type assumed often is interval or ratio, and the strongest results and most readily available software are for this case (Gordon & Henderson 1977; Spaeth 1980: Ch. 3; 1985, 1986a). Klein & Dubes (1989) have recently suggested that the simulated annealing approach to optimization (Kirkpatrick et al 1983; see Dubes 1988; Ripley 1990) might be useful for partitioning, in contrast to negative results for somewhat related problems of data analysis (De Soete et al 1988). In an interesting and novel development, Spaeth (1986b,c) has turned the traditional partitioning problem inside out with "anticlustering," which seeks maximal heterogeneity within clusters and minimal heterogeneity between clusters.

Constrained Clustering

The imposition of a priori constraints on a cluster solution generally makes both the description and implementation of a clustering algorithm more complicated but can sometimes confer major benefits. For example, if objects to be partitioned are already sequenced (i.e. placed on a line), so that all clusterings of them must respect this ordering, then the amount of computation is reduced enough to allow finding a global optimum in circumstances where an unconstrained global optimum would be inconceivable (Gordon 1973a; Hartigan 1975: Ch. 6; Spaeth 1980:61–64).

A commoner constraint is contiguity in a plane, with consequent difficulties in designing corresponding algorithms (as reviewed by Murtagh 1985b). The problem occurs frequently enough to have engendered an impressive literature within clustering (Gordon 1980b; 1981:61–69; Matula & Sokal 1980; Ferligoj & Batagelj 1982, 1983; Perruchet 1983a; DeSarbo & Mahajan 1984; Margules et al 1985; Finden & Gordon 1985; Legendre 1987).

Consensus Clustering

Inventors of spatial models have long disgreed over whether and how one should aggregate data and represent group structure or should instead portray individual differences (among subjects or other sources of data) (Tucker & Messick 1963; Ross 1966; Cliff 1968; Carroll & Chang 1970; Arabie et al 1987a; also see chapters in Law et al 1984 and Coppi & Bolasco 1989). But such discrete structures as dendrograms afford a different approach to this problem of data analysis: representation of the group structure as a consensus over the structures fitted to individuals' data (or dendrograms from other sources—such as different clustering procedures applied to the same data). That is, using the topology of the dendrograms for a common set of objects, and based on each source's data (and in general ignoring the ultrametric or other numerical values associated with levels of the dendrograms), highly formalized rules, often embodying classical approaches to voting and social choice (Mirkin 1979: Ch. 2; Day 1988), are used to construct a best-fitting consensus dendrogram. Excellent reviews and bibliographies of selected areas of this research are available (Barthélemy & Monjardet 1981, 1988; Day 1986b; Leclerc & Cucumel 1987; Barthélemy et al 1986; Leclerc 1988, 1989), and a special issue of the *Journal of Classification* (Day 1986a) was devoted to consensus classifications. In addition to empirically oriented developments (Gordon & Finden 1985; Gordon 1986, 1987b; Faith 1988; Leclerc 1988), numerous axiomatic frameworks have been devised for consensus structures (Barthélemy et al 1984; McMorris 1985; Day & McMorris 1985; Day et al 1986; Barthélemy & Janowitz 1990).

Cluster Validity

So long as the input data are of the appropriate number of modes, ways, etc, most methods of clustering will deterministically produce a clustering solution. Moreover, different methods will generally produce different solutions based on the same input data. The question naturally arises whether the clusters have "reality" or validity vis-à-vis the data (cf Hartigan 1975:202–203; Dubes & Jain 1979). Jain & Dubes (1988: Ch. 4) provide a useful summary of strategies for validation: "External criteria measure performance by matching a clustering structure to a priori information. . . . *Internal criteria* assess the fit between the structure and the data, using only the data themselves. . . . *Relative criteria* decide which of two structures is better in some sense, such as being more stable or appropriate for the data" (emphasis in the original, p. 161). Among the issues most commonly investigated are selection of indices of cluster structure and their distributions (Day 1977; Murtagh 1984c; Milligan & Cooper 1986; also see the section on practical advances, below) and determining the appropriate number of clusters (Dubes 1987; Cooper & Milligan 1988; Critchley 1988; Krzanowski & Lai 1988; Peck et al 1989).

Variable Selection and Weighting

Although we noted above in the section on dissimilarity coefficients that conversion of a two-mode to a one-mode matrix prior to clustering should be regarded as a step separate from the actual cluster analysis, some authors have sought to link the original two-mode matrix more directly to the mechanics of the cluster analysis. DeSarbo et al (1984) devised an approach for "synthesized clustering" in which the variables in a two-mode (objects by variables) matrix were iteratively and differentially weighted according to their relative importance to the emergent K-means (MacQueen 1967) cluster structure. This procedure was extended from partitions to ultrametric trees by De Soete et al (1985), who also sketched details for further extensions to additive, multiple, and three-way trees (discussed below), some of which were implemented later by De Soete & Carroll (1988). De Soete provided both an algorithm (1986) and software (1988) for optimal variable weighting in fitting either an ultrametric or an additive tree to a single two-mode matrix. Fowlkes et al (1988) devised a forward selection procedure for variables in a two-mode matrix intended for complete-link hierarchical clustering as well as other methods (1987).

The practical importance of such approaches quickly led to evaluations. Milligan (1989) reported positive results from a validation study of De Soete's (1986, 1988) techniques, whereas Green et al (1990) obtained disappointing results in evaluating algorithms of DeSarbo et al (1984) and De Soete et al (1985).

Computational Advances

Clustering was among the first areas of data analysis to be influenced by computer scientists' preoccupation with computational complexity (Day 1983a provides a nice overview), and efforts to design clustering algorithms for large data sets are ongoing. Relevant aspects of clustering comprise partitioning (Hansen & Jaumard 1987; Hansen et al 1989), complete-link (Defays 1977; Hansen & Delattre 1978), single-link (Rohlf 1973; Sibson

1973; Hansen & Lehert 1980; Hansen et al 1989), and other forms of hierarchical clustering (Morineau & Lebart 1986; Guénoche et al 1991), including by parallel algorithms (Lerman & Peter 1984).

Substantive Developments

We noted earlier that the enormous literature of applications of clustering could not be covered in this chapter, but we do want to mention two substantive areas that have been especially active in their use and advancement of clustering.

SOCIOMETRY AND SOCIAL PSYCHOLOGY Many articles appearing in the journal *Social Networks* use and include discussions of clustering, even though methodological sophistication is sometimes lacking (e.g. Burt 1988), as Faust & Romney (1985) pointed out. Surveys of this area of research (e.g. Burt 1980; Knoke & Kuklinski 1982; Wasserman & Faust 1991) generally include sections on clustering techniques, and the contributors have greatly expanded the range of problems to which it is applied [e.g. to studying complex economic legislation (Boorman & Levitt 1983) or to legal precedents and structures of communication among state supreme courts (Caldeira 1988); also see abstracts collected in the bibliographic survey *Connections* (A. Wolfe 1990)]. Moreover, the area is increasingly quick to adopt recently developed combinatorial and statistical techniques (e.g. Feger & Bien 1982; Feger & Droge 1984; Noma & Smith 1985; Wasserman & Anderson 1987; Dow & de Waal 1989; Hummon & Doreian 1990).

In the area of social personality and autobiography, Rosenberg's innovative analyses (1988, 1989) of data meticulously extracted from the autobiographical novels of Thomas Wolfe coincide with a greater public demand for autobiographies. Such work might even provide a curative for psychobiographies.

EVOLUTIONARY TREES When *Science* initiated its software review section, the first contribution (Fink 1986) compared programs for reconstructing phylogenetic trees, typically on the basis of molecular data. Further evidence of the surge of interest in the role of clustering in reconstructing evolutionary patterns is given by articles in the *Proceedings of the National Academy of Sciences* (e.g. Cavalli-Sforza et al 1988; Harding & Sokal 1988; Sokal 1988b) and attendant controversies (Cavalli-Sforza et al 1989; Bateman 1990). Methodological contributions from combinatorial data analysis to this substantive area include hierarchical clustering (Corpet 1988), additive trees (Dress et al 1986), computational complexity (Day et al 1986; Day & Sankoff 1986; Barthélemy & Luong 1987), graph theory (Mirkin & Rodin 1984; see Hubert 1984), lattice theory (Estabrook & McMorris 1980), sequence comparison (Kruskal 1983; Sankoff & Kruskal 1983), and statistical analysis

(Astolfi et al 1981; Felsenstein 1983b,c; Barry & Hartigan 1987). A general review is given by Sokal (1985), and numerous chapters in Felsenstein (1983a); Dress & von Haeseler (1990) and Luong (1989) provide a range of current topics of investigation. Holman (1985) provides an important psychological and methodological perspective on some of the basic issues of taxonomy.

ADDITIVE TREES AND OTHER NETWORK MODELS

In graph theory, a tree is a connected graph without cycles. As noted earlier when considering the MST problem, for representing psychological structure, the nodes of the graph correspond to stimuli and the links connecting them have weights whose numerical values are used to reconstruct or predict the input data so that goodness-of-fit can be gauged. In the subsections above concerned with hierarchical clustering, the metric used for predicting the data was usually based on the ultrametric. A different metric, based on a relaxation of the ultrametric inequality and often called the "four-points condition" is a popular alternative and gives representations variously known as free trees, path length trees, or additive trees. In general, we do not repeat the review of the topic given in Carroll & Arabie (1980:623–24) except to note general overviews by Carroll (1976) and Shepard (1980).

All subsections below, until the section on representations of two- and higher-mode data, assume a single input (one-mode two-way) proximities matrix.

Algorithms and Models

Considerable work on algorithms for fitting additive trees has been done recently (Abdi et al 1984; Brossier 1985; Guénoche 1986a; Barthélemy & Guénoche 1988); Guénoche (1987) has compared five algorithms. Some of the strategies of specialization used successfully for hierarchical clustering have also proved useful for additive trees. Specifically, there are versions for binary data (Guénoche 1986b) and for constrained representations (De Soete et al 1987).

Recent advances in devising and fitting more general graph-theoretic models are impressive (Orth 1988, 1989; Hutchinson 1989; Klauer 1989; Klauer & Carroll 1989, 1991). Some of these papers have also provided remarkable substantive results as well (e.g. Hutchinson 1989) while others (e.g. Cooke & McDonald 1987) have not.

Substantive Advances

Friendly (1977, 1979) pointed out the advantages of modeling structure of organization in free recall around the combinatorial framework of the MST (see Hubert 1974a). Combinatorial models have since been devised (Hubert &

Levin 1976, 1977, 1978; Levin & Hubert 1980; Pellegrino & Hubert 1982) to allow testing for a wide range of substantive structural predictions. Results from Hirtle and his colleagues (McKeithen et al 1981; Hirtle 1982; 1987; Hirtle & Crawley 1989) have demonstrated that a tree with seriated nodes can be reconstructed using replicated orderings of a set of objects, as in multi-trial free recall paradigms; Shiina (1986) has attempted the same feat for obtaining MDS solutions.

Representations Based on Two- and Higher-Mode Data

Although for many years ultrametric representations were limited to one mode, Furnas (1980) elegantly generalized the ultrametric inequality to two-mode data, and De Soete et al (1984a,b) provided least squares algorithms for fitting either ultrametric or additive trees to two-way two-mode data (also see contributions by Brossier 1986, 1990). De Soete et al (1986) have also devised an algorithm for fitting ultrametric or additive trees to two-mode data and simultaneously estimating optimal weights for the variables as well during the conversion to one-mode data. Finally, in a development that has seen rapid progress, two- or higher-mode preference data are now suitable for fitting stochastic tree unfolding models (Carroll et al 1988, 1989; Carroll & De Soete 1990).

