Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Mathematical Psychology 49 (2005) 142–154
Journal of Mathematical Psychology
www.elsevier.com/locate/jmp

ConPar: a method for identifying groups of concordant subject proximity matrices for subsequent multidimensional scaling analyses

Michael J. Brusco*, J. Dennis Cradit

Department of Marketing, College of Business, Florida State University, Tallahassee, FL 32306-1110, USA

Received 25 June 2004; received in revised form 7 October 2004
Available online 12 January 2005

Abstract

A common representation of data within the context of multidimensional scaling (MDS) is a collection of symmetric proximity (similarity or dissimilarity) matrices for each of M subjects. There are a number of possible alternatives for analyzing these data, which include: (a) conducting an MDS analysis on a single matrix obtained by pooling (averaging) the M subject matrices, (b) fitting a separate MDS structure for each of the M matrices, or (c) employing an individual differences MDS model. We discuss each of these approaches, and subsequently propose a straightforward new method (CONcordance PARtitioning—ConPar), which can be used to identify groups of individual-subject matrices with concordant proximity structures. This method collapses the three-way data into a subject × subject dissimilarity matrix, which is subsequently clustered using a branch-and-bound algorithm that minimizes partition diameter. Extensive Monte Carlo testing revealed that, when compared to K-means clustering of the proximity data, ConPar generally provided better recovery of the true subject cluster memberships. A demonstration using empirical three-way data is also provided to illustrate the efficacy of the proposed method.

Keywords: Concordance; Multidimensional scaling; Cluster analysis; Pooling across subjects

1. Introduction

Multidimensional scaling (MDS) is an important method for spatial representation of psychological phenomena. The literature is replete with empirical studies that use MDS as the primary method of analysis, and considerable work has also focused on methodological development. A variety of textbooks, edited volumes, and review articles have addressed the importance of MDS models in various branches of psychology (Ashby, 1992; Borg & Groenen, 1997; Carroll & Arabie, 1980, 1998; Nosofsky, 1992). For example, Nosofsky (1992) emphasized the critical role that MDS plays in the development of theories related to cognitive processing models. More recently, Carroll and Arabie (1998) provide an encompassing review associated with the role of MDS in representing the psychological structure underlying perception and judgment. This review article also highlights a number of recent methodological advancements.

The strength of its appeal in psychology arises in large part from the fact that in MDS information about complex competitive relationships between stimuli can be shown compactly in visual, graphic displays, usually taking the form of spaces. These spaces can be generated from proximity matrices developed from a variety of respondent judgments such as similarity, identification confusion, same–different errors, or any other measure of pairwise relations between stimuli. The key assumptions are that these judged stimuli are described by values along a set of dimensions that place the stimuli as points in a multidimensional space and that the similarity between stimuli is inversely related to the distances of the corresponding points within that space.

*Corresponding author. Fax: 850 644 4098.
E-mail address: mbrusco@cob.fsu.edu (M.J. Brusco).

0022-2496/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2004.11.004

A common practice in applications of MDS is for the analyst to apply the selected model to a single proximity matrix produced by "pooling" proximity scores across two or more respondent matrices, the resulting "group" matrix reflecting a simple averaging of the individual scores. Despite the widespread popularity of this seemingly straightforward practice, a number of researchers caution that pooling by averaging across source matrices can distort the psychological structure of the data (Ashby, Maddox, & Lee, 1994; Furnas, 1989; Lee, 2001; Lee & Pope, 2003; Siegler, 1987). For example, Ashby et al. (1994) demonstrate that standard MDS models are apt to provide very good structural fits to similarity matrices obtained via averaging across respondents, regardless of the properties of the individual respondent data. Building on the work of Ashby et al. (1994), Lee and Pope (2003) recently proposed a method that uses the Bayesian Information Criterion (Schwarz, 1978) to determine the viability of the MDS representation for averaged data. The findings of both Ashby et al. (1994) and Lee and Pope (2003) suggest that excellent structural fits to the averaged data are not necessarily indicative of even moderately successful fits to the individual respondent data. It is quite possible that the MDS model fit to the pooled matrix will provide a structural representation inappropriate for at least some (and possibly all) of the matrices used in the pooling process.

If pooling across proximity matrices poses such risks, what options are available to the analyst? One immediate possibility is to develop individual MDS solutions for each of the source matrices. Although some academic studies of scaling use limited respondent sample sizes (e.g., Ashby and Lee (1991) used only two subjects), many research applications of MDS are applied to relatively large samples. Because of the need for larger sample sizes, it seems that individual-subject analyses for many empirical applications of MDS would be prohibitive. In addition, analysts would often be hard pressed to offer a convincing interpretation for the range of differing solutions across the individual matrices, particularly when the number of matrices is large.

One of the earliest attempts to deal with averaging in MDS was a "points of view" (POV) procedure suggested by Tucker and Messick (1963). They sought a compromise between the analysis of individual subject-level data and the scaling of pooled group data in an approach that purported to capture individual differences. In their procedure, similarity judgments were used to form a matrix of correlations among the individuals that was then factor analyzed to obtain a subject space, which in turn, was interpreted to identify individual differences. In the POV model, clusters of subjects were identified such that the individuals within clusters display similar judgment patterns (i.e., are all highly correlated) and individuals in different clusters display different judgment patterns (i.e., have weak correlations). Once identified, the procedure generates sets of hypothetical judgments, one set for each cluster, called a POV. These POV, in effect, represent a weighted average of the judgments made by all subjects, with those subjects in the corresponding cluster getting the most weight (Young & Hamer, 1987). The final step in the POV model is the analysis of each POV by an MDS procedure, culminating in a separate and independent solution for each view.

Despite some initial support, POV was criticized on a number of important points (cf., Carroll & Chang, 1970; Cliff, 1968; Ross, 1966), the most damning of which was that it failed to capture individual differences in a parsimonious manner. Critics argued that it was not really an individual differences MDS, but actually a factor analysis followed by MDS (Young & Hamer, 1987).[1] It is important for our present purpose to note that, in fact, the POV procedure actually involves averaging subjects, albeit a sophisticated weighted average determined by factor analytic techniques (Young & Hamer, 1987). The perils of averaging in this manner were articulated by Ross (1966), who observed that the sum of Euclidean distances is *not* Euclidean.

A third, and more elegant, option is for the analyst to apply an appropriate three-way MDS model (Arabie, Carroll, & DeSarbo, 1987; Carroll & Chang, 1970; Takane, Young, & de Leeuw, 1977). One of the earliest and most widely used three-way MDS methods is Carroll and Chang's (1970) *IN*dividual *D*ifferences *SCAL*ing (INDSCAL). This weighted Euclidean model addresses the pooling issue by explicitly capturing the effect of individual differences as weights, or importance, influencing the derived dimensions of the MDS solution. In this approach, a set of dimension coordinates is established for each stimulus, and each source matrix is linked to this general coordinate space via a set of dimension weights that stretches (or compresses) the coordinates on each dimension. Carroll and Chang (1970, 1972) also proposed the more general IDIOSCAL model, which permits an idiosyncratic rotation of the stimulus coordinate space for each subject in addition to the differential weighting of dimensions. Winsberg and De Soete (1993) developed CLASCAL, which is an extension of the INDSCAL model that incorporates a latent class method for obtaining clusters of subjects and simultaneously linking each cluster to a general coordinate space via a set of dimension weights.