ASSESSING AND COMPARING STRUCTURES

In our introductory characterization of what CDA might legitimately encompass, we mentioned the confirmatory comparison of two (or more) structures definable on some common set of objects. Usually, structures (e.g. input matrices, sequences, partitions, graphs, trees) to be compared are first represented in the form of matrices whose entries numerically gauge some relationship among the common objects; in the simplest case of two structures, one matrix is typically empirical and the second either posited theoretically or also generated empirically. The actual comparison strategy invariably relies on some correlational measure between the entries from the two given matrices (or their suitable transformations); a substantial literature illustrates the procedures using various types of descriptive measures. Depending on the objects and type of matrices involved, this work may be (*a*) axiomatic in attempting to characterize "good" measures in a particular context (Barthélemy 1979; Leclerc 1985a,b; Barthélemy et al 1986), (*b*) specific to certain types of structural representations (Day 1983b; Gower 1983; Leclerc 1982; Rohlf 1974, 1982; Gordon, 1980a, 1981:132–37), and (*c*) perhaps even dependent on solving certain initial (and possibly difficult) optimization tasks (Gordon 1973b, 1982, 1988; Delcoigne & Hansen 1975; Klastorin 1980; Lerman 1988; W. Miller & Myers 1988; ten Berge 1988; Gordon et al 1989).

The most active area of work involving the comparison of structures (through matrices) can be seen as extending a seminal paper by Mantel (1967), which suggests a particular randomization method that allows a correlational measure of association between matrix entries to be assessed for relative size, and does so through a significance test that maintains the integrity of the structures being compared. The actual evaluation is based on the conjecture of no relationship between matrices, and is operationalized by the explicit hypothesis of randomness in the pairing of the objects between the two structures.

The range of applications for this matrix comparison method and associated significance testing strategy is enormous; many of the possibilities, at least as of 1986, are documented by Hubert (1987). The encompassed topics include, among others, almost all methods encountered in classical nonparametric statistics (Hubert 1987), the assessment of spatial autocorrelation for variables observed over a set of geographical locations (Upton & Fingleton 1985; Sokal 1986b; Sokal et al 1987), multivariate analysis of variance (Mielke 1978, 1979; Mielke et al 1976), assessment techniques concerned with various conjectures of combinatorial structure that might be posited for an empirically determined measure of proximity (Dow & de Waal 1989), and the comparison of two empirically generated matrices that might contain rather general measures of proximity (Dow & Cheverud 1985; Cheverud et al 1989) or matrices with very restricted entries (e.g. binary) defining various combinatorial structures (Verhelst et al 1985; Lerman 1987, 1988; Lerman & Peter 1988).

The same general strategy for comparing two matrices has recently been extended to the comparison of sets of matrices through the use of optimally weighted composites. The case of particular interest in the literature thus far compares a single matrix to a set of matrices through the use of a multiple correlation coefficient between the corresponding matrix entries (Smouse et al 1986; Hubert & Arabie 1989)

NONDESTRUCTIVE DATA ANALYSIS

Murtagh (1989) contributed the engaging rubric of "nondestructive data analysis" to a particular class of matrix permutation strategies; we use it here to refer in general to matrix permutation approaches to data analysis. Such methods simply seek to find a permutation or reordering of the rows and columns of matrices so as to reveal interpretable patterns not otherwise apparent; historically these methods are linked to seriation (Katz 1947; see Hubert & Baker 1978; Hubert & Golledge 1981; and Hubert et al 1982). Perhaps because they have been orphaned in most overviews of data analysis, some of these techniques keep being reinvented (e.g. Beum & Brundage's 1950 approach by Deutsch & Martin 1971 and by Lingoes 1968).

Hartigan (1972, 1975: Ch. 14–15; 1976) has shown the advantages of "direct" clustering approaches that address two-mode data directly, without first converting them to a one-mode matrix (see also De Soete et al 1984a,b). Among the other strategies of matrix permutation and/or partitioning that have seen the most activity in recent years is the "bond energy" approach of McCormick et al (1972; also see Lenstra 1974). Reviews are given by Murtagh (1985a: Ch. 1) and Arabie & Hubert (1990). For a sampling of recent work on the problem, see Kusiak et al (1986), Marcotorchino (1986, 1987), Kusiak & Finke (1987), Hilger et al (1989), and Arabie et al (1990).

RELATIONSHIPS BETWEEN DISCRETE VERSUS SPATIAL STRUCTURES

C. J. Jardine et al (1967) provided a continuous transformation relating the triangle and ultrametric inequalities, somewhat in support of the common intuition that the Euclidean metric and ultrametric-based hierarchical clustering were highly compatible vehicles for representing structure in data. Holman's classic (1972; also see the appendix of Gower & Banfield 1975) result shattered this complacency by showing that data conforming perfectly to one metric were somewhat antithetical to the other. But because empirical data rarely ever fit either model without error, the folklore of compatibility between relevant discrete and spatial models is still empirically useful (see Kruskal 1977 for an excellent discussion). Critchley (1986a) aptly decried and undermined the "widespread myths surrounding the work [i.e. result] of Holman (1972)," which are still promulgated by some cognitive psychologists (e.g. McNamara 1990). Such misunderstandings can hardly be blamed on Holman, who stated his results concisely and elegantly.

Strategies of comparing the two classes of representations have included geometric analyses (Tversky & Hutchinson 1986) and computationally based comparative data analyses (Pruzansky et al 1982) which suggested that data from perceptual domains were more compatible with Euclidean spatial representations whereas data from conceptual domains were better suited to discrete representations. Furnas (1989) has provided an innovative graphical approach showing interrelations among families of relevant metrics. Critchley & Heiser (1988) showed that data perfectly conforming to hierarchical trees can also be represented without error unidimensionally, while Brossier (1984) and Diday (1986) have sought to generalize and exploit relationships between these different types of representations (Arabie 1986). Hybrid approaches seeking simultaneously to combine the advantages of MDS and clustering continue to be appealing (Carroll & Pruzansky 1980; Bock 1986; Mirkin 1989b).

REPRESENTATIONS OF THREE- AND HIGHER-WAY DATA

Carroll & Arabie (1980:638) noted that "we see a strong trend toward the development of three-way models with applications of three- and higher-way methods becoming almost as numerous as two-way applications." As reviewed by Arabie & Daws (1988), various substantive developments have helped assure the outcome of this prediction (Snyder et al 1984), and we noted earlier that some recent edited volumes (Law et al 1984; Coppi & Bolasco 1989) are exclusively concerned with representing higher-way data. In addition to the papers cited in the section above on additive trees, examples of such generalizations include hybrid models for three-way data (Carroll & Pruzansky 1983) and ultrametric representations for three-way two-mode (Carroll et al 1984) as well as three-mode data (De Soete & Carroll 1989).

RANKINGS, RELATIONS, AND PARTIALLY ORDERED SETS

As noted above, this congeries has demonstrated the signs of a mature subdiscipline, including its own specialized journal (*Order*, established in 1984), joining numerous others of relevance and amassing a burgeoning literature. We can give these topics only cursory consideration here—a constraint regrettable because too many psychologists are unaware of the enormous strides (some of them of eminently practical use) that have recently taken place in this area of research. For example, Cook and his collaborators (Armstrong et al 1982; Cook & Kress 1984; Cook et al 1986) have provided useful results for obtaining a consensus ordering from a set of ordinal rankings of n entities from a committee of m members. Other applications-oriented developments include those reported by Critchlow (1985) and Fligner & Verducci (1986).

Monjardet and his collaborators (Monjardet 1973, 1979a,b, 1981b; Barthélemy & Monjardet 1981; Barthélemy et al 1982, 1986) have capably surveyed the general area of partially ordered sets, ordered sets, and complete orderings, and have shown their relevance to combinatorial data analysis and the social sciences. Other especially noteworthy contributions include those of Marcotorchino & Michaud (1979), Michaud & Marcotorchino (1979, 1980), Schader (1979), Doignon & Falmagne (1984), Leclerc (1984, 1987, 1989), Gaul & Schader (1988), Saporta (1988), and Fishburn & Falmagne (1989).

PRACTICAL ADVANCES

A considerable lag will undoubtedly precede much of CDA's impact on workaday data analysis. We now wish to consider instead some results that should have more immediate impact. For example, we noted above that converting from a two- to a one-mode matrix is a common prerequisite for many cluster analyses. If the data are a matrix of objects by variables and the analyst wishes to compute Euclidean distances between all pairs of objects, a common problem is whether and how to standardize the variables, prior to using their entries as coordinates for computing inter-object distances. Milligan & Cooper (1988) have provided a result (the superiority of dividing by the range of a variable) that probably merits "written in stone" status.

Another practical problem on which Milligan and his collaborators have made progress is the question of which measure of relatedness between partitions is best for cluster validation. (As noted earlier, the framework of Figure 1 facilitates proliferation of such measures between partitions, just as it does for measures of association for paired variables.) Concluding a comparative study of the coefficients regarded for either theoretical or empirical reasons to be forerunners, Milligan & Cooper (1986:457) stated: "... it would appear that of the five indices the Hubert and Arabie [1985] adjusted Rand measure seems to be the index of choice for clustering validation research."

Most statistical consultants have at some time been badgered by users of clustering who feel unfulfilled and forlorn without some test of significance. This lacuna in general reflects no lack of interest in devising such tests (Perruchet 1983b; Bock 1985; Hartigan 1977, 1978, 1985) but rather the adamantine nature of the problems. It should be noted, however, that inferential procedures are available for testing significance for bimodality (Giacomelli et al 1971) as well as for multimodality (Hartigan & Hartigan 1985; Hartigan 1988; Hartigan & Mohanty 1992).

Another common problem arises as users of techniques try to compare the output from two or more analyses when in fact the substantive theory suggests that a correlation between the input proximity (or other types of) matrices is instead called for (cf Carroll & Arabie 1980:636). Although the inferential problem was solved over a decade ago (Hubert 1978, 1979), no general-purpose software is available for carrying out such analyses. Mehta (1990; personal communication) has informed us of the possibility of including such a capability in StatXact (Mehta 1990). StatXact runs as a stand-alone package but can also be invoked from SYSTAT (Wilkinson 1989). This development should undercut all excuses for doing the wrong analysis.

BIBLIOGRAPHIC CONSIDERATIONS

Authors of *Annual Review* chapters should be allowed the indulgence of observing what makes their work—and presumably that of others in the field—easy or difficult: We have already noted that much of the literature on CDA is found in conference proceedings and other edited volumes. These single volumes often cost three or more times the price of an annual subscription to the most relevant journals, and the chapters are generally not covered by indexing services like the Institute for Scientific Information or their publications. Reviewers' unhappiness has apparently become ritual: "In future publications, I hope editors can be prevailed upon to provide an index and the publisher can match price with quality of production and do greater justice to the contributors' work" (Coxon 1988:298), or "Finally, two 'classical' critical comments on such publications: unfortunately, this volume does not include a subject index and it is very expensive (US $136.75)" (Ferligoj 1990:158). Even when an index is included (Bock 1988), it is found inadequate (Okada 1989). Publishers' increasing reluctance to produce such volumes coincides with librarians' (not to mention private individuals') displeasure over the prices; the problem may be self-terminating.

PROSPECTS

Despite the disciplinary diversity of contributions (both negative and positive), it is clear that the field is coalescing around certain themes: (*a*) types of data and their implications for possible representations; (*b*) the relationships among algebraic, geometric, and logical structures; and (*c*) those relationships' implications for representations of structure in data. Such developments, however, are not buttressed by the software found in statistical packages, and the result is a widening gap between elegant developments in algorithms and models versus access to them by potential users.

ACKNOWLEDGMENTS

This research was supported in part by AT&T Information Systems' Industrial Affiliates Program at the University of Illinois. Of the many colleagues who helpfully pointed out references to us, we are especially indebted to J.-P. Barthélemy, J. Douglas Carroll, Geert De Soete, A. D. Gordon, John C. Gower, Steve Hirtle, Bruno Leclerc, I. C. Lerman, Jacqueline Meulman, Bernard Monjardet, Fionn Murtagh, and Akinori Okada. We are grateful to Francesca Lundström for editorial expertise. For bibliographic assistance, we are indebted to François Bronsard, Tom Merritt, and Nicole Sullivan.