Although three-way MDS methods capture individual differences in an elegant formal framework, some

[1]More recently, Meulman and Verboon (1993) have reassessed the POV perspective with a streamlined, reformulation of the original procedure. Meulman and Verboon (1993) argue that despite the power of an individual differences perspective, many settings still arise in which "group" analyses are important (a point we take up subsequently).

problems nevertheless remain. First, analysts have argued that the INDSCAL common space may not be the best way to display group (as opposed to individual) differences; in fact, the common space does not need to fit any group or individual (Meulman & Verboon, 1993). It is important to recognize that this limitation also pertains to CLASCAL, which relies on a linkage of the clusters produced by the latent class method to a single configuration space. A second problem that arises with INDSCAL is that the analyst is encumbered by the need to provide interpretations for the derived dimension weights. Arabie et al. (1987, p. 32) have argued that when employing INDSCAL, "...the data analyst has a responsibility to attempt to offer a convincing interpretation of these fitted parameters. If they are not interpretable, then the use of two-way MDS may be more appropriate for the particular dataset." However, reverting to the use of two-way MDS begs the previous pooling question as we have defined it. A third limitation of INDSCAL and CLASCAL is that the available software programs typically rely on weighted *Euclidean* models, and are not designed to incorporate alternative distance metrics (e.g., city block). This is not to suggest that three-way MDS models for the city block are not possible (see, e.g., Heiser, 1989), but available software packages are almost exclusively restricted to Euclidean distances.

In the present paper, we propose an alternative approach, one which abandons the search for a single best-fitting proximity matrix in favor of a search for a *set* of best-fitting proximity matrices. We investigate the possibility that the respondent sample can be partitioned into groups of respondents who share comparable proximity matrix structures. The viability of such an approach has recently been recognized by Lee and Pope (2003) who observed the need for a method that can find groups of subjects with the same underlying spatial representation. Our proposed method is a straightforward preprocessing tool that quantitative analysts can apply prior to fitting MDS models. This method is quite flexible and makes no assumptions about dimensionality of the MDS solutions or the appropriate metric (Euclidean, city block, etc.). In fact, our proposed method can be used as a preprocessing tool for both traditional deterministic MDS methods, as well as probabilistic approaches (MacKay, 2001; Zinnes & MacKay, 1983).

Our objective in this paper is to propose a method for identifying groups of subject dissimilarity matrices that have comparable structural properties. To accomplish this task, we develop a measure for collapsing three-way dissimilarity data into a subject × subject pairwise dissimilarity matrix, and subsequently produce groups of subjects using a partitioning algorithm. We refer to this new procedure as CONcordance PARtitioning (ConPar).

Our selected concordance measure draws heavily from Hubert's (1978, 1979) work on comparisons of proximity matrices and the measurement of concordance among those matrices. Most notably, we utilize Hubert's measures for capturing the internal structural properties of the matrices. This principle is an important aspect of our procedure, and helps us to construct plausible measures of dissimilarity between pairs of subjects. A second important aspect of our procedure is the deployment of optimal algorithms for partitioning the subject × subject pairwise dissimilarity matrix. A number of criteria can be employed for conducting this partitioning task, although certain criteria have the propensity to bias the cluster sizes that are obtained. By choosing our partitioning criteria carefully, we are able for the most part to avoid these biases. In subsequent sections, we test the integrity of our model in a series of Monte Carlo simulations. From an experimental standpoint, the emphasis of our analyses is on the ability of the proposed procedure to recover "true" clusters of subjects that possess comparable structural properties.[2]

In Section 2, we first develop ConPar as a method for identifying clusters of concordant dissimilarity matrices. In Sections 3–5 we report the results of extensive Monte Carlo testing of the proposed approach, including a comparison to a direct *K*-means clustering of the subject dissimilarity matrices. We subsequently provide a demonstration using published three-way data in Section 6. The paper concludes in Section 7 with a brief summary and the identification of the limitations and possible extensions of this research.

2. ConPar: a method for finding clusters of subjects

2.1. Model development

We define $\{i = 1,2,\ldots,N\}$ as the indices for a collection of N objects (stimuli) for which proximity data are to be obtained. We also define $C = \{m = 1,2,\ldots,M\}$ as the set of indices corresponding to M individual subjects in the experimental setting. Three-way data are assumed to be available in the form of a set of M individual-subject dissimilarity matrices, $B = \{\mathbf{A}_1,\mathbf{A}_2,\ldots,\mathbf{A}_M\}$, of dimension $N \times N$. Each dissimilarity matrix in B consists of an arbitrary main diagonal and non-negative off-diagonal elements $a_{ijm} = a_{jim}$ that represent the dissimilarity between objects i and j for subject m. From the matrices in B, we create an $M \times M$ dissimilarity matrix among the subjects, $\mathbf{S}$, using a pairwise concordance measure similar to one proposed

[2]The actual fitting of MDS models to these clusters of subjects is not part of the focus of our study. In fact, any number of metric or non-metric MDS models might be applied to the groups obtained by our method.

by Hubert (1987, Chapter 5). This type of measure, which uses within-row and within-column gradient information from the individual-subject matrices, is especially relevant in the contexts of permutation tests of matrix agreement (Brusco, 2004; Hubert, 1978, 1979, 1987) and combinatorial seriation (Brusco, 2002; Hubert, Arabie, & Meulman, 2001). The selected measure is specified as follows:

$$s_{uv} = s_{vu} = \sum_{h=1}^{N-2} \sum_{i=h+1}^{N-1} \sum_{j=i+1}^{N} [|\alpha_{hiju} - \alpha_{hijv}| + |\beta_{hiju} - \beta_{hijv}|]$$
$$\forall u = 1, \ldots, M-1 \text{ and } v = u+1, \ldots, M; \quad (1)$$

$$\text{where, } \alpha_{hiju} = \begin{cases} \text{sign}(a_{hju} - a_{hiu}) & \text{for distinct } h, i, j, \\ 0 & \text{otherwise,} \end{cases}$$

$$\beta_{hiju} = \begin{cases} \text{sign}(a_{jhu} - a_{jiu}) & \text{for distinct } h, i, j, \\ 0 & \text{otherwise,} \end{cases}$$

$$\text{and sign}(x) = \begin{cases} +1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -1 & \text{if } x < 0. \end{cases}$$

Thus, the dissimilarity measure between each pair of subjects (u and v) is a non-metric measure of the extent to which those subjects differ in terms of their dissimilarity orderings among triples of objects.

As pointed out by Hubert (1978), indices such as (1) have at least two advantages as a measure of concordance between two matrices. First, unlike correlation measures, they are invariant to monotone transformations of the matrices. Second, and much more importantly, concordance indices based on gradient information capture internal structural properties of the matrices, as opposed to correlation measures that are based on one-to-one products of matrix elements. One well-known example of an internal structural property is anti-Robinson structure (Brusco, 2002; Hubert et al., 2001, Chapters 3, 4; Robinson, 1951), which is perfectly achieved when matrix elements are always non-decreasing when moving away from the main diagonal within a specific row or column. Within the context of unidimensional scaling, if a perfect Euclidean recovery of the data is achievable, then there exists a reordering of the rows and columns of the proximity matrix that exhibits perfect anti-Robinson structure.

To illustrate relationships among correlation, our selected concordance index (1), and anti-Robinson structure, consider matrices $\mathbf{A}_1$ and $\mathbf{A}_2$ in Table 1. Clearly, both $\mathbf{A}_1$ and $\mathbf{A}_2$ exhibit a perfect anti-Robinson structure, and this is reflected by the concordance dissimilarity index (1) of $s_{12} = 0$. The correlation coefficient for these two matrices, however, is a modest .22. Thus, although the two dissimilarity matrices are weakly correlated, they are perfectly concordant with respect to the patterning of elements within rows and columns.

After producing the dissimilarity matrix $\mathbf{S}$, the next step is to develop a partition of the subjects such that subsequent MDS analyses can be pursued for each group, k ($k = 1,\ldots,K$), of subjects. Hubert et al. (2001, Chapter 3) present a number of alternative indices for partitioning a dissimilarity matrix. One of the most straightforward of these indices is the within-cluster sum of the pairwise dissimilarity measures. Defining C_k as the set of subject indices assigned to group k ($k = 1,\ldots,K$), the resulting optimization problem can be posed as follows:

$$\min: \quad Z_1 = \sum_{k=1}^{K} \sum_{(u<v)\in C_k} s_{uv}, \quad (2)$$

$$\text{subject to:} \quad C_k \neq \{\varnothing\} \quad \forall k = 1, \ldots, K, \quad (3)$$

$$C_f \cap C_k = \{\varnothing\} \; \forall f = 1, \ldots, K-1, \;\; k = f+1, \ldots, K, \quad (4)$$

$$C_1 \cup C_2 \cup \cdots \cup C_K = C. \quad (5)$$

The objective function (2) represents the sum, across all clusters, of the sum of pairwise dissimilarity measures in each cluster. Constraint set (3) guarantees that all clusters contain at least one subject. Constraint set (4) ensures that clusters do not overlap, and constraint (5) requires all subjects to be assigned to a cluster. Together, these constraints ensure that the clusters are mutually exclusive and exhaustive. More succinctly, they guarantee that the clusters form a partition of the subjects.

Table 1
Two small matrices for illustrating the potential differences between correlation and gradient-based concordance indices

	Matrix $\mathbf{A}_1$			
	1	2	3	4
1	—	38	40	42
2	38	—	32	36
3	40	32	—	4
4	42	36	4	—

	Matrix $\mathbf{A}_2$			
	1	2	3	4
1	—	3	40	75
2	3	—	25	31
3	40	25	—	30
4	75	31	30	—

Hansen and Jaumard (1997) identified several optimal solution procedures for solving the optimization problem posed by (2)–(5), including a straightforward implicit enumeration algorithm developed by Klein and Aronson (1991). Brusco (2003) recently strengthened Klein and Aronson's algorithm by incorporating some improved bounding procedures. Although its performance is sensitive to the properties of the dissimilarity matrix, the enhanced branch-and-bound algorithm is capable of solving problems of up to about $M = 50$ and $K = 6$. For larger matrices, the exchange algorithm of Banfield and Bassil (1977) is recommended.

One of the limitations of the within-cluster sum of pairwise dissimilarities index is a tendency to produce clusters of approximately the same size (Brusco, 2003; Hubert et al., 2001). For example, if $M = 20$ subjects are partitioned into $K = 4$ groups of size five, then there are $4(5(5-1)/2) = 40$ pairwise dissimilarities that are summed to compute the index. However, if the 20 subjects are partitioned into three groups of size one and one group of size 17, then there are $17(16)/2 = 136$ pairwise dissimilarity sums used in the index computation. For this reason, we also consider another important index, *partition diameter*, which is represented as follows:

$$\min: \quad Z_2 = \max_{(k=1,\ldots,K)} \left(\max_{(u<v)\in C_k} (s_{uv}) \right), \tag{6}$$

subject to (3)–(5). Following Johnson's (1967, p. 249) description of his "Maximum Method" for hierarchical clustering, the *diameter* of a cluster is the maximum pairwise dissimilarity index among objects in that cluster. The partition diameter is the maximum of the cluster diameters. Eq. (6) represents the minimization of partition diameter. Johnson's Maximum Method, which is also known as complete-link hierarchical clustering (Baker & Hubert, 1976; Hubert, 1974) can provide an approximate solution for a fixed number of clusters, but frequently will not minimize the partition diameter. However, guaranteed minimum-diameter solutions can be obtained using branch-and-bound methods comparable to those used for coloring the nodes of a graph (Brusco, 2003; Hansen & Delattre, 1978).

One of the advantages of the partition diameter index is that it is not predisposed to produce clusters of particular sizes. This is important for our context because of the potential for one fairly large cluster of subjects and a few small clusters. Another advantage is that minimization of the partition diameter is computationally less difficult than minimizing the within-cluster sum of dissimilarities. Using an implicit enumeration algorithm, in conjunction with a strong upper bound from a complete-link algorithm, Brusco (2003) obtained optimal solutions to some fairly sizable dissimilarity matrices (up to 45×45 and $K = 16$ clusters) within 2 min of microcomputer central processing unit (CPU) time. We have incorporated this implicit enumeration scheme in ConPar, and have successfully solved problems with up to 100 subjects and 10 clusters in a reasonable amount of microcomputer CPU time. For problems with hundreds or thousands of subjects, we recommend replacing implicit enumeration with the complete-link algorithm followed by an exchange algorithm from Banfield and Bassil (1977).

2.2. Algorithm implementation

The branch-and-bound algorithm for minimizing the within-cluster sum of pairwise dissimilarity measures was written in Fortran. This program incorporates heuristic exchange procedures (Banfield & Bassil, 1977) prior to the branch-and-bound process, which are designed to establish a good initial upper bound. The branch-and-bound algorithm for minimizing partition diameter was also written in Fortran, and incorporates the biased sampling complete-link algorithm for obtaining the upper bound. All computational results reported in this paper correspond to the implementation of the algorithms on a 2.2 GHz, Pentium IV PC, with 1 GB of random-access-memory.

3. Experiment #1: comparison of objective criteria

3.1. Experimental design

We began our experimental analyses with a computational study designed to assess the relative efficacy of the two partitioning algorithms. Sixty-four test problems were generated by varying six factors at two levels each. The levels for the first factor, the number of objects, were $N = 10$ and 20. The number of true clusters of subjects was the second factor, which was tested at levels of $K = 2$ and 4. The levels of the third factor, cluster density, were: (a) equally sized clusters, and (b) approximately 60% of the subjects in the first cluster with the remaining clusters equally sized. The number of dimensions ($D = 2$ and 3) and the distance metric (Euclidean and city block) comprised the fourth and fifth factors, respectively. The sixth factor was the level of perturbation of the object distances. For the first level of this factor, the raw distances were not perturbed, whereas for the second factor, each distance element was perturbed by $\pm 40\%$ of its raw value based on a uniform distribution.