Literature Cited

Abdi, H., Barthélemy, J.-P., Luong, X. 1984. Tree representations of associative structures in semantic and episodic memory research. In *Trends in Mathematical Psychology*, ed. E. Degreef, J. van Buggenhaut, pp. 3-31. Amsterdam: North-Holland

Ambros-Ingerson, J., Granger, R., Lynch, G. 1990. Simulation of paleocortex performs hierarchical clustering. *Science* 247:1344–48

Arabie, P. 1986. Comments on confounded experimental designs, pyramids, and seriation. In *Multidimensional Data Analysis*, ed. J. de Leeuw, W. Heiser, J. Meulman, F. Critchley, pp. 235–37. Leiden, The Netherlands: DSWO-Press

Arabie, P., Carroll, J. D. 1980a. MAPCLUS: a mathematical programming approach to fitting the ADCLUS model. *Psychometrika* 45:211–35

Arabie, P., Carroll, J. D. 1980b. *How To Use MAPCLUS, a Computer Program for Fit-*

ting the ADCLUS Model. Murray Hill, NJ: AT&T Bell Laboratories

Arabie, P., Carroll, J. D., DeSarbo, W. S. 1987a. *Three-way Scaling and Clustering*. Newbury Park, CA: Sage. Translated into Japanese by A. Okada and T. Imaizumi, 1990. Tokyo: Kyoritsu Shuppan

Arabie, P., Daws, J. 1988. The interface among data analysis, marketing, and knowledge representation. In *Data, Expert Knowledge and Decisions*, ed. W. Gaul, M. Schader, pp. 10-15. Berlin: Springer-Verlag

Arabie, P., Hubert, L. J. 1990. The bond energy algorithm revisited. *IEEE Trans. Syst. Man Cybern*. 20:268-74

Arabie, P., Hubert, L., Hesson-McInnis, M. 1989. Combinatorial optimization approaches to multidimensional scaling in city-block spaces. Paper presented at the 2nd Conf. Int. Fed. Classif. Soc., Univ. Virginia, Charlottesville

Arabie, P., Hubert, L., Schleutermann, S. 1990. Blockmodels from the bond energy algorithm. *Soc. Networks* 12:99-126

Arabie, P., Maschmeyer, C. J. 1988. Some current models for the perception and judgment of risk. *Organ. Behav. Hum. Decis. Process*. 41:300-29

Arabie, P., Maschmeyer, C. J., Carroll, J. D. 1987b. Impact scaling: method and application. *Technol. Forecast. Soc. Change* 32:245-72

Armstrong, R. D., Cook, W. D., Seiford, L. M. 1982. Priority ranking and consensus formation: the case of ties. *Manage. Sci.* 28:638-45

Astolfi, P., Kidd, K. D., Cavalli-Sforza, L. L. 1981. A comparison of methods for reconstructing evolutionary trees. *Syst. Zool*. 30:156-69

Bandelt, H.-J., Dress, A. 1989. Weak hierarchies associated with similarity measures—an additive clustering technique. *Bull. Math. Biol*. 51:133-66

Barbut, M., Monjardet, B. 1970a. *Ordre et Classification, Algèbre et Combinatoire*, Tome I. Paris: Hachette

Barbut, M., Monjardet, B. 1970b. *Ordre et Classification, Algèbre et Combinatoire*, Tome II. Paris: Hachette

Barnes, E. R. 1982. An algorithm for partitioning the nodes of a graph. *SIAM J. Algorithms Disc. Methods* 3:541-50

Barnett, J. A., Gower, J. C. 1971. Selecting tests in diagnostic keys with unknown responses. *Nature* 232:491-93

Baroni-Urbani, C., Buser, M. W. 1976. Similarity of binary data. *Syst. Zool*. 25:251-59

Barry, D., Hartigan, J. A. 1987. Statistical analysis of hominoid molecular evolution. *Stat. Sci*. 2:191-210

Barthélemy, J.-P. 1979. Caractérisations axiomatiques de la distance de la différence symétrique entre des relations binaires. *Math. Sci. hum*. 17(67): 85-113

Barthélemy, J.-P. 1989. Introduction: les arbres, de la matière pour bricoler. See Luong 1989, pp. 3-21

Barthélemy, J.-P., Flament, C., Monjardet, B. 1982. Ordered sets and social sciences. In *Ordered Sets*, ed. I. Rival pp. 721-58). Dordrecht: Reidel

Barthélemy, J.-P., Guénoche, A. 1988. *Les arbres et les représentations des proximités*. Paris: Masson. To be translated into English as *Tree Representation of Proximity Data*, and published by Wiley.

Barthélemy, J.-P., Janowitz, M. F. 1990. A formal theory of consensus. *SIAM J. Disc. Math*. In press

Barthélemy, J.-P., Leclerc, B., Monjardet, B. 1984. Ensembles ordonnés et taxonomie mathématique. *Ann. Disc. Math*. 23:523-48

Barthélemy, J.-P., Leclerc, B., Monjardet, B. 1986. On the use of ordered sets in problems of comparison and consensus of classifications. *J. Classif*. 3:187-224

Barthélemy, J.-P., Luong, X. 1987. Sur la topologie d'un arbre phylogénétique: aspects théoriques, algorithmes et applications à l'analyse de données textuelles. *Math. Sci. hum*. 25(100): 57-80

Barthélemy, J.-P., Monjardet, B. 1981. The median procedure in cluster analysis and social choice theory. *Math. Soc. Sci*. 1:235-67

Barthélemy, J.-P., Monjardet, B. 1988. The median procedure in data analysis: new results and open problems. See Bock 1988, pp. 309-16

Basford, K. E., McLachlan, G. J. 1985. The mixture method of clustering applied to three-way data. *J. Classif*. 2:109-25

Batagelj, V. 1981. Note on ultrametric hierarchical clustering algorithms. *Psychometrika* 46:351-52

Bateman, R. M. 1990. Balancing American linguists [Letter to editor]. *Science* 249:1228

Baulieu, F. B. 1989. A classification of presence/absence based dissimilarity coefficients. *J. Classif*. 6:233-46

Beane, T. P., Ennis, D. M. 1987. Market segmentation: a review. *Eur. J. Mark*. 21:20-42

Belbin, L. 1987. The use of non-hierarchical allocation methods for clustering large sets of data. *Aust. Comput. J*. 19:32-41

Beum, C. O., Brundage, E. G. 1950. A method for analyzing the sociomatrix. *Sociometry* 13:141-45

Blashfield, R. K., Aldenderfer, M. S. 1988. The methods and problems of cluster analysis. In *Handbook of Multivariate Experimental Psychology*, ed. J. R. Nesselroade, R. B. Cattell, pp. 447–73). New York: Plenum. 2nd ed.

Bock, H. H. 1974. *Automatische Klassifikation. Theoretische und praktische Methoden zur Gruppierung und Strukturierung von Daten (Clusteranalyse)*. Goettingen: Vandenhoeck Ruprecht

Bock, H. H. 1985. On some significance tests in cluster analysis. *J. Classif.* 2:77–108

Bock, H. H. 1986. Multidimensional scaling in the framework of cluster analysis. In *Die Klassifikation und Ihr Umfeld*, ed. P. O. Degens, H.-J. Hermes, O. Opitz, pp. 247–58. Frankfurt: Indeks Verlag

Bock, H. H., ed. 1988. *Classification and Related Methods of Data Analysis*. Amsterdam: North-Holland

Boorman, S. A., Levitt, P. R. 1983. Blockmodeling complex statutes: mapping techniques based on combinatorial optimization for analyzing economic legislation and its stress point over time. *Econ. Lett.* 13:1–9

Brennan, T. 1987. Classification: an overview of selected methodological issues. In *Prediction and Classification: Criminal Justice Decision Making*, ed. D. M. Gottfredson, M. Tonry, pp. 201–48. Chicago: Univ. Chicago Press

Brossier, G. 1984. Algorithmes d'ordonnancement des hiérarchies binaires et propriétés. *Rev. Stat. Appl. 23(3): 65–79*

Brossier, G. 1985. Approximation des dissimilarites par des arbres additifs. *Math. Sci. hum.* 23(91): 5–21

Brossier, G. 1986. Etude des matrices de proximité rectangulaires en vue de la classification. *Rev. Stat. Appl.* 25 (4):43–68

Brossier, G. 1990. Piecewise hierarchical clustering. *J. Classif.* 7:197–216

Bryant, P. 1988. On characterizing optimization-based clustering methods. *J. Classif.* 5:81–84

Burt, R. S. 1980. Models of network structure. *Annu. Rev. Sociol.* 6:79–141

Burt, R. S. 1988. Some properties of structural equivalence measures derived from sociometric choice data. *Soc. Networks.* 10:1–28

Buser, M. W., Baroni-Urbani, L. 1982. A direct nondimensional clustering method for binary data. *Biometrics* 38:351–60

Caldeira, G. A. 1988. Legal precedent: structures of communication between state supreme courts. *Soc. Networks* 10:29–55

Carroll, J. D. 1976. Spatial, non-spatial and hybrid models for scaling. *Psychometrika* 41:439–63

Carroll, J. D., Arabie, P. 1980. Multidimensional scaling. *Ann. Rev. Psychol.* 31:607–49

Carroll J. D., Arabie, P. 1982. *How to Use INDCLUS, a Computer Program for Fitting the Individual Differences Generalization of the ADCLUS Model*. Murray Hill, NJ: AT&T Bell Laboratories

Carroll, J. D., Arabie, P. 1983. INDCLUS: an individual differences generalization of the ADCLUS model and the MAPCLUS algorithm. *Psychometrika* 48:157–69

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

Carroll, J. D., Clark, L. A., DeSarbo, W. S. 1984. The representation of three-way proximities data by single and multiple tree structure models. *J. Classif.* 1:25–74

Carroll, J. D., De Soete, G. 1990. Fitting a quasi-Poisson case of the GSTUN (General Stochastic Tree Unfolding) model and some extensions. In *Knowledge, Data and Computer-Assisted Decisions*, ed. M. Schader, W. Gaul. Berlin: Springer-Verlag

Carroll, J. D., DeSarbo, W., De Soete, G. 1988. Stochastic tree unfolding (STUN) models: theory and application. See Bock 1988, pp. 421–30

Carroll, J. D., DeSarbo, W., De Soete, G. 1989. Two classes of stochastic tree unfolding models. In *New Developments in Psychological Choice Modeling*, ed. G. De Soete, H. Feger, C. Klauer, pp. 161–76. Amsterdam: North-Holland

Carroll, J. D., Pruzansky, S. 1980. Discrete and hybrid scaling models. In *Similarity and Choice*, ed. E. D. Lantermann, H. Feger, pp. 108–39. Bern: Hans Huber

Carroll, J. D., Pruzansky, S. 1983. Representing proximities data by discrete, continuous or "hybrid" models. See Felsenstein 1983a, pp. 229–48

Cavalli-Sforza, L. L., Piazza, A., Menozzi, P., Mountain, J. 1988. Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *Proc. Natl. Acad. Sci. USA* 85:6002–6

Cavalli-Sforza, L. L., Piazza, A., Menozzi, P., Mountain, J. 1989. Genetic and linguistic evolution [Letter to the editor]. *Science* 244:1128–29

Cheetham, H. L., Hazel, J. E. 1969. Binary (presence-absence) similarity coefficients. *J. Paleontol.* 43:1130–36

Cheverud, J. M., Wagner, G. P., Dow, M. M. 1989. Methods for the comparative analysis of variation patterns. *Syst. Zool.* 38:201–13

Cliff, N. 1968. The "idealized individual" interpretation of individual differences in multidimensional scaling. *Psychometrika* 33: 225–32