The generation of the test problems was initiated by reading in the factor levels. For each group of subjects, a random permutation of the objects was obtained for each dimension. The $N \times N$ dissimilarity matrix for each subject was subsequently produced by randomly generating a distance between each pair of objects on each dimension using a uniform distribution, and then

computing the distance between each pair of objects using the appropriate metric. These distances either served directly as the dissimilarities or were subsequently perturbed, depending on the perturbation factor level. We produced $M = 20$ individual $N \times N$ dissimilarity matrices using this process, and subsequently used Eq. (1) to generate the $M \times M$ subject dissimilarity matrix, **S**. This matrix, along with the true cluster memberships for each subject, were written to separate files for subsequent processing.

For each of the 64 experimental conditions, the subject dissimilarity matrix, **S**, was partitioned using both the within-cluster sum of dissimilarities algorithm and the partition diameter algorithm. The cluster memberships produced by these algorithms were subsequently compared to the true cluster memberships using Hubert and Arabie's (1985) adjusted Rand index as the measure of agreement. This index is well-recognized as an effective measure for evaluating the level of agreement between partitions (Brusco & Cradit, 2001; Krieger & Green, 1999; Milligan, 1996; Milligan & Cooper, 1986).

3.2. Experimental results

Table 2 presents a summary of the results for the two partitioning algorithms. Although both algorithms performed well, the minimization of partition diameter was particularly effective. The partition diameter algorithm perfectly recovered each of the 64 true cluster structures, thus never misplacing a subject for any of the datasets. The algorithm for minimizing the within-cluster sum of pairwise dissimilarities provided perfect recovery for only 45 of the 64 datasets (70%), with rather poor adjusted Rand indices for some of the datasets. Most notably, this algorithm exhibited tremendous difficulty for problems with unequal cluster density, providing perfect recovery for only 13 of 32 (41%) test problems under such conditions. The partition diameter algorithm was also appreciably more efficient, requiring a maximum CPU time of .02 s. When minimizing the within-cluster sum of dissimilarities, the median CPU time was .91 s with a maximum of 54 s. In summary, the findings of this first experiment unequivocally supported the superiority of the partition diameter algorithm, which was both more efficient and less sensitive to cluster density properties. Our remaining experiments, therefore, only utilize the partition diameter algorithm.

4. Experiment #2: comparison to *K*-means clustering of subjects

4.1. Experimental design

Our second experiment focused on a comparison of ConPar with a direct *K*-means clustering of subjects based on their dissimilarity values. This comparison focused on the ability of the algorithms to recover true cluster structure across a broad range of test conditions. Test problems were generated in the same manner as described for Experiment 1; however, the number of levels for each factor was increased from 2 to 3. A seventh factor, number of subjects, was also added with levels of $M = 24$, 36, and 48. The level added for the number of objects was $N = 30$, and the level added for the number of clusters was $K = 6$. The third level of cluster density consisted of 10% of subjects in the first cluster with an equal number of subjects in each of the remaining clusters. The level added for the number of dimensions was $D = 4$, the level added for the distance metric was a Minkowski metric with a coefficient of 1.5 (augmenting the coefficients of 2 and 1 for Euclidean and city-block distances, respectively). The perturbation level added to the study was $\pm 20\%$.

Our implementation of *K*-means clustering as a competitive method for ConPar began by stringing out the $N(N-1)/2$ dissimilarity elements for each subject and considering these elements as variables. A partition of subjects based on these variable measures was subsequently obtained using a convergent version of a *K*-means algorithm (MacQueen, 1967). Ten replications of the *K*-means algorithm were completed. Nine of the

Table 2
Experiment 1 results: cluster recovery and computational effort for the two partitioning algorithms

	Within-cluster sum of dissimilarities	Partition diameter
Cluster recovery		
Number of perfect recoveries (out of 64)	45	64
Average adjusted Rand index (across 64 problems)	.8620	1.0000
Smallest adjusted Rand index (across 64 problems)	.4701	1.0000
Computational effort		
Median CPU time	.91	.02
Minimum CPU time	.79	<.01
Maximum CPU time	54.00	.02

Table 3
Experiment 2 results: cluster recovery

Factor	Factor level	Mean adjusted Rand index		% of perfect recoveries	
		K-means	ConPar	*K*-means	ConPar
Number of subjects	$M = 24$	.9592	.9966	86.15	98.90
	$M = 36$	.9775	.9976	91.08	98.22
	$M = 48$	.8745	.9946	75.58	97.39
Number of objects	$N = 10$	.9420	.9935	82.58	96.71
	$N = 20$	.9779	1.0000	92.32	100.00
	$N = 30$	.8914	.9953	77.91	97.81
Number of clusters	$K = 2$	.9614	1.0000	95.06	100.00
	$K = 4$	.9547	.9977	89.30	98.49
	$K = 6$	.8951	.9912	68.45	96.02
Cluster density	Equal	.9616	.9996	92.59	99.73
	60% in cluster 1	.8884	.9979	69.00	98.77
	10% in cluster 1	.9613	.9913	91.22	96.02
Number of dimensions	$D = 2$	.9639	.9967	88.75	98.35
	$D = 3$	.9340	.9969	80.66	98.22
	$D = 4$	.9133	.9952	83.40	97.94
Minkowski metric	2 (Euclidean)	.9309	.9947	83.13	97.67
	1.5	.9382	.9971	84.64	98.35
	1 (city-block)	.9421	.9970	85.05	98.49
Level of perturbation	None	.9747	.9998	92.73	99.86
	$\pm 20\%$	.9402	.9975	84.64	98.77
	$\pm 40\%$	.8963	.9915	75.45	95.88
OVERALL		.9371	.9963	84.27	98.17

replications used random initial seed points, but one of the replications used seed points obtained by applying Ward's minimum variance algorithm (Ward, 1963) and cutting the tree at K clusters. The use of random seed points, as well as those produced by a hierarchical clustering algorithm, is consistent with practice in the psychometric literature (Brusco & Cradit, 2001; Lattin, Carroll, & Green, 2003; Milligan, 1996).

4.2. Experimental results

ConPar provided perfect recovery for 2147 (98.17%) of the 2187 datasets, and yielded an average adjusted Rand index of .9963. These results were markedly superior to the *K*-means implementation, which produced a perfect recovery for 1843 (84.27%) of the 2187 datasets and resulted in an average adjusted Rand index of .9371. ConPar yielded a larger adjusted Rand index than the *K*-means procedure for 339 of the datasets, whereas the reverse was true for only 13 datasets. ConPar was also more efficient than the *K*-means implementation. The minimum, median, and maximum CPU times for ConPar were ($<.01$), .07, and .15 CPU seconds, respectively. The corresponding figures for the *K*-means procedure were ($<.01$), 1.0, and 4.49 CPU seconds, respectively.

The mean adjusted Rand indices and percentage of perfect recoveries provided by *K*-means and ConPar are presented, for each level of each factor, in Table 3. This table shows that, relative to the *K*-means implementation, ConPar produced more perfect recoveries and a larger adjusted Rand index for each level of each factor. Across the 21 factor levels in the table, the minimum average adjusted Rand index for ConPar was .9912, and the minimum percentage of perfect recoveries was 95.88%.

The recovery provided by ConPar was relatively consistent across the different levels for the number of subjects, whereas the recovery performance of the *K*-means procedure dipped noticeably for the $M = 48$ test problems. A similar result was observed for the number of objects, where the percentage of perfect recoveries for the *K*-means algorithm fell to 77.91% for $N = 30$. The recovery performances of ConPar and *K*-means both decreased as the number of clusters increased; however, the decline was much more profound for the *K*-means algorithm. For $K = 2$, ConPar produced perfect recovery of true cluster structure for 100% of the test problems, whereas the corresponding figure for *K*-means

was 95.06%. At $K = 6$, the percentage of perfect recoveries provided by ConPar fell to only 96.02%, but the corresponding percentage for K-means dropped sharply to 68.45%.

Cluster density also had a more profound impact on the K-means algorithm than it did on ConPar. This is perhaps not surprising because of the propensity of the K-means algorithm to produce clusters of approximately equal size. With respect to cluster density, perfect recovery was obtained by ConPar (K-means) for 99.73% (92.59%) of the equal cluster size test problems. For test problems with 60% of the subjects in the first cluster and an equal distribution among the remaining clusters, ConPar provided perfect recovery for 98.77% of the problems, whereas the corresponding figure for K-means was only 69%. Neither the number of dimensions nor the distance metric had a profound impact on the recovery performances of ConPar or K-means.

The effect of error perturbation on true cluster structure recovery was also readily apparent, a finding that is clearly consistent with the literature (Brusco & Cradit, 2001; Milligan, 1989, 1996). Perfect recovery was achieved by ConPar for all but one (99.86%) of the error-free datasets, whereas recovery dipped to 98.77% and 95.88% for the $\pm 20\%$ and $\pm 40\%$ perturbation levels, respectively. The corresponding percentages for the K-means implementation for the error free, $\pm 20\%$, and $\pm 40\%$ perturbation levels were 92.73%, 84.64%, and 75.45%, respectively. Some follow-up results suggested that true cluster recovery deteriorated significantly at $\pm 60\%$ and $\pm 80\%$ levels. However, as observed by Brusco and Cradit (2001), large perturbation levels effectively decimate any true cluster structure in the data. Thus, the fact that recovery is poor for such problems is not especially noteworthy.

5. Experiment #3: identifying the number of subject clusters

5.1. Experimental design

A pervasive problem in cluster analysis is the identification of the appropriate number of clusters, and our situation is no exception. In a practical situation, it would be necessary to run the algorithm for a varying number of clusters and perhaps use a scree diagram procedure to identify an elbow for partition diameter. The quantitative analyst would hope to identify some number of clusters, K', such that partition diameter improves significantly when moving from $K'-1$ to K' clusters, but only modestly (or not at all) when going from K' to $K'+1$ clusters. To evaluate the propensity for this situation to occur in synthetic datasets, we completed a follow-up experiment on a subset of the datasets from Experiment 2.

A subset of the datasets from Experiment 2 was obtained by considering only two of the three levels for each factor. The levels eliminated from the respective factors were $M = 36$, $N = 30$, $K = 6$, density of 10% for the first cluster, $D = 4$, Minkowski metric of 1.5, and perturbation level of $\pm 20\%$. For each of the $2^7 = 128$ datasets, we obtained minimum-diameter partitions for $K-1$, K, and $K+1$ clusters. We subsequently computed the percentage reductions in partition diameter when moving from $K-1$ to K clusters and from K to $K+1$ clusters.

5.2. Experimental results

Table 4 presents statistics pertaining to average partition diameter and CPU time for the $K-1$, K, and $K+1$ solutions. For 127 of the 128 datasets, increasing the number of clusters from $K-1$ to K resulted in a larger reduction in partition diameter than the increase from K to $K+1$ clusters. Moreover, the average partition diameter for $K-1$ clusters was more than double the corresponding average for K clusters, whereas the partition-diameter average for $K+1$ clusters was only 9.9% less than the average for K clusters. For the dataset where $K+1$ clusters were slightly preferred, the result of adding another cluster was to split one of the clusters from the K-cluster solution. Overall, the results of the experiment suggested that using incremental reductions in partition diameters is a viable approach for determining the number of subject clusters in the dataset.

The results in Table 4 also reveal that the CPU times for K clusters exhibit much less variability than the corresponding times for $K-1$ or $K+1$ clusters. Across the 128 test problems, the maximum CPU time for K clusters was only .15 s, whereas several of the $K-1$ and $K+1$ solutions required 20 s or more. Although it would certainly be risky to base selection of the number of clusters on computational effort alone, this measure might provide at least some supporting evidence for the appropriate number of clusters.

Table 4
Experiment 3 results: recovery of the number of clusters

	Assumed number of clusters		
	$K-1$	K	K + 1
Mean partition diameter	1284	619	558
Computational effort			
Median CPU time	.08	.07	.08
Minimum CPU time	<.01	<.01	<.01
Maximum CPU time	26.61	.15	28.92

6. A demonstration using empirical data

To demonstrate ConPar with a practical application, we applied the methodology to a dataset originally presented by Aaker, Kumar, and Day (1995, pp. 655–57). Aaker et al. (1995) collected similarity data from a representative sample of 64 respondents who evaluated 10 brands of soft drinks: Pepsi, Diet 7Up, Calistoga Natural Orange, 7Up, Slice, Schweppes Sparkling Water, Dr. Pepper, Diet Coke, Coke, and Pepsi Light. For each of these 10 brands, respondents provided pairwise dissimilarities and preferences, as well as subjective perceptions regarding each brand along six separate product attributes: (1) refreshing, (2) sweet tasting, (3) fruity, (4) full bodied, (5) young and active, and (6) fattening.

We initially focused our attention on the pooled (across all subjects) dissimilarity matrix, fitting a non-metric Euclidean model to these data using KYST3 (Lattin et al., p. 225) with Kruskal's (1964a, b) original Stress formula (also known as Stress1) as the criterion. Fig. 1 presents the MDS solution for the pooled ($M = 64$) dissimilarity matrix, which served as a benchmark for ConPar. As in past evaluations of these dissimilarity data, a two-dimensional solution achieved a very good fit (Stress = .069) and the configuration of the 10 brands within the two-dimensional space is consistent with published solutions. The horizontal dimension separates cola soft drinks from non-cola soft drinks, while the vertical dimension appears to reveal a separation based on diet versus non-diet drinks. This reproduces the solution published by Aaker et al. (1995, Chapter 21).

We applied the ConPar procedure to this dataset and investigated two, three, and four-cluster solutions using

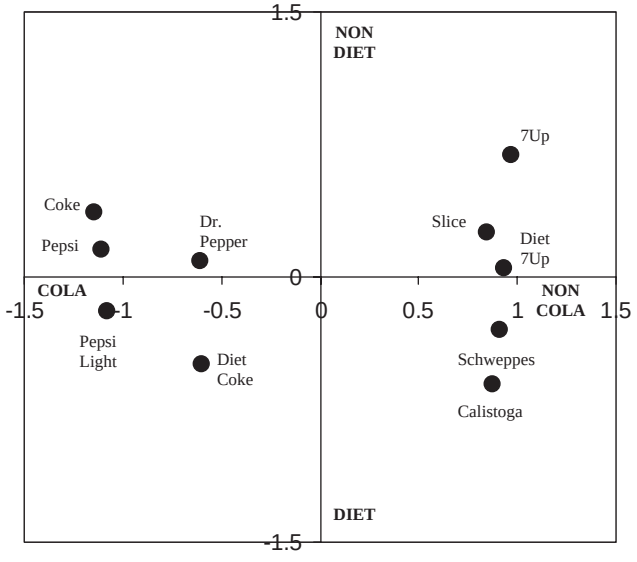

Fig. 1. Two-dimensional Euclidean MDS solution from KYST3 for the pooled dissimilarity data from Aaker et al. (1995, pp. 655–657).

the Euclidean metric. The best solution appeared to be a two-dimensional solution partitioned into three clusters. The number of subjects for clusters 1, 2, and 3 are denoted $|C_1|$, $|C_2|$, and $|C_3|$. Figs. 2a–c show the derived configurations for this three-cluster solution. Fig. 2a presents the two-dimensional space for Cluster 1 ($|C_1| = 33$, Stress = .070). This cluster of subjects reproduces the traditional pooled solution with little or no change in brand locations. The horizontal dimension once again clearly separates colas from non-colas, while the vertical dimension separates diet brands from non-diet brands. Coke and Pepsi are the closest brands in terms of proximity.