Cliff, N., McCormick, D. J., Zatkin, J. L., Cudek, R. A., Collins, L. M. 1986. BINCLUS: nonhierarchical clustering of binary data. *Multivar. Behav. Res.* 21:201–27

Cole, A. J., Wishart, D. 1970. An improved algorithm for the Jardine-Sibson method of generating overlapping clusters. *Computer J.* 13:156–63

Cook, W. D., Kress, M. 1984. Relationships between l^1 metrics on linear ranking spaces. *SIAM J. Appl. Math.* 44:209–20

Cook, W. D., Kress, M., Seiford, L. M. 1986. Information and preference in partial orders: a bimatrix representation. *Psychometrika* 51:197–207

Cooke, N. M., Durso, F. T., Schvaneveldt, R. W. 1986. Recall and measures of memory organization. *J. Exp. Psychol.: Learn. Mem. Cognit.* 12:538–49

Cooke, N. M., McDonald, J. E. 1987. The application of psychological scaling techniques to knowledge elicitation for knowledge-based systems. *Int. J. Man-Machine Stud.* 26:533–50

Cooper, M. C., Milligan, G. W. 1988. The effect of measurement error on determining the number of clusters in cluster analysis. See Arabie and Daws 1988, pp. 319–28

Coppi, R., Bolasco, S., eds. 1989. *Multiway Data Analysis*. New York: North-Holland

Corpet, F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* 16:10881–890

Corter, J., Tversky, A. 1986. Extended similarity trees. *Psychometrika* 51:429–51

Cox, D. R. 1970. *The Analysis of Binary Data.* London: Chapman & Hall

Coxon, A. P. M. 1982. *The User's Guide to Multidimensional Scaling*. Portsmouth, NH: Heinemann

Coxon, A. P. M. 1988. Review of *Classification as a Tool of Research,* edited by W. Gaul, M. Schader. *J. Classif.* 5:297–98

Critchley, F. 1986a. Dimensionality theorems in multidimensional scaling and hierarchical cluster analysis. See Diday et al 1986, pp. 45–70

Critchley, F. 1986b. Some observations on distance matrices. See Arabie 1986, pp. 53–60

Critchley, F. 1988. The Euclidean structure of a dendrogram, the variance of a node and the question: "How many clusters really are there?" See Bock 1988, pp. 75–84

Critchley, F., Heiser, W. 1988. Hierarchical trees can be perfectly scaled in one dimension. *J. Classif.* 5:5–20

Critchley, F., Van Cutsem, B. 1989. Predissimilarities, prefilters and ultrametrics on an arbitrary set. Res. Rep. No. 171. Univ. Warwick, Dept. Stat.

Critchlow, D. E. 1985. *Metric Methods for Analyzing Partially Ranked Data.* New York: Springer-Verlag

Czekanowski, J. 1909. Zur Differential diagnose der Neandertalgruppe. Korrspondenz-Blatt der Deutschen Gesellschaft für Anthroplogie. *Ethnol. Urgeschichte* 40:44–47

Dale, M. B. 1985. On the comparison of conceptual clustering and numerical taxonomy. *IEEE Trans. Pattern Anal. Machine Intell.* PAMI-7:241–44

Day, W. H. E. 1977. Validity of clusters formed by graph-theoretic cluster methods. *Math. Biosci.* 36:299–317

Day, W. H. E. 1983a. The relevance of computational complexity to classification research. Presented at Meet. Classif. Soc. N. Am., Philadelphia

Day, W. H. E. 1983b. The role of complexity in comparing classifications. *Math. Biosci.* 66:97–114

Day, W. H. E. 1984. Efficient algorithms for agglomerative hierarchical clustering methods. *J. Classif.* 1:7–24

Day, W. H. E., ed. 1986a. Consensus classification. *J. Classif.* 3(2) (Special issue)

Day, W. H. E. 1986b. Foreword: comparison and consensus of classifications. *J. Classif.* 3:183–85

Day, W. H. E. 1988. Consensus methods as tools for data analysis. See Bock 1988, pp. 317–24

Day, W. H. E., ed. 1990. *Classif. Lit. Automat. Search Serv.* 19

Day, W. H. E., Edelsbrunner, H. 1985. Investigation of proportional link linkage clustering methods. *J. Classif.* 2:239–54

Day, W. H. E., Johnson, D. S., Sankoff, D. 1986a. The computational complexity of inferring rooted phylogenies by parsimony. *Math. Biosci.* 81:33–42

Day, W. H. E., McMorris, F. R. 1985. A formalization of consensus index methods. *Bull. Math. Biol.* 47:215–29

Day, W. H. E., McMorris, F. R., Meronk, D. B. 1986b. Axioms for consensus functions based on lower bounds in posets. *Math. Soc. Sci.* 12:185–90

Day, W. H. E., Sankoff, D. 1986. Computational complexity of inferring phylogenies by compatibility. *Syst. Zool.* 35:224–29

De Boeck, P., Rosenberg, S. 1988. Hierarchical classes: model and data analysis. *Psychometrika* 53:361–81

Defays, D. 1977. An efficient algorithm for a complete link method. *Computer J.* 20:364–66

Degens, P. O. 1983. Hierarchical cluster methods as maximum likelihood estimators. See Felsenstein 1983a, pp. 249–53

Degens, P. O. 1985. Ultrametric approximation to distances. *Comput. Stat. Q.* 2:93–101

Delcoigne, A., Hansen, P. 1975. Sequence comparison by dynamic programming. *Biometrika* 62:661–64

de Leeuw, J., Heiser, W. 1977. Convergence of correction-matrix algorithms for multidimensional scaling. In *Geometric Representations of Relational Data: Readings in Multidimensional Scaling* ed. J. C. Lingoes, pp. 735–52. Ann Arbor, MI: Mathesis

Denning, P. J. 1989. The science of computing: Bayesian learning. *Am. Sci.* 77:216–18

DeSarbo, W. S. 1982. GENNCLUS: new models for general nonhierarchical clustering analysis. *Psychometrika* 47:446–49

DeSarbo, W. S., Carroll, J. D., Clark, L. A., Green, P.E. 1984. Synthesized clustering: a method for amalgamating alternative clustering bases with differential weighting of variables. *Psychometrika* 49:57–78

DeSarbo, W. S., Mahajan, V. 1984. Constrained classification: the use of a priori information in cluster analysis. *Psychometrika* 49:187–216

De Soete, G. 1986. Optimal variable weighting for ultrametric and additive tree clustering. *Qual. Quant.* 20:169–80

De Soete, G. 1988. Tree representations of proximity data by least squares methods. See Bock 1988, pp. 147–56

De Soete, G., Carroll, J. D. 1988. Optimal weighting for one-mode and two-mode ultrametric tree representations of three-way three-mode data. In *The Many Faces of Multivariate Data Analysis,* ed. M. G. H. Jansen, W. H. van Schuur, pp. 16–29. Groningen: RION

De Soete, G., Carroll, J. D. 1989. Ultrametric tree representations of three-way three-mode data. See Coppi & Bolasco 1989, pp. 415–26

De Soete, G., Carroll, J. D., DeSarbo, W. S. 1986. Alternating least squares optimal variable weighting algorithms for ultrametric and additive tree representations. See Gaul & Schader 1986, pp. 97–103

De Soete, G., Carroll, J. D., DeSarbo, W. S. 1987. Least squares algorithms for constructing constrained ultrametric and additive tree representations of symmetric proximity data. *J. Classif.* 4:155–73

De Soete, G., DeSarbo, W. S., Carroll, J. D. 1985. Optimal variable weighting for hierarchical clustering: an alternating least-squares algorithm. *J. Classif.* 2:173–92

De Soete, G., DeSarbo, W. S., Furnas, G. W., Carroll, J. D. 1984a. The estimation of ultrametric and path length trees from rectangular proximity data. *Psychometrika* 49:289–310

De Soete, G., DeSarbo, W. S., Furnas, G. W., Carroll, J. D. 1984b. Tree representations of rectangular proximity matrices. See Abdi et al 1984, pp. 377–92

De Soete, G., Hubert, L., Arabie, P. 1988. On the use of simulated annealing for combinatorial data analysis. See Arabie & Daws 1988, pp. 329–40

Deutsch, S. B., Martin, J. J. 1971. An ordering algorithm for analysis of data arrays. *Oper. Res.* 19:1350–62

Dickinson, J. R. 1990. *The Bibliography of Marketing Research Methods.* Lexington, MA: Lexington. (3rd ed.)

Diday, E. 1986. Orders and overlapping clusters by pyramids. See Arabie 1986, pp. 201–34

Diday, E., Escoufier, Y., Lebart, L., Pagès, J. P., Schektman, Y., Tomassone, R., eds. 1986. *Data Analysis and Informatics, IV.* New York: North-Holland

Diday, E., Jambu, M., Lebart, L., Pagès, J., Tomassone, R., eds. 1984. *Data Analysis and Informatics, III.* Amsterdam: North-Holland

Doignon, J.-P., Falmagne, J.-C. 1984. Matching relations and the dimensional structure of social choices. *Math. Soc. Sci.* 7:211–29

Doignon, J.-P., Monjardet, B., Roubens, M., Vincke, P. 1986. Biorder families, valued relations, and preference modelling. *J. Math. Psychol.* 30:435–80

Dow, M. M., Cheverud, J. M. 1985. Comparison of distance matrices in studies of population structure and genetic microdifferentiation: quadratic assignment. *Am. J. Phys. Anthropol.* 68:367–73

Dow, M. M., de Waal, F. B. M. 1989. Assignment methods for the analysis of network subgroup interactions. *Soc. Networks* 11:237–55

Dress, A., von Haeseler, A., eds. 1990. *Trees and Hierarchical Structures (Lecture Notes in Biomathematics 84).* New York: Springer

Dress, A., von Haeseler, A., Krueger, M. 1986. Reconstructing phylogenetic trees using variants of the "four-point-condition." See Bock 1986, pp. 299–305

Dubes, R. C. 1987. How many clusters are best? —An experiment. *Pattern Recognit.* 20:645–63

Dubes, R. C. 1988. Review of *Simulated Annealing: Theory and Practice* by P. J. M. van Laarhoven, E. H. L. Aarts. *J. Classif.* 5:126–28

Dubes, R., Jain, A. K. 1979. Validity studies in clustering methodologies. *Pattern Recognit.* 11:235–54

Duran, B. S., Odell, P. L. 1974. *Cluster Analysis: A Survey.* New York: Springer-Verlag