Fig. 2b corresponds to a very small cluster with a less-clear solution ($|C_2| = 4$, Stress = .117). The horizontal dimension in this two-dimensional solution appears to maintain the cola/non-cola distinction. Note that the proximities between Pepsi, Coke, Pepsi Light, and Diet-Coke are heightened, while the differences among many of the non-colas are not as pronounced. The vertical dimension is less clear, however, and we have opted not to label this axis. To a limited extent, the vertical dimension maintains a diet/non-diet distinction, although the positions of 7Up and Schweppes are contradictory to this interpretation. It is important to note that this cluster is small (6.25% of the total sample), and the Stress level is significantly higher than for either the first cluster or the benchmark solution for the full sample. It is possible that this smaller cluster represents a small set of outliers with a confused perception of the brands, or an inability to make consistent similarity judgments. What is important is that the ConPar procedure has successfully identified them, allowing the analyst to study the other segments in a clearer context.

The surprising result of this application of ConPar is the emergence of the third cluster presented in Fig. 2c. This segment ($|C_3| = 27$, Stress = 0), constituting 42% of the sample, achieved a perfect fit. In contrast to the previous two clusters, this cluster of subjects was best represented by a one-dimensional solution in which cluster members cleanly separated colas from non-colas, completely ignoring any distinction between diet brands and non-diet brands. What this representation suggests is that respondents in this cluster simplified the pairwise-similarity data collection task and focused their attention on a very obvious, simplistic categorization scheme. It is important to note that the existence of this simplified categorization scheme would not be apparent if a cluster approach had not been applied.

Finally, for comparative purposes, we produced an individual differences MDS solution using Pruzansky's (1975) SINDSCAL program, which is an efficient procedure for fitting Carroll and Chang's (1970) INDSCAL model. The common space for a two-dimensional fit to the soft drink data is displayed in

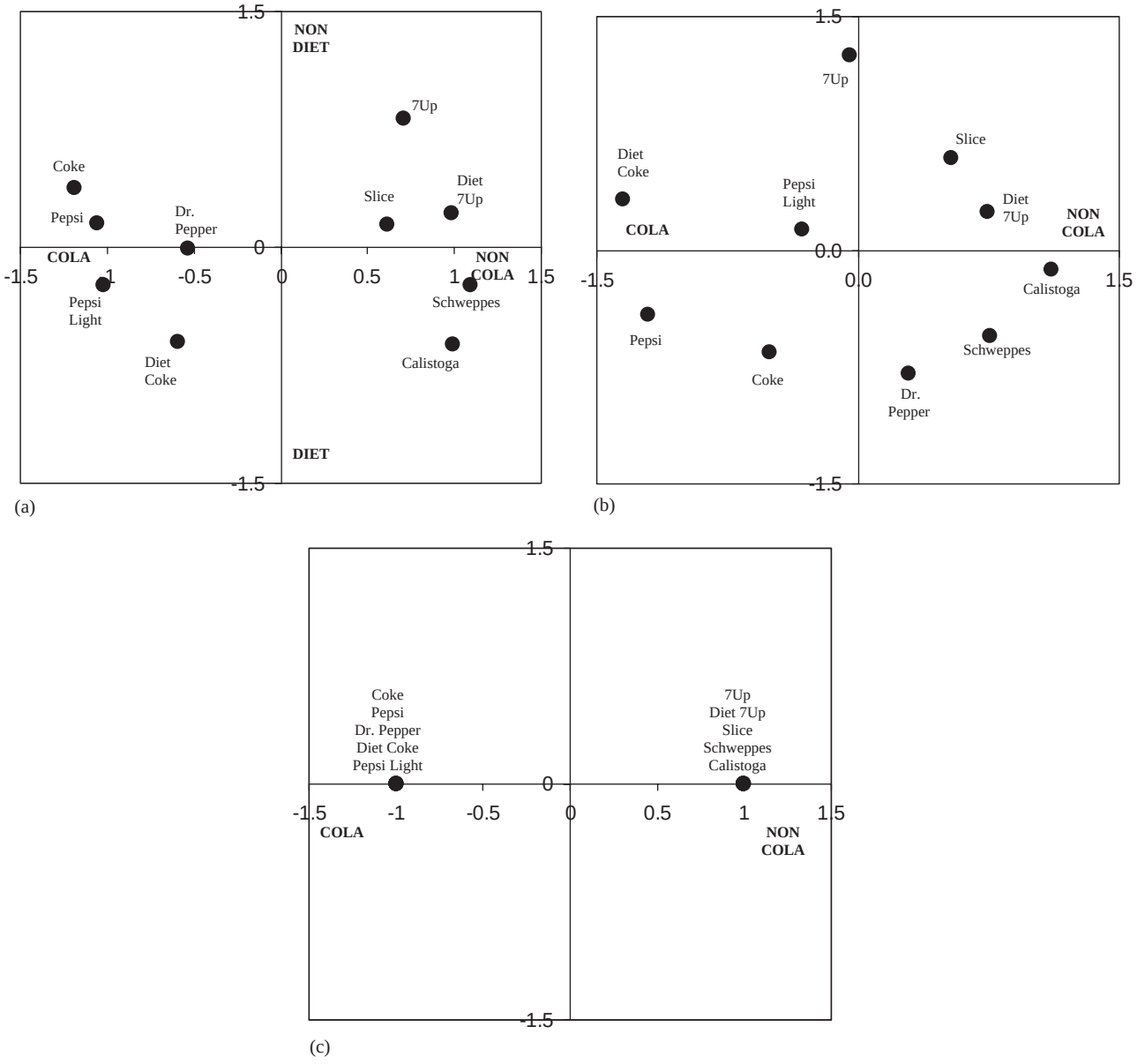

Fig. 2. (a) Two-dimensional Euclidean MDS solution from KYST3 for cluster 1 of the three-cluster ConPar solution, (b) two-dimensional Euclidean MDS solution from KYST3 for cluster 2 of the three-cluster ConPar solution, and (c) two-dimensional Euclidean MDS solution from KYST3 for cluster 3 of the three-cluster ConPar solution.

Fig. 3. Although the basic interpretation of the individual differences common space in Fig. 3 is consistent with Fig. 1, there are a few noticeable differences. For example, in Fig. 1, Coke occupies a more extreme location than Pepsi on both axes, but this is reversed in Fig. 3. Similarly, Slice is slightly to the left of Calistoga and Schweppes in Fig. 1, but slightly to the right in Fig. 3. It is also interesting to observe the relative vertical placements of Pepsi Light and Dr. Pepper in the two figures.

The SINDSCAL solution has the advantage of linking all of the subjects to the common space in Fig. 3 via dimension weights, whereas ConPar produces independent solutions for different clusters. An indication that the individual differences model is recovering the same information as ConPar would manifest itself through the dimension weights. In other words, for the 27 subjects in cluster 3 of the ConPar solution, we might expect larger (smaller) weights on dimension 1 (dimension 2) relative to the other clusters. Such a finding would indicate that subjects in cluster 3 were making only the cola vs. non-cola distinction and ignoring the diet vs. non-diet distinction. Table 5 reports average SINDSCAL dimension weights for the subjects in each of the three clusters of the ConPar solution. These results indicate that, relative to ConPar, the SINDSCAL

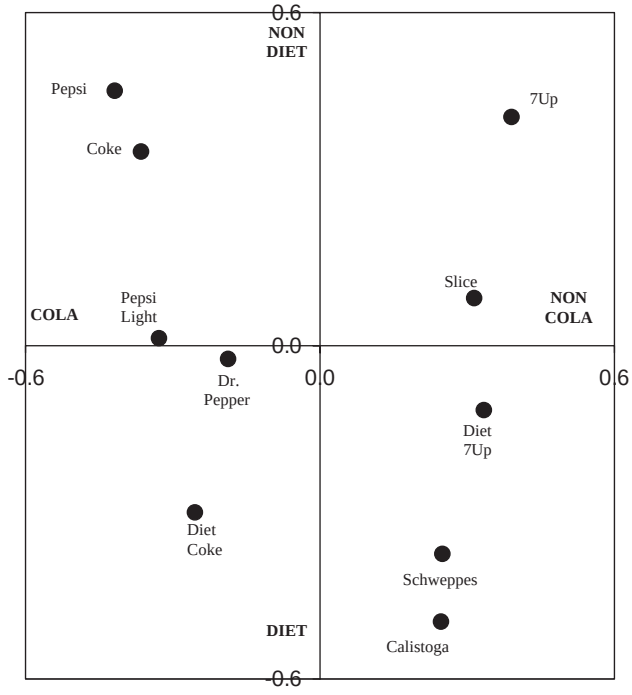

Fig. 3. Two-dimensional SINDSCAL common space solution for the dissimilarity data from Aaker et al. (1995, pp. 655–657).

Table 5
Average SINDSCAL dimension weights for objects in the ConPar clusters

	Cluster size	Average weights	
		Dimension 1	Dimension 2
ConPar cluster 1	33	.673	.377
ConPar cluster 2	4	.543	.284
ConPar cluster 3	27	.681	.346

solution does not as clearly reveal the cluster of subjects using a one-dimensional perception of soft drinks. This important finding suggests that INDSCAL and ConPar each have their own unique advantages and do not necessarily draw out the same information.

7. Discussion

7.1. The problem with averaging

At first glance, it seems entirely reasonable to average individual dissimilarity measures across a number of subjects to obtain a "pooled" proximity matrix. The main result of such an approach would appear to be positive in that the averaging process will undoubtedly result in a minimization of any measurement error present in the dissimilarity judgments. But what if the averaging operation affected the underlying structure of the judgment data? In that case, at the least, one might anticipate the averaging operation to result in a poor model fit. At the extreme, however, a much more serious result might arise if the averaging operation resulted in the successful fitting of a spurious model. "The worst possible effect of averaging would be to alter the underlying psychological structure of the data in such a way that an invalid model appears valid" (Ashby et al., 1994, p. 144). Ashby et al. argue that such a scenario is not only possible, but highly likely.

What actions does an analyst take if the representation of averaged data is suspect? We reviewed a number of options, finding each wanting in one respect or another. We introduced ConPar as a simple method for identifying groups of individual-subject matrices with concordant proximity structures. This method collapses the three-way data into a subject × subject dissimilarity matrix, which is subsequently partitioned using a branch-and-bound algorithm. Extensive Monte Carlo testing of the method reveals that a partition-diameter index is especially effective for recovery of the true subject cluster memberships present in the data.

We compared ConPar to a *K*-means procedure for identifying clusters across more than 2000 synthetic datasets. ConPar provided better recovery than *K*-means across all factor levels, and was especially dominant under certain conditions. The *K*-means algorithm was particularly sensitive to cluster density and the number of true clusters in the dataset, whereas ConPar was far more robust across the levels of these factors.

7.2. Distance metrics, dimensionality, and model selection

It is important to recognize that ConPar is not designed for identifying the appropriate MDS model for one or more groups of subjects. The procedure captures information pertaining to the concordance among pairs of the individual-subject dissimilarity matrices, and subsequently uses this information to form subject groups. The value of this process is that the individual-subject matrices within the selected groups can more safely be pooled for further analysis. The fitting of an MDS model to each group is a separate and subsequent process, and it is possible that the appropriate dimensionality, or even the appropriate distance metric, might vary among the identified groups. As an alternative to pooling within the subsets identified by our proposed method, the research analyst might wish to obtain INDSCAL solutions for one or more of the groups.

7.3. Selecting the number of subject groups

We have found that recovery of the number of subject groups is facilitated by using information regarding the reduction in partition diameter that occurs from incremental increases in K. Although this process is subjective, problems associated with selection of the wrong number of clusters might not be especially severe, especially if one errs on the side of selecting too many groups. In our Monte Carlo experiments, we observed that choosing a slightly larger number of groups typically resulted in the splitting of one or more of the larger (true) groups. Once the individual MDS analyses were run for the selected groups, the analyst is apt to identify the fact that the split groups were, indeed, all members of the same larger group.

7.4. Limitations of the partitioning algorithms

In this paper, we have presented branch-and-bound algorithms for optimizing the selected partitioning indices. Alternative optimal solution procedures include dynamic programming (Hubert et al., 2001) and integer linear programming (Hansen & Jaumard, 1997). For example, using the recent software published by Hubert et al. (2001), the quantitative research analyst could use dynamic programming to provide partitions based on a broad range of heterogeneity indices, but would be limited to roughly 25 subjects. Integer programming methods that incorporate column-generation methods and cutting planes, in addition to branch-and-bound, might be capable of solving quite sizable problems for some partitioning indices; however, they require considerable sophistication on the part of the analyst as well as mathematical programming software.

We have concluded that partition diameter is an effective index for identifying groups of subjects, and our branch-and-bound procedure for optimizing this index can handle 50 or more subjects and 10 or more clusters. Although the algorithm is typically quite efficient, we have occasionally observed some excessive computation times (30 min or more) for problems with $M > 50$ and $K > 5$. For problem instances with hundreds or thousands of subjects, we recommend obtaining an initial solution by applying complete-link hierarchical clustering and cutting the tree at K clusters. An exchange algorithm (Banfield & Bassil, 1977) should subsequently be applied to improve on this starting solution. By replacing branch-and-bound with this heuristic approach, ConPar would be scalable for problems with hundreds or thousands of subjects. Moreover, a heuristic approach can easily be adapted to select from among the set of all minimum-diameter partitions when alternative optima are present.