Edgington, E. S. 1980. *Randomization Tests.* New York: Marcel Dekker

Edgington, E. S. 1987. *Randomization Tests*. New York: Marcel Dekker. (2nd ed.)
Estabrook, G. F., McMorris, F. R. 1980. When is one estimate of evolutionary relationships a refinement of another? *J. Math. Biol.* 10:367–73
Everitt, B. 1980. *Cluster Analysis*. New York: Wiley-Halsted. (2nd ed.)
Faith, D. P. 1984. Patterns of sensitivity of association measures in numerical taxonomy. *Math. Biosci.* 69:1–9
Faith, D. P. 1985. A model of immunological distances in systematics. *J. Theor. Biol.* 114:511–26
Faith, D. P. 1988. Consensus applications in the biological sciences. See Bock 1988, pp. 325–32
Faust, K., Romney, A. K. 1985. Does STRUCTURE find structure? A critique of Burt's use of distance as a measure of structural equivalence. *Soc. Networks* 7:77–103
Feger, H., Bien, W. 1982. Network unfolding. *Soc. Networks* 4:257–83
Feger, H., Droge, U. 1984. Ordinale Netzwerkskalierung [Ordinal network scaling]. *Kölner Z. Soziol. Sozialpsychol.* 3:417–23
Felsenstein, J. , ed. 1983a. *Numerical Taxonomy*. Berlin: Springer-Verlag
Felsenstein, J. 1983b. Methods for inferring phylogenies: a statistical view. See Felsenstein 1983a, pp. 315–34
Felsenstein, J. 1983c. Statistical inference of phylogenies (with discussion). *J. R. Stat. Soc. Ser. A* 146:246–72
Ferligoj, A. 1990. Review of *Data Analysis and Informatics V*, edited by E. Diday. *J. Classif.* 7:153–58
Ferligoj, A., Batagelj, V. 1982. Clustering with relational constraint. *Psychometrika* 47:413–26
Ferligoj, A., Batagelj, V. 1983. Some types of clustering with relational constraints. *Psychometrika* 48:541–52
Fichet, B. 1986. Distances and Euclidean distances for presence-absence characters and their application to factor analysis. See Arabie 1986, pp. 23–46
Finden, C. R., Gordon, A. D. 1985. Obtaining common pruned trees. *J. Classif.* 2:255–76
Fink, W. L. 1986. Microcomputers and phylogenetic analysis. *Science* 234:1135–39
Fishburn, P. C., Falmagne, J.-C. 1989. Binary probabilities and rankings. *Econ. Lett.* 31:113–17
Fligner, M. A., Verducci, J. S. 1986. Distance based ranking models. *J. R. Stat. Soc. Ser. B* 48:359–69
Florek, K., Łukaszewicz, J., Perkal, J., Steinhaus, H., Zubrzycki, S. 1951. Sur la liaison et la division des points d'un ensemble fini. *Colloq. Math.* 2:282–85
Fowlkes, E. B., Gnanadesikan, R., Kettenring, J. R. 1987. Variable selection in clustering and other contexts. In *Design, Data, and Analysis*, ed. C. L. Mallows, pp. 13–34. New York: Wiley
Fowlkes, E. B., Gnanadesikan, R., Kettenring, J. R. 1988. Variable selection in clustering. *J. Classif.* 5:205–28
Frank, O. 1978. Inferences concerning cluster structure. In *COMPSTAT 1978 Proceedings in Computational Statistics*, ed. L. C. A. Corsten, J. Hermans, pp. 259–65) Vienna: Physica-Verlag
Friendly, M. L. 1977. In search of the M-gram: the structure of organization in free recall. *Cognit. Psychol.* 9:188–249
Friendly, M. L. 1979. Method for finding graphic representations of associative memory structures. In *Memory Organization and Structure*, ed. C. R. Puff, pp. 85–129. New York: Academic
Furnas, G. W. 1980. *Objects and their features: the metric analysis of two-class data.* PhD thesis. Stanford Univ., Stanford, CA
Furnas, G. W. 1989. Metric family portraits. *J. Classif.* 6:7–52
Ganesalingam, S. 1989. Classification and mixture approaches to clustering via maximum likelihood. *Appl. Stat. (J. R. Stat. Soc. Ser. C)* 38:455–66
Gaul, W., Schader, M., eds. 1986. *Classification as a Tool of Research*. Amsterdam: North-Holland
Gaul, W., Schader, M. 1988. Clusterwise aggregation of relations. *Appl. Stoch. Models Data Anal.* 4:273–82
Giacomelli, F., Wiener, J., Kruskal, J. B., Pomeranz, J. V., Loud, A. V. 1971. Subpopulations of blood lymphocytes demonstrated by quantitative cytochemistry. *J. Histochem. Cytochem.* 19:426–33
Giakoumakis, V., Monjardet, B. 1987a. Coefficients d'accord entre deux preordres totaux: comparaison ordinale des coefficients. *Math. Sci. hum.* 25(98):69–87
Giakoumakis, V., Monjardet, B. 1987b. Coefficients d'accord entre deux préordres totaux. *Stat. Anal. Données* 12:46–99
Godehardt, E. 1988. *Graphs as Structural Models: The Application of Graphs and Multigraphs in Cluster Analysis*. Wiesbaden: Vieweg
Gordon, A. D. 1973a. Classification in the presence of constraints. *Biometrics* 29:821–27
Gordon, A. D. 1973b. A sequence-comparison statistic and algorithm. *Biometrika* 60:197–200
Gordon, A. D. 1980a. On the assessment and comparison of classifications. In *Analyse de*

Données et Informatique, ed. R. Tomassone, pp. 149–60. Le Chesnay: Inst. Natl. Rech. Inform. Automat.

Gordon, A. D. 1980b. Methods of constrained classification. See Gordon 1980a, pp. 161–71

Gordon, A. D. 1981. *Classification*. London: Chapman and Hall

Gordon, A. D. 1982. An investigation of two sequence-comparison statistics. *Aust. J. Stat.* 24:332–42

Gordon, A. D. 1986. Consensus supertrees: the synthesis of rooted trees containing overlapping sets of labeled leaves. *J. Classif.* 3:335–48

Gordon, A. D. 1987a. A review of hierarchical classification. *J. R. Stat. Soc. Ser. A* 150:119–37

Gordon, A. D. 1987b. Parsimonious trees. *J. Classif.* 4:85–101

Gordon, A. D. 1988. Sequence comparison statistics. In *Encyclopedia of Statistical Sciences*, cd. S. Kotz, N. L. Johnson, C. B. Read, 8:375–77. New York: Wiley

Gordon, A. D., Finden, C. R. 1985. Classification of spatially-located data. *Comput. Stat. Q.* 2:315–28

Gordon, A. D., Henderson, J. T. 1977. An algorithm for Euclidean sum of squares classification. *Biometrics* 33:355–62

Gordon, A. D., Jupp, P. E., Byrne, R. W. 1989. The construction and assessment of mental maps. *Br. J. Math. Stat. Psychol.* 42:169–82

Govaert, G. 1988. Classification binaire et modéle. Rep. No. 949. Le Chesnay: Inst. Natl. Rech. Inform. Automat.

Govaert, G. 1989. Clustering model and metric with continuous data. In *Data Analysis, Learning Symbolic and Numeric Knowledge*, ed. E. Diday, pp. 65–71. Commack, NY: Nova Science

Gower, J. C. 1971. A general coefficient of similarity and some of its properties. *Biometrics* 27:857–71

Gower, J. C. 1973. Classification problems. *Bull. Int. Stat. Inst.* 45:471–77

Gower, J. C. 1974. Maximal predictive classification. *Biometrics* 30:643–54

Gower, J. C. 1983. Comparing classifications. See Felsenstein 1983a, pp. 137–55

Gower, J. C. 1985. Measures of similarity, dissimilarity and distance. In *Encyclopedia of Statistical Sciences*, ed. S. Kotz, N. L. Johnson, C. B. Read, 5:307–405). New York: Wiley

Gower, J. C. 1986a. Euclidean distance matrices. See Arabie 1986, pp. 11–22, 61–63

Gower, J. C. 1986b. Reply to discussants. See Arabie 1986, pp. 61–63

Gower, J. C., Banfield, C. F. 1975. Goodness-of-fit criteria for hierarchical classification and their empirical distributions. In *Proc. 8th Int. Biometric Conference*, ed. L. C. A. Corsten, T. Dostelnicu, pp. 347–61). Bucharest: Editura Academiei Republicii Socialiste Romania

Gower, J. C., Legendre, P. 1986. Metric and Euclidean properties of dissimilarity coefficients. *J. Classif.* 3:5–48

Gower, J. C., Payne, R. W. 1975. A comparison of different criteria for selecting binary tests in diagnostic keys. *Biometrika* 62:665–72

Gower, J. C., Ross, G. J. S. 1969. Minimum spanning trees and single linkage cluster analysis. *Appl. Stat.* 18:54-64

Graham, R. L., Hell, P. 1985. On the history of the minimum spanning tree problem. *Ann. Hist. Comput.* 7:43–57

Green, P. E., Carmone, F. J. Jr., Smith, S. M. 1989. *Multidimensional Scaling: Concepts and Applications*. Boston: Allyn and Bacon

Green, P. E., Carmone, F. J., Kim, J. 1990. A preliminary study of optimal variable weighting in k-means clustering. *J. Classif.* 7:271–85

Grover, R., Srinivasan, V. 1987. A simultaneous approach to market segmentation and market structuring. *J. Market. Res.* 24:139–53

Guénoche, A. 1985. Classification using dilemma functions. *Comput. Stat. Q.* 2:103–8

Guénoche, A. 1986a. Représentations arborées des classifications. *Rech. Opér./Oper. Res.* 20:341–54

Guénoche, A. 1986b. Graphical representation of a Boolean array. *Comput. Hum.* 20:277–81

Guénoche, A. 1987. Cinq algorithmes d'approximation d'une dissimilarité par des arbres à distances additives. *Math. Sci. hum.* 25(98): 21–40

Guénoche, A., Monjardet, B. 1987. Méthodes ordinales et combinatoires en analyse des données. *Math. Sci. hum.* 25(100):5–47

Guénoche, A., Hansen, P., Jaumard, B. 1991. Efficient algorithms for divisive hierarchical clustering with the diameter criterion. *J. Classif.* 8:5–30

Halperin, D. 1989. Towards deciphering the Ugaritic musical notation. Tel-Aviv Univ., Dept. Musicol.

Hand, D. J. 1981. Branch and bound in statistical data analysis. *Statistician* 30:1–13

Hansen, P., Delattre, M. 1978. Complete-link cluster analysis of graph coloring. *J. Am. Stat. Assoc.* 73:397–403

Hansen, P., Jaumard, B. 1987. Minimum sum of diameters clustering. *J. Classif.* 4:215–26

Hansen, P., Jaumard, B., Frank, O. 1989.

Maximum sum-of-splits clustering. *J. Classif.* 6:177–93

Hansen, P., Lehert, P. 1980. Clustering by connected components large data sets with Minkowsky distances between entities. In *Data Analysis and Informatics*, ed. E. Diday, L. Lebart, J. P. Pagès, R. Tomassone, pp. 561–67. Amsterdam: North-Holland

Harding, M. R., Sokal, R. R. 1988. Classification of the European language families by genetic distance. *Proc. Natl. Acad. Sci. USA* 85:9370–72

Hartigan, J. A. 1967. Representation of similarity matrices by trees. *J. Am. Stat. Assoc.* 62:1140–58

Hartigan, J. A. 1972. Direct clustering of a data matrix. *J. Am. Stat. Assoc.* 67:123–29

Hartigan, J. A. 1973. Clustering. *Annu. Rev. Biophysics* 2:81–101

Hartigan, J. A. 1975. *Clustering Algorithms.* New York: Wiley

Hartigan, J. A. 1976. Modal blocks in dentition of West Coast mammals. *Syst. Zool.* 25:149–60

Hartigan, J. A. 1977. Distribution problems in clustering. See Van Ryzin 1977, pp. 45–71

Hartigan, J. A. 1978. Asymptotic distributions for clustering criteria. *Ann. Stat.* 6:117–31

Hartigan, J. A. 1985. Statistical theory in clustering. *J. Classif.* 2:63–76

Hartigan, J. A. 1988. The span test for unimodality. See Bock 1988, pp. 229–36

Hartigan, J. A., Hartigan, P. M. 1985. The dip test of unimodality. *Ann. Stat.* 13:70–84

Hartigan, J. A., Mohanty, S. 1992. The runt test for multimodality. *J. Classif.* 9

Heiser, W. J. 1986. Distances and their approximation. See Arabie 1986, pp. 47–52

Heiser, W. J. 1989. The city-block model for three-way multidimensional scaling. See Coppi & Bolasco 1989, pp. 395–404

Heltshe, J. F. 1988. Jackknife estimate of the matching coefficient of similarity. *Biometrics* 44:447–60

Herden, G. 1984. Some aspects of clustering functions. *SIAM J. Algorithms Disc. Methods* 5:101–16

Hilger, J., Harhalakis, G., Proth, J.-M. 1989. Generalized cross-decomposition method: algorithm and implementation Res. Rep. No. 1055. Le Chesnay: Inst. Natl. Rech. Inform. Automat.