References

Aaker, D. A., Kumar, V., & Day, G. S. (1995). *Marketing research* (5th ed). New York: Wiley.

Arabie, P., Carroll, J. D., & DeSarbo, W. S. (1987). *Three-way scaling and clustering*. Newbury Park, CA: Sage.

Ashby, F. G. (Ed.). (1992). *Multidimensional models of perception and cognition*. Hillsdale, NJ: Erlbaum.

Ashby, F. G., & Lee, W. W. (1991). Predicting similarity and categorization from identification. *Journal of Experimental Psychology: General*, *120*, 150–172.

Ashby, F. G., Maddox, W. T., & Lee, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychological Science*, *5*, 144–151.

Baker, F. B., & Hubert, L. J. (1976). A graph-theoretic approach to goodness-of-fit in complete-link hierarchical clustering. *Journal of the American Statistical Association*, *71*, 870–878.

Banfield, C. F., & Bassil, L. C. (1977). A transfer algorithm for nonhierarchical classification. *Applied Statistics*, *26*, 206–210.

Borg, I., & Groenen, P. (1997). *Modern multidimensional scaling: theory and applications*. New York: Springer.

Brusco, M. J. (2002). A branch-and-bound algorithm for fitting anti-Robinson structures to symmetric dissimilarity matrices. *Psychometrika*, *67*, 459–471.

Brusco, M. J. (2003). An enhanced branch-and-bound algorithm for a partitioning problem. *British Journal of Mathematical and Statistical Psychology*, *56*, 83–92.

Brusco, M. J. (2004). On the concordance of empirical confusion matrices for visual and tactual letter recognition. *Perception & Psychophysics*, *66*, 392–397.

Brusco, M. J., & Cradit, J. D. (2001). A variable-selection heuristic for K-means clustering. *Psychometrika*, *66*, 249–270.

Carroll, J. D., & Arabie, P. (1980). Multidimensional scaling. *Annual Review of Psychology*, *31*, 607–649.

Carroll, J. D., & Arabie, P. (1998). Multidimensional scaling. In Birnbaum, M. H. (Ed.), *Management, judgment, and decision making* (pp. 179–250). San Diego: Academic Press.

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., & Chang, J. J. (1972). *IDIOSCAL (Individual Differences In Orientation SCALing): a generalization of INDSCAL allowing IDIOsyncratic reference systems as well as an approximation to INDSCAL*. Unpublished manuscript, Murray Hill, NJ: AT&T Bell Laboratories. Presented at a meeting of the Psychometric Society, Princeton, NJ.

Cliff, N. (1968). The 'idealized individual' interpretation of individual differences in multidimensional scaling. *Psychometrika*, *33*, 225–232.

Furnas, G. W. (1989). Metric family portraits. *Journal of Classification*, *6*, 7–52.

Hansen, P., & Delattre, M. (1978). Complete-link cluster analysis by graph coloring. *Journal of the American Statistical Association*, *73*, 397–403.

Hansen, P., & Jaumard, B. (1997). Cluster analysis and mathematical programming. *Mathematical Programming*, *79*, 191–215.

Heiser, W. J. (1989). The city-block model for three-way multidimensional scaling. In Coppi, R., & Belasco, S. (Eds.), *Multiway data analysis* (pp. 395–404). Amsterdam: North-Holland.

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

Hubert, L. J. (1978). Generalized proximity function comparisons. *British Journal of Mathematical and Statistical Psychology*, *31*, 179–192.

Hubert, L. J. (1979). Generalized concordance. *Psychometrika*, *44*, 135–142.

Hubert, L. J. (1987). *Assignment methods in combinatorial data analysis*. New York: Marcel Dekker.

Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*, 193–218.

Hubert, L., Arabie, P., & Meulman, J. (2001). *Combinatorial data analysis: optimization by dynamic programming*. Philadelphia: Society for Industrial and Applied Mathematics (SIAM).

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

Klein, G., & Aronson, J. E. (1991). Optimal clustering: a model and method. *Naval Research Logistics*, *38*, 447–461.

Krieger, A. M., & Green, P. E. (1999). A generalized Rand-index method for consensus clustering of separate partitions of the same data base. *Journal of Classification*, *16*, 63–89.

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*, 115–128.

Lattin, J. J., Carroll, J. D., & Green, P. E. (2003). *Analyzing multivariate data*. Pacific Grove, CA: Thomson.

Lee, M. D. (2001). Determining the dimensionality of multidimensional scaling representations for cognitive modeling. *Journal of Mathematical Psychology*, *45*, 149–166.

Lee, M. D., & Pope, K. J. (2003). Avoiding the dangers of averaging across subjects when using multidimensional scaling. *Journal of Mathematical Psychology*, *47*, 32–46.

MacKay, D. B. (2001). Probabilistic multidimensional scaling using a city-block metric. *Journal of Mathematical Psychology*, *45*, 249–264.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. In Le Cam, L. M., & Neyman, J. (Eds.), *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Vol. 1 (pp. 231–297). Berkeley: University of California Press.

Meulman, J. J., & Verboon, P. (1993). Points of view analysis revisited: fitting multidimensional structures to optimal distance components with cluster restrictions on the variables. *Psychometrika*, *58*, 7–35.

Milligan, G. W. (1989). A validation study of a variable-weighting algorithm for cluster analysis. *Journal of Classification*, *6*, 53–71.

Milligan, G. W. (1996). Clustering validation: results and implications for applied analyses. In Arabie, P., Hubert, L. J., & De Soete, G. (Eds.), *Clustering and classification* (pp. 341–375). River Edge, NJ: World Scientific Publishing.

Milligan, G. W., & Cooper, M. C. (1986). A study of the comparability of external criteria for hierarchical cluster analysis. *Multivariate Behavioral Research*, *21*, 441–458.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology*, *43*, 25–53.

Pruzansky, S. (1975). *How to use SINDSCAL: a computer program for individual differences in multidimensional scaling*. Murray Hill, NJ: Bell Telephone Laboratories (Unpublished manuscript).

Robinson, W. S. (1951). A method for chronologically ordering archaeological deposits. *American Antiquity*, *16*, 293–301.

Ross, J. (1966). A remark on Tucker and Messick's "Points of view" analysis. *Psychometrika*, *31*, 27–32.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461–464.

Siegler, R. S. (1987). The perils of averaging data over strategies: an example from children's addition. *Journal of Experimental Psychology: General*, *116*, 250–264.

Takane, Y., Young, F. W., & de Leeuw, J. (1977). Nonmetric individual differences multidimensional scaling: an alternating least squares method with optimal scaling features. *Psychometrika*, *42*, 7–67.

Tucker, L. R., & Messick, S. J. (1963). An individual differences model for multidimensional scaling. *Psychometrika*, *28*, 333–367.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*, 236–244.

Winsberg, S., & De Soete, G. (1993). A latent class approach to fitting the weighted Euclidean model, CLASCAL. *Psychometrika*, *58*, 315–330.

Young, F. W., & Hamer, R. M. (1987). *Multidimensional scaling: history, theory, and applications*. Hillsdale, NJ: Erlbaum.

Zinnes, J. L., & MacKay, D. B. (1983). Probabilistic multidimensional scaling: complete and incomplete data. *Psychometrika*, *48*, 27–48.