Hirtle, S. C. 1982. Lattice-based similarity measures between ordered trees. *J. Math. Psychol.* 25:206–25

Hirtle, S. C. 1987. On the classification of recall strings using lattice-theoretic measures. *J. Classif.* 4:227–42

Hirtle, S. C., Crawley, E. 1989. The use of ordered trees for uncovering strategies in free-recall. See Luong 1989, pp. 125–38

Hojo, H. 1983. A maximum likelihood method for additive clustering and its applications. *Jpn. Psychol. Res.* 25:191–201

Holman, E. W. 1972. The relation between hierarchical and Euclidean models for psychological distances. *Psychometrika* 37: 417–23

Holman, E. W. 1985. Evolutionary and psychological effects in pre-evolutionary classifications. *J. Classif.* 2:29–39

Hubálek, Z. 1982. Coefficients of association and similarity, based on binary (presence-absence) data: an evaluation. *Biol. Rev.* 57:669–89

Hubert, L. J. 1974a. Spanning trees and aspects of clustering. *Br. J. Math. Stat. Psychol.* 27:14–28

Hubert, L. J. 1974b. Some applications of graph theory to clustering. *Psychometrika* 39:283–309

Hubert, L. J. 1977a. A set-theoretical approach to the problem of hierarchical clustering. *J. Math. Psychol.* 15:70–88

Hubert, L. J. 1977b. Nominal scale response agreement as a generalized correlation. *Br. J. Math. Stat. Psychol.* 30:98–103

Hubert, L. J. 1978. Generalized proximity function comparisons. *Br. J. Math. Stat. Psychol.* 31:179–92

Hubert, L. J. 1979. Generalized concordance. *Psychometrika* 44:135–42

Hubert, L. J. 1983. Hierarchical cluster analysis. In *Encyclopedia of Statistical Sciences*, ed. S. Kotz, N. Johnson, C. Read, 3:623–30. New York: Wiley

Hubert, L. J. 1984. Review of *Graphs and Genes* by B. G. Mirkin and S. N. Rodin. *J. Classif.* 1:275–77

Hubert, L. J. 1987. *Assignment Methods in Combinatorial Data Analysis.* New York: Marcel Dekker

Hubert, L. J., Arabie, P. 1985. Comparing partitions. *J. Classif.* 2:193–218

Hubert, L. J., Arabie, P. 1986. Unidimensional scaling and combinatorial optimization. See Arabie 1986, pp. 181–96

Hubert, L. J., Arabie, P. 1988. Relying on necessary conditions for optimization: unidimensional scaling and some extensions. See Bock 1988, pp. 463–72

Hubert, L., Arabie, P. 1989. Combinatorial data analysis: confirmatory comparisons between sets of matrices. *Appl. Stoch. Models Data Anal.* 5:273–325

Hubert, L. J., Baker, F. B. 1976. Data analysis by single-link and complete-link hierarchical clustering. *J. Educ. Stat.* 1:87–111

Hubert, L. J., Baker, F. B. 1978. Applications of combinatorial programming to data analysis: the traveling salesman and related problems. *Psychometrika* 43:81–91

Hubert, L. J., Baker, F. B. 1979. Identifying a migration effect in complete-link hierarchical clustering. *J. Educ. Stat.* 4:74–92

Hubert, L. J., Golledge, R. G. 1983. Rater agreement for complex assessments. *Br. J. Math. Stat. Psychol.* 36:207–16

Hubert, L. J., Golledge, R. G. 1981. Matrix reorganization and dynamic programming: applications to paired comparison and unidimensional seriation. *Psychometrika* 46:429–41

Hubert, L. J., Golledge, R. G., Richardson, G. D. 1982. Proximity matrix reorganization and hierarchical clustering. *Environ. Plan. Ser. A* 14:195–203

Hubert, L. J., Levin, J. R. 1976. A general statistical framework for assessing categorical clustering in free recall. *Psychol. Bull.* 83:1072–80

Hubert, L. J., Levin, J. R. 1977. Inference models for categorical clustering. *Psychol. Bull.* 84:878–87

Hubert, L. J., Levin, J. R. 1978. Evaluating priority effects in free recall. *Br. J. Math. Stat. Psychol.* 31:11–18

Hubert, L. J., Schultz, J. R. 1976. Quadratic assignment as a general data analysis strategy. *Br. J. Math. Stat. Psychol.* 29:190–241

Hummon, N. P., Doreian, P. 1990. Computational methods for social network analysis. *Soc. Networks* 12:273–88

Hutchinson, J. W. 1989. NETSCAL: a network scaling algorithm for nonsymmetric proximity data. *Psychometrika* 54: 25–51

Imaizumi, T., Okada, A. 1990. *Three-way Scaling and Clustering: How to Use the Software*. Tokyo: Kyoritsu Shuppan

Jain, A. K., Dubes, R. C. 1988. *Algorithms for Clustering Data.* Englewood Cliffs, NJ: Prentice-Hall

Jambu, M. 1978. *Classification Automatique pour l'Analyse des Données,* Tome I. Paris: Dunod

Janowitz, M. F. 1978. An order theoretic model for cluster analysis. *SIAM J. Appl. Math.* 34:55–72

Janowitz, M. F. 1979. Monotone equivariant cluster methods. *SIAM J. Appl. Math.* 37:148–65

Janowitz, M. F., Schweizer, B. 1989. Ordinal and percentile clustering. *Math. Soc. Sci.* 18:135–86

Jardine, C. J., Jardine, N., Sibson, R. 1967. The structure and construction of taxonomic hierarchies. *Math. Biosci.* 1:173–79

Jardine, N., Sibson, R. 1968. The construction of hierarchic and non-hierarchic classifications. *Comput. J.* 11:177–84

Jardine, N., Sibson, R. 1971. *Mathematical Taxonomy*. London: Wiley

Johnson, S. C. 1967. Hierarchical clustering schemes. *Psychometrika* 32:241–54

Jones, L. V., Appelbaum, M. I. 1989. Psychometric methods. *Annu. Rev. Psychol.* 40:23–43

Jones, W. P., Furnas, G. W. 1987. Pictures of relevance: a geometric analysis of similarity measures. *J. Am. Soc. Inform. Sci.* 38:420–42

Kamensky, V. 1990. Review of *Klasternyi Analyz* by I. D. Mandel. *J. Classif.* 7:119–23

Katz, L. 1947. On the matric analysis of sociometric data. *Sociometry* 10:233–41

Kim, K. H., Roush, F. W. 1978. Ultrametrics and matrix theory. *J. Math. Psychol.* 18:195–203

Kirkpatrick, S., Gelatt, C. D. Jr., Vecchi, M. P. 1983. Optimization by simulated annealing. *Science* 220:671–80

Klastorin, T. D. 1980. Merging groups to maximize object partition comparison. *Psychometrika* 45:425–33

Klastorin, T. D. 1982. An alternative method for hospital partition determination using hierarchical cluster analysis. *Oper. Res.* 30:1134–47

Klauer, K. C. 1989. Ordinal network representation: representing proximities by graphs. *Psychometrika* 4:737–50

Klauer, K. C., Carroll, J. D. 1989. A mathematical programming approach to fitting general graphs. *J. Classif.* 6:247–70

Klauer, K. C., Carroll, J. D. 1991. A comparison of two approaches to fitting directed graphs to nonsymmetric proximity measures. *J. Classif.* 8

Klein, R. W., Dubes, R. C. 1989. Experiments in projection and clustering by simulated annealing. *Pattern Recognit.* 22:213–20

Knoke, D., Kuklinski, J. H. 1982. *Network Analysis*. Newbury Park, CA: Sage

Krippendorf, K. 1980. *Content Analysis, an Introduction to Its Methodology*. Newbury Park, CA: Sage

Krippendorff, K. 1987. Association, agreement, and equity. *Qual. Quant.* 21:109–23

Kruskal, J. B. 1956. On the shortest spanning subtree of a graph and the traveling salesman problem. *Proc. Am. Math. Soc.* 7:48–50

Kruskal, J. B. 1964a. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika* 29:1–27

Kruskal, J. B. 1964b. Nonmetric multidimensional scaling: a numerical method. *Psychometrika* 29:115–29

Kruskal, J. B. 1977. The relationship between multidimensional scaling and clustering. See Van Ryzin 1977, pp. 17–44

Kruskal, J. B. 1983. An overview of sequence comparison. In *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*, ed. D. Sankoff, J. B. Kruskal, pp. 1–44. Reading, MA: Addison-Wesley

Krzanowski, W. J., Lai, Y. T. 1988. A criterion for determining the number of groups in a data set using sum-of-squares clustering. *Biometrics* 44:23–4

Kusiak, A., Finke, G. 1987. Hierarchical approach to the process planning problem. *Disc. Appl. Math.* 18:145–84

Kusiak, A., Vannelli, A., Kumar, K. R. 1986. Clustering analysis: models and algorithms. *Control Cybern.* 15:139–54

Lance, G. N., Williams, W. T. 1967. A general theory of classificatory sorting strategies. I. Hierarchical systems. *Comput. J.* 9:373–80

Law, H. G., Snyder, C. W. Jr., Hattie, J. A., McDonald, R. P., eds. 1984. *Research Methods for Multimode Data Analysis*. New York: Praeger

Leclerc, B. 1979. Semi-modularité des treillis d'ultramétriques. *C. R. Acad. Sci.* 288:575–77

Leclerc, B. 1981. Description combinatoire des ultramétriques. *Math. Sci. hum.* 19 (73):5–37

Leclerc, B. 1982. Description, évaluation et comparaison de hiérarchies de parties. Cent. Anal. Math. Soc., Paris

Leclerc, B. 1984. Efficient and binary consensus functions on transitively valued relations. *Math. Soc. Sci.* 8:45–61

Leclerc, B. 1985a. La comparaison des hiérarchies: indices et metriques. *Math. Sci. hum.* 23(92):5–40

Leclerc, B. 1985b. Les hiérarchies de parties et leur demi-treillis. *Math. Sci. hum.* 23(89):5–34

Leclerc, B. 1986. Caractérisation, construction et dénombrement des ultramétriques supérieures minimales. *Stat. Anal. Données* 11(2):26–50

Leclerc, B. 1987. Arbres minimums communs et compatibilité de données de types variés. *Math. Sci. hum.* 23(98):41–67

Leclerc, B. 1988. Consensus applications in the social sciences. See Bock 1988, pp. 333–340

Leclerc, B. 1989. Consensus approaches for multiple categorical or relational data. See Coppi & Bolasco 1989, pp. 65–75

Leclerc, B., Cucumel, G. 1987. Consensus en classification: une revue bibliographique. *Math. Sci. hum.* 25(100):109–28

Leclerc, B., Monjardet, B. 1987a. Avant-propos. *Math. Sci. hum.* 25(98):5–7

Leclerc, B., Monjardet, B. 1987b. Combinatorics and data analysis, Part I *Math. Sci. hum.* 25(98) (Special issue)

Leclerc, B., Monjardet, B. 1987c. Combinatorics and data analysis, Part II *Math. Sci. hum.* 25(100) (Special issue)

Lee, H. B. 1980. A K-means cluster analysis computer program with cross-tabulations and next-nearest-neighbor analysis. *Educ. Psychol. Meas.* 40:133–38

Lee, H. B., MacQueen, J. B. 1980. Control parameters for program KMEANS. Program documentation, UCLA

Legendre, P. 1987. Constrained clustering. In *Developments in Numerical Ecology* [*NATO Adv. Stud. Inst. Ser. G (Ecol. Sci.)*], ed. P. Legendre, L. Legendre, pp. 289–307. Berlin: Springer-Verlag

Lenstra, J. K. 1974. Clustering a data array and the traveling-salesman problem. *Oper. Res.* 22:413–14

Lerman, I. C. 1980. Combinatorial analysis in the statistical treatment of behavioral data. *Qual. Quant.* 14:431–69

Lerman, I. C. 1981. *Classification et analyse ordinale des données ata classification and ordinal analysis*. Paris: Dunod

Lerman, I. C. 1983a. Indices d'association partielle entre variables "qualitatives nominales". *Rech. Opér./Oper. Res.* 17:213–59

Lerman, I. C. 1983b. Indices d'association partielle entre variables "qualitatives" ordinales. *Publ. Stat. Univ. Paris* 28:7–46

Lerman, I. C. 1987. Construction d'un indice de similarité entre objets décrits par des variables d'un type quelconque. Application au problème du consensus en classification (1). *Rev. Stat. Appl.* 25(2):39–60

Lerman, I. C. 1988. Comparing partitions: mathematical and statistical aspects. See Bock 1988, pp. 121–31

Lerman, I. C., Hardouin, M., Chantrel, T. 1980. Analyse de la situation relative entre deux classifications floues. See Hansen & Lehert 1980, pp. 523–52

Lerman, I. C., Peter, P. 1984. Analyse d'un algorithme de classification hiérarchique "en parallele" pour le traitement de gros ensembles. Res. Rep. No. 339. Le Chesnay: Inst. Natl. Rech. Inform. Automat.

Lerman, I. C., Peter, P. 1988. Structure maximale pour la somme des carrés d'une contingence aux marges fixées; une solution algorithmique programmée. *Rech. Opér./Oper. Res.* 22:83–136

Levin, J. R., Hubert, L. J. 1980. Measuring clustering in free recall. *Psychol. Bull.* 87:59–62

Li, X., Dubes, R. C. 1989. A probabilistic measure of similarity for binary data in pat-

tern recognition. *Pattern Recognit.* 22:397–409
Ling, R. F. 1973. A probability theory of cluster analysis. *J. Am. Stat. Assoc.* 68:159–64
Ling, R. F. 1975. An exact probability distribution on the connectivity of random graphs. *J. Math. Psychol.* 12:90–98
Lingoes, J. C. 1968. The multivariate analysis of qualitative data. *Multivar. Behav. Res.* 3:61–94
Lorr, M. 1983. *Cluster Analysis for Social Scientists*. San Francisco: Jossey-Bass
Luce, R. D., Krumhansl, C. L. 1988. Measurement, scaling, and psychophysics. In *Stevens' Handbook of Experimental Psychology,* ed. R. C. Atkinson, R. J. Herrnstein, G. Lindzey, R. D. Luce, pp. 3–74. New York: Wiley
Luong, X., ed. 1989. *Analyse arborée des Données textuelles* [*Tree Analysis of Textual Data*]. Nice: Cent. Natl. Rech. Sci.
MacQueen, J. 1967. Some methods for classification and analysis of multivariate observations. In *Proceedings Fifth Berkeley Symposium on Mathematical Statistics and Probability*, ed. L. M. Le Cam, J. Neyman, 1:281–97). Berkeley: Univ. Calif. Press
Mandel, I. D. 1988. *Klasternyi analyz* [*Cluster Analysis*]. Moscow: Finansy and Statistika
Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27:209–20
Marcotorchino, F. 1986. Cross association measures and optimal clustering. In *Compstat,* ed. F. DeAntoni, N. Lauro, A. Rizzi, pp. 188–94. Heidelberg: Physica-Verlag
Marcotorchino, F. 1987. Block seriation problems: a unified approach. *Appl. Stoch. Models Data Anal.* 3:73–91
Marcotorchino, F., Michaud, P. 1979. *Optimisation en Analyse ordinale des Données*. Paris: Masson
Margules, C. R., Faith, D. P., Belbin, L. 1985. An adjacency constraint in agglomerative hierarchical classifications of geographic data. *Environ. Plan. A* 17:397–412
Marriott, F. H. C. 1982. Optimization methods of cluster analysis. *Biometrika* 69:417–21
Matula, D. W. 1977. Graph theoretic techniques for cluster analysis algorithms. See Van Ryzin 1977, pp. 95–129
Matula, D. W., Sokal, R. R. 1980. Properties of Gabriel graphs relevant to geographic variation research and the clustering of points in the plane. *Geograph. Anal.* 12:205–22
McCormick, W. T. Jr., Schweitzer, P. J., White, T. W. 1972. Problem decomposition and data reorganization by a clustering technique. *Oper. Res.* 20:993–1009
McKeithen, K. B., Reitman, J. S., Rueter, H. R., Hirtle, S. C. 1981. Knowledge organization and skill differences in computer programmers. *Cognit. Psychol.* 13:307–25
McLachlan, G. J. 1982. Classification and mixture maximum likelihood approaches to cluster analysis. In *Handbook of Statistics 2: Classification, Pattern Recognition, and Reduction of Dimensionality,* ed. P. R. Krishnaiah, L. N. Kanal, pp. 199–208. Amsterdam: North-Holland
McLachlan, G. J., Basford, K. E. 1988. *Mixture Models: Inference and Applications to Clustering*. New York: Marcel Dekker
McMorris, F. R. 1985. Axioms for consensus functions on undirected phylogenetic trees. *Math. Biosci.* 74:17–21
McNamara, T. P. 1990. Memory's view of space. Paper presented at Dept. Psychol., Univ. Illinois, Champaign
McQuitty, L. L. 1960. Hierarchical syndrome analysis. *Educ. Psychol. Meas.* 20:293–304
Meehl, P. E., Golden, R. R. 1982. Taxometric methods. In *Handbook of Research Methods in Clinical Psychology*. ed. P. C. Kendall, J. N. Butcher, pp. 127–81. New York: Wiley
Mehta, C. R. 1990. StatXact: a statistical package for exact nonparametric inference. *J. Classif.* 7:111–4
Michalski, R. S., Stepp, R. E. 1983. Automated construction of classifications: conceptual clustering versus numerical taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* PAMI-5:396–410
Michaud, P., Marcotorchino, F. 1979. Modéles d'optimisation en analyse des données relationnelles. *Math. Sci. hum.* 17(67):7–38
Michaud, P., Marcotorchino, F. 1980. Optimisation en analyse des données relationnelles: application à l'agrégation des préférences et a la recherche de partitions optimales. See Hansen & Lehert 1980, pp. 655–70
Mielke, P. W. 1978. Clarification and appropriate inference for Mantel & Valand's nonparametric multivariate analysis technique. *Biometrics* 34:277–82
Mielke, P. W. 1979. On asymptotic nonnormality of null distributions of MRPP statistics. *Comm. Stat.: Theory and Methods* A8:1541–50. Errata: 1981. A10:1795 and 982. A11:847
Mielke, P. W., Berry, K. J., Johnson, E. S. 1976. Multi-response permutation procedures for a priori classifications. *Comm. Stat.: Theory and Methods* A5:1409–24
Miller, K. F. 1987. Geometric methods in developmental research. In *Formal Methods in Developmental Psychology,* ed. J.

Bisanz, C. J. Brainerd, R. Kail, pp. 216–62. New York: Springer-Verlag

Miller, W., Myers, E. W. 1988. Sequence comparison with concave weighting functions. *Bull. Math. Biol.* 50:97–120

Milligan, G. 1979. Ultrametric hierarchical clustering algorithms. *Psychometrika* 44: 343–46

Milligan, G. 1989. A validation study of a variable weighting algorithm for cluster analysis. *J. Classif.* 6:53–71

Milligan, G. W., Cooper, M. C. 1986. A study of the comparability of external criteria for hierarchical cluster analysis. *Multivar. Behav. Res.* 21:441–58

Milligan, G. W., Cooper, M. C. 1988. A study of standardization of variables in cluster analysis. *J. Classif.* 5:181–204

Mirkin, B. G. 1979. Group Choice, ed. P. C. Fishburn, transl. Y. Oliker. Washington, DC: V. H. Winston. (Original work published 1974)

Mirkin, B. G. 1986. Additive clustering and qualitative factor analysis methods. See Diday et al 1986, pp. 71–82

Mirkin, B. G. 1987. Additive clustering and qualitative factor analysis methods for similarity matrices. *J. Classif.* 4:7–31 (Erratum, 1989, 6:271–72)

Mirkin, B. G. 1989a. Erratum. *J. Classif.* 6:271–72

Mirkin, B. G. 1989b. Bilinear clustering algorithms. See Govaert 1989, pp. 65–71

Mirkin, B. G. 1990. A sequential fitting procedure for linear data analysis models. *J. Classif.* 7:167–95

Mirkin, B. G., Rodin, S. N. 1984. *Graphs and Genes*. Heidelberg: Springer-Verlag

Mkhadri, A. 1989. Pondération des variables pour la classification binaire. Res. Rep. No. 1079. Le Chesnay: Inst. Natl. Rech. Inform. Automat.

Monjardet, B. 1973. Tournois et ordres médians pour une opinion. *Math. Sci. hum.* 11(43):55–70

Monjardet, B. 1978. *Théorie des Graphes et Classification: Bibliographie et Index matière*. Paris: Cent. Math. Soc. Univ. Paris-V

Monjardet, B. 1979a. Metriques et relations: avant-propos. *Math. Sci. hum.* 17(67):5–6

Monjardet, B. 1979b. Relations à "éloignement minimum" de relations binaires. Note bibliographique. *Math. Sci. hum.* 17(67): 115–22

Monjardet, B. 1981a. Théorie des graphes et taxonomie mathématique. In *Regards sur la Théorie des Graphes,* ed. P. Hansen, D. de Werra, pp. 111–25. Lausanne: Presses Polytechniques Romandes

Monjardet, B. 1981b. Metrics on partially ordered sets: a survey. *Disc. Math.* 35:173–84

Morgan, B. J. T. 1973. Cluster analyses of two acoustic confusion matrices. *Percept. Psychophys.* 13:13–24

Morin, A. M. 1990. Review of *Mixture Models: Inferences and Applications to Clustering* by G. J. McLachlan, K. E. Basford. *Psychometrika* 55:167–68

Morineau, A., Lebart, L. 1986. Specific clustering algorithms for large data sets and implementation in SPAD software. See Gaul & Schader 1986, pp. 321–29

Muchnik, I. B., Chkuaseli, N. F., Shvartser, L. V. 1986. Modeling of behavior and intelligence: linguistic analysis of 0–1 matrices using monotone systems. *Automat. Remote Control* 4:533–60

Murtagh, F. 1983. A survey of recent advances in hierarchical clustering algorithms. *Comput. J.* 26:354–59

Murtagh, F. 1984a. Counting dendrograms: a survey. *Disc. Appl. Math.* 7:191–99

Murtagh, F. 1984b. Complexities of hierarchic clustering algorithms: state of the art. *Comput. Stat. Q.* 1:101–13

Murtagh, F. 1984c. An empirical study of coefficients for measuring the structure of hierarchic classifications. See Diday et al 1984, pp. 385–93

Murtagh, F. 1985a. *Multidimensional Clustering Algorithms*. Vienna: Physica-Verlag

Murtagh, F. 1985b. A survey of algorithms for contiguity-constrained clustering and related problems. *Comput. J.* 28:82–88

Murtagh, F. 1989. Review of *Data, Expert Knowledge and Decisions,* edited by W. Gaul, M. Schader. *J. Classif.* 6:129–32

Noma, E., Smith, D. R. 1985. Benchmark for the blocking of sociometric data. *Psychol. Bull.* 97:583–91

Ohsumi, N., Nakamura, N. 1989. Space distorting properties in agglomerative hierarchical clustering algorithms and a simplified method for combinatorial method. See Govaert 1989, pp. 103–8

Okada, A. 1989. Review of *Classification and Related Methods of Data Analysis,* edited by H. H. Bock. *J. Classif.* 6:121–25

Orth, B. 1988. Representing similarities by distance graphs: monotonic network analysis (MONA). See Bock 1988, pp. 489–94

Orth, B. 1989. Graph theoretical representations of proximities by monotonic network analysis (MONA). In *Mathematical Psychology in Progress,* ed. E. E. Roskam, pp. 299–308. New York: Springer-Verlag

Panel on Discriminant Analysis, Classification and Clustering. 1988. *Discriminant Analysis and Clustering*. Washington, DC: Natl. Academy Press

Payne, R. W., Preece, D. A. 1980. Identification keys and diagnostic tables: a review. *J. R. Stat. Soc. Ser. A* 143:253–92

Peck, R., Fisher, L., Van Ness, J. 1989. Approximate confidence intervals for the number of clusters. *J. Am. Stat. Assoc*. 84:184–91

Pellegrino, J., Hubert, L. J. 1982. The analysis of organization and structure in free recall. In *Handbook of Research Methods in Human Memory and Cognition*, ed. C. Richard Puff, pp. 129–72. New York: Academic

Perruchet, C. 1983a. Constrained agglomerative hierarchical classification. *Pattern Recognit*. 16:213–17

Perruchet, C. 1983b. Significance tests for clusters: overview and comments. See Felsenstein 1983a, pp. 199–208

Pliner, V. M. 1984. A class of metric scaling models. *Automat. Remote Control* 6:122–28

Popping, R. 1988. On agreement indices for nominal data. In *Sociometric Research*. Vol. 1. *Data Collection and Scaling*, ed. W. E. Saris I. N. Gallhofer, pp. 90–105. London: Macmillan

Pruzansky, S., Tversky, A., Carroll, J. D. 1982. Spatial versus tree representations of proximity data. *Psychometrika* 47:3–24

Punj, G., Stewart, D. W. 1983. Cluster analysis in marketing research: review and suggestions for application. *J. Market. Res*. 20:134–48

Rammal, R., Toulouse, G., Virasoro, M. A. 1986. Ultrametricity for physicists. *Rev. Mod. Phys*. 58:765–88

Rao, C. R. 1982. Diversity and dissimilarity coefficients: a unified approach. *Theor. Popul. Biol*. 21:24–43

Ripley, B. D. 1990. Review of *Simulated Annealing (SA) & Optimization* edited by M. E. Johnson *J. Classif*. 7:287–90

Rohlf, F. J. 1973. Hierarchical clustering using the minimum spanning tree. *Comput. J*. 16:93–5

Rohlf, F. J. 1974. Graphs implied by the Jardine-Sibson overlapping clustering models, B_k. *J. Am. Stat. Assoc*. 69:705–10

Rohlf, F. J. 1975. A new approach to the computation of the Jardine-Sibson B_k clusters. *Comput. J*. 18:164–68

Rohlf, F. J. 1982. Single-link clustering algorithms. See McLachlan 1982, pp. 267–84

Rosenberg, S. 1988. Self and others: studies in social personality and autobiography. *Adv. Exp. Soc. Psychol*. 21:57–95

Rosenberg, S. 1989. A study of personality in literary autobiography: an analysis of Thomas Wolfe's "Look Homeward, Angel". *J. Pers. Soc. Psychol*. 56 416–430

Ross, J. 1966. A remark on Tucker & Messick's "points of view" analysis. *Psychometrika* 31:27–31

Rouanet, H., Bernard, J.-M., Lecoutre, B. 1986. Nonprobabilistic statistical inference: a set-theoretic approach. *Am. Statistician* 40:60–65

Sabucedo, J., Ekehammar, B., Arce, C., Wendelheim, A. 1990. The cognitive representation of European countries: a comparison between Swedish and Spanish samples. Tech. Rep. No. 722. Univ. Stockholm, Dept. Psychol.

Sackin, M. J., Sneath, P. H. A. 1988. Choosing the smallest number of binary characters to distinguish the greatest number of OTUs. See Bock 1988, pp. 141–46

Sankoff, D., Kruskal, J. B., eds. 1983. *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*. Reading, MA: Addison-Wesley

Saporta, G. 1988. About maximal association criteria in linear analysis and in cluster analysis. See Bock 1988, pp. 541–50

Sattath, S., Tversky, A. 1987. On the relation between common and distinctive feature models. *Psychol. Rev*. 94:16–22

Schader, M. 1979. Distance minimale entre partitions et préordonnances dans un ensemble fini. *Math. Sci. hum*. 17(67):39–48

Shepard, R. N. 1962a. The analysis of proximities: Multidimensional scaling with an unknown distance function. I. *Psychometrika* 27:125–40

Shepard, R. N. 1962b. Analysis of proximities: multidimensional scaling with an unknown distance function. II. *Psychometrika* 27:219–46

Shepard, R. N. 1972. A taxonomy of some principal types of data and of multidimensional methods for their analysis. In *Multidimensional Scaling: Theory and Applications in the Behavioral Sciences*. Vol. I: *Theory*, ed. R. N. Shepard, A. K. Romney, S. B. Nerlove, pp. 24–47. New York: Seminar Press

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* 210:390–98

Shepard, R. N. 1988. George Miller's data and the development of methods for representing cognitive structures. In *The Making of Cognitive Science: Essays in Honor of George A. Miller*, ed. W. Hirst, pp. 45–70. Cambridge: Cambridge Univ. Press

Shepard, R. N., Arabie, P. 1979. Additive clustering: representation of similarities as combinations of discrete overlapping properties. *Psychol. Rev*. 86:87–123

Shiina, K. 1986. A maximum likelihood nonmetric multidimensional scaling procedure for word sequences obtained in free-recall experiments. *Jpn. Psychol. Res*. 28(2):53–63

Sibson, R. 1973. SLINK: an optimally efficient algorithm for the single-link cluster method. *Comput. J*. 16:30–45

Smith, J. E. K. 1976. Analysis of qualitative data. *Annu. Rev. Psychol.* 27:487–99

Smouse, P. E., Long, J. C., Sokal, R. R. 1986. Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Syst. Zool.* 35:627–32

Sneath, P. H. A., Sokal, R. R. 1973. *Numerical Taxonomy*. San Francisco: W. H. Freeman

Snijders, T. A. B., Dormaar, M., van Schuur, W. H., Dijkman-Caes, C., Driessen, G. 1990. Distribution of some similarity coefficients for dyadic binary data in the case of associated attributes. *J. Classif.* 7:5–31

Snyder, C. W. Jr., Law, H. G., Hattie, J. A. 1984. Overview of multimode analytic methods. See Law et al 1984, pp. 2–35

Sokal, R. R. 1985. The continuing search for order. *Am. Nat.* 126:729–49

Sokal, R. R. 1986a. Phenetic taxonomy: theory and methods. *Annu. Rev. Ecol. Syst.* 17:423–42

Sokal, R. R. 1986b. Spatial data analysis and historical processes. See Diday et al 1986, pp. 29–43

Sokal, R. R. 1988a. Unsolved problems in numerical taxonomy. See Bock 1988, pp. 45–56

Sokal, R. R. 1988b. Genetic, geographic, and linguistic distances in Europe. *Proc. Natl. Acad. Sci. USA* 85:1722–26

Sokal, R. R., Lengyel, I. A., Derish, P. A., Wooten, M. C., Oden, N. L. 1987. Spatial autocorrelation of ABO serotypes in mediaeval cemeteries as an indicator of ethnic and familial structure. *J. Archaeol. Sci.* 14:615–33

Spaeth, H. 1980. *Cluster Analysis Algorithms*, transl. U. Bull. Chichester, England: Ellis Horwood. (Original work published 1977)

Spaeth, H. 1985. *Cluster Dissection and Analysis: Theory, FORTRAN Programs, Examples*, transl. J. Goldschmidt. Chichester: Ellis Horwood. (Original work published 1983)

Spaeth, H. 1986a. Maximizing partitioning cluster criteria for quantitative data. See Bock 1986, pp. 221–28

Spaeth, H. 1986b. Anticlustering: maximizing the variance criterion. *Control Cybern.* 15:213–18

Spaeth, H. 1986c. Homogenous and heterogeneous clusters for distance matrices. See Bock 1988, pp. 157–64

Sriram, N. 1990. Clique optimization: a method to construct parsimonious ultrametric trees from similarity data. *J. Classif.* 7:33–52

Tarjan, R. E. 1982. A hierarchical clustering algorithm using strong components. *Inf. Process. Lett.* 14:26–29

ten Berge, J. M. F. 1988. Generalized approaches to the maxbet problem and the maxdiff problem, with applications to canonical correlations. *Psychometrika* 53:487–94

Tucker, L. R 1964. The extension of factor analysis to three-dimensional matrices. In *Contributions to Mathematical Psychology*, ed. N. Frederiksen, H. Gulliksen, pp. 109–27. New York: Holt, Rinehart, and Winston

Tucker, L. R 1983. Searching for structure in binary data. In *Principals of Modern Psychological Measurement*. ed. H. Wainer, S. Messick, pp. 215–35. Hillsdale, NJ: Erlbaum

Tucker, L. R, Messick, S. J. 1963. An individual difference model for multi-dimensional scaling. *Psychometrika* 28:333–67

Tversky, A., Hutchinson, J. W. 1986. Nearest neighbor analysis of psychological spaces. *Psychol. Rev.* 93:3–22

Upton, G., Fingleton, B. 1985. *Spatial Data Analysis by Example*. Vol. I. *Point Pattern and Quantitative Data*. New York: Wiley

Van Mechelen, I., De Boeck, P. 1990. Projection of a binary criterion into a model of hierarchical classes. *Psychometrika* 55:677–94

Van Ryzin, J., ed. 1977. *Classification and Clustering*. New York: Academic

Vegelius, J., Janson, S. 1982. Criteria for symmetric measures of association for nominal data. *Qual. Quant.* 16:243–50

Verhelst, N. D., Koppen, M. G. M., Van Essen, E. P. 1985. The exact distribution of an index of agreement between partitions. *Br. J. Math. Stat. Psychol.* 38:44–57

Walker, M. 1989. An analysis of auditory alphabet confusions. *Percept. Psychophys.* 45:315–22

Waller, N. G., Lykken, D. T., Tellegen, A. 1992. Occupational interests, leisure time interests and personality: three domains or one? Findings from the Minnesota Twin Registry. In *Assessing Individual Differences in Human Behavior: New Methods, Concepts, and Findings*, ed. R. V. Dawis, D. Lubinski. Minneapolis: Univ. Minnesota Press. In press

Wasserman, S., Anderson, C. 1987. Stochastic *a posteriori* blockmodels: construction and assessment. *Soc. Networks* 9:1–36

Wasserman, S., Faust, K. 1991. *Social Network Analysis: methods and Applications*. New York: Cambridge Univ. Press

Weisberg, H. F. 1974. Dimensionland: an excursion into spaces. *Am. J. Polit. Sci.* 18:743–76

Weiss, M. J. 1988. *The Clustering of America*. New York: Harper Row

Whaley, C. P. 1982. Interactive clustering. *Behav. Res. Meth. Instrument.* 14:170–75

Wilkinson, L. 1989. *SYSTAT: The System for Statistics*. Evanston, IL: Systat, Inc.
Willett, P. 1988. Recent trends in hierarchic document clustering: a critical review. *Inf. Process. Manage.* 24:577–97
Wind, Y. 1978. Issues and advances in segmentation research. *J. Market. Res.* 15:317–37
Windham, M. P. 1985. Numerical classification of proximity data with assignment measures. *J. Classif.* 2:157–72
Windham, M. P. 1987. Parameter modification for clustering criteria. *J. Classif.* 4:191–214
Windham, M. P. 1988. Review of *Mixture Models: Inference and Applications to Clustering* by G. J. McLachlan, K. E. Basford. *J. Classif.* 5:105–7
Wolfe, A. , ed. 1990. *Connections* 13
Wolfe, J. H. 1970. Pattern clustering by multivariate mixture analysis. *Multivar. Behav. Res.* 5:329–50
Wright, P. 1987. *Spycatcher.* New York: Viking
Zegers, F. E. 1986. Two classes of elementwise transformations preserving the psd nature of coefficient matrices. *J. Classif.* 3:49–53